

ISSN 2237-2083

REVISTA DE ESTUDOS DA LINGUAGEM

Faculdade de Letras da UFMG

JAN./MAR. 2026

V. 34 – N. 1



REVISTA DE ESTUDOS DA
LINGUAGEM

Universidade Federal de Minas Gerais

REITORA: Alessandro Fernandes Moreira; VICE-REITOR: Alamanda Kfoury Pereira

Faculdade de Letras

DIRETORA: Sandra Gualberto Bianchet; VICE-DIRETOR: Lorenzo Teixeira Vitral

Editora-chefe

Janayna Maria da Rocha Carvalho (UFMG)

Editoras-associadas

Ana Regina Vaz Calindro (UFRJ)

Maria Mendes Cantoni (UFMG)

Conselho Editorial

Alejandra Vitale (UBA, Ciudad Autónoma de Buenos Aires, Argentina), Didier Demolin (Université de la Sorbonne Nouvelle Paris 3, Paris, França), Ieda Maria Alves (USP, São Paulo/SP, Brasil), Jairo Nunes (USP, São Paulo/SP, Brasil), Scott Schwenter (OSU, Columbus, Ohio, Estados Unidos), Shlomo Izre'el (TAU, Tel Aviv, Israel), Stefan Gries (UCSB, Santa Barbara/CA, Estados Unidos), Teresa Lino (NOVA, Lisboa, Portugal), Tjerk Hagemeijer (ULisboa, Lisboa, Portugal)

Editor de Arte

Emerson Eller

Projeto Gráfico

Stéphanie Paes

Secretaria

Ana Clara Marques

Revisão e normalização

Ana Regina Vaz Calindro (UFRJ), Janayna Maria da Rocha Carvalho (UFMG), Maria Mendes Cantoni (UFMG)

Diagramação

Izabelly Silva Duque

REVISTA DE ESTUDOS DA LINGUAGEM



Rev. Est. Ling. | Belo Horizonte | v. 34 | n. 1 | jan./mar. 2026 | 457 p. | e-ISSN 2237-2083



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Os conceitos emitidos em artigos assinados são de responsabilidade exclusiva de seus autores.

Indexadores

Diadorim [Brazil]
DRJI (Directory of Research Journals Indexing) [India]
EBSCO [USA]
EuroPub [England]
JournalSeek [USA]
Latindex [Mexico]
Linguistics & Language Behavior Abstracts [USA]
MIAR (Matriu d'Informació per a l'Anàlisi de Revistes) [Spain]
MLA Bibliography [USA]
OAJI (Open Academic Journals Index) [Russian Federation]
Portal CAPES [Brazil]
REDIB (Red Iberoamericana de Innovación y Conocimiento Científico) [Spain]
Sindex (Scientific Indexing Services) [USA]
Web of Science [USA]
WorldCat / OCLC (Online Computer Library Center) [USA]
ZDB (Elektronische Zeitschriftenbibliothek) [Germany]

REVISTA DE ESTUDOS DA LINGUAGEM, v. 1 - 1992 - Belo Horizonte, MG,
Faculdade de Letras da UFMG

Histórico:

1992 ano 1, n.1 (jul/dez)
1993 ano 2, n.2 (jan/jun)
1994 Publicação interrompida
1995 ano 4, n.3 (jan/jun); ano 4, n.3, v.2 (jul/dez)
1996 ano 5, n.4, v.1 (jan/jun); ano 5, n.4, v.2; ano 5, n. esp.
1997 ano 6, n.5, v.1 (jan/jun)

Nova Numeração:

1997 v.6, n.2 (jul/dez)
1998 v.7, n.1 (jan/jun)
1998 v.7, n.2 (jul/dez)

1. Linguagem - Periódicos I. Faculdade de Letras da UFMG, Ed.

CDD: 401.05

Faculdade de Letras da UFMG
Seção de Periódicos, sala 2017
Av. Pres. Antônio Carlos, 6627 – Pampulha
31270-901 – Belo Horizonte, MG – Brasil
Tel.: (31) 3409-6009
www.lettras.ufmg.br/periodicos
periodicosfaleufmg@gmail.com

Sumário

- 8 *Per aspera ad astra*: Improving the Automatic Evaluation of a Universal Dependencies Treebank for a Low-Resource Language
Per aspera ad astra: melhorando a avaliação automática de um treebank no modelo Dependências Universais para uma língua com poucos recursos
Leonel Figueiredo de Alencar; Hélio Leonam Barroso Silva; Juliana Lopes Gurgel; Dominick Maia Alexandre
- 51 Code-mixing, code-switching e empréstimos no Hunsrückisch: uma análise de contato linguístico
Code-Mixing, Code-Switching and Borrowing in Hunsrückisch: An Analysis of Language Contact
Leonardo Alves
- 81 A iconicidade no Português e sua relação com idade de aquisição e classes de palavras
Iconicity in Portuguese and its Relationship with Age of Acquisition and Parts of Speech
Thayná Cristina Ananias; Mahayana Godoy
- 106 Sobre as propagandas dos governos Dilma e Bolsonaro: uma análise dos sentidos atribuídos à escola (2013-2019)
On the Government Campaigns of Dilma and Bolsonaro: An Analysis of the Meanings Attributed to School (2013-2019)
Nilton César Ferreira; Francisco Renato Lima

- 130** Building the Terminology of an Audiovisual Saga: The Example of Star Wars
Elaborando a terminologia de uma saga audiovisual: o exemplo de Star Wars
Guilherme Fromm; Guilherme Rodrigues Ferreira
- 154** Tipos de entidades semânticas dos referentes que são tópicos discursivos em descrições, narrativas de experiência e relatos de opinião
Types of Semantic Entities of Referents that are Discourse Topics in Descriptions, Experience Narratives and Opinion Reports
Aline Gomes Garcia
- 181** Variação diacrônica do grau de transparência linguística do português brasileiro: o impacto do aumento percentual do uso de “a gente” em lugar de “nós”
Diachronic Variation of Linguistic Transparency in Brazilian Portuguese: The Role of the Percentage Increase in the Use of “a gente” (“we”) Instead of “nós” (“we”)
Alessandra Regina Guerra
- 209** Ensino de língua materna e edição de livros didáticos: políticas educacionais e linguísticas do Estado Novo brasileiro
Mother Tongue Teaching and Textbook Publishing: Educational and Linguistic Policies in Brazil
Cristian Imbruniz
- 229** Educação bilíngue para surdos em Portugal e Brasil: testemunhos de professores de Língua Gestual Portuguesa e Língua Brasileira de Sinais
Bilingual Education for Deaf People in Portugal and Brazil: Testimonies of Portuguese Sign Language and Brazilian Sign Language Teachers
Lodenir Becker Karnopp; Cristina Isabel Caciones Gil; Paulo Vaz de Carvalho
- 249** A tipologia de tópicos no português brasileiro na interface entre sintaxe e estrutura informacional
The Topic Typology in Brazilian Portuguese at the Interface Between Syntax and Information Structure
Renato Lacerda
- 271** Um estudo das representações linguísticas de descendentes germânicos em Niterói
A Study of the Linguistic Representations of German Descendants in Niterói
Anderson Lucas da Silva Macedo; Mônica Maria Guimarães Savedra

- 291 A alteridade na aquisição da língua materna: um olhar retrospectivo
Alterity in the Acquisition of the Mother Tongue: A Retrospective Look
Giovane Fernandes Oliveira
- 321 As interrogativas polares e a falha de pressuposição em português brasileiro: uma análise sob a perspectiva da semântica e da pragmática formais
Polar Interrogatives and Presupposition Failure in Brazilian Portuguese: An Analysis from the Perspective of Formal Semantics and Pragmatics
Fernanda Rosa da Silva; Elisa de Oliveira Silva
- 348 Construções insubordinadas causais no português brasileiro: um estudo da fala espontânea
Insubordinated Constructions in Brazilian Portuguese Speech: A Study of Spontaneous Speech
Alessandra Emanuelle Macieira Silva; Luana Lopes Amaral
- 371 Sobre o papel das relações dialógicas em artigos científicos da área das Ciências Humanas
On the Role of Dialogic Relations in Scientific Articles in the Humanities
Fabrício José da Silva
- 393 “Strikingly”, an Academic Writer’s Style Can Be Interactional: Towards Characterizing Metadiscourse Style in Applied Linguistics Research Articles
“Surpreendentemente”, o estilo de um autor acadêmico pode ser interacional: rumo à caracterização do estilo metadiscursivo em artigos de pesquisa da linguística aplicada
Janailton Mick Vitor da Silva
- 422 Desbravando O Continente: a Estilística de Corpus aplicada à literatura regionalista brasileira
Unveiling O Continente: Corpus Stylistics Applied to Brazilian Regionalist Literature
Geovani Henrique Santos de Souza; Rozane Rodrigues Rebechi

Per aspera ad astra: Improving the Automatic Evaluation of a Universal Dependencies Treebank for a Low-Resource Language

Per aspera ad astra: melhorando a avaliação automática de um treebank no modelo Dependências Universais para uma língua com poucos recursos

Leonel Figueiredo de Alencar

Universidade Federal do Ceará (UFC)
Fortaleza | CE | BR
leonel.de.alencar@ufc.br
<https://orcid.org/0000-0001-8148-6994>

Hélio Leonam Barroso Silva

Universidade Federal do Ceará (UFC)
Fortaleza | CE | BR
heliolbs@alu.ufc.br
<https://orcid.org/0000-0003-2752-9871>

Juliana Lopes Gurgel

Universidade Federal do Ceará (UFC)
Fortaleza | CE | BR
julianagurgel@letras.ufc.br
<https://orcid.org/0009-0005-4640-454X>

Dominick Maia Alexandre

Universidade Federal do Ceará (UFC)
Fortaleza | CE | BR
dominick@letras.ufc.br
<https://orcid.org/0009-0000-7749-7762>

Abstract: Until recently, advancements in digital humanities have favored majority languages. The construction of treebanks for 14 Brazilian Indigenous languages within the Universal Dependencies (UD) framework marks a significant step toward bridging the digital divide affecting these minority languages. However, aside from the Nheengatu treebank, these corpora are small and/or receive low quality ratings. This paper details our recent efforts to enhance the Nheengatu treebank. We examined UD's automatic evaluation methodology to pinpoint the areas we should focus on. Among other improvements, we corrected nearly all 2,726 errors flagged by the Udaapi framework, a core component of the UD rating system. As a result, the treebank's rating advanced from 2.0 stars in UD v2.14 to 3.5 stars in v2.15 and 4.0 stars in the upcoming v2.16. It is now the largest and best evaluated among the 21 UD treebanks for Amerindian languages. A corollary contribution of the annotation revision was the identification of discrepancies between the UD documentation and the Udaapi algorithm. Specifically, while the documentation permits (i) assigning the Degree feature to nouns and (ii) using the ExtPos feature to mark an unexpected POS tag on the head of an exocentric MWE, Udaapi systematically treats both configurations as annotation errors.

Keywords: Nheengatu; Universal Dependencies; treebank; morphosyntactic annotation; computational linguistics.



Resumo: Até recentemente, os avanços nas humanidades digitais privilegiaram as línguas hegemônicas. A construção de *treebanks* para 14 línguas indígenas brasileiras no modelo Dependências Universais (UD) representa uma iniciativa importante para reduzir a exclusão digital que afeta essas línguas minoritárias. No entanto, com exceção do *treebank* de nheengatu, esses *corpora* são pequenos e/ou recebem notas baixas de avaliação. Neste artigo, detalhamos nossos esforços recentes para aprimorar o *treebank* de nheengatu. Analisamos a metodologia de avaliação automática de UD a fim de identificar os pontos que exigiam maior atenção. Entre outras melhorias, corrigimos quase todos os 2.726 erros apontados pelo *framework* Udapi, um componente central do sistema de pontuação de UD. Como resultado, a avaliação do *treebank* avançou de 2,0 estrelas na UD v2.14 para 3,5 estrelas na v2.15 e 4,0 estrelas na próxima v2.16. Atualmente, esse é o maior e mais bem avaliado entre os 21 *treebanks* de línguas ameríndias da coleção UD. Uma contribuição corolária da revisão das anotações foi revelar discrepâncias entre a documentação de UD e o algoritmo do Udapi. Especificamente, enquanto a documentação permite (i) a atribuição do traço *Degree* a substantivos e (ii) o uso do traço *ExtPos* para marcar uma classe de palavra inesperada no núcleo de uma MWE exocêntrica, o Udapi trata sistematicamente ambas as configurações como erros de anotação.

Palavras-chave: Nheengatu; Dependências Universais; *treebank*; anotação morfossintática; linguística computacional.

1 Introduction

Rodrigues (1966, p. 5) defined the documentation, description, and comparison of Indigenous languages as “the greatest task of linguistics in Brazil.” More than half a century later, it is evident that Indigenous linguistics never reached such magnitude in the country, despite the considerable accumulation of publications over this period (Oliveira; Camacho, 2013; Rodrigues, 1986; Seki, 1999; Storto, 2019). On the other hand, natural language processing, computational linguistics, and corpus linguistics have expanded in Brazil at an increasing pace since the 1990s, completely detached from the Indigenous languages. Thus, Seki’s (1999, p. 285) proposal to create a unified database compiling data from various studies on these

languages, especially the “corpora that mediate between spoken languages and the produced works”, was ignored.

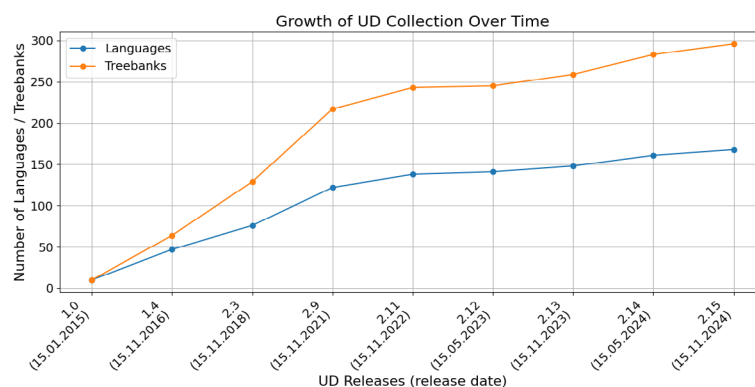
Fortunately, the marginalization of Indigenous languages from digital humanities begins to be reversed. We highlight some recent initiatives. Galves *et al.* (2017) and Sandalo and Galves (2023) propose an annotation scheme for a constituency treebank of Kadiwéu. Rueter *et al.* (2021) and Santos, Aragon, and Gerardi (2024) address the development of treebanks in the Universal Dependencies framework (Marneffe *et al.*, 2021) for the Apurinã language and various languages of the Tupian language family, respectively. D’Angelis, Oliveira, and Schwade (2021) deals with the translation of a smartphone operating system into Nheengatu and Kaingang. Cavalin *et al.* (2023), Pinhanez, Cavalin, and Nogima (2024), and Silva and Pardo (2024) report on experiments in the field of artificial intelligence involving, respectively, language identification, machine translation, and grammar induction for various Brazilian Indigenous languages.

The Universal Dependencies (UD) model underpins the largest collection of syntactically annotated corpora (Marneffe *et al.*, 2021; Zeman, 2023). Launched in 2015 with 10 treebanks for 10 languages, this collection has grown exponentially, reaching, in less than a decade, 296 treebanks for 168 languages (Figure 1). This represents, according to Equation 1, a compound annual growth rate (CAGR) of 41.15%, reflecting the increasingly widespread adoption of the UD model and the growing interest in multilingual morphosyntactic annotation and parsing.

$$CAGR = \left(\left(\frac{\text{final value}}{\text{initial value}} \right)^{\frac{1}{\text{years}}} - 1 \right) \times 100\% = \left(\left(\frac{296}{10} \right)^{\frac{1}{9.83}} - 1 \right) \times 100\% \approx 41.15\%$$

(1)

Figure 1 – Growth of the UD collection in terms of treebanks and languages since 2015



Treebanks have been used for language documentation, morphosyntactic research—particularly in linguistic typology—and the development of computational tools such as syntactic parsers. We have identified, in release 2.15 of this collection, a total of 21 treebanks for 20 languages of the Americas, 14 of which are from Brazil, including Nheengatu. First introduced in the November 15, 2022 release as the seventh-largest Amerindian language treebank in terms of token count, UD_Nheengatu-CompLin became the largest in the UD collection by this criterion in the November 15, 2023 release (Alencar 2024a, 2024b).

Once widespread in Amazonia due to its adoption as *lingua franca* during the colonial period, Nheengatu—also known, *inter alia*, as Modern Tupi and Amazonian General Language (*Língua Geral Amazônica* in Portuguese)—has rapidly shrunk with the advancement of Western civilization in the region since the second half of the 19th century (Borges, 1996; Freire, 2011; Navarro; Ávila; Trevisan, 2017; Rodrigues, 1986; Rodrigues, 1993; Rodrigues, 1996; Rodrigues; Cabral, 2011). Reduced to an estimated 6,000 speakers in Brazil and 8,000 in Colombia, Nheengatu faces extinction, ranking at level 6b on the Expanded Graded Intergenerational Disruption Scale (EGIDS) (Eberhard; Simons; Fennig, 2025). The recent digitalization efforts have increased its Digital Language Support (DLS) from 0.07 (Eberhard; Simons; Fennig, 2023) to 0.21 (Eberhard; Simons; Fennig, 2025). However, Nheengatu remains threatened by the digital divide that marginalizes Indigenous languages. In contrast, majority languages such as Portuguese, Spanish, and English reach the maximum DLS score of 1.00.

Over five releases of the collection, UD_Nheengatu-ComplIn has grown not only quantitatively but also qualitatively. In the UD project's automatic rating system, which ranges from zero to five stars, UD_Nheengatu-ComplIn was, as of the May 15, 2024 version, one of only four Amerindian language treebanks rated at 2.0 stars, whereas the ratings of the other 17 treebanks of this language group ranged between 0 and 1.5 stars. The 2.0-star rating, however, is only half the number of stars assigned to treebanks of majority languages such as UD_Portuguese-Porttinari (Duran *et al.*, 2023) and UD_Portuguese-CINTIL (Branco *et al.*, 2022).

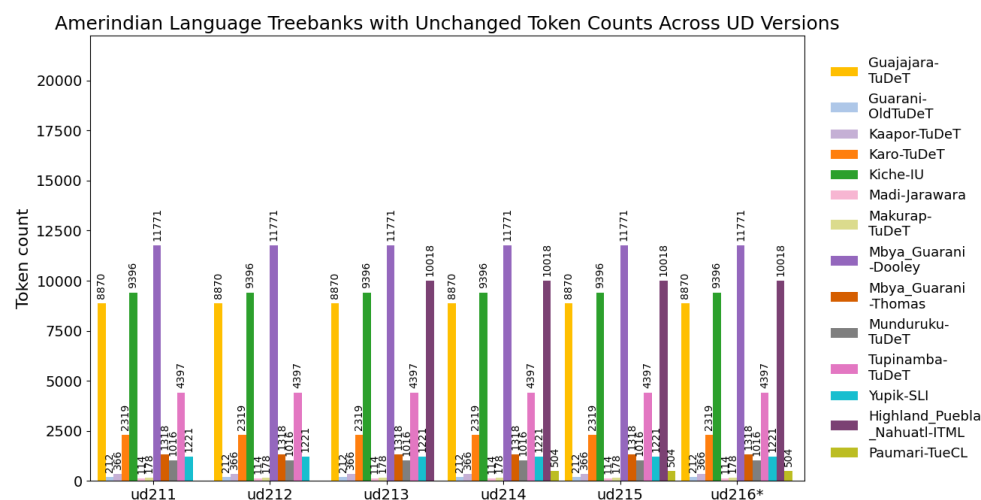
This paper describes our latest efforts to enhance UD_Nheengatu-ComplIn, which led to a 1.5-star increase in the UD v2.15 version, released on November 15, 2024, making it the only Amerindian language treebank with more than two stars. The primary factor behind this rating boost was the systematic correction of 98.42% of the errors identified by the Udapi framework (Popel; Žabokrtský; Vojtek, 2017). This revision involved a reworking of our approach to Nheengatu's verb valency and verbal inflectional morphology. It also revealed inconsistencies between Udapi's constraints and the UD documentation. For the upcoming May 15, 2025 release of the UD collection, we have continued improving UD_Nheengatu-ComplIn. By dividing the treebank into training and test sets that meet the minimum word count thresholds established in the scoring algorithm, we increased its rating to 4.0 stars in the development version of February 12, 2025. Meanwhile, other Amerindian treebanks have either maintained or lost stars.

In the next section, we compare UD_Nheengatu-ComplIn with other Amerindian language treebanks based on statistical parameters such as token count, sentence count, and annotation richness across different UD versions. In Section 3, we present the two evaluation methodologies applied to all UD treebanks: validation and rating. We also highlight UD_Nheengatu-ComplIn's recent achievements in this domain. Since joining the UD collection in the November 15, 2022 v2.11 release as a fully validated treebank, UD_Nheengatu-ComplIn has maintained this status across all subsequent releases. In UD v2.15, it became the only Amerindian treebank ranked within the top 33.5% according to the star rating system. In Section 4, we discuss the errors detected in UD_Nheengatu-ComplIn by the Udapi framework, one of the key components of the rating system, as well as the annotation scheme revisions that successfully resolved most of these errors, thereby increasing the treebank's star rating. In the final section, we summarize our main contributions and outline strategies to reach the top of the rating system, which currently consists of 4.5 stars.

2 Overview of UD_Nheengatu-CompLin

The UD_Nheengatu-CompLin treebank joined the UD collection on November 15, 2022, in release 2.11, ranking seventh and tenth among the then 18 treebanks of Amerindian languages in terms of token and sentence count, respectively¹. A year later, in version 2.13, it became the largest Amerindian language treebank in terms of tokens, and in version 2.14, it also took first place in sentence count. In the 2.15 release of the UD collection, it remains the leader in both criteria, with 61.7% more tokens than UD_Mbya_Guarani-Dooley (the second-largest treebank in token count) and 27.7% more sentences than UD_Kiche-IU (the second-largest in sentence count). In the development snapshot of January 20, 2025, these differences have increased to 75% and 40.6%, respectively. Figure 2 displays the treebanks whose token counts remained unchanged between UD 2.11 and this snapshot, whereas Figure 3 shows those whose token counts changed over this period. Figure 4, by contrast, tracks across these UD versions only the treebanks that rank among the ten largest by sentence count in the January 20, 2025 snapshot.

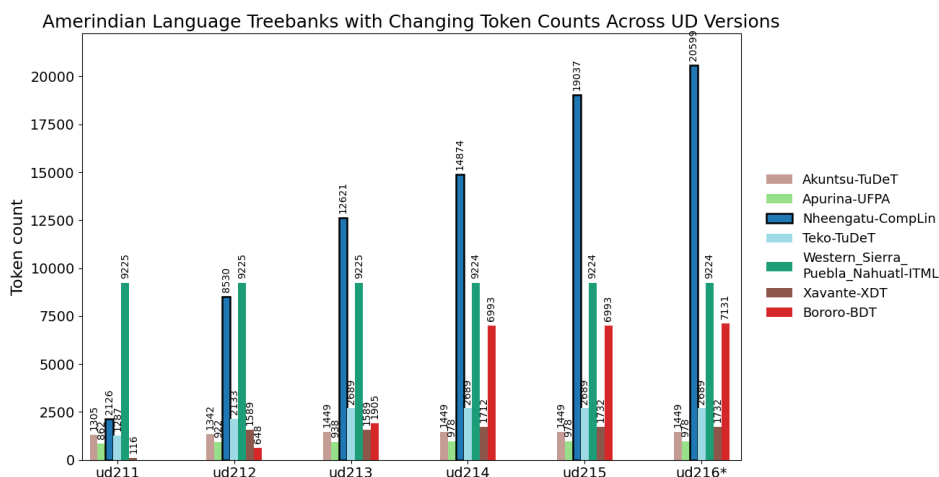
Figure 2 – Token counts of Amerindian language treebanks whose token counts remained unchanged between UD release 2.11 and the pre-release UD 2.16* development snapshot of January 20, 2025. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

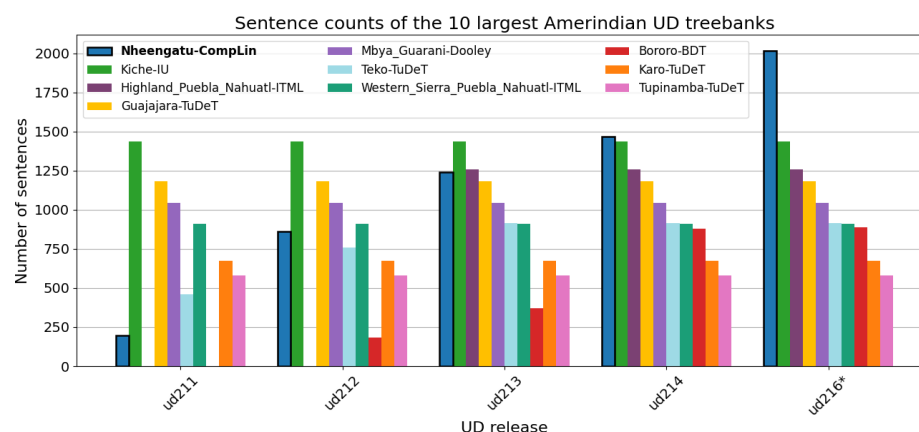
¹ All UD versions are stored in the digital library LINDAT/CLARIAH-CZ, available at <<https://lindat.mff.cuni.cz/repository/xmlui/>>.

Figure 3 – Token counts of Amerindian language treebanks whose token counts changed between UD release 2.11 and the pre-release UD 2.16* development snapshot of January 20, 2025. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

Figure 4 – Sentence counts across different UD versions for the ten largest Amerindian language treebanks by sentence count, as determined in the pre-release UD 2.16* development snapshot of January 20, 2025. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

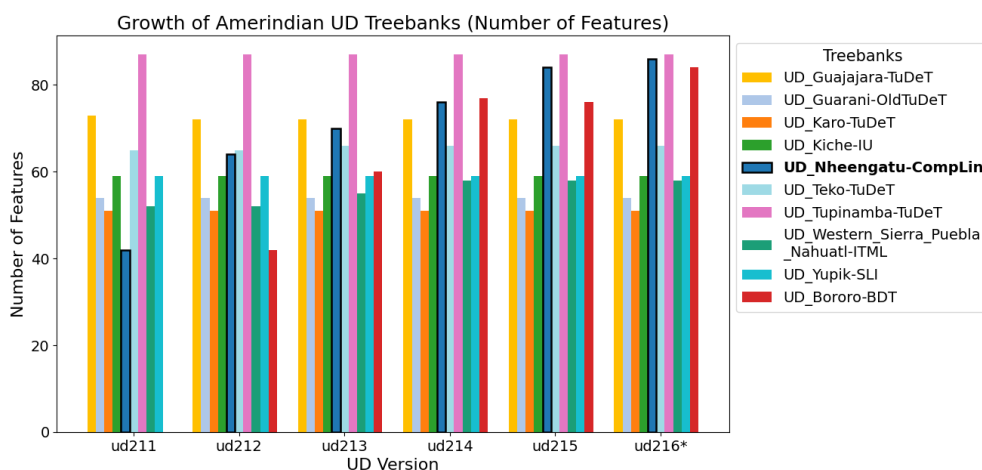
Given the disparity in growth rates between UD_Nheengatu-CompLin and the other treebanks of Amerindian languages, it is likely that the gap between them will widen even further in the next UD version. As shown in Figures 2 and 3, UD_Nheengatu-CompLin is the only corpus in this group of languages that has substantially grown with each UD release since its launch in UD v2.11. We observe that 15 treebanks have seen no increase in size during this period². Only the Bororo and Xavante language treebanks have had token increments since UD v2.13. The UD_Teko-TuDeT treebank experienced significant growth only in the two versions following its release. In UD v2.12, its token count increased by 65.7%, and in the next

² Included in Figure 3 among the treebanks with variation in token counts, UD_Western_Sierra_Puebla_Nahuatl-ITML did not increase in size over the period considered; instead, its token count decreased by one token between UD v2.13 and v2.14.

release, it grew by another 26.1%. However, since v2.14, no further increases have been recorded, indicating a stabilization in treebank size. The Apurinã, Akuntsu, and Xavante treebanks also seem to have reached their final size plateau.

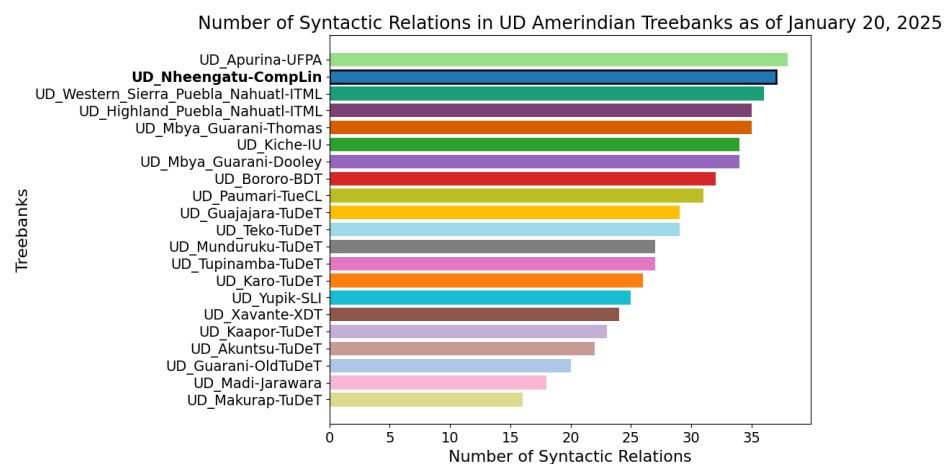
Regarding the different layers of UD annotation, UD_Nheengatu-CompLin ranks among the Amerindian language treebanks with the highest counts of part-of-speech tags, morphological features, and syntactic relations. As shown in Figure 5, the Nheengatu treebank has more than doubled its inventory of morphological features between its first version and the present, now ranking second in this category, the same position it holds for syntactic relations (Figure 6). On the other hand, UD_Nheengatu-CompLin is among the five Amerindian language treebanks with the highest number of part-of-speech tags, totaling 16 out of the maximum of 17 (Figure 7).

Figure 5 – Evolution of the number of features in the 10 treebanks with the highest feature counts in the pre-release UD 2.16* development snapshot of January 20, 2025, tracked across the successive versions since 2.11



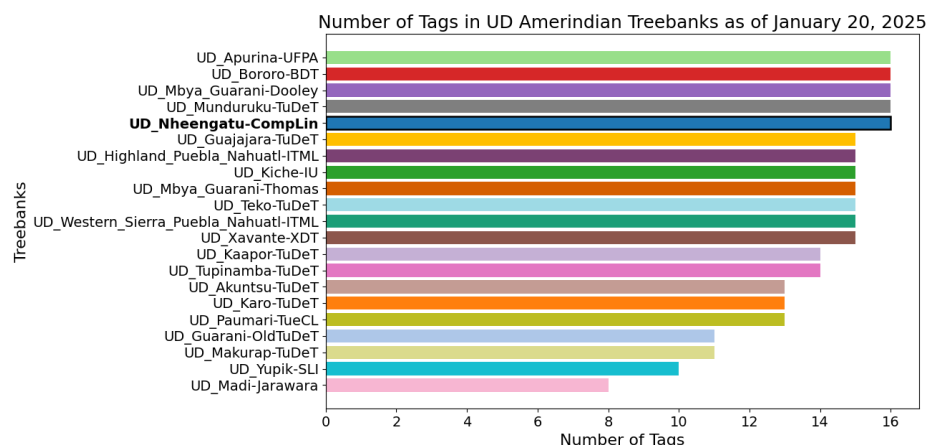
Fonte: elaboração própria.

Figure 6 – Number of dependency relations per Amerindian language treebank in the pre-release UD 2.16* development snapshot of January 20, 2025. The data was computed by UD project's conllu-stats.pl script, including subtyped relations.



Fonte: elaboração própria.

Figure 7 – Number of part-of-speech tags per Amerindian language treebank in the pre-release UD 2.16* development snapshot of January 20, 2025



Fonte: elaboração própria.

In summary, UD_Nheengatu-Complin stands out in release 2.15 of the UD collection as the largest of the 21 treebanks of Amerindian languages. It surpasses UD_Mbya_Guarani-Doolley by 61.7% in the number of tokens and UD_Kiche-IU by 40.6% in the number of sentences, which rank second in these two criteria, respectively. In the next section, we will see how UD_Nheengatu-Complin, which in UD 2.14 was among the group of only four treebanks of these languages with 2 stars, advanced by 1.5 in this ranking in UD v2.15, reaching 4 stars in the development version of February 12, 2025.

3 The automatic evaluation of UD treebanks

The exponential growth of the UD collection in just under a decade would have been impractical if manual validation were required for all 296 treebanks. Without this validation, the usability of these resources would be drastically reduced, if not entirely infeasible, for the development of natural language processing applications and linguistic investigations. On the other hand, the increasing coverage of the collection, with a growing number of languages represented by more than one treebank, has made it necessary to compare the treebanks in qualitative terms to guide the selection of the most appropriate resources for specific scenarios.

To address these requirements, every treebank is subject to an automatic evaluation that assesses both quantitative and qualitative aspects across various levels of annotation, in addition to analyzing textual genre diversity and the structure of its GitHub repository. This evaluation utilizes two distinct methodologies: one focused on validating the treebank and the other on ranking it on a scale from 0 to 5 stars. In the following two subsections, we will explore each process in detail.

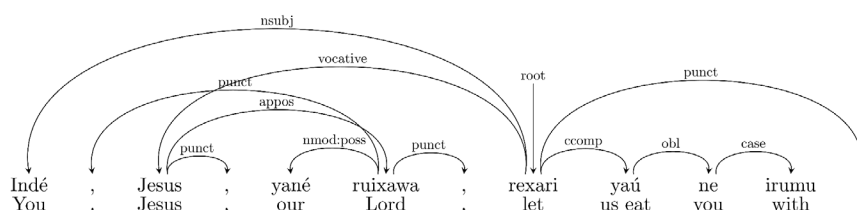
3.1 The automatic validation

The validation process employs the programs `check_files.pl` and `validate.py`. While the former verifies whether the treebank repository's file structure meets a series of requirements, the latter examines, for each sentence, whether it constitutes a tree-like graph, whether its annotation conforms to the CoNLL-U format, and whether it adheres to certain constraints of the UD theory. The CoNLL-U format consists of a table with ten columns, each specifying a particular piece of information about a word in a sentence: (i) ID, (ii) form, (iii) lemma, (iv) universal part-of-speech (UPOS), (v) treebank-specific part-of-speech, (vi) morphological features (FEATS), (vii) governing word (HEAD), (viii) universal dependency relation (DEPREL), (ix) enhanced dependency, and (x) other annotation (Marneffe *et al.*, 2024a). At a minimum, the fields UPOS, HEAD, and DEPREL must be filled with categories specified by the model (Marneffe *et al.*, 2024b). The FEATS column must be either underspecified with an underscore or consist of attribute-value pairs in the prescribed format.

Additionally, the validity of specific combinations of word classes with syntactic relations is checked. For example, `advmod` requires one of the labels from the set {ADV, ADJ, CCONJ, DET, PART, SYM}. Similarly, only a pronoun (PRON) or a determiner (DET) can serve as a syntactic determiner (`det`). Certain tree configurations are identified as invalid, such as a single node governing two subjects or a punctuation mark being attached to a governor in a non-projective manner, as shown in Figure 8, an error corrected in Figure 9. Violations of these and other constraints are classified as errors, preventing the treebank from being deemed valid.

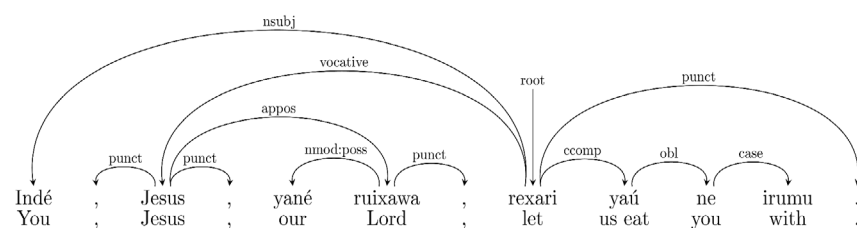
- 1) Indé, Jesus, yané r-uixawa, re-xari ya-ú ne irumu.
 you Jesus our CONT-Lord 2SG.ACT-let 1PL.ACT-eat you with
 ‘You, Jesus, our Lord, let us eat with you.’ (Avila, 2021, p. 777)

Figure 8 – Dependency graph of (1) with punctuation attached non-projectively



Fonte: elaboração própria.

Figure 9 – Dependency graph of sentence Avila2021:0:0:2 from UD_Nheengatu-Complin with all punctuation attached projectively

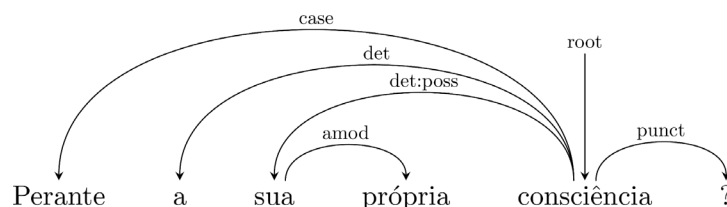


Fonte: elaboração própria.

Other requirements are less strict, and when violated, they generate messages classified as warnings, which do not affect the validity of the treebank. One example of a configuration that triggers such a message is when a determiner governs another node, as shown in Figure 10, which contradicts the UD principle that only content words can function as governors (Marneffe *et al.*, 2021).

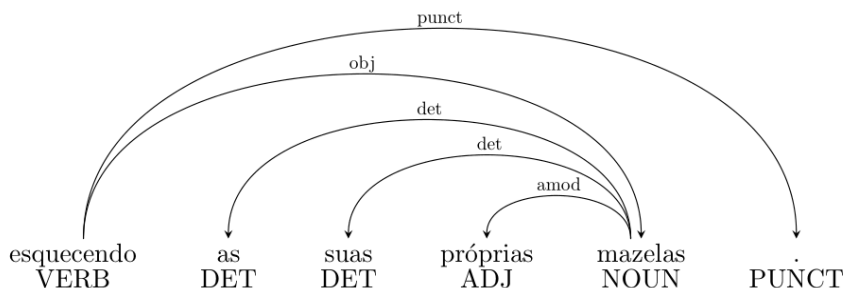
Apparently, classifying a deviation from annotation guidelines as undesirable rather than an error stems from the lack of sufficient grounds to consider a given phenomenon as universal, even though it is hypothesized to be so. The warning against configurations like the one in Figure 10, which is currently ignored by the maintainers of UD_Portuguese-CINTIL, seems to point to an analysis that is indeed unlikely—so much so that it has not been adopted by UD_Portuguese-Porttinari, as evidenced by Figure 11.

FIGURE 10 – Example of a determiner governing an adjective in the UD_Portuguese-CINTIL treebank



Fonte: elaboração própria.

FIGURE 11 – Excerpt from sentence FOLHA_DOC000115_SENT004 of UD_Portuguese-Porttinari with an alternative analysis of the construction of Figure 10



Fonte: elaboração própria.

As a free and open-source software, the program `validate.py` is easy to run from the command line of operating systems like Linux. By detecting errors that often go unnoticed by annotators and human reviewers, it plays a key role in the quality control of annotations. Thus, it is integrated into the development pipeline of UD_Nheengatu-CompLin, being executed periodically after changes to the treebank. This has ensured that successive development versions of UD_Nheengatu-CompLin consistently rank among the treebanks with the highest validation status.

The validity status, as determined by `check_files.pl` and `validate.py`, is a necessary condition for the admission of a new treebank into a release of the UD collection. On the other hand, veteran treebanks that become invalid between one release and the next must address the issues detected within four years, after which they are removed from the collection.

The results of the `check_files.pl` and `validate.py` programs are consolidated into an online report that classifies the treebanks in the UD collection into various categories based on their validation status, depending on the types of problems detected. This report is always updated every time the development version of a treebank is uploaded to the UD repository. Table 1 shows the number of treebanks and languages in each of these categories based on data collected from the report on January 18, 2025³.

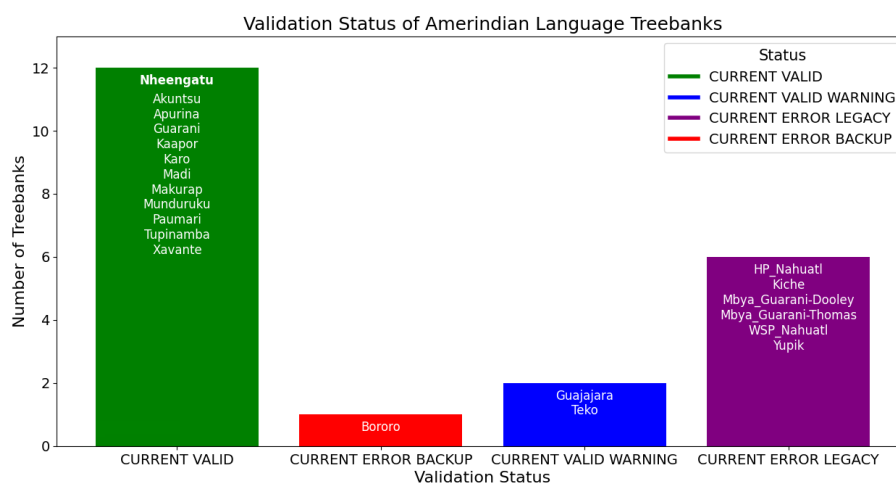
Table 1 – Frequencies of the different validation categories of the development versions of the UD treebanks as of January 18, 2025

Category	Treebanks	Languages
CURRENT VALID WARNING	48	40
CURRENT VALID	51	44
SAPLING EMPTY	71	61
CURRENT ERROR LEGACY	181	96
SAPLING ERROR	24	24
CURRENT ERROR BACKUP	11	9
RETIRED ERROR	6	5
CURRENT ERROR NEGLECTED	5	5
TOTAL VALID/BACKUP/LEGACY/NEGLECTED	296	168

Fonte: elaboração própria.

Version 2.15, the most recent, includes treebanks in the categories CURRENT VALID, CURRENT VALID WARNING, CURRENT ERROR BACKUP, CURRENT ERROR LEGACY, and CURRENT ERROR NEGLECTED, totaling 296 corpora of 168 languages. The UD_Nheengatu-CompLin is included in the first category, along with eleven other treebanks of Amerindian languages, representing the highest adherence to the principles of UD (Figure 12).

Figure 12 – Validation status of the treebanks for Indigenous languages of the Americas as of January 18, 2025. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

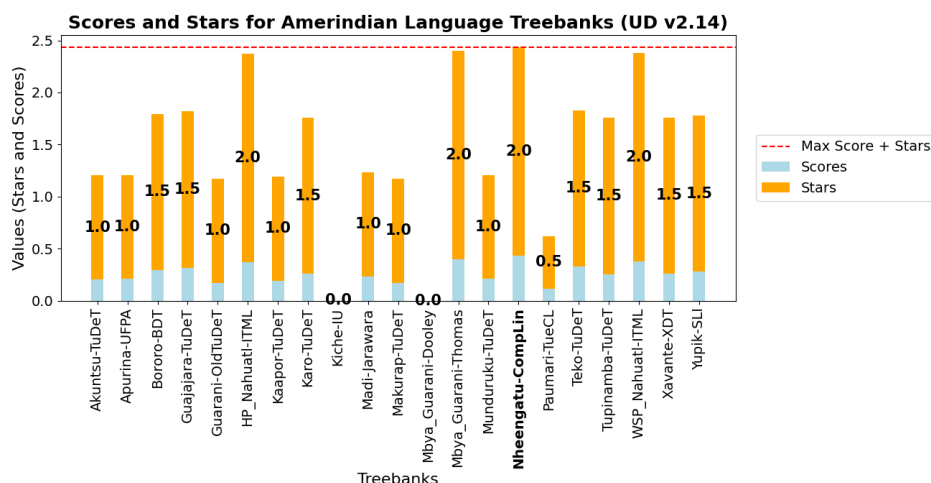
³ <https://quest.ms.mff.cuni.cz/udvalidator/cgi-bin/unidep/validation-report.pl>

The remaining categories reflect decreasing levels of validation. The categories with the keywords WARNING and ERROR characterize the treebanks whose validation generated warning and error messages, respectively. The keywords BACKUP, LEGACY, and NEGLECTED reflect an increasing period, counted from the last UD release, during which errors persist in the treebanks. BACKUP indicates errors produced after the most recent release, while LEGACY and NEGLECTED refer to errors persisting up to and after three years, respectively. RETIRED designates treebanks that were excluded from the latest release of the collection for not correcting the detected errors in the prescribed time frame. The categories starting with the term SAPLING identify treebanks that have never been included in any UD release, among which those marked additionally with VALID and VALID WARNING are eligible for admission in the next release.

3.2 The stars classification system

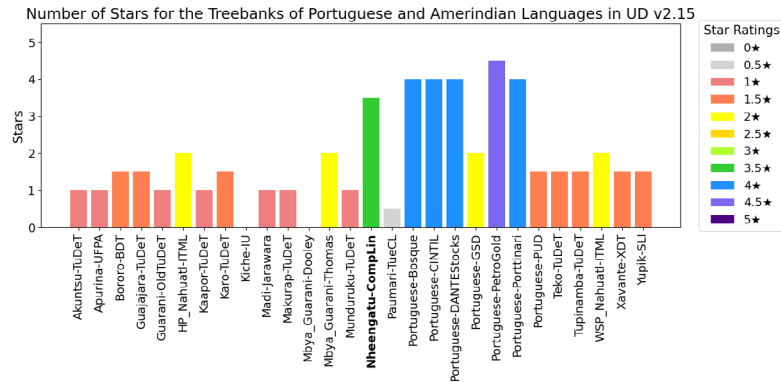
The second evaluation system uses the program `evaluate_treebank.pl` to compute a score between 0 and 1 for each treebank in the UD collection, considering a series of criteria, which will be explained later. This score serves as the basis for ranking the treebanks on a discrete scale of 0 to 5 stars, with increments of 0.5 (i.e., the possible values are 0, 0.5, 1.0, 1.5, ..., 5.0). Figure 13 shows the scores and star quantities of the treebanks of Amerindian languages in version 2.14 of the UD collection. Only four of these corpora achieve a two-star rating, among which is UD_Nheengatu-CompLin, which outperforms all others in the sum of score and star quantity. In UD v2.15, the number of stars for UD_Nheengatu-CompLin increased by 75% to 3.5 (Figure 14), making it the only Amerindian language treebank to enter the third-highest rating range (Figure 15), close to the major language treebanks such as Portuguese.

Figure 13 – Total scores and number of stars of the treebanks for Amerindian languages in UD v2.14. Treebank identifiers are abbreviated in the figure for readability.



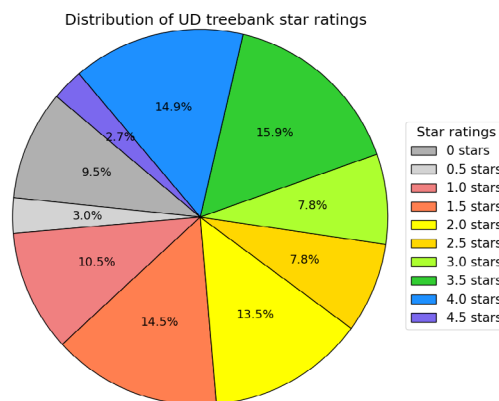
Fonte: elaboração própria.

Figure 14 – Number of stars of Portuguese and Amerindian language treebanks in UD v2.15. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

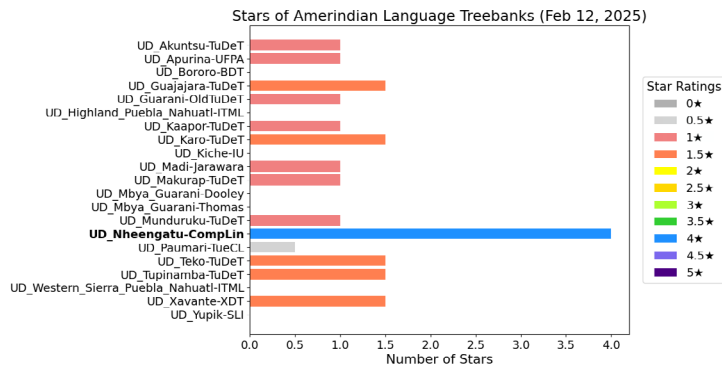
Figure 15 – Frequency distribution of star ratings in UD v2.15. Treebank identifiers are abbreviated in the figure for readability.



Fonte: elaboração própria.

In the development version of January, 20, 2025, UD_Nheengatu-Complin has reached the 4-star rating of a treebank like UD_Portuguese-Porttinari in UD 2.15, while the number of stars for four Indigenous language treebanks has dropped to zero, with the others maintaining the rating from UD v2.15 (Figure 16).

Figure 16 – Number of stars of the development versions of the treebanks for Amerindian languages as of February 12, 2025



Fonte: elaboração própria.

To assign a score from 0 to 1 for a treebank, eight dimensions are evaluated, namely: features, genres, lemmas, size, split, tags, udapi, udeprels. Each dimension has a weight and a score. The same weights, as shown in Table 2, apply to all treebanks, while the scores are calculated for each treebank.

Table 2 – Weights assigned to different dimensions in the scoring system

Dimension	Weight
features	3
genres	3
lemmas	3
size	10
split	2
tags	3
udapi	12
udeprels	3

Fonte: elaboração própria.

The scores for genres, size, split, and udapi are calculated according to Equation (2), while the scores for features, lemmas, tags, and udeprels are calculated according to Equation (3), where FS_{dim} represents the final score of the dimension, w_{dim} is the weight assigned to the dimension, f_{dim} is the additional multiplicative factor (when applicable), and s_{dim} is the raw score of the dimension.

$$FS_{dim} = w_{dim} \cdot s_{dim} \quad (2)$$

$$FS_{dim} = w_{dim} \cdot f_{dim} \cdot s_{dim} \quad (3)$$

The multiplicative factor depends on information contained in the <pre> element, which constitutes the last part of the README.md file of the treebank repositories on GitHub⁴. Table 3 reproduces the specifications for these four dimensions in the README.md file of four treebanks in the UD collection. For each dimension **considered independently**, the multiplicative factor reaches the maximum value of 1.0 when its annotation is natively manual, whereas the automatic conversion of a manual annotation is assigned a lower factor of 0.8. In these four dimensions, UD_Nheengatu-CompLin achieves the maximum multiplicative factor.

Table 3 – Comparison of Four UD Treebanks on Key Dimensions

Treebank	Relations	Lemmas	UPOS	Features
Nheengatu	manual native	manual native	manual native	manual native
Paumari	manual native	not available	manual native	not available
CINTIL	converted from manual	converted from manual	converted from manual	converted from manual
Porttinari	manual native	manual native	manual native	manual native

Fonte: elaboração própria.

The features score reflects the richness and reliability of the morphological information in the FEATS column, taking into account the number of features and the nature of the annotation, as shown in Table 3. Based on the percentage of words with at least one feature, a value between 0.01 and 1.0 is computed. The higher the percentage, the higher the value, which is then multiplied by the corresponding factor for the nature of the annotation. The UD_Paumari-TueCL, for example, has 504 words, none of which have morphological features, resulting in the minimum score value of 0.01. Just under half of the words in UD_Apurina-UFPA have some feature, resulting in the value 0.5, which is multiplied by the factor 0.8, giving a score of 0.4. UD_Nheengatu-CompLin has more than two-thirds of its words annotated morphologically, all of which are done manually, achieving the maximum score of 1.0.

The genres score takes into account the diversity of textual genres in the treebank. Currently, the program `evaluate_treebank.pl` lists 17 different genres, although the UD documentation specifies 18, including an additional government genre, used, for example, by UD_English-GUM (Müller-Eberstein; Goot; Plank, 2021). To calculate this score, one simply takes the number of textual genres in the treebank, which is provided in the aforementioned block delimited by the <pre> and </pre> tags, and divides it by the maximum number accepted by UD. With five different genres, UD_Nheengatu-CompLin surpasses the seven Portuguese treebanks and only loses in this regard, among the Amerindian language treebanks, to UD_Kiche-IU, which identifies 6 genres.

The lemmas score, like the previous ones, ranges from 0.01 to 1. At the lower end are treebanks that are predominantly or entirely without lemmatization, meaning that the LEMMAS column is mostly or entirely filled with underscores (“_”). At the other end, we have treebanks like UD_Nheengatu-CompLin, with natively manual lemmatization. With a score of 0.8, UD_Portuguese-CINTIL occupies an intermediate position due to the automatic conversion of manual lemmatization, as shown in Table 3.

⁴ <https://github.com/UniversalDependencies>

The size score takes into account the number of words in the treebank. To compute it, minimum and maximum reference values were defined: 1,000 and 1,000,000 words, respectively. Any treebank with fewer than 1,000 words receives a score of 0. This applies to treebanks of the following Indigenous languages: Apurinã, Old Guarani, Kaapor, Jarawara, Makurap, and Paumari. Conversely, any treebank with 1,000,000 words or more is assigned the maximum score value.

Since treebank sizes vary across several orders of magnitude, normalization on a logarithmic scale was necessary. To this end, a size coefficient (SC) was defined as the natural logarithm of the square of one-thousandth of the number of words N in the treebank, as shown in Equation 4:

$$SC = \ln\left[\left(\frac{N}{1,000}\right)^2\right] \quad (4)$$

Substituting the maximum word count into the equation yields:

$$SC = \ln\left[\left(\frac{1,000,000}{1,000}\right)^2\right] = \ln[(1,000)^2] = \ln[1,000,000] = 13.815511 \quad (5)$$

These maximum and minimum logarithmic limits allow for a linear comparison of treebank sizes. The final size score is calculated by dividing the result of Equation 4 by the maximum value from Equation 5. For example, UD_Portuguese-Porttinarí contains $N = 168,080$ words. Substituting this value into Equation 4, we obtain $\ln[(168,080/1,000)^2] = 10.248880$. The size score is then $10.248880/13.815511 = 0.741838$. Similarly, version 2.14 of UD_Nheengatu-CompLin, with 15,036 words, received a size score of 0.392377. In version 2.15, with 19,278 words, the score increased to 0.428353.

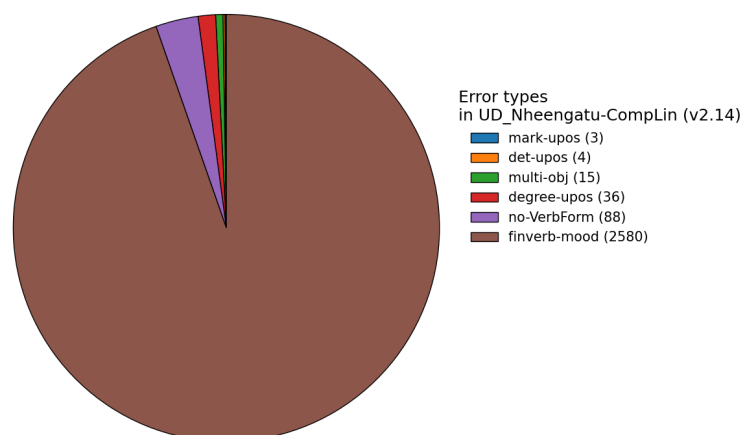
The split score takes into account the suitability of the treebank for training a parser using machine learning algorithms, which typically require a division into three sets: training, development, and testing (Jurafsky; Martin, 2009). For a treebank to obtain the maximum score of 1, the first set must contain more than 10,000 words, while the others must consist of at least 10,000 words. Each of these conditions is worth 0.33. To the sum of these values, 0.01 is added, which is the value assigned to treebanks that do not meet any of the conditions. This situation applies to most of the 21 Indigenous language treebanks in UD v2.15, with only five achieving a split score of 0.34, including UD_Nheengatu-CompLin in versions 2.14 and 2.15. In the development version of January 20, 2025, UD_Nheengatu-CompLin, as the only one among the 21 Amerindian language treebanks, achieved a split score of 0.64, while six of the seven Portuguese treebanks reached the maximum value.

The tags score takes into account the use of the part-of-speech tags provided by UD in the treebank. The approach is similar to that of textual genres. Currently, UD accepts 17 different tags. To calculate this score, simply take the number of tags that occur in the treebank and divide by the 17 provided by UD. UD_Nheengatu-CompLin, for example, uses 15 tags in version 2.14, resulting in a score of $15/17 = 0.882352$. With 16 tags, version 2.15 attains a score of $16/17 = 0.941176$. UD_Portuguese-Porttinarí, on the other hand, uses 16 labels, reaching a score of $16/17 = 0.941176$.

The udapi score takes into account the results of the `ud.MarkBugs` block of the Udapi framework (Popel; Žabokrtský; Vojtek, 2017). This program checks whether the annotations violate a set of restrictions, generating a report with the types and frequencies of detected

errors (referred to as “bugs”). Figure 17 displays the number of errors for UD_Nheengatu-CompLin in version 2.14 of the UD collection, which we will address in detail in Section 4.

Figure 17 – Frequencies of types of the 2726 errors detected by Udati



Fonte: elaboração própria.

The udapi score is calculated similarly to the lemma score, comparing the number of errors B with the number N of words in the treebank. The worst case is one error for every 10 words. If the ratio of errors to words, i.e., N/B , for $B > 0$, is equal to or smaller than this threshold, the score will be the minimum value of 0.01. Otherwise, Equation 6 is applied, by which a treebank with no errors achieves the maximum score of 1.

$$US = 1 - \frac{B \times 10}{N} \quad (6)$$

Version 2.14 of UD_Nheengatu-CompLin has 2,726 errors for 15,036 words, which equates to one error every 5.5 words, resulting in the minimum score of 0.01. In version 2.15 of this treebank, we managed to reduce the number of errors to 58, while the total number of words increased to 19,278, catapulting the score to 0.969913. The corresponding version of UD_Portuguese-Porttinari, in turn, has 67 errors for 168,080 words, which equates to one error every 2,626 words, yielding a score of 0.996013. Among the 21 Amerindian language treebanks, only the two Mbyá Guaraní treebanks, being error-free, achieve the maximum score of 1. On the other hand, in version 2.15 of the UD collection, none of the seven Portuguese treebanks reach the score of 1. Two of these treebanks achieve the minimum value of 0.01, with the scores of the others varying between 0.996013 and 0.988368, values obtained by UD_Portuguese-Porttinari and UD_Portuguese-Bosque, respectively.

Finally, the udeprels score reflects the use of the inventory of universal dependency relations by the treebank. This metric only considers the set of 37 main relations, such as subject (nsubj), direct object (obj), and nominal modifier (nmod), excluding subtypes like passive subject (nsubj:pass) and possessive modifier (nmod:poss). The calculation is analogous to the textual genre and part-of-speech tag scores, dividing the number of universal relations in the treebank by the maximum number of 37 universal relations. Version 2.15 of UD_Nheengatu-CompLin uses 33 of these 37 relations, as does UD_Portuguese-Porttinari, resulting in a score of $33 / 37 = 0.891891$. Among the 21 Amerindian language treebanks in version 2.15 of the UD collection, only the two Nahuatl treebanks surpass Nheengatu in this

respect. With 35 items each, only UD_Portuguese-Bosque and UD_Portuguese-DANTEStocks, among the Portuguese treebanks, have a larger set of dependency relations.

To calculate the total score of a treebank, the sum of the partial scores in the eight dimensions is multiplied by the factors availability $\in \{0.01, 1\}$ and validity $\in \{0.1, 1\}$. The first factor is applied at its minimum value to a treebank lacking sentence text; otherwise, the value is the maximum. In UD v2.15, among the 21 Amerindian language treebanks, only UD_Mbya_Guarani-Dooley falls into the first condition. The second factor takes the maximum value if the treebank is valid according to the validate.py program; otherwise, it takes the minimum value, a condition that applies to both UD_Mbya_Guarani-Dooley and UD_Kiche-IU. The number of stars of a treebank is computed using Equation 7.

$$stars = \frac{[score \times 10 + 0.5]}{2} \quad (7)$$

Summing up, version 2.15 of UD_Nheengatu-CompLin attains the individual dimension scores reported in Table 4. Each dimension score is multiplied by its corresponding weight, yielding the weighted scores shown in the last column. The weights shown in Table 4 are the normalized versions of the raw dimension weights listed in Table 2, obtained by dividing each raw weight by their total sum 39, so that they sum to 1 and can be used to compute a weighted average score between 0 and 1. The resulting total score is 0.743181144774127, which—since both availability and validity factors equal 1—is converted into stars using Equation 7, yielding an overall rating of 3.5 stars.

Table 4 – Scores computed by evaluate-treebank.pl for version 2.15 of UD_Nheengatu-CompLin the eight evaluation dimensions

Dimension	Weight	Score	Weighted score
features	0.0769230769230769	1	0.0769230769230769
genres	0.0769230769230769	0.294117647058824	0.0226244343891403
lemmas	0.0769230769230769	1	0.0769230769230769
size	0.256410256410256	0.428353991978387	0.109834356917535
split	0.0512820512820513	0.34	0.0174358974358974
tags	0.0769230769230769	0.941176470588235	0.0723981900452489
udapi	0.307692307692308	0.969913891482519	0.298435043533083
udeprels	0.0769230769230769	0.891891891891892	0.0686070686070686
TOTAL			0.743181144774127

Fonte: elaboração própria.

4 Analysis and correction of the Udapi errors

In this section, we detail each error type shown in Figure 17 and explain how we resolved the majority of them. Subsections 4.1, 4.2, and 4.3 address the corrected errors, which account for 98.4% of the total, while subsection 4.4 discusses the small remaining percentage that has yet to be corrected.

4.1 Underspecified verbal mood

The finverb-mood error occurs when verb forms annotated as finite lack the Mood feature, as illustrated in Figure 18. This type makes up 94.65% of the 2,726 errors flagged by Udaapi in version 2.14 of UD_Nheengatu-CompLin. In the following, we summarize different accounts of Nheengatu's verb inflection. These generally converge on the observation that Nheengatu verb forms, abstracting away from historical or irregular imperative forms, only encode person and number. Tense and mood are conveyed by other means, typically particles and auxiliaries. This motivated our previous decision to specify mood as a morphological feature in the treebank only if the verb form is, regardless of context, unambiguously imperative. To address these errors, we have revised our annotation policy to always specify the Mood feature for finite verbs. In version 2.15 of the treebank, the number of such errors was reduced to zero.

- 2) Re-rikú será mukawa?
 2SG.ACT-have PQ shotgun
 'Do you have the shotgun?' (Magalhães, 1876, p. 14)

FIGURE 18 – Analysis of example (2) with a finverb-mood error detected by Udaapi

```
# sent_id = Magalhaes1876:1-1-2:1:103
# text = Rerikú será mukawa?
└─ Rerikú rikú VERB V Number=Sing|Person=2|VerbForm=Fin root Bug=finverb-mood|TokenRange=0:6
   └─ será será PART PQ PartType=Int advmod TokenRange=7:11
      └─ mukawa mukawa NOUN N Number=Sing obj SpaceAfter=No|TokenRange=12:18
         └─ ? ? PUNCT PUNCT _ punct SpaceAfter=No|TokenRange=18:19
```

Fonte: elaboração própria.

Following Hartt (1872), Magalhães (1876, p. 10) observes that, in Nheengatu, “tense roots are not yet incorporated into the verb or the attributive root, as occurs in inflectional languages”⁵. For both authors, the “undefined present”, exemplified in (2), is decomposed into the personal prefix *re-* and the root *rikú* (“have”), forming other verb tenses through the auxiliary *ikú* (“be”) and certain particles, such as *kurí* and *ana*⁶. Similarly, Sympson (1877, p. 11) states that “[t]he verbs of the *Brasilica* language never change their endings.”⁷ Except for the present indicative and the imperative, other moods, tenses, and verb forms are expressed through auxiliaries and “signs”, which roughly correspond to the particles described by Magalhães (1876). Compare (2) in the “undefined present” with sentences (3)–(5) in the “defined present”, future, and past tenses, respectively, according to Magalhães (1876).

- 3) Maã taá re-munhã re-ikú?
 what CQ 2SG.ACT-do 2SG.ACT-be
 ‘What are you doing?’ (Magalhães, 1876, p. 28)

⁵ Our translation from Portuguese: “as raízes de tempo ainda não estão incorporadas ao verbo, ou à raiz atributiva, como sucede nas línguas de flexão.”

⁶ Magalhães (1876, p. 10) includes the subordinator *ramé* (“when”) among the particles forming verb tenses, whereas Hartt (1872) classifies this word as an adverb.

⁷ Our translation from Portuguese: “[o]s verbos da língua brasílica nunca mudam de terminação.”

- 4) Re-maã kurí ixé wirandé.
2SG.ACT-see FUT me tomorrow
'You will see me tomorrow.' (Magalhães, 1876, p. 73)

- 5) Mairamé taá re-maã ana ixé?
when PQ 2SG.ACT-see PFV me
'When did you see me?' (Magalhães, 1876, p. 73)

From the perspective of the grammatical exposition by Magalhães (1876), the notion of verbal mood, characteristic of languages like Portuguese, does not participate in the inflectional mechanism of Nheengatu. The same second-person singular personal prefix constitutes the only inflection of the main verb, regardless of whether the sentence is a question, statement, or command, as demonstrated in examples (2), (6), and (7). The illocutionary force of these sentences is determined by the syntactic structure, prosody, or discourse context rather than by verb inflection.

- 6) re-puká sé nhaã kunhã-etá r-enundé
2SG.ACT-laugh happy those woman-PL CONT-in_front_of
'you laughed happily in front of those women'
- 7) Re-tirika, yautí, kurumú xa-pirú indé.
2SG.ACT-get_out, tortoise, or_else 1SG.ACT-step_on you
'Get out, tortoise, or else I'll step on you.'

The term imperative occurs only once in Magalhães (1876, p. 94): "The verb *ço*, to go, forms the imperative *cóĩ*, which is read as: *cóin*."⁸ Sentences with this form, such as (8), are unequivocally imperative.

- 8) Re-kūi apigawa úkupi.
2SG.ACT-go[IMP] man house:LOC
'Go to the man's house.'

Unlike Magalhães (1876), imperative forms are part of the conjugation paradigms of both Hartt (1872) and Simpson (1877), as in examples (9)–(11).

- 9) saisú iné
[2SG.IMP]love 2SG
'You love!' (Simpson, 1877, p. 43)
- 10) pe-saisú penhẽ
2SG.ACT-love 2PL
'You guys love!' (Simpson, 1877, p. 43)
- 11) E-kūi e-sikináu nhaã kawasúyurú.
2SG.IMP-go 2SG.IMP-cover that bucket mouth
'Go cover the mouth of the bucket.' (Hartt, 1938, p. 320)

⁸ Our translation from Portuguese: "O verbo *ço*, *ir*, *faz no imperativo cóĩ*, *que se lê: cóin*."

Hartt (1872) contrasts the prefix *e-* of the second-person singular imperative, represented according to Simpson (1877) by the absence of a personal prefix, with the corresponding prefix *pe-* of the second-person plural. This prefix, according to Simpson (1877, p. 43), is distinguished from the homonymous form of the present indicative “by usually adding the particle *penhê*”⁹. Hartt (1872) and Simpson (1877) discuss the irregular formation of the imperative of *sú* (“go”), exemplified in the second-person singular by *ekóin* and *icúen*, respectively. Hartt (1872) also highlights the prohibitive particle *teñé*, which Simpson (1877, p. 11) spells as *ten* and calls an imperative “sign”, used “when the verb is conjugated negatively”¹⁰.

Following Magalhães (1876), Stradelli (1929, p. 40) states that “the conjugation paradigm of Nheengatu is reduced to the paradigm of the present tense”¹¹. The same personal prefixes are used in all tenses and moods, rarely being omitted. The different tenses are formed either with the suffix *ana* or with adverbs such as *kwerá*, *kurí*, etc., which are not integral parts of the verb. In his grammatical synopsis, Stradelli (1929, p. 41), differing from Magalhães (1876) but aligning with Simpson (1877), specifically addresses the formation of the imperative, which “is obtained with the present indicative used without the personal pronoun and lacking the first person”¹². He exemplifies this with *rikú* (*recô*, “you have”) and *urikú* (*orecô*, “he has”) ¹³. He highlights both the regional variation between the irregular imperative forms *rekūi* (*recoin*) and *kūi* (*coin*), on one hand, and the “imperfect” form *resú ana* (*resoana*), on the other, as well as the *i-* allomorph of the second-person singular prefix, which he observed in localities along the Solimões River.

Corroborating the observations of Simpson (1877) and Stradelli (1929), Avila (2021) includes *i-*, *u-*, and *ø-* as variants of *e-*, the historical lemma corresponding to *re-*, used in contemporary Rio Negro Nheengatu in both the indicative and imperative moods. Avila (2021) lists a large number of historical records of these variants, attested, for example, in the forms *eruri*, *ururi*, and *munhã* from examples (12)–(15). The second form coincides with the third-person indicative form, while the third corresponds to the basic verb form, devoid of inflection. Avila (2021) describes in specific entries different irregular imperative forms, such as *rekūi* or *ekūi*, covered by Hartt (1872), Magalhães (1876), Simpson (1877), and Stradelli (1929), also dedicating a separate entry to *yuri*, the irregular second-person imperative form of the verb *yuri* (“come”). These entries list diverse citations and historical spelling variants, most of which we incorporated into UD_Nheengatu-CompLin.

⁹ The original Portuguese passage reads as follows: “A segunda pessoa do plural do imperativo diferença-se da segunda pessoa do presente do indicativo, em todos os verbos, por se lhe acrescentar usualmente a partícula — *penhê*.”

¹⁰ Our translation from Portuguese: “quando o verbo é conjugado negativamente.”

¹¹ Our translation from Portuguese: “o paradigma da conjugação nheengatu se resume no paradigma do tempo presente.”

¹² The original passage reads as follows: “O imperativo se obtém com o presente indicativo usado sem o pronome pessoal e com a falta da primeira pessoa.”

¹³ The original examples read as follows: “*Recô* — tenha; *Orecô* — tenha.” There is probably a typo in the first example. The second-person singular prefix *re-* is missing. The form should be *rerikú* (*rerecô*), which appears on the same page in a past tense example. An Italian naturalized as a Brazilian, Stradelli (1929) translates the imperative second-person singular form of *rikú* (“have”) into Portuguese *tenhas* instead of *tem*. The former is restricted to the negative imperative (Cunha; Cintra, 2017).

- 12) Mitú e-ruri se igara a-sú arama
 curassow 2SG.IMP-bring my canoe 1SG.ACT-go for
 a-ú tuyuka.
 1SG.ACT-eat clay
 ‘Curassow, bring my canoe so I can go eat clay.’ (Rodrigues, 1890, p. 159)
- 13) I-ruri xa-ú ne piá.
 2SG.IMP-bring 1SG.ACT-eat your heart
 ‘Give me your heart to eat.’ (Rodrigues, 1890, p. 36)
- 14) U-ruri Karã se muirakitã!
 2SG.IMP-bring limpkin my talisman
 ‘Limpkin, bring my talisman!’ (Rodrigues, 1890, p. 132)
- 15) Munhã tatá, siusí-etá u-sema ikú.
 [2SG.IMP]make fire, Pleiades 3SG.ACT-go.out[INF]be
 ‘Make the fire, the Pleiades are rising.’ (Rodrigues, 1890, p. 174)

The specific morphological markings of the affirmative imperative have not survived in contemporary Rio Negro Nheengatu, as described by Casasnovas (2006), Cruz (2011), Navarro (2016), and Avila (2021). Only the marking of the negative imperative has been preserved, through *tenhẽ* or *te* [te], the latter form classified by Cruz (2011) as a clitic and by Avila (2021) as a particle, as in example (16).

- 16) Te re-kiri, Aline!
 PROH 2SG.ACT-sleep, Aline
 ‘Don’t sleep, Aline!’ (Cruz, 2011, p. 407)

Since the Nheengatu verb, as a rule, does not express verbal mood by itself, the morphological analysis of examples such as (2)–(7) in UD_Nheengatu-CompLin as of UD v2.14 does not include this feature. It limits itself to the Number, Person, and VerbForm features. The latter allows distinguishing between finite and non-finite forms, such as infinitive, participle, and gerund (see Section 4.2). Only specific imperative forms, such as in (8) or (11), are annotated with the Mood feature. In the case of the negative imperative with the particle *te* or the historical variant *tenhẽ*, the Mood feature applies to the particle, while the verb remains underspecified regarding mood. Table 5 displays the frequencies of mood specifications by word classes. Only a tiny fraction of auxiliaries and full verbs, totaling 14 occurrences, have the Mood feature, in this case with the imperative value, while 18 particles are specified for the imperative mood and 14 for the conditional mood, expressed in Nheengatu by the *amú* and *maã* particles. The latter is exemplified in (17).

Table 5 – Parts of speech and mood frequencies in UD_Nheengatu-CompLin v2.14. The dash indicates that the feature is absent.

Part of Speech	Mood	Frequency
AUX	Imp	2
PART	Cnd	14
PART	Imp	18
VERB	Imp	12
AUX	—	356
VERB	—	2403

Fonte: elaboração própria.

In the analysis of (17), we follow Navarro (2016) and Avila (2021). The former distinguishes, in Nheengatu, three types of conditional mood, namely, (i) real hypothesis, (ii) possible hypothesis, as in the example at hand, and (iii) unreal hypothesis. According to Avila (2021), *maã* in this example functions as a conditional particle, applying to both the protasis and the apodosis. Cruz (2011, p. 502, 385), on the other hand, analyzes this word as a subordinator, assigning it the function of a “marker of hypothetical modality”¹⁴, a classification that does not seem to align with her definition of subordinators as “specialized marks for indicating the syntactic dependency between two clauses.”¹⁵

- 17) mira ramé maã indé, indé intí maã re-xari ixé xa-manú
human if COND you, you NEG COND 2SG.ACT-let me 1SG.ACT-die
‘if you were human, you wouldn’t let me starve.’ (Amorim, 1928, p. 28)

To fix the finverb-mood errors flagged by Udapi, we automatically included the Mood feature in all verb forms. In structurally unambiguous cases, such as in interrogative sentences, adverbial subordinate clauses, or with the negative imperative particle *te* (or *tenhê*), we used the values Ind and Imp for indicative and imperative, respectively. Otherwise, we used the value Imp,Ind to indicate that the verbal form may be either imperative or indicative, requiring manual disambiguation, something to be done in a next version of the treebank.

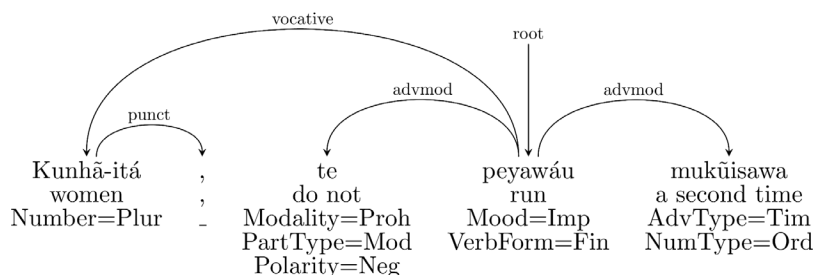
On the other hand, we removed the Mood feature in the FEATS field of the negative imperative and conditional particles, while maintaining the other features. The former is additionally annotated with features specifying it as a negative polarity negation particle, while the latter is characterized as a modal particle. Instead of Mood=Imp and Mood=Cnd, we now annotate these two particles with Modality=Proh, as exemplified in Figure 19, and Modality=Cond, respectively. The abbreviation *proh* is widely used in linguistic typology to designate the prohibitive modality (Bybee; Perkins; Pagliuca, 1994; Palmer, 2001), a practice that seems to underlie Hartt (1872)’s classification of *teñé* as a prohibitive particle. The substitution of Mood by Modality aligns with the UD documentation, which restricts the application of the Mood feature to verbs, defining Mood as “a feature that expresses modality and

¹⁴ Our translation from Portuguese: “marcador de modalidade ‘hipotética’ [sic].”

¹⁵ Our translation from Portuguese: “marcas especializadas na função de marcar a dependência sintática entre duas orações.”

subclassifies finite verb forms” (Marneffe *et al.*, 2024a), although the validator does not prohibit its application to particles.

Figure 19 – Annotation of the prohibitive particle *te* and an imperative verb in version 2.15 of the UD_Nheengatu-ComPLin treebank. The example is a fragment from a sentence of Amorim (1928, p. 23–24).



Fonte: elaboração própria.

The Modality feature is an innovation of UD_Nheengatu-ComPLin version 2.15. The treebanks of Egyptian and Polish limit themselves to annotating modal particles with PartType=Mod, while the Bulgarian and Czech treebanks leave the FEATS field empty for the epistemic modality particle and the deontic modality particle, respectively, that, according to UD documentation, express sentential modality.

4.2 Missing verb form specification

The no-VerbForm error consists of a verb without the VerbForm feature, which characterizes finite forms, infinitives, participles, etc. This type makes up 3.23% of the bugs flagged by Udaapi in UD_Nheengatu-ComPLin version 2.14. The rationale underpinning our annotation practice in this regard was similar to that behind the finverb-mood errors. Analogously, we only annotated a verb with the VerbForm feature when it formally encodes this property, as is the case in (18) with *yamburi*, but not *membeka*.

- 18) Ya-mburi maniáka paranã upé i
1PL.ACT-put maniocriver in 3SG.INACT
membeka arama.
become_soft to
'We put the manioc in the river to soften it.' (Moore; Facundes; Pires, 1994, p. 105)

Hartt (1872), Magalhães (1876), and Stradelli (1929) highlight a restricted group of auxiliary verbs that do not take personal prefixes when following another verb in the same sentence. An example of this particularity is *putari* ("want") in (19). Magalhães (1876, p. 11) explains: "The verb *putári* 'want' has a very peculiar way of appearing in the sentence; whenever it comes together with another verb, it is the other verb that receives the pronominal prefix, while *putári* remains unchanged [...]".¹⁶

¹⁶ Our translation from Portuguese: "O verbo *putári* querer, [sic] tem um mui singular modo de figurar na oração; sempre que ele vem junto com outro verbo, é esse outro verbo que recebe o prefixo pronominal, ao passo que

- 19) Re-mixiri-putari será pirá?
 2SG.ACT-grill-want[NFIN] PQ fish
 ‘Do you want to grill fish?’ (Magalhães, 1876, p. 96)

The verb *putari* (“want”) is conjugated when functioning as the main verb, as in (20). Based on the contrast in pairs like *putari* and *xaputari* in (19) and (20), the morphological analysis of active verbs in UD_Nheengatu-CompLin version 2.14 included the VerbForm feature, with the values Inf and Fin for non-finite and finite forms, respectively.

- 20) Intí mã; xa-putari yepé panera xa-memúi arama.
 no thing 1SG.ACT-want one pot 1SG.ACT-cook to
 ‘No; I want a pot to cook in.’ (Magalhães, 1876, p. 96–97)

The spelling of inactive prefixes is one of the many aspects in which Nheengatu orthographies differ. In this regard, we follow the approach of Navarro (2016) and Avila (2021), adopted by Yamã *et al.* (2021), treating these elements as second-class pronouns, spelled separately from the verbal root, as in (18)¹⁷. As a consequence of this decision, the FEATS field of inactive verbs was left without any morphological features since these verbs are invariable. This led to 88 instances of the no-VerbForm error detected by Udapi (Figure 20).

Figure 20 – Analysis of example (20) with no-VerbForm errors identified and highlighted by the Udapi tool

```
# sent_id = MooreFP1994:0:0:2
# text = Yamburi maniáka paranã upé i membeka arama.
Yamburi m-buri VERB V Number=Plur|Person=1|VerbForm=Fin root Bug=finverb-mood|TokenRange=0:7
maniáka maniáka NOUN N Number=Sing obj TokenRange=8:15
paranã paranã NOUN N Number=Sing obl TokenRange=16:22
upé upé ADP ADP AdpType=Post case TokenRange=23:26
i i PRON PRON2 Case=Gen|Number=Sing|Person=3|PronType=Prs nsubj TokenRange=27:28
membeka membeka VERB V2 advcl Bug=no-VerbForm|TokenRange=29:36
arama arama SCONJ SCONJ _ mark SpaceAfter=No|TokenRange=37:42
. . PUNCT PUNCT _ punct SpaceAfter=No|TokenRange=42:43
```

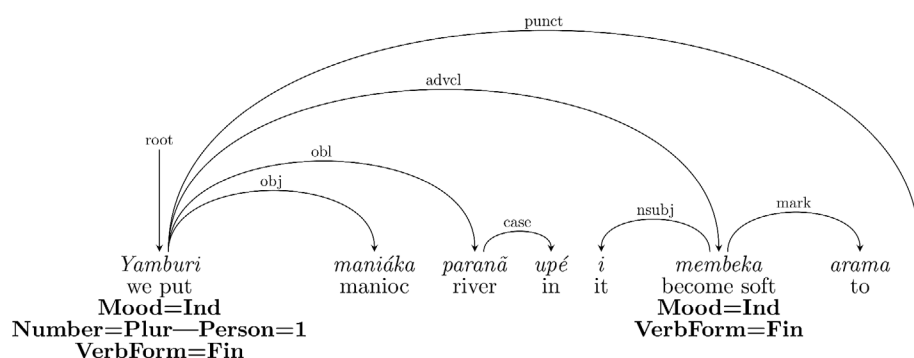
Fonte: elaboração própria.

In the UD_Nheengatu-CompLin version 2.15, all these errors were corrected, providing the missing Mood and VerbForm features, as exemplified in Figure 21. We revised our annotation policy to ensure that the VerbForm feature is consistently included in the FEATS field of all verbs.

ele fica invariável [...].”

¹⁷ Cruz (2011), on the other hand, treats these elements as agreement morphemes prefixed directly to the verbal roots. This spelling appears to be more widespread in the Rio Negro region, e.g., Casasnovas (2006).

Figure 21 – Corrected analysis of the example in Figure 20



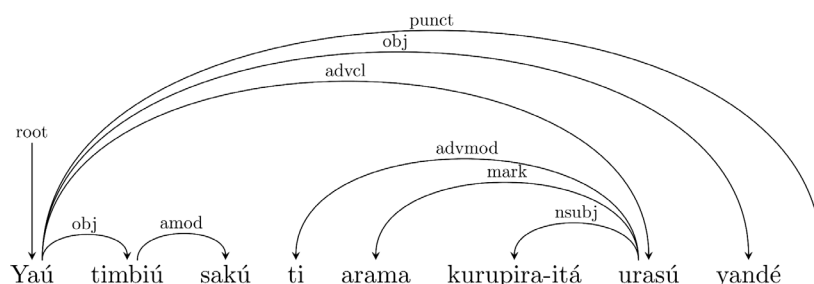
Fonte: elaboração própria.

4.3 Multiple objects

Udapi detected 15 multi-obj errors in version 2.14 of UD_Nheengatu-CompLin. These errors fall into two subtypes. In the first subtype, a verb governs two direct objects within the same sentence, as shown in Figure 22. The error in this tree resulted from the oversight of the human annotator, who incorrectly identified the head verb of the sentence as the governor of the embedded verb's object. The remaining errors of this subtype arose from analogous mistakes in identifying the governor or the syntactic relation of a nominal, which, for example, was analyzed as a direct object instead of a nominal complement when preceding another in a possessive genitive relation. We eliminated all errors involving double direct objects by modifying the governor or the syntactic relation of the spurious direct object.

- 21) Ya-ú t-imbiúsakú ti arama kurupira-itá
 1PL.ACT-eat ABS-food hot not to cupurira-PL
 u-rasú yandé.
 3.ACT-take_away us
 'We would eat hot food for the curupiras not to take us away.' (Moore; Facundes; Pires, 1994, p. 108)

Figure 22 – Dependency tree of (21) with a verb incorrectly governing two direct objects

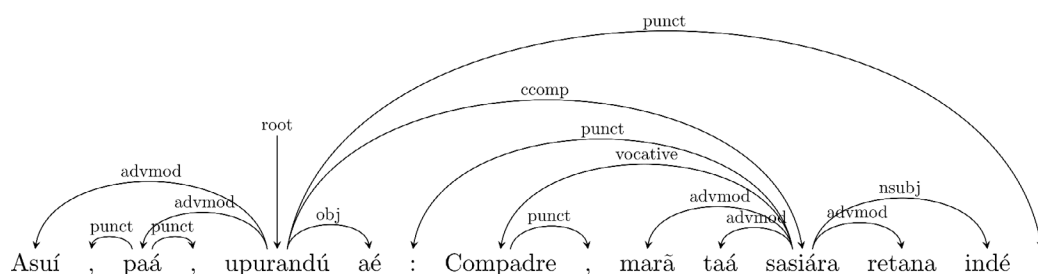


Fonte: elaboração própria.

In the second subtype, illustrated in Figure 23, the same verb governs both a direct object and a ccomp clausal complement. This subtype accounts for about two-thirds of the multi-obj instances in the Udapi report. It involves sentences with communication verbs such as *purandú* (“ask”) and *suaxara* (“answer”), where the recipient semantic role is realized by a bare nominal, that is, without an adposition, and the patient by the ccomp clausal complement.

- 22) Asuí, paá, u-purandú aé: Compadre, marã taá
 and RPR 3SG.ACT-asked him: Compadre, why CQ
 sasiára retana indé.
 sad very you
 ‘And he asked him: Compadre, why are you so sad?’ (Casasnovas, 2006, p. 66)

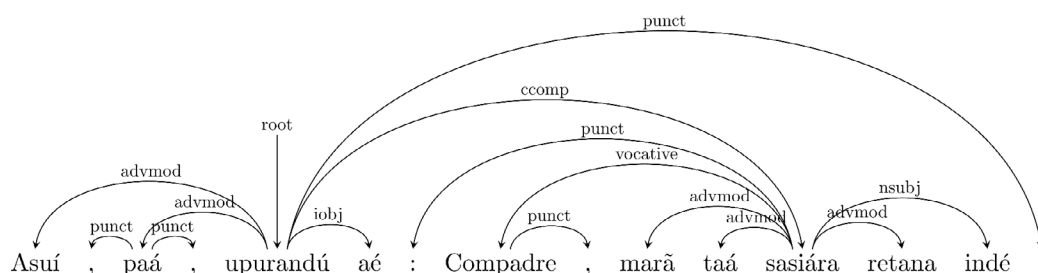
Figure 23 – Dependency tree of (22) with a verb incorrectly governing a direct object and a complement clause



Fonte: elaboração própria.

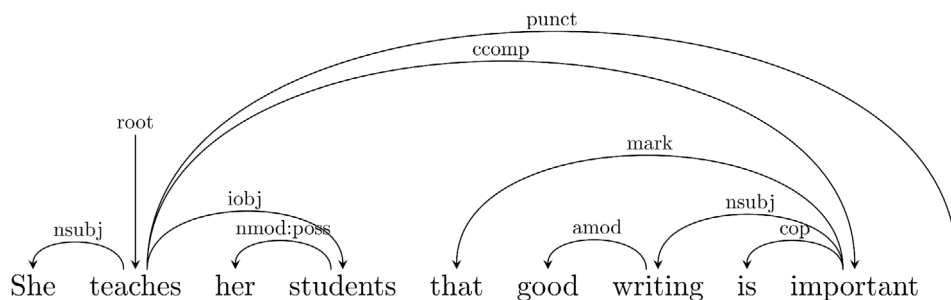
The solution in this case was to analyze the nominal that realizes the recipient not as a direct object, but as an indirect object, as shown in Figure 24. This solution coheres with the analysis in the UD documentation for an analogous example in English (Figure 25).

Figure 24 – Corrected dependency tree of (22) with the main verb governing an indirect object and a complement clause



Fonte: elaboração própria.

Figure 25 – Dependency tree for an example from the UD documentation with a verb governing an iobj and a ccomp generated by UDPipe 2 using the english-ewt-ud-2.15-241121 model



Fonte: elaboração própria.

The type of indirect object exemplified in Figure 24 was not part of the annotation scheme of UD_Nheengatu-Complin before the revision triggered by the errors discussed in this section. The assignment of the iobj syntactic relation was limited to dative case pronouns, as in (23), or nominals governing the postpositions *arama* and *supé*, which Cruz (2011) classifies as denoting the “prospective” (or “intralocutive”) dative and the “extralocutive”, respectively, as illustrated in (24) – (26).

- 23) E-purú ne kiwawaixé-u.
2SG.IMP-purú your comb 1SG-DAT
'Lend me your comb.' (Hartt, 1938, p. 319)
- 24) E-purú xinga ixé arama.
2SG.IMP-purú some me to
'Lend me some.' (Magalhães, 1876, p. 201)
- 25) “Mamãe”, u-nheẽ-wera ixé arã nhaã se kurumĩ.
mom 3SG.ACT-say-FREQ me to that my boy
“Mom,” that little boy of mine always used to say to me.' (Cruz, 2011, p. 217)
- 26) “Kuĩri-ta?” u-nheẽ i mú supé.
now=PQ 3SG.ACT-say his brother to
“‘What now?’ he said to his brother.’ (Cruz, 2011, p. 217)

The UD documentation on the obj and ccomp syntactic relations does not explicitly forbid a verb to govern an obj and a ccomp simultaneously. Similarly, the validate.py program does not flag an annotation error in this case, which only Udapi does¹⁸. The reason for excluding this valency frame lies in the parallelism between obj and ccomp, which is manifested in alternations such as in (27)–(29) (Levin, 1993, p. 205). From the perspective of Frame Semantics, in all of

¹⁸ Udapi's ud.MarkBugs block was based on constraints listed on a legacy page at <https://universaldependencies.org/svalidation.html> (Popel; Žabokrtský; Vojtek, 2017). We are not aware of any direct reference to this page in the main UD annotation guidelines. However, the main UD documentation includes Nivre *et al.* (2017) among the UD project publications, and this tutorial mentions the URL. According to the tutorial, a previous version of the validation system used these constraints. One of them states: “No predicate can have more than one direct object. Ccomp counts as [a] direct object.” It is unclear to us why subsequent versions of the validator discarded this constraint.

these sentences, the verb expresses the same elements of the Communication_manner frame: SPEAKER, ADDRESSEE, and MESSAGE (Ruppenhofer *et al.*, 2016), differing only in the syntactic expression of the latter. These elements correspond to the more general semantic roles of Agent, Recipient (or Goal), and Theme (Levin, 1993; Perini, 2015, 2019). In (27), the MESSAGE element is realized by an NP, whereas in the other two examples, it is expressed through a subordinate clause introduced by the complementizer *that* and through direct speech, respectively. Zwicky (1971) classifies all three of these syntactic realizations as direct objects of the verb, constituting mutually exclusive forms of expression of the same semantic element.

- 27) Susan whispered the news to Rachel.
- 28) Susan whispered to Rachel that the party would be tonight.
- 29) Susan whispered to Rachel, "Leave the room."

Cruz (2011, p. 424) denies the existence of ditransitive verbs in Nheengatu, arguing that the Beneficiary semantic role is compatible with any predicate. However, this claim lacks sufficient justification when examined through the lens of semantic role theory and valency theory. First, it is essential to distinguish between the semantic roles of Recipient or Goal, as in (27), and Beneficiary, as in the following example (Levin, 1993, p. 49):

- 30) Martha carved a toy for the baby.

Second, in communication and possession-transfer verbs, the syntactic relation that realizes the recipient of the message or the transferred entity functions as an argument of the verb's logical-semantic structure, alongside the arguments corresponding to the Agent and the Theme (Helbig, 1992). Similarly, in Frame Semantics, the three elements SPEAKER, ADDRESSEE, and MESSAGE are classified as core arguments.

4.4 Unsolved errors

Of the 2726 bugs flagged by Udapi in version 2.14 of UD_Nheengatu-CompLin, we failed to correct only 43. In 4.4.1, we address the 36 errors triggered by nouns with the diminutive or augmentative degree. In 4.4.2, we discuss the det-upos and mark-upos errors, identified in a total of seven sentences. These cases involve the incompatibility between the part-of-speech tag of the head of an exocentric multiword expression and the syntactic relation assigned to that expression.

4.4.1 Errors involving degree features

With 36 occurrences, the degree-upos error results from assigning the degree feature to nouns. The program only accepts this feature for adjectives and adverbs, as shown in Listing 1:

Listing 1 – Validation of the Degree feature against UPOS values

```
1 if feats['Degree'] and upos not in ('ADJ', 'ADV'):
2     self.log(node, 'degree-upos',
3             'Degree=%s, upos!=ADJ|ADV, (but %s)' % (feats['Degree'], upos))
```

Fonte: elaboração própria.

This rule appears to contradict the UD guidelines, which state:

Degree of comparison is typically an inflectional feature of some adjectives and adverbs. A different flavor of degree is diminutives and augmentatives, which often apply to nouns but are not restricted to them (Marneffe *et al.*, 2024a).

In compliance with this explanation, the Degree feature includes, among others, the values Dim and Aug, which apply to morphologically derived forms of nouns that express, respectively, “small size, or, metaphorically, affection towards the entity described by the noun” and “large size or force”, such as *appeltje* (“little apple” in Dutch) and *apartamento* (“big apartment” in Portuguese) (Marneffe *et al.*, 2024a). The UD documentation emphasizes that this feature also applies to verbs and adjectives in some languages. According to Rio-Torto (2015), diminutives and augmentatives are part of the formation of evaluatives, which in Portuguese is both isocategorical and pluricategorical, meaning it preserves the category of the base, which may belong to different categories, namely nouns, adjectives, adverbs, verbs, and pronouns.

According to Cruz (2011, p. 242), the derivational suffixes “*miri* ‘diminutive’ and *wasu* ‘augmentative’ allow the creation of new nouns from nominal bases” to refer to “entities with larger or smaller dimensions than the prototypical entity designated by the simple base noun”¹⁹, as in (31) and (32).

- 31) U-munhã kurusa-mirĩ-etá Kurupira r-apé upé.
3.ACT-make cross-DIM-PL Curupira CONT-path on
‘They made little crosses on Curupira’s path.’ (RODRIGUES, 1890, p. 78)
- 32) Ape ana tẽ paá tipusi-wasú u-mungiri aé.
then PFV FOC RPRT sleep-AUG 3.ACT-make_sleep her
‘At the same moment, they say, a great sleep put her to sleep.’ (Amorim, 1928, p. 175)

Following Cruz (2011), Avila (2021) classifies *wasú* as an augmentative suffix but diverges from her regarding the status of *mirĩ*. Avila (2021) classifies it as an adjective or adverb, writing it separately from the word it modifies, although acknowledging that it is in the process of grammaticalization as a suffix, given its combination with the plural particle *itá* (or *etá*), as in (31).

Cruz (2011) documents the use, albeit limited, of *miri* as an independent word, as does Avila (2021). The latter observes the incipient nominalization of *wasú* as an adjective. Previously, Stradelli (1929, p. 33) noted the ambivalent nature of these two morphemes, func-

¹⁹ The original Portuguese text reads as follows: “Os morfemas derivativos *miri* ‘diminutivo’ e *wasu* ‘augmentativo’ permitem criar novos nomes a partir de bases nominais [...] esses novos nomes designam entidades com dimensões maiores ou menores do que a entidade prototípica, designada pelo nome simples de base.”

tioning either as a “qualifying adjective” or as an adverbial modifier of another adjective²⁰, or as diminutive and augmentative suffixes, as in (33)–(36).

- 33) Se s-etama ne r-etama mirĩ piri.
my NCONT-country your CONT-country small CMPR
‘My homeland is smaller than yours.’ (Stradelli, 1929, p. 28)
- 34) Se t-uixawa amú t-uixawa-etá wasú piri.
my NCONT-chief other NCONT-chief-PL big CMPR
‘My chief is bigger than the other chiefs.’ (Stradelli, 1929, p. 28)
- 35) parana-miri-etá
river-DIM-PL
‘tributaries’ (Stradelli, 1929, p. 31)
- 36) pira-usu-etá
fish-AUG-PL
‘wales’ (Stradelli, 1929, p. 31)

Stradelli (1929) also highlights the reduction of *wasú* to *asú* in certain contexts, such as in *cunhãmucuasú* (Port. “mocetona”, Engl. “big girl”) and *cáuasú* (Port. “vespa grande”, Engl. “large wasp”), the augmentative forms of *cunhãmucú* (Port. “moça”, Engl. “girl”) and *cáua* (Port. “vespa”, Engl. “wasp”)²¹. In our view, this allomorphy further supports the affixal nature of the element.

Regardless of the status of the morphemes *mirĩ* and *wasú*, the bound morpheme *-í* productively suffixes to nouns in 19th-century Nheengatu (Stradelli, 1929), as in (37), corroborating the need to include the degree feature in the morphological analysis of the respective derived words. For this reason, we prefer to maintain this aspect of our annotation scheme, which, as we have shown, aligns with UD guidelines, despite the errors flagged by Udapi.

- 37) aresé pe-yeréu makaka-í aramapanhẽ ara upé
therefore 2PL.ACT-turn monkey-DIM to all life LOC
‘therefore, you will become little monkeys forever’ (Magalhães, 1876, p. 171)

4.4.2 Errors involving exocentric multiword expressions

Among the six error types identified by the Udapi tool, there were only four occurrences of *det-upos* and three of *mark-upos*. These two types of errors involve the multiword expressions *mayé waá* and *waá upé*, which function as *det* and *mark*, respectively. These syntactic relations are deemed incompatible with the part-of-speech tags of the heads of the corresponding multiword expressions (MWEs). Below, we first address the *mark-upos* errors, triggered by the classification of *waá*, the head of the MWE *waá upé*, as a relative pronoun, an analysis that contradicts the approaches of Cruz (2011) and Avila (2021), requiring a more in-depth discussion.

²⁰ The original Portuguese text reads as follows: “O adjetivo qualificativo admite, tal como o substantivo, três graus de qualificação: diminutiva, aumentativa e superlativa. Aumentativo, superlativo, diminutivo e adjetivo que modifica outro adjetivo é sempre posposto a este.”

²¹ In the original text by Stradelli (1929, p. 30), the first augmentative lacks the acute accent that marks the final syllable of the other examples of these formations.

The mark-upos errors occur in sentences such as (38) and (39), where the relative pronoun *waá* is governed by a verb under the syntactic relation mark, which is typically assigned to conjunctions.

- 38) Sampaio u-sendú nheengatú puapuãmu waá upé
 Sampaio 3.ACT-hear Nheengatu wanderREL LOC
 Barcelos upé.
 Barcelos LOC
 ‘Sampaio heard Nheengatu while wandering in Barcelos.’ (Avila, 2021, p. 802)
- 39) Tiwaá upé yawaraté u-sarú, mairamé u-saã
 not REL LOC jaguar 3.ACT-expect, when 3.ACT-feel
 u-suú ana [...].
 3.ACT-bite PRF
 ‘Unexpectedly, the jaguar felt itself being bitten [...].’ (Casasnovas, 2006, p. 70)

According to the code in Udapi’s `ud.MarkBugs` block in Listing 2, mark-upos errors are flagged exactly when PRON and mark occupy the UPOS and DEPREL columns of the annotation for a word, i.e., when a pronoun functions as a conjunction.

Listing 2 – Validation of the mark relation against UPOS value

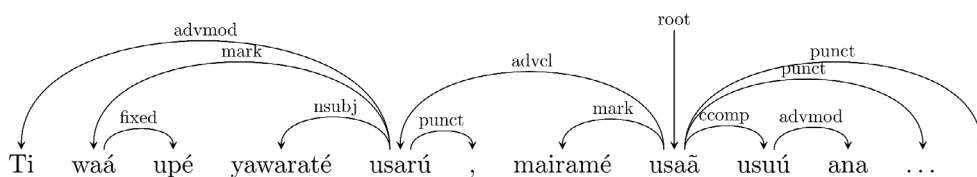
```
1 if udeprel == 'mark' and upos == 'PRON':
2     self.log(node, 'mark-upos', 'deprel=mark_upos=PRON')
```

Fonte: elaboração própria.

Following Avila (2021), the combination of the relative *waá* with the locative postposition *upé* in (38) and (39) makes up a MWE equivalent to the Portuguese conjunction *quando* (“when”). This expression exhibits the typical behavior of postpositional subordinating conjunctions in Nheengatu, which precede the predicate of the clause when it falls under the scope of negation.

In UD, the first member of an MWE functions as the head, to which the following members are attached via the fixed syntactic relation (Figure 26). It is in the very nature of these constructions to engage in syntactic relations that are incompatible with the grammatical class of the expression’s head (Constant *et al.*, 2017). To ensure compatibility between determined word classes and specific syntactic relations, UD introduced, after the release of version 2.14 on 15/05/2024, the `ExtPos` attribute to encode, in the FEATS field, the “external” tag of an exocentric MWE. Following this change, we began annotating *waá upé* with the feature `ExtPos=SCONJ`, without which the annotation of sentences like (38) and (39) triggers a validation error in the current validator version. However, Udapi continues to flag mark-upos errors for occurrences of *waá upé* in UD_Nheengatu-CompLin.

Figure 26 – Dependency tree of example (39)

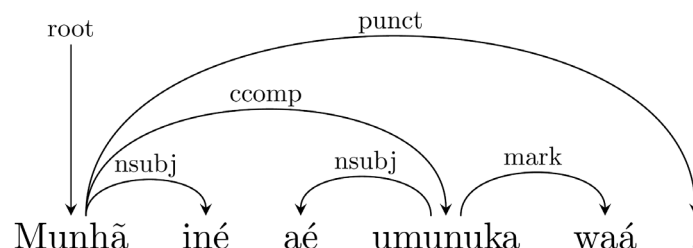


Fonte: elaboração própria.

This mismatch between the validation tool and Udaapi suggests the need for an update to the latter's code, something we hope will be addressed soon. Otherwise, one way to resolve the mark-upos errors would be to annotate the relative not as a pronoun but as a subordinating conjunction. Indeed, instances of *waá* as a subordinator occur in UD_Nheengatu-CompLin, as shown in Figure 27. However, this use of *waá* as a complementizer is not documented by Avila (2021). It appears in UD_Nheengatu-CompLin only in this example and another from Casasnovas (2006, p. 40), seemingly constituting a calque from Portuguese.

- 40) Munhã iné aé u-munuka waá.
 [2SG.IMP]make you he 3.ACT-cut that
 'You make him cut it.' (Seixas, 1853, p. XII)

Figure 27 – Dependency tree of example (40)



Fonte: elaboração própria.

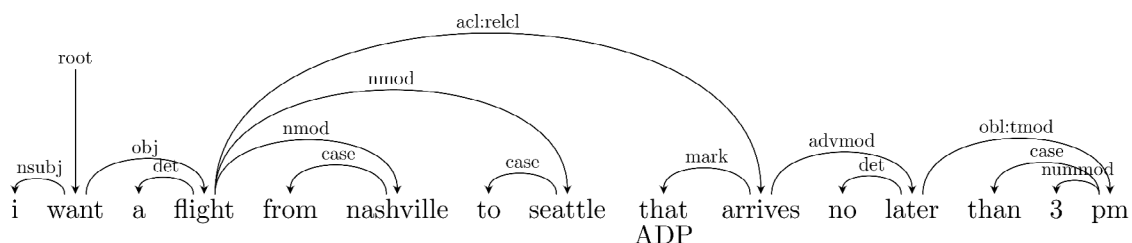
According to the UD documentation (Marneffe *et al.*, 2024a), “one needs to distinguish between relative clause markers, which are mark, from relative pronouns such as [en] *who* or *that*, which fill a regular verbal argument or modifier grammatical relation.” In UD, relative pronouns are generally annotated with the PRON tag and the feature PronType=Rel, as in UD_Portuguese-Porttinarari (Duran, 2021).

However, the classification of *that* as a relative pronoun is not unanimous in examples like (41), extracted from UD_English-Atis. As the only one among the six largest English treebanks in version 2.15 of the UD collection, UD_English-Atis classifies *that* in this type of example as an adposition linked via mark to the verb of the relative clause. According to Andrews (2007, p. 218), “[i]t is generally assumed *that* in relative clauses is not a relative pronoun.”

- 41) i want a flight from nashville to seattle that arrives no later than 3 pm

In UD_English-GUM, we found 690 occurrences of *that* as a relative pronoun, including examples analogous to (41). However, there are 6 occurrences of *that* as SCONJ in relative clauses, such as (42).

Figure 28 – Dependency tree of (41)



Fonte: elaboração própria.

42) We've seen mass strikes all around the world, in countries that we wouldn't expect it.

The treatment of the relative *waá* in Nheengatu descriptions, regarding the dichotomy established in UD between relative clause markers and pronouns, is not uniform. While Hartt (1872) and Simpson (1877) classify this element as a relative pronoun and a “relative adjective or pronoun”, respectively, Stradelli (1929) refers to it as a “conjunctive adjective.”

According to Moore, Facundes e Pires (1994), *waá* belongs to the set of subordinating particles, alongside elements such as *ramé* and *arama*, which mark temporal and purposive subordinate clauses, respectively. Relative clauses with *waá* conform to the structural model of subordinate clauses that Moore, Facundes e Pires (1994) designate as “indigenous”, in which the subordinating particle immediately follows the head of the VP, which, in example (43), is the adjective *nharú* (“wild”).

Similarly, Cruz (2011) classifies *waá* as a relativizing particle, used to construct clauses with the same degree of subordination as subordinators. Avila (2021) refers to *waá* both as a “relativizing particle” and a “nominalizing particle.” Both highlight that it admits the plural suffix *-itá* and precedes the predicate of the clause when under the scope of a negation particle. Navarro (2016), in turn, adopts the traditional label of relative pronoun. Finally, Melgueiro, Cabral e Martins (2019), despite calling *waá* a relativizing particle, acknowledge that it is responsible for the “retrieval of the relativized nominal”, implementing the “relative pronoun” strategy described by Givón (1990, p. 187), in which the relative pronoun also functions as a subordinating morpheme.

43) Asuí aikwé yuíri makú-itá nharú waá-itá kaá rupí.
and EXST also Indian-PL fierce REL-PL forest PERL
'And there were also fierce Indians in the woods.' (Magalhães, 1876, p. 171)

According to Cruz (2011, p. 514), in headless or free relative clauses, “the head of the modified noun phrase is not explicit, or, depending on the adopted theory, the noun phrase is expressed by a zero”²², as in (44). In the analysis she proposes for this example, the relativ-

²² Our translation from Portuguese: “o núcleo do sintagma nominal modificado não é explícito, ou, dependendo da teoria adotada, o sintagma nominal é expresso por um zero.”

zing particle is reduced to the clitic *wa*, which adjoins to *itá*²³. Cruz (2011) analyzes this host as a plural particle with scope over the noun phrase. The relative clause with the plural particle functions as a genitive complement of the noun *ara* (“day”).

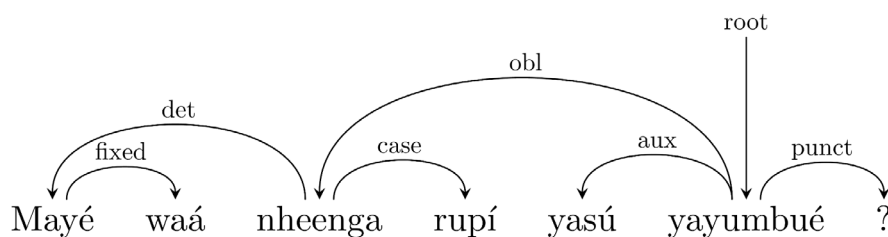
- 44) uwiara [u-manu wa]=ita ara
 today 3.ACT-die REL=PL day
 ‘Today is the day of those who died.’ (Cruz, 2011, p. 515)

In UD_Nheengatu-CompLin, we prefer to maintain the analysis of the relativizer *waá* as a relative pronoun rather than a relativizing particle, despite the errors pointed out by Udaapi in sentences with *waá upé*. Since UD is a lexicalist theory that tends to reject empty categories, this analysis seems to align better with the role of *itá* as a pluralizer of the noun phrase. Otherwise, one must postulate an empty nominal head in (44) to assign a noun phrase status to the expression in brackets. Instead, we argue that the head of this phrase is the relativizer, which, like the proximal and distal demonstrative pronouns *kwá* and *nhaã* and nouns, can be pluralized by *itá*, which, following Moore, Facundes e Pires (1994), among others, we treat as a plural suffix.

The det-upos errors result from assigning the syntactic relation det to *mayé* as head of the MWE *mayé waá*, as in (45), whose analysis is depicted in Figure 29. This MWE functions as a determiner or pronoun (Avila, 2021). Since *mayé* as an independent word is an interrogative adverb, a conflict arises with a specification in Udaapi’s code constraining det to pronouns and determiners. Similar to the ExtPos=SCONJ feature that specifies the external part-of-speech tag of *waá upé*, Udaapi ignores the ExtPos=DET feature that we inserted in the FEATS field for examples with *mayé waá* to comply with the current version of the validator.

- 45) Mayé waá nheenga rupíya-sú ya-yu-mbué?
 how REL language PERL1PL.ACT-go 1PL.ACT-MID-study
 ‘Through which language are we going to study?’ (Avila, 2021, p. 465)²⁴

Figure 29 – Dependency tree of example (45)



Fonte: elaboração própria.

²³ This morpheme is spelled *itá* in the orthography of Avila (2021).

²⁴ *Apud* Oliveira and Schwade (2012, p. 82).

5 Final remarks

In just a decade, the UD collection has grown to 296 treebanks covering 168 languages, as of the 2.15 release. This unprecedented milestone for such a resource highlights the growing interest in multilingual parsing and the increasing adoption of UD as a standard for morphosyntactic annotation. Brazilian Indigenous languages, which were initially neglected in natural language processing, corpus linguistics, and computational linguistics—fields traditionally focused on major languages like Portuguese—now represent 4.7% of the treebanks and 8.3% of the languages in UD. Additionally, the collection includes seven more Amerindian language treebanks.

Despite these percentages, the vast linguistic diversity of Brazil consisting of circa 150 living languages according to Storto (2019) remains underrepresented in UD, as is the case with other natural language processing domains. Another challenge faced by these minority languages, most of which are endangered, is data scarcity. Most of the 14 treebanks for Brazilian Indigenous languages are characterized by their small size: five contain fewer than 1,000 tokens, another five range between 1,000 and 5,000, and two between 5,000 and 10,000. Only UD_Mbya_Guarani-Dooley and UD_Nheengatu-CompLin exceed 10,000 tokens, the latter surpassing the former by 61.7% in size.

Another major discrepancy between Brazilian Indigenous language treebanks and those for majority languages like Portuguese results from their evaluation. Except for UD_Nheengatu-CompLin, these treebanks fall within the lower end of the star ranking system, receiving ratings between 0 and 1.5. Ironically, UD_Mbya_Guarani-Dooley is the only treebank in this group with a 0-star rating. In UD's most recent release, ratings range from 0 to 4.5 stars. A treebank's star rating is based on a score derived from multiple factors, including its size, annotation richness, textual genre diversity, and validation status.

This paper has detailed our efforts to improve UD_Nheengatu-CompLin's rating. In UD v2.14, the treebank received the highest score among all 21 Amerindian language treebanks, followed by UD_Mbya_Guarani-Thomas. Along with two modern Nahuatl treebanks, these two formed a group of four treebanks rated at 2.0 stars. Between UD v2.14 and v2.15, we raised UD_Nheengatu-CompLin's rating to 3.5 stars. Since UD v2.13, it has been the largest of all 21 treebanks for Amerindian languages. With the v2.15 update, it also became the only one in this group to surpass 2 stars.

To achieve this result, we focused on correcting errors detected by the Udapi framework. The number of words per error is the most heavily weighted factor in computing a treebank's score. We managed to reduce these errors drastically. While in UD v2.14, there was one error every 5.5 words, in UD v2.15, the ratio improved to one error per 332.38 words. As a result, the Udapi score increased from a minimum of 0.01 to 0.969913. Between versions 2.14 and 2.15, not only did we reduce errors, but we also continued expanding UD_Nheengatu-CompLin, as in previous update cycles. The word count grew by 78%, resulting in a 9.17% increase in the size score and a 4.87% increase in the part-of-speech tag score. Together, these improvements boosted UD_Nheengatu-CompLin's rating by 71.2%, elevating it to 3.5 stars.

As noted by 19th-century grammarians, including Hartt (1872), Magalhães (1876), and Sympson (1877), the Nheengatu verb in general does not express in itself tense or mood distinctions, only inflecting for person and number. Except for certain irregular forms and the

now-obsolete second-person singular imperative prefixes, interpreting a verb form as indicative or imperative depends on immediate syntactic context, prosody, or pragmatic factors. In addition, finite verbs occur in constructions where languages like Portuguese employ participles and infinitives. Only a few auxiliaries incorporate uninflected into the main verb, establishing a formal contrast with the corresponding inflected forms as full verbs.

A total of 97.87% of the errors flagged by Udapi resulted from our initial treatment of Nheengatu's verbal inflectional morphology. These errors involve the absence of Mood and VerbForm features in verb forms without a corresponding formal marking. These errors were systematically corrected by adding the missing features. Our annotation policy was revised accordingly. In addition to these corrections, we also eliminated the multi-obj errors, which made up 0.55% of the total error instances. Some of these cases resulted from human annotation oversights. Another part led to a revision of the annotation policy, whereby bare nominals expressing recipients of communication verbs with complement clauses are now treated as indirect objects. The remaining errors represent a mismatch between the constraints implemented in Udapi and UD's annotation guidelines. Therefore, we have postponed handling them until further discussion clarifies this matter.

In preparation for UD v2.16, we have raised the rating of the February 12, 2025 development version of UD_Nheengatu-CompLin to 4 stars, primarily due to a 97.06% increase in the split score. In UD v2.15, only 44 treebanks, or 14.9% of the UD collection, achieved a 4-star rating, covering 23 languages²⁵. All these languages are European. Except for Catalan and Scottish Gaelic, they are the majority languages of sovereign states, with the highest vitality index, i.e. 1, on the EGIDS scale. Catalan and Scottish Gaelic are statutory provincial languages in Spain and the United Kingdom, respectively, ranking at level 2 on the EGIDS scale (Eberhard; Simons; Fennig, 2025).

An even greater challenge will be raising UD_Nheengatu-CompLin's rating to 4.5 stars. Only eight treebanks for eight languages achieved this rating in UD v2.15, including UD_Portuguese-PetroGold. While five languages enjoy the highest vitality status, two—Western Armenian and Belarusian—are endangered, ranked at levels 6b and 7 on the EGIDS scale, respectively. However, their respective treebanks are among the largest in the UD collection, with approximately 122,000 and 305,000 tokens each, underscoring the scale of our challenge. Indeed, size is the second most heavily weighted factor in a treebank's rating. To achieve 4.5 stars, UD_Nheengatu-CompLin will need to reach approximately 59,000 words, assuming improvements in split, gender, POS tagging, syntactic relations, and Udapi scores, as outlined below. This growth represents nearly tripling its current size, making UD_Nheengatu-CompLin, *ceteris paribus*, the only treebank of a minority and non-European language at the top of the UD ranking system.

In the short to medium term, we will gradually increase the size score by adding new examples to UD_Nheengatu-CompLin. In this expansion, we will enrich UD_Nheengatu-CompLin in various dimensions. First, we will broaden the range of textual genres, increasing the respective score by 60% by incorporating excerpts from the Nheengatu translation of the Brazilian Constitution (Lucchesi *et al.*, 2023), songs, and abstracts, representing the legal, poetic, and academic genres. Once the corpus surpasses 30,000 words, we will establish a 10,000-

²⁵ This group includes two extinct languages: Latin and Old East Slavic, the latter being the ancestor of Russian, Ukrainian, and Belarusian (Simone, 2018).

word test set, thereby achieving the maximum split score. Second, we will seek to increase POS tagging and syntactic relation scores. Currently, UD_Nheengatu-CompLin does not use one of UD's 17 POS tags, namely SYM, which is applied to mathematical operators and currency symbols such as \$ (dollar), € (euro), and ¥ (yen). To incorporate this tag and achieve the maximum tag score, we could ask Nheengatu speakers to translate examples from other treebanks or create analogous ones. Of UD's 37 syntactic relations, three are not used in the current development version of UD_Nheengatu-CompLin: *clf*, *list*, and *orphan*. The first applies to nominal classifiers, a phenomenon common in Asian languages such as Chinese but absent in Nheengatu. The *list* relation applies to list elements, while *orphan* links a dependent word to an alternative head when the original head is elided, as in (46). Both relations seem possible in Nheengatu, potentially increasing the syntactic relation score from 0.94 to 0.97.

46) Marie won gold and Peter bronze.

At the same time, we will strive to eliminate all errors detected by Udati to maximize this score. To this end, we will explore two strategies. On the one hand, we will engage with the Udati maintenance team to address cases where sanctioned UD annotation configurations are flagged as errors. The first of these configurations is the marking of noun degree. The second is assigning a POS tag to the head of an exocentric MWE in violation of the set of syntactic relations licensed for that tag. Udati ignores the UD v2.15 *ExtPos* feature, which encodes the POS tag of these MWEs. On the other hand, we will assess the feasibility of tagging these MWE heads based on the *ExtPos* value, though this does not seem linguistically satisfactory at present.

By raising these questions, we hope this paper, in addition to contributing to greater inclusion of Nheengatu in digital humanities, also helps refine UD's rating system and inspires other treebank developers, particularly those working with minority and endangered languages, to improve the evaluation of their resources.

Authorship contribution statement

Leonel Figueiredo de Alencar: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Hélio Leonam Barroso Silva: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review & editing. Juliana Lopes Gurgel: Data curation, Formal analysis, Visualization, Writing – review & editing. Dominick Maia Alexandre: Data curation, Formal analysis, Writing – review & editing.

Acknowledgments

We are grateful to the following institutions for providing scholarships and other financial assistance to transcribers and annotators involved in the construction of UD_Nheengatu-CompLin: The Ceará State Foundation to Support Scientific and Technological Development

(Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico – Funcap), the Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES), and, last but not least, the São Paulo State Research Support Foundation (Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP), under Grant No. 22/09158-5, awarded to the DACILAT project at the Department of Linguistics of the University of Campinas (UNICAMP). We are also very grateful to the two anonymous reviewers for their numerous and extremely valuable comments and suggestions. We acknowledge the use of AI-based tools, including Grammarly, Quillbot, and ChatGPT, to assist with spelling, grammar, vocabulary, and style, as well as to facilitate the drafting of Python and Bash scripts used in data processing. We carefully examined, tested, and, when necessary, corrected the AI suggestions, taking full responsibility for the form and content of the paper.

References

ALENCAR, L. F. de. Aspectos da construção de um corpus sintaticamente anotado do nheengatu no modelo Dependências Universais. *Texto Livre*, Belo Horizonte, v. 17, p. e52653, 2024a. DOI: <https://doi.org/10.1590/1983-3652.2024.52653>.

ALENCAR, L. F. de. A Universal Dependencies treebank for Nheengatu. In: GAMALLO, P. et al. (Eds.). *Proceedings of the 16th International Conference on Computational Processing of Portuguese*. Santiago de Compostela, Galicia, Spain: Association for Computational Linguistics, 2024b. v. 2, p. 37–54. Available at: <https://aclanthology.org/2024.propor-2.8.pdf>. Accessed on: Apr. 2, 2025.

AMORIM, A. B. de. Lendas em Nheêngatu e em Portuguese. *Revista do Instituto Historico e Geographico Brasileiro*, Rio de Janeiro, v. 154, t. 100, p. 9–475, 1928.

ANDREWS, A. D. Relative clauses. In: SHOPEN, T. (ed.). *Language typology and syntactic description*. v. 2 — *Complex constructions*. Cambridge, UK: Cambridge University Press, 2007. p. 206–236.

AVILA, Marcel Twardowsky. *Proposta de dicionário nheengatu-português*. 2021. Tese (Doutorado em Estudos da Tradução) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2021. doi:10.11606/T.8.2021.tde-10012022-201925. Acesso em: 2026-04-12.

BRANCO, A. et al. Universal grammatical dependencies for Portuguese with CINTIL data, LX processing and CLARIN support. In: CALZOLARI, N. et al. (Eds.). *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 2022. p. 5617–5626. Available at: <https://aclanthology.org/2022.lrec-1.603/>. Accessed on: Apr. 2, 2025.

BORGES, L. C. O nheengatú: uma língua amazônica. *Papia*, Brasília, v. 4, n. 2, p. 44–55, 1996. Available at: https://etnolinguistica.wdfiles.com/local--files/artigo/borges-1996/borges_1996_nheengatu.pdf. Accessed on: Apr. 2, 2025.

BYBEE, J. L.; PERKINS, R.; PAGLIUCA, W. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press, 1994.

CASASNOVAS, A. *Noções de língua geral ou nheengatú: gramática, lendas e vocabulário*. 2nd ed. Manaus: Editora da Universidade Federal do Amazonas; Faculdade Salesiana Dom Bosco, 2006.

- CAVALIN, P. *et al.* Understanding native language identification for Brazilian indigenous languages. In: *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*. Toronto, Canada: Association for Computational Linguistics, 2023. p. 12–18. Available at: <https://aclanthology.org/2023.americasnlp-1.3>. Accessed on: Apr. 2, 2025.
- CONSTANT, M. *et al.* Multiword expression processing: A survey. *Computational Linguistics*, Cambridge, v. 43, n. 4, p. 837–892, dez. 2017. Available at: <https://aclanthology.org/J17-4005/>. Accessed on: Apr. 2, 2025.
- CRUZ, A. da. *Fonologia e gramática do nheengatú: a língua falada pelos povos Baré, Warekena e Baniwa*. Utrecht: LOT, 2011.
- CUNHA, C.; CINTRA, L. *Nova gramática do português contemporâneo*. 7. ed. Rio de Janeiro: Lexicon, 2017.
- D'ANGELIS, W. da R.; OLIVEIRA, M. C. de; SCHWADE, M. C. de D. L. Acesso ao mundo digital ou acesso digital ao mundo? *Revista Digital de Políticas Linguísticas*, Córdoba, v. 15, p. 134–158, 2021.
- DURAN, M. *et al.* The dawn of the Porttinari multigenre treebank: Introducing its journalistic portion. In: *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*. Porto Alegre: SBC, 2023. p. 115–124. Available at: <https://sol.sbc.org.br/index.php/stil/article/view/25443>. Accessed on: Apr. 2, 2025.
- DURAN, M. S. *Manual de anotação de POS Tags: orientações para anotação de etiquetas morfossintáticas em língua portuguesa, seguindo as diretrizes da abordagem Universal Dependencies (UD)*. São Carlos, SP: Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, 2021. Available at: <https://sites.google.com/icmc.usp.br/poetisa/publications>. Accessed on: Apr. 2, 2025.
- EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (Eds.). *Ethnologue: Languages of the World*. 26. ed. Dallas: SIL International, 2023. Available at: <http://www.ethnologue.com>. Accessed on: Jul. 30, 2023.
- EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (Eds.). *Ethnologue: Languages of the World*. 28. ed. Dallas: SIL International, 2025. Available at: <http://www.ethnologue.com>. Accessed on: Apr. 2, 2025.
- FREIRE, J. R. B. *Rio Babel: A história das línguas na Amazônia*. 2nd ed. Rio de Janeiro: EdUERJ, 2011.
- GALVES, C. *et al.* Annotating a polysynthetic language: From Portuguese to Kadiwéu. *Cadernos de Estudos Linguísticos*, Campinas, v. 59, n. 3, p. 631–648, 2017. DOI: <https://doi.org/10.20396/cel.v59i3.8651003>
- GIVÓN, T. *Syntax: A functional-typological introduction*. Amsterdam: John Benjamins, 1990. v. 2.
- HARTT, C. F. Notes on the Lingoa Geral or Modern Tupi of the Amazonas. *Transactions of the American Philological Association*, Baltimore, v. 3, p. 58–76, 1872. Available at: <https://www.jstor.org/stable/310258>. Accessed on: Apr. 2, 2025.
- HARTT, C. F. Notas sobre a língua geral, ou tupí moderno do Amazonas. *Anais da Biblioteca Nacional do Rio de Janeiro*, Rio de Janeiro, vol. 51, p. 305–390, 1938.
- HELBIG, G. *Probleme der Valenz- und Kasustheorie*. Tübingen: Max Niemeyer Verlag, 1992.
- JURAFSKY, D.; MARTIN, J. H. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 2nd ed. London: Pearson, 2009.
- LEVIN, B. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press, 1993.

LUCCHESI, M. *et al.* (Eds.). *Mundu Sa Turusu Waá: Ubêuwa Mayé Míra Itá Uikú Arãma Purãga Iké Braziu Upé*. Brasília: Supremo Tribunal Federal, Conselho Nacional de Justiça, 2023. Available at: <https://bibliotecadigital.cnj.jus.br>. Accessed on: Apr. 2, 2025.

MAGALHÃES, J. V. C. de. *O selvagem*. Rio de Janeiro: Typographia da Reforma, 1876.

MARNEFFE, M.-C. de *et al.* Universal Dependencies. *Computational Linguistics*, Cambridge, v. 47, n. 2, p. 255–308, 2021. Available at: <https://aclanthology.org/2021.cl-2.11>. Accessed on: Apr. 2, 2025.

MARNEFFE, M.-C. de *et al.* *Universal Dependencies Guidelines*. [S. n.]: 2024a. Available at: <<https://universaldependencies.org/guidelines.html>>. Accessed on: Apr. 2, 2025.

MARNEFFE, M.-C. de *et al.* *UD Validation since release 2.5*. [S. n.]: 2024b. Available at: <https://universaldependencies.org/validation-rules.html>. Accessed on: Apr. 2, 2025.

MELGUEIRO, E. M.; CABRAL, A. S. A. C.; MARTINS, M. F. Orações relativas em nheengatú ou ingatú. *Revista Brasileira de Linguística Antropológica*, Brasília, v. 11, n. 2, p. 151–166, 2019. DOI: <https://doi.org/10.26512/rbla.v11i02.28115>

MOORE, D.; FACUNDES, S.; PIRES, N. Nheengatu (Língua Geral Amazônica), its history, and the effects of language contact. In: *Proceedings of the Meeting of the Society for the Study of the Indigenous languages of the Americas, July 2–4, 1993 and the Hokan-Penutian Workshop, July 3, 1993*. Berkeley, CA: [University of California], 1994. p. 93–118. Available at: <https://escholarship.org/uc/item/7tb981s1>. Accessed on: Jul. 26, 2024.

MÜLLER-EBERSTEIN, M.; GOOT, R. van der; PLANK, B. How universal is genre in Universal Dependencies? In: DAKOTA, D.; EVANG, K.; KÜBLER, S. (Eds.). *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*. Sofia, Bulgaria: Association for Computational Linguistics, 2021. p. 69–85. Available at: <https://aclanthology.org/2021.tlt-1.7/>. Accessed on: Apr. 2, 2025.

NAVARRO, E. d. A. *Curso de Língua Geral (nheengatu ou tupi moderno): a língua das origens da civilização amazônica*. 2nd ed. São Paulo: Centro Angel Rama da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, 2016.

NAVARRO, E. d. A.; ÁVILA, M. T.; TREVISAN, R. G. O Nheengatu, entre a vida e a morte: a tradução literária como possível instrumento de sua revitalização lexical. *Revista Letras Raras*, Campina Grande, v. 6, n. 2, p. 9–29, 2017. DOI: <https://dx.doi.org/10.35572/rlr.v6i2.768>

NIVRE, J.; ZEMAN, D.; GINTER, F.; TYERS, F. M. Tutorial on Universal Dependencies: Infrastructure, resources and tools for UD. [S. l.: s. n.], 2017. Tutorial presented at the *15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017)*. Available at: <https://universaldependencies.org/each17tutorial/infrastructure.pdf>. Accessed on: Apr. 3, 2025.

OLIVEIRA, G. M. de; CAMACHO, R. G. *Estratégias de relativização e construções alternativas nas línguas indígenas do Brasil*. São Paulo: Cultura Acadêmica, 2013. Available at: <http://hdl.handle.net/11449/109292>. Accessed on: Apr. 2, 2025.

PALMER, F. R. *Mood and modality*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2001.

PERINI, M. A. *Describing verb valency: Practical and theoretical issues*. Cham, Switzerland: Springer, 2015. DOI: <https://doi.org/10.1007/978-3-319-20985-2>

PERINI, M. A. *Thematic relations: A study in the grammar-cognition interface*. Cham, Switzerland: Springer, 2019. DOI: <https://doi.org/10.1007/978-3-030-28538-8>

PINHANEZ, C.; CAVALIN, P.; NOGIMA, J. Human evaluation of the usefulness of fine-tuned English translators for the Guarani Mbya and Nheengatu indigenous languages. In: GAMALLO, P. et al. (eds.). *Proceedings of the 16th International Conference on Computational Processing of Portuguese*. Santiago de Compostela, Galicia/Spain: Association for Computational Linguistics, 2024. v. 2, p. 32–36. Available at: <https://aclanthology.org/2024.propor-2.7/>. Accessed on: Apr. 2, 2025.

PEPEL, M.; ŽABOKRTSKÝ, Z.; VOJTEK, M. UdaPI: Universal API for Universal Dependencies. In: *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*. Gothenburg, Sweden: Association for Computational Linguistics, 2017. p. 96–101. Available at: <https://aclanthology.org/W17-0412>. Accessed on: Apr. 2, 2025.

RIO-TORTO, G. Formação de avaliativos. In: RIO-TORTO, G. et al. (Eds.). *Gramática derivacional do Português*. 2nd ed. Coimbra: Imprensa da Universidade de Coimbra, 2015. p. 357–389. Available at: <http://dx.doi.org/10.14195/978-989-26-0864-8>. Accessed on: Apr. 2, 2025.

RODRIGUES, A. D. Tarefas da lingüística no Brasil. *Estudos Lingüísticos (Revista Brasileira de Lingüística Teórica e Aplicada)*, Rio de Janeiro, v. 1, n. 1, p. 4–15, 1966. Available at: <http://www.etnolinguistica.org/biblio:rodrigues-1966-tarefas>. Accessed on: Apr. 2, 2025.

RODRIGUES, A. D. *Línguas brasileiras: para o conhecimento das línguas indígenas*. São Paulo: Loyola, 1986.

RODRIGUES, A. D. Línguas indígenas: 500 anos de descobertas e perdas. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, São Paulo, v. 9, n. 1, p. 83–103, 1993. Available at: <https://revistas.pucsp.br/index.php/delta/article/view/45596>. Accessed on: Apr. 2, 2025.

RODRIGUES, A. D. As línguas gerais sul-americanas. *Papia*, São Paulo, v. 4, n. 2, p. 6–18, 1996. Available at: https://etnolinguistica.wdfiles.com/local--files/artigo%3Arodrigues-1996/rodrigues_1996_linguas_gerais.pdf. Accessed on: Apr. 2, 2025.

RODRIGUES, A. D.; CABRAL, A. S. A. C. A contribution to the linguistic history of the Língua Geral Amazônica. *ALFA: Revista de Linguística*, São José do Rio Preto, v. 55, n. 2, 12 2011. DOI: <https://doi.org/10.1590/S1981-57942011000200012>

RODRIGUES, J. B. *Poranduba amazonense ou kochiyima-uara porandub*: 1872–1887. Rio de Janeiro: Typ. de G. Leuzinger & Filhos, 1890.

RUETER, J. et al. Apurinã Universal Dependencies treebank. In: MAGER, M. et al. (eds.). *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*. Online: Association for Computational Linguistics, 2021. p. 28–33. Available at: <https://aclanthology.org/2021.americasnlp-1.4>. Accessed on: Apr. 2, 2025.

RUPPENHOFER, J. et al. *FrameNet II: Extended theory and practice*. [Berkeley: International Computer Science Institute]: 2016. Revised version. Available at: https://akb89.github.io/myValencer/framenet_book.pdf. Accessed on: Apr. 2, 2025.

SANDALO, M. F. S.; GALVES, C. M. C. Anotando sintaticamente uma língua originária do Brasil: O problema de anchieta. *Cadernos de Estudos Lingüísticos*, Campinas, v. 65, n. 00, p. e023007, 2023. DOI: <https://doi.org/10.20396/cel.v65i00.8673592>

- SANTOS, L. L.; ARAGON, C. C.; GERARDI, F. Línguas minoritárias e anotações sintáticas de corpora: experiências de pesquisa na iniciação científica. *Letras de hoje*, Porto Alegre, v. 59, n. 1, p. 1–9, 2024. DOI: <https://doi.org/10.15448/1984-7726.2024.1.44734>
- SEIXAS, M. J. d. *Vocabulário da lingua indigena geral para o uso do Seminario Episcopal do Pará*. Pará: Typ. de Mattos e Comp^a, 1853.
- SEKI, L. A lingüística indígena no Brasil. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, São Paulo, v. 15, n. especial, p. 257–290, 1999.
- SILVA, D. P. G. da; PARDO, T. A. S. Grammar induction for Brazilian indigenous languages. In: CAMALLO, P. et al. (Eds.). *Proceedings of the 16th International Conference on Computational Processing of Portuguese*. Santiago de Compostela, Galicia/Spain: Association for Computational Linguistics, 2024. v. 2, p. 64–72. Available at: <https://aclanthology.org/2024.propor-2.10/>. Accessed on: Apr. 2, 2025.
- SIMONE, L. R. Uma breve introdução ao idioma eslavo oriental antigo. *Slovo – Revista de Estudos em Eslavística*, Rio de Janeiro, v. 1, n. 1, p. 16–17, 2018. Available at: <https://revistas.ufrj.br/index.php/slovo/article/view/17473/11271>. Accessed on: Apr. 2, 2025.
- STORTO, L. R. *Línguas indígenas: tradição, universais e diversidade*. Campinas: Mercado de Letras, 2019.
- STRADELLI, E. Vocabulários da lingua geral portuguez-nheêngatú e nheêngatú-portuguez, precedidos de um esboço de Grammatica nheênga-umbuê-sáua mirî e seguidos de contos em lingua geral nheêngatú poranduuua. *Revista do Instituto Historico e Geographico Brasileiro*, Rio de Janeiro, v. 158, n. 104, p. 9–768, 1929.
- SYMPSON, P. L. *Grammatica da lingua brazilica geral, fallada pelos aborigines das provincias do Pará e Amazonas*. Manaus: Typographia do Commercio do Amazonas, 1877.
- YAMÃ, Y. et al. *Dicionário e estudo de nheengatu tradicional*. 2nd ed. São Paulo: Cintra, 2021.
- ZEMAN, D. *Cross-Language Harmonization of Linguistic Resources*. Prague: Institute of Formal and Applied Linguistics (ÚFAL), 2023. Habilitation thesis. Available at: <https://chres.is.cuni.cz/media/documents/2024/02/25/thesis-without-papers.pdf>. Accessed on: Apr. 2, 2025.
- ZWICKY, A. M. In a manner of speaking. *Linguistic Inquiry*, Cambridge, v. 2, n. 2, p. 223–233, 1971.

Code-mixing, code-switching e empréstimos no Hunsrückisch: uma análise de contato linguístico

Code-Mixing, Code-Switching and Borrowing in Hunsrückisch: An Analysis of Language Contact

Leonardo Alves

Universidade Federal de Santa Catarina
(UFSC) | Florianópolis | SC | BR
dadoleo91@gmail.com
<https://orcid.org/0009-0009-8933-7489>

Resumo: O Hunsrückisch, língua de imigração germânica com mais de dois séculos de presença no Brasil, constitui um laboratório privilegiado para o estudo de fenômenos de contato linguístico. Este artigo examina, de maneira integrada, *code-switching*, *code-mixing* e empréstimos lexicais no repertório hunsriqueano, articulando-os a processos de *koineização*, formação de variedades intermediárias e práticas contemporâneas de bilinguismo e multilinguismo. A análise combina revisão de literatura, dados lexicográficos (Boll, 2021; Langenscheidt, 2010), registros espontâneos de escrita digital (Lindenfelser, 2024; Limberger, 2021) e descrições sociolinguísticas recentes (Altenhofen; Morello, 2018; Mackedanz; Spinassé, 2022; Spinassé, 2016, 2017). O arcabouço teórico mobiliza modelos de contato e processamento bilíngue (Matras, 2020; Riehl, 2009, 2014), bem como perspectivas fonológico-morfológicas sobre empréstimos e adaptações (Nübling *et al.*, 2017). Os resultados mostram que o repertório hunsriqueano contemporâneo opera num contínuo entre alternância de códigos, integração lexical e formas híbridas, mais do que numa oposição rígida entre dois sistemas. Esses fenômenos revelam tanto a permeabilidade estrutural quanto a vitalidade adaptativa da língua, cuja continuidade depende de ecologias multilíngues dinâmicas e de políticas linguísticas que reconheçam práticas *in vivo*, especialmente em contextos marcados por repressão histórica, estigmatização social e pressões de monolinguismo institucional. Percebe-se que a vitalidade do Hunsrückisch emerge justamente da negociação cotidiana entre português, variedades germânicas e reper-



tórios locais, reafirmando a centralidade do *code-mixing* e dos empréstimos como mecanismos de preservação e renovação linguística.

Palavras-chave: *code-mixing*; *code-switching*; contato linguístico; empréstimo lexical; Hunsrückisch.

Abstract: Hunsrückisch, a German-immigration language with more than two centuries of presence in Brazil, offers a privileged laboratory for the study of linguistic contact phenomena. This article examines, in an integrated manner, code-switching, code-mixing, and lexical borrowing within the Hunsrückisch repertoire, linking them to processes of koineization, the formation of intermediate varieties, and contemporary practices of bilingualism and multilingualism. The analysis combines literature review, lexicographic data (Boll, 2021; Langenscheidt, 2010), spontaneous digital writing records (Limberger, 2021; Lindenfelser, 2024), and recent sociolinguistic descriptions (Altenhofen; Morello, 2018; Mackedanz; Spinassé, 2022; Spinassé, 2016, 2017). The theoretical framework draws on models of contact and bilingual processing (Matras, 2020; Riehl, 2009, 2014), as well as phonological–morphological perspectives on borrowing and adaptation (Nübling *et al.*, 2017). The results show that the contemporary Hunsrückisch repertoire operates along a continuum between code alternation, lexical integration, and hybrid forms, rather than through a rigid opposition between two systems. These phenomena reveal both the structural permeability and the adaptive vitality of the language, whose continuity depends on dynamic multilingual ecologies and on language policies that recognize *in vivo* practices—particularly in contexts marked by historical repression, social stigmatization, and institutional monolingual pressures. The study concludes that the vitality of Hunsrückisch emerges precisely from the everyday negotiation among Portuguese, German varieties, and local repertoires, reaffirming the centrality of code-mixing and borrowing as mechanisms of linguistic preservation and renewal.

Keywords: code-mixing; code-switching; language contact; lexical borrowing; Hunsrückisch.

1 Introdução

O Brasil é amplamente reconhecido por sua diversidade linguística e cultural, que antecede o período colonial e tem origem nas centenas de línguas indígenas que se distribuíam pelo território muito antes do século XVI. Esse mosaico foi posteriormente ampliado pela chegada de povos africanos durante a colonização escravagista e, mais tarde, pelos diversos fluxos migratórios europeus e asiáticos, que trouxeram novas línguas e tradições culturais. Dentre as línguas de imigração que hoje compõem o cenário sociolinguístico brasileiro, o Hunsrückisch se destaca como a variedade de origem alemã mais amplamente difundida no país, sendo língua de referência para mais de um milhão de brasileiros, com presença consolidada sobretudo nos estados do Rio Grande do Sul, Santa Catarina, Paraná e Espírito Santo, além de ramificações no Mato Grosso do Sul, no Paraguai e na Argentina (Altenhofen; Morello, 2018, p. 18). Introduzido por imigrantes provenientes da região do Hunsrück e do Palatinado, na atual Renânia-Palatinado, a partir de 1824, o Hunsrückisch consolidou-se como língua de uso familiar e comunitário em diferentes regiões do sul do país. Ao longo de sua história, porém, enfrentou desafios significativos – entre eles, a repressão linguística durante o Estado Novo (1937-1945), quando o Decreto-Lei n.º 1.545/1939 proibiu o uso público de idiomas estrangeiros, resultando em prisões, agressões físicas, censura e silenciamento (Gälzer; Magalhães, 2024, p. 7; Oliveira, 2009, p. 22). Esse processo foi intensificado pelo contexto da Segunda Guerra Mundial, quando o Brasil, após romper a neutralidade em 1942 e declarar guerra aos países do Eixo, adotou medidas rígidas para reforçar a identidade nacional, incluindo a proibição de idiomas associados às nações inimigas (Gälzer; Magalhães, 2024, p. 4). Essa política de nacionalização buscava homogeneizar a população por meio da imposição do português como língua exclusiva da vida pública e da educação, o que produziu um impacto duradouro sobre as línguas de imigração no Brasil. Paralelamente, a hegemonia do português foi reforçada por políticas de valorização do monolinguismo nacional, contribuindo para o deslocamento gradual das línguas de herança e para a marginalização de seus falantes. Este artigo busca explorar a trajetória dessa língua, com ênfase em seu contato com o português brasileiro, investigando os fenômenos de *code-mixing*, *code-switching* e de empréstimos lexicais.

Este estudo analisa o *code-mixing* entre o Hunsrückisch e o português brasileiro como um mecanismo de vitalidade linguística, responsável por renovar e expandir o repertório da língua de imigração. Parte-se da hipótese de que os empréstimos lexicais decorrentes desse contato não representam perda, mas sim um processo adaptativo e criativo que assegura a continuidade do Hunsrückisch como língua viva e funcional em diferentes domínios sociais. A pesquisa adota uma abordagem qualitativa, de caráter empírico-documental, fundamentada nos estudos clássicos sobre contato linguístico e variação propostos por Weinreich, Labov e Herzog (2006), Ferguson (1972), Fishman (1964, 1965a, 1965b), Gumperz (1982) e Labov (1994, 2008). O corpus é constituído por registros lexicográficos e descritivos que documentam as línguas hunsriqueana e alemã – o *Dicionário hunsriqueano riograndense – português* (Boll, 2021) e o *Langenscheidt Universal-Wörterbuch Brasilianisches Portugiesisch: Portugiesisch–Deutsch / Deutsch–Portugiesisch* (Langenscheidt, 2010) – revisados por falantes de Hunsrückisch e de alemão padrão, que colaboraram na validação e interpretação de equivalências lexicais. Também se incorporam análises e descrições de Vilela (2004), Spinassé (2016; 2017), Altenhofen (2013), Altenhofen e Morello (2018) e Oliveira (2018), que situam o Hunsrückisch

no contexto sociolinguístico brasileiro e Weber (2020) e Gregory (2013) na perspectiva histórica. As ocorrências de *code-mixing* e de empréstimos lexicais foram sistematizadas em uma tabela analítica, evidenciando padrões de integração lexical e morfológica de vocábulos portugueses, dialogando com contribuições de Riehl (2009, 2014), Matras (2020), Limberger (2021), Spinassé (2016, 2017) e Lindenfelser (2024). Dessa forma, o estudo evidencia que os empréstimos lexicais constituem parte essencial do *code-mixing* que emerge do contato entre o Hunsrückisch e o português brasileiro. Esses empréstimos revelam processos de adaptação linguística que refletem o modo como as duas línguas se entrelaçam nas práticas comunicativas cotidianas, preservando a funcionalidade e o valor simbólico do Hunsrückisch em contextos multilíngues.

O artigo organiza-se em quatro movimentos analíticos articulados. Primeiro, apresentou-se, na Introdução, a pergunta de pesquisa e o enquadramento teórico, formulando a hipótese de que *code-mixing* e empréstimos lexicais operam como mecanismos de vitalidade no repertório hunsriqueano em ecologias multilíngues. Em seguida, discute-se a constituição histórica da língua de imigração, da formação comunitária à repressão do Estado Novo e às políticas recentes de mapeamento e cooficialização, situando o Hunsrückisch entre denominações concorrentes e propostas normativas distintas – notadamente o modelo associado ao ALMA-H e a ortografia Hunsrik/Taytx proposta pela SIL. A terceira parte examina os fenômenos de contato linguístico, diferenciando *code-switching*, *code-mixing*, empréstimos lexicais, *Lehnprägungen*, processos de *koineização* e a formação de variedades intermediárias. Por fim, a conclusão retoma os principais achados, mostra como eles respondem à hipótese inicial e aponta implicações para a análise dos repertórios de línguas de imigração em contextos multilíngues contemporâneos.

2 A língua da imigração alemã no Brasil: o caso do Hunsrückisch

O Hunsrückisch é uma abstração linguística que representa uma variedade do alemão originada da região homônima na Renânia Central, na Alemanha. Historicamente, é associado às comunidades de imigração que preservaram essa língua após migrarem, especialmente para o Brasil (Gregory, 2013). Esse modo de falar varia significativamente entre gerações, classes sociais, e situações de uso, refletindo as características dinâmicas de todas as línguas. Apesar disso, a delimitação dessa variedade é essencial para identificar e descrever o que se entende como Hunsrückisch, considerando aspectos linguísticos e históricos comuns que caracterizam essa língua.

De acordo com Gregory (2013, p. 11), é possível situar o Hunsrückisch no interior de um processo migratório mais amplo, marcado pela chegada de camponeses de diferentes regiões de língua alemã ao Brasil, sobretudo a partir do século XIX. As colônias do Sul – em especial no Rio Grande do Sul, Santa Catarina e Paraná – foram povoadas por grupos oriundos de áreas diversas, mas a região do Hunsrück destacou-se como importante fornecedora de imigrantes camponeses, majoritariamente católicos e falantes do dialeto Hunsrück (Gregory, 2013, p. 15). Esses contingentes foram assentados em núcleos coloniais relativamente homogêneos e, ao longo do tempo, expandiram-se por meio de novas frentes de colonização e migrações internas, formando redes de descendentes que se estendem hoje por diferentes regiões do território brasileiro. Nesse contexto, a variedade que aqui se denomina Hunsrückisch se enraíza em

trajetórias históricas específicas de ocupação do espaço, organização comunitária e negociação identitária, nas quais a língua se articula a práticas religiosas, escolares e associativas que deixaram marcas duradouras na paisagem sociocultural teuto-brasileira.

Essa diversidade de destinos e fluxos migratórios também apresenta diferentes concepções identitárias e linguísticas que refletem até mesmo nas tentativas de uma grafia para a língua e seu nome. Como mostram Von Mühlen e Kersch (2021, p. 327), a discussão sobre a escrita do hunsriqueano insere-se em um quadro mais amplo de multilinguismo e de hierarquização de línguas no Brasil. As autoras lembram que, segundo a UNESCO, a vitalidade de uma língua depende também de sua capacidade de se abrir a novos âmbitos e meios de circulação, o que torna a modalidade escrita um fator decisivo para a sobrevivência de línguas minoritárias em contextos crescentemente urbanizados. No caso do hunsriqueano – variedade alóctone de origem alemã, ainda hoje falada por cerca de 1,2 milhão de pessoas sobretudo no Rio Grande do Sul, em Santa Catarina e no Espírito Santo –, trata-se de uma língua majoritariamente oral, marcada por situação de diglossia em relação ao alemão-padrão e ao português, mas que já conta com uma pequena, embora importante, tradição escrita em jornais, livros e, mais recentemente, na internet (Von Mühlen; Kersch, 2021, p. 332). Diante das pressões sociais e simbólicas que favorecem o uso das línguas oficiais e de prestígio, o desenvolvimento e a ampliação dos usos sociais da escrita em hunsriqueano – ainda que por meio de diferentes estratégias gráficas e antes mesmo de um consenso ortográfico – constituem um instrumento central de visibilização, valorização e manutenção dessa língua de imigração no Brasil.

Existe uma perspectiva que adota o Hunsrückisch como denominação buscando manter e descrever uma prática de escrita já existente entre os falantes, como defendido pelo Projeto Atlas Linguístico-Contatual das Minorias Alemãs na Bacia do Prata: Hunsrückisch (ALMA-H) e amplamente aceita pela academia. A proposta desenvolvida pelo grupo ESCRITHU (Sistema de Escrita do Hunsrückisch), vinculado ao ALMA-H e coordenado por Cléo V. Altenhofen, baseia-se em critérios linguísticos, históricos e pedagógicos, procurando equilibrar traços do alemão-padrão com a autonomia adquirida pela língua no contexto brasileiro. A normatização proposta por esse grupo visa preservar o Hunsrückisch como patrimônio cultural e também fomentar sua educação e reflexão linguística, promovendo sua integração em contextos acadêmicos e comunitários (Altenhofen *et al.*, 2007).

Nessa direção, Altenhofen *et al.* (2007, p. 74) argumentam que a fixação de uma escrita para o Hunsrückisch parte do reconhecimento da língua como sistema autônomo, dotado de direito pleno a uma forma gráfica própria, e não apenas como “variante baixa” subordinada ao alemão-padrão. A escrita é concebida como convenção que precisa ser aprendida e testada na prática, devendo garantir, ao mesmo tempo, legibilidade para os falantes do Hunsrückisch e inteligibilidade mínima para leitores familiarizados com outras variedades do alemão. Os autores ressaltam ainda que a proposta não pretende apagar a variação interna, mas organizá-la, reconhecendo pelo menos três grandes perfis de Hunsrückisch (com traços mais mose-lanos, mais renanos ou mais próximos do padrão). Ao articular essa base ortográfica com objetivos pedagógicos mais amplos – como a elaboração futura de um dicionário trilingue Hunsrückisch–alemão-padrão–português e a transliteração sistemática de etnotextos coletados pelo ALMA-H –, o grupo projeta a escrita como ferramenta de registro, comparação e educação linguística, abrindo caminho para que o Hunsrückisch circule em novos domínios e consolide seu estatuto de língua de imigração no Brasil (Altenhofen *et al.*, 2007, p. 76).

Em contraposição, o Projeto Hunsrik propõe uma abordagem voltada à autonomia do idioma enquanto língua germânica distinta, adaptada à realidade sul-americana, denominando-a Hunsrik. Essa proposta defende o uso de um sistema ortográfico transparente, fundamentado na correspondência direta entre grafemas e fonemas, de modo que cada som seja representado por um único símbolo. Inspirado no inventário gráfico do português brasileiro, o modelo proposto pela SIL Global busca facilitar o aprendizado e a codificação linguística pelas novas gerações. Um exemplo ilustrativo é a forma Taytx para “alemão” (Deutsch), em que ay representa o ditongo [aɪ], como em pai, e tx o som [tʃ], como em tchau, tornando a escrita previsível para leitores alfabetizados em português. Desenvolvido desde 2004 no município de Santa Maria do Herval (RS), o projeto adota essa ortografia “abrasileirada” para valorizar a evolução do idioma no Brasil e promover uma revitalização mais acessível e desvinculada de padrões externos ou históricos, sob o argumento de que o aprendizado do alemão-padrão é mais difícil (Rio Grande Do Sul, [s.d.]; Maselko, 2013).

A proposta de Wiesemann (2008) insere essa ortografia “abrasileirada” em um escopo mais abrangente, no qual a codificação do Hunsrik depende tanto de uma análise fonológica rigorosa – que seleciona apenas os sons contrastivos para compor o alfabeto – quanto do reconhecimento do contexto social e escolar em que a língua é usada. Em vez de pressupor o domínio do alemão standard, tratado como “ciência quase perdida” entre muitos descendentes de imigrantes, a autora defende que não há qualquer necessidade de saber ler em alemão-padrão para ler e escrever em Hunsrik, e que a escolha de um sistema regular, com um símbolo para cada fonema, facilita a alfabetização de crianças e adultos já letrados em português. Nessa perspectiva, a ortografia proposta foi pensada para dialogar com o inventário gráfico do português brasileiro, mas sem subordinar o Hunsrik a ele, em que se opta, por exemplo, por não marcar substantivos com maiúscula nem aglutinar longos compostos, buscando uma escrita funcional, legível e apropriada ao uso escolar, projetando essa ortografia como base possível para uma “nova escola brasileira” que reconheça também a língua dos colonos como língua primeira de ensino, convertendo o bilinguismo Hunsrik–português em recurso e não em obstáculo (Wiesemann, 2008, p. 3).

Mais recentemente, Lindenfelser (2024) introduz uma terceira via ao discutir práticas efetivas de escrita do Hunsrückisch em contextos digitais. Em um estudo grafemático com 350 postagens e comentários de falantes em redes sociais, o autor demonstra que a escrita espontânea da língua revela uma grande variação nas correspondências grafema-fonema, frequentemente influenciada pelo português brasileiro e pelos usos cotidianos dos falantes, e não necessariamente pelos modelos normativos propostos. Fenômenos como a vocalização do /r/ em <a> (fa ‘für’), a alternância entre <y>, <i> e <j> para /j/, ou a adoção de grafias como <ai> e <oi> em lugar das correspondências alemãs <ei> e <eu> indicam um processo ativo de grafematização de base comunitária.

Esses resultados sugerem que a vitalidade e o próprio uso do Hunsrückisch não se limita às iniciativas institucionais de padronização, mas também se manifesta na criatividade e na adaptabilidade das práticas escritas cotidianas. A escrita “de baixo para cima” descrita por Lindenfelser (2024, p. 73) reforça a ideia de que a preservação linguística ocorre tanto nas esferas acadêmicas e pedagógicas quanto nas dinâmicas sociais e digitais, onde os falantes constroem – intuitiva e coletivamente – uma ortografia viva, que reflete sua condição bi/multilíngue e o contínuo diálogo entre o português e o alemão no Brasil contemporâneo (Lindenfelser, 2024).

Há diferentes níveis de complexidade ao se analisar o Hunsrückisch – desde a própria nomeação da língua até os processos de normatização e as dimensões identitárias envolvidas. Diante dessa diversidade, marcada por contatos múltiplos e trajetórias históricas distintas, torna-se inviável propor uma descrição uniforme dos empréstimos e de suas ocorrências. As interações linguísticas não se limitam ao português, mas também envolvem outras línguas de imigração e contextos regionais variados, o que faz com que o vocabulário e as práticas discursivas assumam características específicas em cada comunidade. Assim, a denominação Hunsrückisch e a seleção dos exemplos analisados neste estudo devem ser entendidas como representativas de um recorte, e não como abarcando a totalidade das variedades em circulação no Brasil e em outros países da América do Sul.

A distinção proposta por Calvet (2007, p. 68) entre os fenômenos *in vivo* e *in vitro* permite compreender a tensão entre o uso espontâneo das línguas e as tentativas de regulá-las por meio de planificação. Fenômenos *in vivo* nascem do uso cotidiano, das práticas comunicativas e da criatividade coletiva dos falantes, refletindo a vitalidade e a adaptabilidade das línguas diante das transformações sociais. Já os fenômenos *in vitro* resultam de processos planejados, conduzidos por instituições ou pesquisadores que observam e descrevem o funcionamento linguístico a partir de uma perspectiva externa. Essa distinção teórica se articula com a reflexão de Oliveira e Altenhofen (2009, p. 211), para quem o desafio da política linguística no Brasil consiste justamente em criar pontes entre essas duas dimensões – reconhecer o dinamismo comunitário das práticas *in vivo* sem negligenciar a importância dos instrumentos *in vitro* de documentação, ensino e salvaguarda. No caso do Hunsrückisch, isso significa reconhecer que o processo de padronização não deve ser apenas um exercício técnico de normatização, mas uma construção coletiva, enraizada na vitalidade cotidiana da língua. A escrita espontânea nas redes sociais, os modos locais de nomear a língua e as diferentes formas de interação bilíngue exemplificam esse caráter *in vivo* da diversidade linguística, que convive e dialoga com esforços *in vitro* de registro e ensino – ambos necessários, mas em constante tensão produtiva.

De acordo com Limberger (2025, p. 491), o contato entre o Hunsrückisch e o português brasileiro manifesta-se em processos cognitivos e metalinguísticos que evidenciam a complexidade do repertório bilíngue dos falantes. Seus estudos demonstram que o conhecimento do Hunsrückisch pode favorecer o reconhecimento de palavras e a leitura em português e em alemão-padrão, indicando que há uma inter-relação entre as representações lexicais das línguas no cérebro e o chamado efeito cognato. Essa constatação reforça a ideia de que o *code-mixing* e os empréstimos lexicais configuram estratégias de adaptação e flexibilidade que mantêm o sistema vivo e funcional.

Além de representar a herança linguística e cultural de seus falantes, o Hunsrückisch passou a constituir uma identidade própria, moldada pela convivência entre diferentes tradições dialetais germânicas e pelo intenso contato com o português brasileiro e outras línguas de imigração. O termo *Brasildeutsch* foi proposto por Heye (1981) para designar, de forma abrangente, as variedades de alemão faladas no Brasil, considerando-as como uma suposta variedade comum resultante da imigração germânica. Contudo, Altenhofen e Morello (2018, p. 24) observam que essa denominação constitui uma classificação *in vitro*, isto é, criada a partir de uma perspectiva externa e teórica, sem correspondência direta no uso real das comunidades.

Mais além, Oliveira e Altenhofen (2009, p. 188) argumentam que termos como *Brasildeutsch* “são resultado da decisão individual (*in vitro*) de pesquisadores dessas línguas, portanto sem intervenção de membros diretos da comunidade de fala, isto é, configuraram-se como termos técnicos atribuídos artificialmente de fora para dentro do contexto de uso da língua”. Em contraste, denominações *in vivo* – como *Hunsrückisch*, *Hunsriqueano* ou *Hunsrick*, – refletem autodenominações efetivamente utilizadas pelos falantes e, portanto, possuem maior legitimidade sociolinguística e valor identitário. Por não designar um sistema uniforme, o *Hunsrückisch* se manifesta de maneiras diversas nas comunidades em que foi preservado, variando conforme o contexto histórico, o grau de contato interétnico e a política linguística local. Como destaca Habel (2018, p. 321), as autodenominações dos falantes – *Deitsch*, *Platt*, *Dialekt*, *Hunsrickisch*, *Hunsbucklisch*, *Heckedeitsch*, *Heckesprach*, *Lajeademeeessisch*, *Batata-Deitsch*, *verbrochne Deutsch*, *normale Deutsch*, *unser Deitsch*, alemão do mato, alemão batata, alemão misturado, entre tantas outras – expressam tanto a vitalidade quanto as tensões simbólicas dessa língua em um espaço multilíngue. A própria noção de “*Hunsrückisch*” é, portanto, uma abstração sociolinguística (Altenhofen; Morello, 2018, p. 23), resultado de múltiplas histórias de contato e de processos de *koineização* entre dialetos germânicos como o pomerano, o bávaro, o suábio, o tirolês, o suíço e o boêmio, bem como com outras línguas de imigração, como o italiano e o polonês, por exemplo.

A *koineização*, conforme explica Riehl (2009, p. 138), é o processo pelo qual diferentes dialetos, ao entrarem em contato prolongado, passam por um nivelamento e simplificação de suas formas, dando origem a uma nova variedade comum. Em outras palavras, é como se, do encontro entre múltiplas maneiras de falar – cada uma com suas peculiaridades regionais –, se formasse uma “língua média”, capaz de ser compreendida por todos os grupos envolvidos, preservando traços de cada origem, mas com maior uniformidade estrutural para fins de comunicação.

Nesse sentido, o Inventário Nacional da Diversidade Linguística (Altenhofen; Morello, 2018, p. 129) reconhece que as denominações *Hunsrückisch*, *alemão daqui*, *Hochdeutsch* e *Platt(deutsch)* podem remeter ao mesmo objeto de referência, sublinhando que a autodenominação comunitária é elemento constitutivo da política de reconhecimento e da própria vitalidade da língua. Essa diversidade denominal e estrutural reflete o percurso histórico de adaptação das comunidades teuto-brasileiras e confirma que o *Hunsrückisch* não se define pela fixidez, mas pela sua capacidade de renovação linguística. Desse modo, os empréstimos lexicais, o *code-mixing* e as práticas de escrita híbrida não devem ser vistos como desvios, mas como formas de atualização da língua que emergem das condições reais de bi/multilinguismo e de suas diferentes histórias de contato.

A frequência e o domínio do uso do alemão variam conforme a geração, o grau de escolaridade e as redes de convívio dos falantes: enquanto alguns grupos mantêm o idioma como língua principal nas interações cotidianas, outros o utilizam de modo mais restrito, alternando-o com o português conforme o contexto comunicativo. É notável que, em várias localidades do Rio Grande do Sul, o *Hunsrückisch* ainda desempenha o papel de língua principal nas interações familiares e comunitárias (Vilela, 2004).

A diglossia também permite compreender, de forma generalizada, como o *Hunsrückisch* pode ser marginalizado em relação ao português, devido à hierarquia linguística estabelecida e às atitudes linguísticas associadas à língua de imigração. Como observado por Ferguson (1972, p. 234), em situações diglósicas, a variedade “alta” tende a ser percebida

como mais prestigiada e útil, enquanto a variedade “baixa” é relegada a contextos informais. Essa ideia também evidencia a visão de uma “língua de colonos” ou até a denominação “alemão batata” que acaba sendo internalizado e utilizado pelos próprios falantes. Parte considerável dos empréstimos lexicais do português no Hunsrückisch provém de termos ligados à “roça” e à lida cotidiana da agricultura familiar. Contudo, esse processo está longe de se restringir ao universo rural. Como veremos adiante, a língua recorre a diferentes tipos de empréstimos para se manter atualizada à medida que a sociedade se transforma e surgem novos objetos, práticas e esferas de interação que demandam nomeação.

A política linguística monoglôssica instaurada historicamente pelo Estado brasileiro consolidou o português como língua dominante, atribuindo-lhe o estatuto de variedade legítima e institucional. Como observa Altenhofen (2013, p. 94), o critério central para definir o que é “minoritário” não é o número de falantes, mas o status político das línguas – e, nesse sentido, o português foi erigido como língua “alta”, símbolo da unidade nacional. Essa assimetria, entretanto, não se traduz em simples subordinação. A convivência entre o português e o Hunsrückisch opera em uma zona de contato onde o empréstimo lexical emerge como prática de mediação simbólica. Calvet (2007, p. 65) lembra que o empréstimo é uma das formas mais vitais de adaptação das línguas, expressão da “gestão *in vivo*” que os próprios falantes exercem sobre seus repertórios. Nessa perspectiva, os empréstimos não configuram a submissão da variedade minoritária à dominante, mas a reconfiguração constante de seus limites de significação – uma forma de autopoiese linguística pela qual o Hunsrückisch absorve, traduz e reinterpreta o português, preservando-se ao modificar-se.

O contato linguístico entre o Hunsrückisch e o português brasileiro, além de outras línguas de imigração e de fronteira, desempenhou um papel determinante na trajetória sociolinguística dessa língua de imigração. Nas colônias germânicas do Sul, o Hunsrückisch (e outras variedades do alemão) predominava nas interações familiares, nas redes comunitárias rurais e mesmo em instituições locais, como escolas e igrejas étnicas (Gregory, 2013, p. 22), enquanto o português se vinculava progressivamente às instâncias estatais – escola pública, imprensa nacionalizada e administração –, sobretudo a partir das campanhas de nacionalização da Era Vargas. O bilinguismo resultante desse convívio produziu influências recíprocas entre as duas línguas, ainda que de naturezas distintas. O português exerceu forte influência estrutural e lexical sobre o Hunsrückisch, introduzindo novos vocábulos e adaptações sintáticas; já o Hunsrückisch deixou marcas mais pontuais no português regional do Sul do país, especialmente em termos ligados à culinária, às festas e à vida comunitária – como nomes de pratos, bebidas e expressões típicas. Fishman (1989) observa que o deslocamento linguístico, em contextos de contato desigual, tende a ocorrer quando as funções sociais e culturais de uma língua minoritária são gradualmente absorvidas pela dominante, em consonância com as mudanças nas relações de poder e na organização social. No entanto, no caso do Hunsrückisch, o processo descrito por Fishman (1989) não se cumpre como simples substituição. O contato com o português produz uma zona de fricção e criação, em que a língua minoritária metaboliza o léxico dominante – reelaborando-o segundo sua própria lógica fonética, sintática e afetiva. Os empréstimos, assim, funcionam como operações de tradução interna: palavras portuguesas passam a circular com outros acentos, outras valências, outro pertencimento. Nessa dinâmica, o Hunsrückisch transforma o português em matéria de continuidade, reinventando-se no próprio gesto de incorporar o alheio e apropriar-se dele.

Durante o Estado Novo (1937–1945), as línguas de imigração—entre elas o Hunsrückisch—tornaram-se alvo direto da política de nacionalização conduzida pelo governo Vargas. O regime, ao buscar consolidar uma identidade nacional monolíngue, impôs o português como única língua legítima de expressão e puniu severamente o uso público das demais (Imagem 1 e 2). Como documenta Oliveira (2009, p. 22), a repressão atingiu com especial intensidade as regiões coloniais do Rio Grande do Sul e de Santa Catarina, onde a estrutura minifundiária e a colonização homogênea haviam garantido, por décadas, a reprodução do alemão e do italiano como línguas de uso cotidiano. Nesses territórios, o Estado Novo transformou a fala em campo de suspeita, em que escolas comunitárias foram ocupadas, jornais e gráficas foram fechados, e cidadãos brasileiros foram presos e até torturados sob a acusação de “crime idiomático”. O impacto foi devastador nesse período—as línguas de imigração perderam seu espaço escrito e urbano, recuando para a oralidade doméstica e para o meio rural, em contextos comunicativos cada vez mais restritos.

Aprofundando-se nesse contexto histórico, Weber (2020, p. 32) evidencia que medidas como o Decreto-Lei nº 406/1938 e a atuação do Departamento de Imprensa e Propaganda (DIP) institucionalizaram a censura e a vigilância sobre o uso das línguas estrangeiras, legitimando juridicamente a política de nacionalização. Esse conjunto de ações produziu um silenciamento linguístico duradouro. No caso do Hunsrückisch, essa ruptura explica a perda de suas práticas escritas e a dificuldade posterior de consolidação de um léxico ortográfico, como demonstra Limberger (2021), ao relacionar a ausência de alfabetização bilíngue à exclusão histórica da língua dos espaços formais de ensino e letramento.

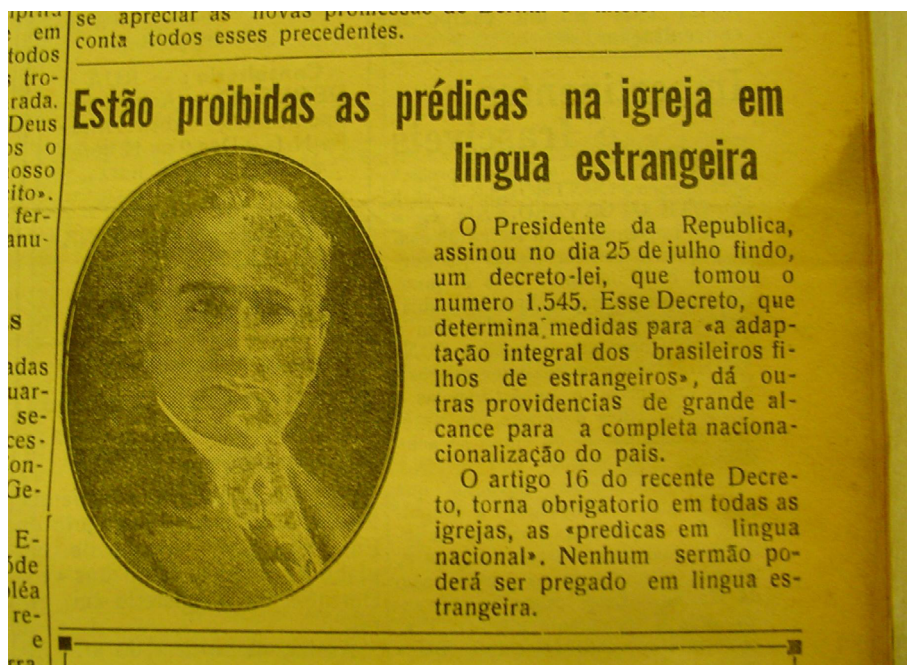
Limberger (2021) realizou um experimento psicocognitivo para investigar como falantes bilíngues constroem o léxico ortográfico do tipo *Deitsch* do Hunsrückisch, isto é, como reconhecem visualmente palavras escritas em uma variedade essencialmente oral. A oposição *Deitsch/Deutsch* da mesma variedade do Hunsrückisch seria para diferenciar uma variedade mais próxima do alemão padrão (tipo *Deutsch*) e outra mais dialetal (tipo *Deitsch*). O estudo envolveu três grupos de participantes: o primeiro grupo (GHA) reuniu falantes de *Deitsch* com proficiência avançada em alemão-padrão (*Hochdeutsch*); o segundo (GH) foi composto por falantes de *Deitsch* que não possuíam conhecimento significativo de alemão-padrão (nível máximo A1); e o terceiro (GA), usado como grupo de controle, incluiu falantes de português e alemão-padrão sem contato com línguas minoritárias germânicas. A tarefa experimental consistiu em uma decisão lexical, na qual os participantes deveriam identificar se as palavras exibidas numa tela “existiam” em sua língua. Foram utilizadas três condições de estímulos: (a) HR-AS, palavras cognatas ao alemão-padrão, (b) HR, palavras típicas do *Deitsch*, e (c) PseudoHR, pseudopalavras criadas conforme as convenções propostas por Altenhofen *et al.* (2007). Os resultados mostraram que os falantes bilíngues com conhecimento de alemão-padrão (GHA) apresentaram maior acurácia e menor tempo de resposta, o que sugere transferência das regras grafema-fonema do alemão-padrão para o *Deitsch*. O grupo GH também demonstrou aprendizagem ao longo das quatro rodadas, indicando a possibilidade de construção de um léxico ortográfico mesmo sem instrução formal. Já o grupo GA tratou as palavras em *Deitsch* como pseudopalavras, evidenciando ausência de representações ortográficas prévias. O estudo confirma que o processamento da leitura em línguas minoritárias pode ocorrer de modo eficiente entre adultos alfabetizados, desde que haja exposição sistemática e oportunidades de prática, reforçando a importância de iniciativas educativas voltadas à escrita e leitura do *Deitsch*.

Imagem 1 – Aviso em Repartição Central de Polícia de São Lourenço do Sul (RS), março de 1942



Fonte: Foto Acervo de Edilberto Luiz Hammes. Publicado em “Folha Pomerana” N° 231, 2018 – 17 de março de 2018. (Disponível em: <https://jornalggn.com.br/historia/e-proibido-falar-italiano-alemao-e-japones/>. Acesso em: 17 abr. 2026)

Imagem 2 – Notícia em jornal de Santa Catarina sobre o Decreto de proibição do uso das línguas de imigração



Fonte: Jornal Diário da Tarde. Florianópolis-SC, Ano V, nº 1.221, datado de 31 de agosto de 1939.

A repressão às línguas de imigração durante o Estado Novo representou um marco na história sociolinguística brasileira, configurando um cenário em que o português foi imposto como única língua legítima no espaço público e institucional. Esse contexto histórico, contudo, não impediu que houvesse resistência. Gälzer e Magalhães (2024, p. 9) mostram que, embora o Estado Novo tenha imposto severas restrições ao uso público do alemão – enquadrando-o como ameaça à unidade nacional e vinculando seu emprego à suspeita de traição à pátria – o idioma de imigração permaneceu central na construção da identidade étnica sobretudo no âmbito doméstico. As autoras evidenciam que a combinação entre legislação repressiva, vigilância policial, delações e um clima difuso de medo levou muitos descendentes de alemães a evitar o uso do idioma em espaços públicos, mas não impediu sua manutenção na família, especialmente entre os mais idosos, responsáveis pela transmissão intergeracional da língua. Nesse contexto, o alemão passa a ser renegociado como marcador de uma “brasileiridade étnica”, em que os sujeitos se afirmam brasileiros, adotam o português como língua de visibilidade nacional e adequação cívica, mas preservam o alemão como patrimônio afetivo, cultural e religioso. Ao analisar relatos de fiéis, pastores e moradores do Vale do Rio dos Sinos, Gälzer e Magalhães (2024, p. 13) demonstram que a política linguística estatal também produziu formas complexas de resistência e acomodação, como o uso seletivo das línguas conforme o interlocutor e o espaço, a tradução de prédicas, o esconderijo de livros em alemão e a persistência de práticas linguísticas que desafiavam, de maneira discreta, o ideal monolíngue projetado pelo Estado.

Altenhofen e Morello (2018, p. 133-134) observam que a continuidade do Hunsrückisch está intrinsecamente ligada à presença de interlocutores falantes no ambiente familiar. O levantamento realizado revela que, em casamentos mistos, onde um dos cônjuges não fala a língua, é costume que o português assuma o papel de “língua comum”, reduzindo o espaço do Hunsrückisch no cotidiano. Mesmo em lares onde ambos os cônjuges falam a língua, isso não garante que ela será transmitida às gerações seguintes. Dados como os apresentados na Tabela 1 mostram um declínio significativo na transmissão intergeracional: enquanto 260 respondentes afirmam usar o Hunsrückisch com o cônjuge, apenas 191 mencionam o uso com os filhos, e esse número cai drasticamente para 32 no caso dos netos.

Tabela 1 – Uso do Hunsrückisch nos lares

Língua respondida	SC06	SC03	SC01	SC02	ES01	ES02	Total
Cônjuge	88	63	53	18	29	9	260
Filhos	74	35	43	8	23	8	191
Netos	13	5	10	1	2	1	32

Fonte: Altenhofen; Morello (2018, p. 133).

Não analisaremos o perfil dos respondentes, apenas para contextualizar ao leitor, as amostras foram coletadas em comunidades falantes de línguas de imigração em nos estados de Santa Catarina (SC01, SC02, SC03, SC06) e do Espírito Santo (ES01, ES02). Os entrevistados contemplam diferentes faixas etárias, gênero, grau de estudo e domínio da língua. Nossa análise, com base na Tabela 1, recai sobre como esse padrão de perda geracional é agravado por transformações sociais e culturais que afastam o processo de formação linguística do núcleo familiar.

Em comunidades rurais, onde as interações sociais são mais restritas a círculos familiares e comunitários, o uso da língua tende a ser mais preservado. Contudo, em áreas urbanas, onde as oportunidades de interação ocorrem em contextos mais amplos e diversificados, o português brasileiro predomina, dificultando a transmissão intergeracional do Hunsrückisch. Pesquisas realizadas como parte do Inventário Nacional da Diversidade Linguística (INDL) revelaram que a língua ainda desempenha um papel importante como marcador identitário em algumas comunidades. No entanto, atitudes linguísticas negativas em relação ao Hunsrückisch, frequentemente visto como uma “língua quebrada” ou “língua de mato”, contribuem para estigmas que favorecem a língua padrão e reforçam a hegemonia do português. Esses fatores, aliados à falta de políticas linguísticas eficazes para sua salvaguarda, resultam em uma diminuição de sua vitalidade linguística (Altenhofen; Morello, 2018, p. 33).

A iniciativa do Inventário do Hunsrückisch, promovida pelo Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN), em parceria com o Instituto de Investigação e Desenvolvimento em Política Linguística (IPOL), é uma tentativa de documentar, valorizar e proteger essa língua como um bem cultural imaterial do Brasil. Por meio de pesquisas de campo e coleta de dados, o inventário busca compreender a diversidade interna da língua, sua extensão geográfica e as práticas linguísticas de seus falantes. O reconhecimento do Hunsrückisch como uma referência cultural brasileira pelo INDL tem sido fundamental para reposicionar a língua no cenário sociopolítico nacional, promovendo sua inclusão em debates sobre políticas de preservação linguística e cultural (Altenhofen; Morello, 2018, p. 97).

3 Histórico do Hunsrückisch no Brasil e o contato linguístico com o português

O contato linguístico emerge quando grupos com repertórios distintos passam a interagir de forma continuada – quase sempre em cenários atravessados por assimetrias históricas como guerras, colonialismo e migrações (Thomason, 2001). Os efeitos desse convívio variam conforme intensidade e duração do contato, tipo de relação social, econômica e política, funções comunicativas envolvidas e grau de parentesco estrutural entre as línguas. Em muitos contextos, minorias linguísticas perdem falantes ou domínios de uso; em outros, reconfiguram práticas e valores para sustentar a circulação da língua.

As mudanças decorrentes do contato podem ocorrer em todos os níveis. No plano fonético, bi/multilíngues frequentemente assimilam sons da L2 às categorias da L1; no gramatical, elementos de uma língua são incorporados e acomodados à morfossintaxe da outra (por exemplo, atribuição de gênero em empréstimos em variedades do alemão da diáspora). Na sintaxe, padrões de ordem e concordância podem ser reinterpretados com base na língua dominante; no léxico, surgem empréstimos diretos, *loanshifts* (ampliação semântica), *calques* (traduções literais como *arranha-céu* do inglês *skyscraper*) e híbridos morfológicos (p. ex., compostos teuto-lusos como *Milhebrot*). Esses processos são estratégias de integração que estabilizam a comunicação e ampliam repertórios (Edwards, 2004; Nübling *et al.*, 2017).

Já o *code-switching* envolve a mescla de materiais de dois códigos na mesma conversa – entre sentenças (inter) ou dentro da sentença (intra) – e pode incluir *tag/discourse-marker switching* com funções discursivas específicas (Thomason, 2001; Muysken, 2007). Já a *code*

alternation refere-se a escolhas separadas por domínios: uma língua em casa, outra no trabalho. Tudo isso se apoia numa visão não binária do bi/multilinguismo: não se exige proficiência idêntica, mas um repertório distribuído por domínios, com graus, funções, alternância e interferência variáveis (Riehl, 2014). O contato gera tanto mudança estrutural quanto ecologias de uso, e é nessa zona de fricção – não no laboratório da língua ideal – que as comunidades constroem continuidades.

No caso do Hunsrückisch no Brasil, o contato prolongado com o português produziu exatamente o pacote de efeitos descritos pela literatura. Já são mais de duzentos anos desde a chegada do Hunsrückisch ao Brasil. Em 1824, o governo brasileiro, interessado em promover a ocupação e o desenvolvimento agrícola de terras no sul do país, deu início a um programa de incentivo à imigração, especialmente de europeus. Os primeiros grupos de colonos alemães, oriundos da região do Hunsrück, na Renânia Central, fixaram-se na então recém-fundada colônia de São Leopoldo, no Rio Grande do Sul. Posteriormente, outros grupos se estabeleceram em Santa Catarina, com destaque para a colônia de São Pedro de Alcântara, fundada em 1829 (Altenhofen; Morello, 2018).

Como uma variedade do alemão médio, o Hunsrückisch rapidamente se consolidou como a “língua de casa” (*Deutsch von dehemm*) e foi usado amplamente no comércio local e nas interações comunitárias (Altenhofen; Morello, 2018, p. 141). A geografia isolada de muitas colônias, aliada à forte coesão social entre os imigrantes, contribuiu para a preservação da língua. Entretanto, os falantes também precisaram interagir com o português, que desempenhava um papel essencial nas interações institucionais e na educação. Essa convivência fomentou um bilinguismo dinâmico, caracterizado pela constante negociação entre os dois sistemas linguísticos.

Com o passar do tempo, a comunidade de falantes enfrentou desafios ainda mais significativos, especialmente durante a Era Vargas (1930-1945). Entre 1937 e 1945, a campanha de nacionalização empreendida pelo Estado Novo restringiu de modo sistemático as expressões públicas de germanidade no Brasil. A partir de 1938, registram-se a imposição do português na escolarização, o declínio da imprensa em língua alemã e a redução de sociabilidades associativas, com o consequente desestímulo ao uso público do alemão e de suas variedades (entre elas o Hunsrückisch) (Oliveira, 2009, p. 23; Weber, 2020, p. 30). Como sintetizam Gregory (2013, p. 25) e Oliveira (2009, p. 22), esse processo resultou na retração da língua para o âmbito doméstico, no arrefecimento de atividades de clubes e sociedades e na produção de um ambiente de vigilância e silenciamento, com efeitos diretos sobre a transmissão intergeracional e as práticas letradas.

Fishman (1989) destaca a extrema vulnerabilidade das línguas minoritárias em cenários marcados por repressão cultural e processos de assimilação forçada, onde essas línguas frequentemente perdem espaço e funcionalidade em favor de línguas dominantes. Ele observa que a manutenção de uma língua e de uma identidade étnica requer apoio institucional forte e uma clara compartimentalização de funções linguísticas e culturais. Sem essa estrutura, as línguas minoritárias tendem a ser rapidamente absorvidas, levando a transições linguísticas e culturais que resultam em assimilação. A ausência de complementação e distribuição funcional entre as línguas e culturas, como exemplificado por contextos em que o bilinguismo é transitório, expõe ainda mais as línguas minoritárias à perda de relevância e continuidade, reforçando a necessidade de políticas de preservação linguística e cultural.

Apesar das adversidades, a língua sobreviveu, embora com diferentes graus de vitalidade dependendo da região e do engajamento das comunidades locais. A redemocratização do Brasil e a crescente valorização da diversidade cultural nas últimas décadas possibilitaram um novo olhar sobre o Hunsrückisch. A política linguística de cooficialização de línguas a nível municipal possibilitou novos caminhos para a preservação das línguas nativas e de imigração no país, bem como projetos como o Inventário do Hunsrückisch, promovido pelo IPHAN, em parceria com o IPOL, foram fundamentais para documentar a língua e destacar sua importância como patrimônio cultural imaterial. Sob diferentes alcunhas, seja Hunsrückisch, Hunsrik ou Hunsrick, a língua é cooficial ao lado da língua portuguesa em Antônio Carlos (SC), Ipumirim (SC), Barão (RS), Horizontina (RS), Santa Maria do Herval (RS) e Harmonia (RS). Além disso, esforços acadêmicos, como os conduzidos pelo projeto ALMA-H, e os centros de pesquisa de diferentes universidades nacionais e internacionais forneceram bases sólidas para a compreensão do impacto histórico e social do Hunsrückisch no Brasil.

Romano (2020, p. 20 e 34) destaca que os estudos de geolinguística no Brasil evoluíram consideravelmente desde seus primórdios, especialmente a partir do Projeto ALiB (Atlas Linguístico do Brasil). Este projeto introduziu uma abordagem metodológica unificada para coletar dados sobre a variação linguística em nível nacional. O movimento de mapeamento linguístico expandiu-se para incluir atlas regionais e de pequeno domínio, os quais documentam aspectos específicos da língua portuguesa em diferentes regiões do Brasil. Esses esforços resultaram em contribuições valiosas, como a cartografia linguística detalhada e a introdução de novas dimensões, como variáveis sociolinguísticas e contextos de contato linguístico, incluindo comunidades bilíngues e multilíngues. Tais iniciativas não apenas aprofundaram a compreensão da diversidade linguística no Brasil, mas também serviram de referência metodológica para atlas e projetos de documentação subsequentes, que passaram a contemplar línguas de imigração e minoritárias, contribuindo, assim, para o ensino e a valorização de variedades como o Hunsrückisch.

Como destaca Severo (2025, p. 20), a vitalidade de uma língua não pode ser reduzida à sua presença estatística, mas deve ser compreendida a partir das práticas afetivas, culturais e educativas que lhe conferem sentido. A autora propõe repensar a transmissão intergeracional como um processo expandido, que ultrapassa o domínio familiar e envolve novos agentes – escolas, associações culturais, universidades, igrejas e coletivos locais – capazes de reativar a língua em contextos plurais e contemporâneos. No caso das línguas de imigração, como o Hunsrückisch, esse deslocamento da transmissão doméstica para a comunitária redefine as formas de continuidade linguística: as línguas deixam de ser apenas heranças para se tornarem projetos compartilhados de futuro. A partir dessa perspectiva, o Hunsrückisch é reconfigurado por práticas agentivas e criativas que, ao mesmo tempo em que reafirmam laços identitários, inserem a língua em novos circuitos culturais e educativos. Assim, entende-se a importância das pesquisas acadêmicas não se distanciarem do uso real e cotidiano das comunidades que falam as línguas minoritárias.

Reconhecer e valorizar uma língua não é apenas um ato de justiça linguístico-cultural, mas também uma oportunidade de promover uma educação plurilíngue que respeite e celebre a multiplicidade de vozes que compõem o patrimônio linguístico do Brasil. Entretanto, cabe mencionar que ao passo que as políticas linguísticas de cooficialização de línguas a nível municipal ganham espaço, diferentes concepções sobre como preservar as línguas emergem

e a falta de coesão nos esforços de preservação de uma língua podem criar correntes contrárias que mais dificultam do que contribuem para sua revitalização.

Mais além, o bi/multilinguismo emergente das comunidades de imigração alemã no Brasil resultou em fenômenos linguísticos complexos. Weinreich, Labov e Herzog (2006, p. 95) destacam que o contato prolongado entre línguas frequentemente transforma o repertório lexical e influencia profundamente as relações culturais e sociais. Como apontam os autores, “o estudo de línguas em contato confirma a ideia de que a coexistência estável de longo prazo é muito frequentemente uma ilusão, promovida talvez pela existência de um léxico e de uma morfofonêmica relativamente estáveis (ou até dessemelhantes) (...)”. No caso do Hunsrückisch, esses fenômenos evidenciam a convivência dinâmica entre o português e o alemão, marcando não apenas o uso da língua, mas também sua adaptação às demandas culturais e linguísticas do contexto brasileiro.

Dentre esses fenômenos, o *code-switching*, ou alternância de códigos, é um dos mais marcantes e amplamente estudados. Este processo reflete tanto as adaptações linguísticas necessárias para a convivência das línguas em contextos bi/multilíngues quanto as dinâmicas culturais e sociais subjacentes às práticas linguísticas dos falantes. Gumperz (1982) defende que a alternância de códigos em contextos bilíngues reflete contrastes culturais e identitários, com a língua minoritária funcionando simbolicamente como o “código nós” em atividades informais, enquanto a língua majoritária é o “código eles”, associada a interações formais. No entanto, essas associações não determinam necessariamente o uso real, que depende de contextos discursivos e pressupostos sociais. Diferenciar a alternância de códigos dos empréstimos linguísticos é fundamental, ainda que essa distinção nem sempre se manifeste de forma abrupta. Em termos gerais, os empréstimos tendem a ser incorporados ao sistema gramatical e fonológico de uma língua, tornando-se elementos relativamente estáveis de seu léxico. Já a alternância de códigos envolve a justaposição de estruturas pertencentes a dois sistemas linguísticos distintos, normalmente sem adaptação morfológica imediata.

De acordo com Bullock e Toribio (2009, p. 2), o *code-switching* refere-se à capacidade dos falantes bilíngues de alternarem fluentemente entre suas línguas, motivados por uma variedade de fatores, como suprir lacunas linguísticas, expressar identidade cultural e étnica, ou alcançar objetivos comunicativos específicos. Em Riehl (2009), a alternância implica a co-ocorrência de dois sistemas gramaticais em uma mesma sequência, preservando suas regras, ao passo que Matras (2020, p. 107) descreve um contínuo funcional-estrutural no qual *switches* podem ser “alternacionais” (entre sentenças/orações) ou “insercionais” (itens encaixados em uma língua-base), frequentemente com metas distintas. Nessa chave, o fenômeno não é binário: ele oscila entre escolhas localmente motivadas e graus variados de integração, com classes mais propensas à inserção (sobretudo substantivos) e outras mais restritas (verbos, por ancorarem a predicação e a língua-base). Além disso, certos operadores – conjunções, partículas, marcadores – são particularmente vulneráveis e podem generalizar-se por economia cognitiva, levando à “fusão” de inventários descrita por Matras (2020, p. 147), em que um mesmo conjunto de modificadores passa a operar independentemente da língua ativada.

Já o *code-mixing* – ou mistura de códigos – designa situações em que elementos de duas ou mais línguas se combinam dentro de uma mesma estrutura linguística, de modo que já não é possível identificar fronteiras nítidas entre os sistemas em contato. Diferentemente do *code-switching*, que pressupõe alternância entre códigos distintos, o *code-mixing* envolve a integração de unidades lexicais, morfológicas e até sintáticas de línguas diversas em um

mesmo enunciado, frequentemente de forma inconsciente ou automatizada. Segundo Riehl (2009, p. 141), esse fenômeno é particularmente recorrente em comunidades bilíngues estáveis, nas quais as línguas coexistem de maneira funcional e simbólica, permitindo a criação de repertórios híbridos que expressam identidades locais e estratégias comunicativas próprias.

Matras (2020, p. 77) observa que o *code-mixing* não deve ser entendido como um sinal de “contaminação linguística”, mas como resultado natural de uma competência multilíngue ativa, na qual os falantes mobilizam recursos de diferentes sistemas para construir sentido. Essa mistura reflete o que ele chama de *linguistic fluidity* – a fluidez característica de contextos de contato prolongado, em que as fronteiras entre línguas se tornam permeáveis e a criatividade linguística emerge como prática cotidiana. No caso do Hunsrückisch, por exemplo, o *code-mixing* manifesta-se na criação de formas híbridas e na incorporação de elementos do português brasileiro, fenômenos que não indicam perda, mas adaptação e vitalidade dentro do repertório linguístico.

Riehl (2009, p. 21), percebe que a diferença entre o *code-switching* ou o *code-mixing* do empréstimo lexical situa-se num continuum mais do que numa dicotomia rígida. A autora explica que nem todo uso isolado de um item de outra língua configura *code-switching*: quando uma palavra aparece apenas *ad hoc* em uma enunciação – como *otpusk* em uma frase majoritariamente alemã – alguns autores classificam esse fenômeno como *nonce borrowing* ou “empréstimo pontual”, enquanto outros o incluem no escopo do *code-switching*. Do ponto de vista estritamente linguístico, tais ocorrências não se distinguem claramente de empréstimos já integrados ao léxico, pois ambos podem assumir funções morfológicas ou sintáticas da língua base. Riehl (2009, p. 22) ressalta, porém, que a diferença relevante não está apenas na estrutura, mas no status sociolinguístico que a comunidade atribui ao item – se ele já pertence ao repertório da variedade ou permanece como elemento “convidado” (*guest-word*). Como critérios formais são insuficientes, a recomendação seria em conceber o fenômeno como um gradiente cuja fronteira depende do grau de integração fonológica, morfológica e funcional do item, bem como da percepção dos próprios falantes. Nesse quadro, *code-switching* pode envolver desde palavras isoladas até sentenças completas, ao passo que o empréstimo, ainda que integrado, permanece restrito ao nível lexical.

A distinção entre *code-switching* e empréstimo lexical, embora conceitualmente clara, nem sempre se manifesta de forma abrupta nas práticas linguísticas. O sujeito bi/multilíngue decide, em primeiro lugar, qual idioma servirá de base, e, apenas em uma segunda etapa, se recorrerá à alternância de códigos – algo que ocorre preferencialmente quando o interlocutor também é bilíngue (Grosjean, 1982 apud Mackedan; Spinassé, 2022). Ziegler (1996) propõe que a diferença essencial entre *switching* e *borrowing* reside no grau de integração do elemento estrangeiro à língua base: enquanto a alternância mantém cada sistema autônomo, o empréstimo já foi fonológica e morfológicamente incorporado. Esse princípio se aplica às línguas de imigração germânicas no Brasil – como o Pomerano e o Hunsrückisch –, nas quais o português funciona simultaneamente como fonte de empréstimos e como língua de alternância. Spinassé (2016, 2017) mostra que o contato com o português gerou tanto hibridismos e criações lexicais quanto eventos de *code-switching* ocasionais, motivados por fatores identitários, temáticos ou expressivos. Assim, a variação lexical nessas comunidades resulta de um contínuo entre alternância e integração, refletindo tanto a vitalidade quanto a permeabilidade do repertório bi/multilíngue.

À luz desse contínuo, os empréstimos lexicais podem ser introduzidos como o ponto em que inserções recorrentes se estabilizam no repertório da comunidade: deixam de depender do controle inibitório típico do bilinguismo em performance e passam a apresentar regularidade de uso, integração fonológica e/ou morfossintática e distribuição como forma “padrão” para um dado conceito. Riehl (2009) destaca os *nonce borrowings* como etapa intermediária – itens ainda não plenamente lexicalizados, mas já moldados ao sistema da língua receptora –, enquanto Matras (2020, p. 150) interpreta o empréstimo como estratégia de gerenciamento de repertório: a generalização de um item, antes alternado, até tornar-se selecionável em qualquer contexto. A tabela a seguir permite-nos observar os empréstimos lexicais do português ao Hunsrückisch e sua correspondência equivalente em alemão padrão:

Tabela 2 – Empréstimos lexicais: Português-Hunsrückisch-Alemão

Português	Hunsrückisch	Alemão (Hochdeutsch)
Ameixa	Amesch	Pflaume
Amiga(o)	Amiga(o)	Freund/Freundin
Anúncio	Anunsio	Veröffentlichung
Aposentadoria	Aposentatorie	Ruhestand
Autoridade	Autoriteet	Bürgermeister
Avião	Aviong	Flugzeug
Avisar	Aviseere	Warnen
Balcão	Barkong	Theke
Bandeja	Bandeesch	Tablett
Banhado	Banhado	Sumpf
Barraca	Barack	Zelt
Barulho	Barulje	Lärm
Bicicleta	Bisiklett	Fahrrad
Bilhete	Biljett	Ticket
Cachimbo	Kaschimb	Pfeife
Caminhão	Kamiong	Transporter
Caneta	Kanett	Stift
Carreta	Karett	Lastkraftwagen (LKW)
Carroça	Kaross	Holz-Karren
Caxumba	Kaschumbe	Zeigenpeter
Churrasco	Schuraske	Grill/Grillen/Grillfleisch
Facão	Fakong	Buschmesser/Machete
Guri(s)	Guri(e)	Junge(n)
Loja	Loja	Geschäft
Macaco (ferramenta)	Makack	Wagenheber
Marreco(a)	Marreck	Ente
Mula	Mule	Maultier
Município	Munizipp	Gemeinde

Neném	Neneche	Baby
Picolé	Pikolee	Eis am Stiel
Pilha	Bilja	Batterie
Potreiro	Potrea	Weide
Puxar	Pusche	Ziehen
Quantia	Quantie	Menge
Relho	Relhe	Peitsche
Roubo	Raab	Diebstahl
Televisão	Tēlevisong	Fernsehen
Venda (pequeno mercado)	Vend	Einkaufsladen

Fonte: elaborado pelo autor com base em Boll (2021) e Langenscheidt (2010).

Calvet (2002, p. 27) observa que “o plurilinguismo faz com que as línguas estejam constantemente em contato. O lugar desses contatos pode ser o indivíduo (bi/multilíngue, ou em situação de aquisição) ou a comunidade” e o resultado desses contatos é um dos objetos centrais dos estudos sociolinguísticos. Como podemos observar com base na Tabela 2, palavras híbridas ou adaptadas do português são amplamente presentes no Hunsrückisch, refletindo o contato linguístico entre o português e o alemão ao longo da história de imigração no Brasil. Exemplos marcantes incluem adaptações como *Fakong* (de “facão”), *Kamiong* (de “caminhão”), *Kaross* (de “carroça”) e *Potrea* (de “potreiro”), que passaram por transformações fonéticas para se encaixar no sistema sonoro do Hunsrückisch.

Outras palavras foram incorporadas diretamente, como *Tēlevisong* (de “televisão”), *Aviong* (de “avião”), *Bilja* (de “pilha”) e *Pikolee* (de “picolé”), que representam conceitos inexistentes na época da imigração ou que ganharam maior utilidade quando já estavam sob maior influência do português. Essas palavras ilustram a necessidade de nomear objetos ou conceitos novos usando a língua majoritária do país. Da mesma forma, termos como *Barack* (de “barraca”), *Schuraske* (de “churrasco”), *Kaschumb* (de “caxumba”), *Autoriteet* (de “autoridade”), *Anunsio* (de “anúncio”), *Raab* (de “roubo”), *Aposentadorie* (de “aposentadoria”) e *Relhe* (de “relho”) demonstram como o vocabulário cotidiano foi influenciado pelo português. No caso de *Schuraske*, também é interessante perceber as preferências intergeracionais, pois o termo é relativamente novo e vem substituindo *Spissfleisch* (literalmente “carne de espeto”), usada pelas gerações mais velhas.

Da mesma forma, os casos de *Amiga(o)*, *Banhado* e *Loja* no Hunsrückisch não correspondem a lacunas lexicais ou à necessidade de nomear realidades inéditas no Brasil, mas sim ao efeito direto do contato intenso e cotidiano com falantes de português. São empréstimos lexicais integrados que se estabelecem não porque o dialeto carecia de termos equivalentes – afinal, o alemão dispõe de formas para “amigo”, “pântano/banhado” e “loja” –, mas porque a convivência interétnica, a mobilidade social e as redes comunicativas bilíngues favoreceram a adoção de certas palavras portuguesas como formas de maior circulação local. Uma vez incorporados, esses vocábulos passam a funcionar como parte natural da língua. Assim, o processo revela a permeabilidade sociolinguística própria de comunidades bi/multilíngues, em que o uso compartilhado e a pragmática da interação determinam o que se estabiliza no repertório hunsriqueano.

Há também fenômenos de formação híbrida e derivação. Exemplos incluem *Curie* (guri + sufixo alemão *-e* para plural), *Neneche* (neném + sufixo diminutivo alemão *-chen*) e *Miljemehl* (milho + farinha). Em *Miljemehl*, nota-se que o Hunsrückisch usa *Milje* no lugar de *Mais*, do alemão padrão, e isso também ocorre no correspondente para farinha de milho, ao que em Hunsrückisch é *Miljemehl* e em alemão padrão é *Maismehl*. Essas combinações preservam elementos das duas línguas, criando palavras únicas que refletem o contexto bi/multilíngue (Spinassé, 2017, p. 104 e 107). Spinassé (2015, n.p.) apresenta outro caso emblemático dessa hibridização: *Balkonjunge*, formado pelo lexema português “balcão” — assimilado com pronúncia alemã como *balkón* — e o substantivo alemão *Junge* (“menino”), para designar “o menino que trabalha no balcão”. Essa combinação segue a lógica morfológica do alemão, estruturada pelo princípio determinativo — o primeiro elemento (balcão) qualifica o segundo (*Junge*), que funciona como núcleo do composto. Essa organização, típica das línguas germânicas, mostra que o Hunsrückisch mantém padrões estruturais do alemão mesmo incorporando elementos lexicais do português. Esses exemplos demonstram o caráter inovador do Hunsrückisch e revelam como seus falantes criaram um repertório linguístico próprio, que preserva a herança germânica enquanto dialoga com a realidade brasileira. Cabe destacar, ainda, que os falantes de Hunsrückisch do Rio Grande do Sul adotaram o termo *Curie* (“menino”), mas também usam *Junge* como no termo *Balkonjunge*.

O Hunsrückisch pode ser compreendido, conforme Steffen e Altenhofen (2014), como parte de um vasto arquipélago linguístico germânico na América Latina — um conjunto de “ilhas” formadas por comunidades que compartilham uma herança linguística comum, mas que se desenvolveram em contextos de contato distintos. A longa separação em relação ao território de origem levou ao enfraquecimento do alemão-padrão e ao fortalecimento das variedades dialetais, que passaram a ocupar domínios antes reservados ao *Hochdeutsch*, inclusive na escrita informal. Dentro desse arquipélago, o contato contínuo com diferentes variedades do português brasileiro — e, em certos contextos, com outras línguas do multilinguismo nacional — gerou repertórios próprios, com empréstimos, calques e inovações lexicais que variam conforme cada região.

Além disso, observa-se uma significativa apropriação do português no uso de verbos no Hunsrückisch, que frequentemente são conjugados segundo as regras do alemão. Verbos como lembrar, namorar e ofender tornam-se *lembrieren*, *namorieren* e *ofendieren*, incorporando terminações alemãs típicas, como *-ieren* (Vilela, 2004, n.p.). Outros exemplos incluem *pusche* (de puxar) e *aviseere* (de avisar), que recebem adaptações semelhantes, com desinências alemãs como *-ie* e *-ere*. Esse rico repertório de palavras adaptadas, híbridas e emprestadas evidencia o dinamismo linguístico que preserva elementos culturais e linguísticos de sua matriz alemã enquanto se adapta ao contexto brasileiro.

Esses empréstimos lexicais frequentemente se ajustam aos padrões fonológicos e morfológicos do Hunsrückisch, incorporando-se ao sistema da língua por meio de adaptações de sufixos, ajustes prosódicos e assimilação de fonemas, o que os torna mais familiares aos falantes da língua de imigração. Spinassé (2017, p. 95) observa que a presença desses itens de origem portuguesa passou a ser percebida, por muitos falantes e pela sociedade ao redor, como indício de um suposto “empobrecimento” da variedade, embora essa leitura se baseie em juízos de valor e não em critérios linguísticos. A autora demonstra que nem todas as diferenças entre o Hunsrückisch e o alemão-padrão decorrem do contato com o português, já que diversos traços remontam aos dialetos germânicos de origem. Ainda assim, o prestígio

da língua majoritária e as políticas de nacionalização – especialmente a proibição do ensino e do uso público do alemão durante o Estado Novo – favoreceram a expansão de empréstimos e interferências, contribuindo para a retração do alemão no espaço público e para a consolidação do português como língua de circulação cotidiana, sobretudo em contextos urbanos.

Riehl (2009, p. 21) demonstra que nem toda inserção de elementos de outra língua constitui *code-switching*, muitas ocorrências são *Entlehnungen* (empréstimos) ou *nonce-borrowings*, isto é, itens usados *ad hoc* que não se fixaram no léxico. A diferença estrutural é decisiva: empréstimos são integrados à morfologia e à fonologia da língua que os recebe – como a adoção de sufixos, adaptações de pronúncia ou encaixe sintático –, enquanto o *code-switching* envolve a alternância entre sistemas linguísticos mantendo-se a forma original do item inserido, sem plena integração. No caso do Hunsrückisch, expressões como *alles gut?* ilustram bem esse limite. Vilela (2004) descreve que a forma não corresponde ao alemão corrente utilizado na Alemanha nos dias atuais (*wie geht's?*) nem ao uso gramatical esperado de *alles Gute* que seria “tudo de bom” como um desejo de felicidades em aniversário e não “como vai”, mas também não constitui alternância de código: trata-se de uma *Lehnprägung* influenciada pelo português *tudo bem?*, reinterpretada dentro do sistema hunsriqueano. Podemos assumir com base em Nübling *et al.* (2017, p. 184) que *Lehnprägung* é uma criação por empréstimo, adota-se a ideia de uma palavra e se “forja” um novo signo com material nativo, como *skyscraper* que virou arranha-céu em português e *Wolkenkratzer* em alemão. Desse modo, o *alles gut?* é uma transferência lexical e semântica motivada pelo contato prolongado com o português, e não uma troca efetiva de códigos. Essa distinção é fundamental para compreender por que muitos elementos portugueses presentes no Hunsrückisch – mesmo quando perceptíveis como “estranhos” ao alemão – não configuram *code-switching*, mas sim processos de integração lexical que acompanham transformações sociais e geracionais, como mostram Mackedanz e Spinassé (2022) para o pomerano.

Isso pode ser observado no Hunsrückisch com o uso de termos como *Tëlevisong* (televisão) e *Pikolee* (picolé), que mesclam a morfologia e a fonologia do alemão com léxicos emprestados do português. No exemplo citado por Vilela (2004, n.p.), a apropriação da expressão *alles gut?* ao invés de *wie geht's?* ou *alles gute* revela não apenas o impacto do português no repertório linguístico dos falantes, mas também a capacidade do Hunsrückisch de integrar novos elementos sem perder sua essência. Esses fenômenos reforçam a ideia de que a língua é um sistema vivo, em constante adaptação às necessidades sociais e culturais de seus falantes.

Calvet (2002) destaca que tanto a alternância de código quanto a mistura de línguas não são meras interferências, mas refletem significados sociais e culturais profundos. No caso do Hunsrückisch, essas práticas frequentemente marcam a dualidade identitária de seus falantes, que transitam entre a herança alemã e a vivência brasileira. A escolha de expressões como *alles gut?*, por exemplo, ancorada no modelo português *tudo bem?*, pode ser vista como uma estratégia para facilitar a comunicação em comunidades bilíngues, mas também pode reforçar estereótipos que associam o Hunsrückisch a uma língua simplificada ou “incorreta”. Contudo, como observa Calvet (2007, p. 63), a capacidade de uma língua de incorporar elementos externos e criar novos significados é um sinal de vitalidade, e não de fragilidade, pois ela precisa ser “equipada” para cumprir determinadas funções.

Altenhofen e Morello (2018, p. 29) mostram que, no Sul do Brasil, o Hunsrückisch rio-grandense passa por um processo de reconfiguração interna em que marcas típicas do francônio-moselano foram, em grande medida, substituídas por variantes do francônio-re-

nano, perceptualmente mais próximas ao Hochdeutsch. No caso do Rio Grande do Sul, os autores observam que o Hunsrückisch atual se aproxima do Pfälzisch, justamente porque formas como *das/was*, associadas ao francônio-renano e ao alemão-padrão, tendem a prevalecer sobre *dat/wat*, mais típicas do francônio-moselano (Altenhofen; Morello, 2018, p. 35). Esse quadro pode ser descrito como um deslocamento de traços em direção a variedades percebidas como mais próximas da norma standard, mas não implica, por si só, o apagamento completo de outras tradições dialetais de origem.

Prediger e Habel (2024, p. 157) mostram que, entre descendentes de imigrantes nordestinos no Rio Grande do Sul, o repertório atual não pode ser descrito como uma simples substituição de uma variedade por outra, mas como a formação de uma *intermediate variety* (“variedade intermediária”). Essa variedade resulta do encontro de quatro fontes estruturais distintas: (1) formas de contato com o português já fonologicamente adaptadas, como *fosfo/fosfa* (de “fósforo”), *isquero* (de “isqueiro”) e *batate* (de “batata”); (2) variantes próximas ao alemão-padrão, como *Kartoffel* (“batata”), *Pferd* (“cavalo”) e *fliegen/flieche* (formas concorrentes de *pflügen*, “arar”); (3) formas do Hunsrückisch sul-brasileiro que atuam como variedade convergente regional, como *Fixfeuer* (“fósforo”), *Gaul* (“cavalo”, marcado estilisticamente) e *zackre* (“lavar”, “arar”); e (4) resíduos de variantes nordestinas hoje sobretudo receptivas, como *Erdapfel* (“batata”) e *ackre* (“arar”).

O que emerge desse conjunto não é uma *variety shift* linear, mas um processo de nivelamento e recomposição em que cada item lexical se distribui segundo padrões diatópicos e geracionais. Assim, a variedade resultante não é simplesmente hunsriqueana, nem simplesmente boêmia, nem apenas mais próxima do alemão-padrão: ela é produto da sobreposição histórica de recursos linguísticos diversos que coexistem, competem e se especializam no interior das comunidades.

Por outro lado, regiões como o leste de Santa Catarina e o Espírito Santo apresentaram maior resistência a essa assimilação, mantendo uma convivência entre diferentes variedades linguísticas. Nessas áreas, configurou-se uma espécie de “arquipélago de ilhas linguísticas”, no qual grupos de diferentes origens preservaram suas variedades originais. Essa diversidade demonstra a influência de fatores sociais e geográficos no nível de substituição linguística e na manutenção das variedades locais (Altenhofen; Morello, 2018, p. 35). Fishman (2003, p. 360) destaca que a manutenção de uma língua depende de sua funcionalidade em contextos específicos (ou “domínios”), como família, religião e comunidade. No caso das regiões mencionadas, a manutenção das variedades linguísticas pode ser explicada pela preservação de domínios onde essas línguas continuaram a ser usadas, como no ambiente doméstico e em interações comunitárias. Do mesmo modo, o modelo de Labov, Herzog e Weinreich (2006) sobre o contato linguístico enfatiza que fatores sociais, como isolamento geográfico e coesão comunitária, desempenham um papel central na preservação de línguas minoritárias. As “ilhas linguísticas” descritas podem ser entendidas como exemplos concretos desse processo, em que o isolamento físico e social permitiu a manutenção dessas características linguísticas, enquanto áreas mais integradas à sociedade majoritária experimentaram maior assimilação.

Weinreich, Labov e Herzog (2006, p. 96) situam os sistemas linguísticos coexistentes como uma condição histórica para o contato, e Riehl (2014) mostra que os efeitos concretos desse contato só se realizam nas escolhas que os bilíngues fazem ao falar. A autora demonstra que, em situação de bilinguismo, as duas línguas permanecem simultaneamente ativas no cérebro, e que a produção linguística envolve constantes mecanismos de seleção e inibição

– o que ela descreve como controle atencional e processamento em modo bilíngue. É nesse jogo fino entre ativação concorrente, escolhas discursivas e demandas situacionais que surgem os fenômenos característicos do contato: inserções lexicais pontuais (*insertion*), alternâncias mais amplas entre segmentos (*alternation*), integrações morfológicas e fonológicas que resultam em empréstimos plenamente incorporados (*Entlehnungen*), e até zonas híbridas de *congruent lexicalization*, quando as duas línguas compartilham uma mesma moldura estrutural (Riehl, 2014, p. 13). Assim, mais do que a presença de duas línguas em uma comunidade, é o manejo interacional desses recursos – o modo como falantes coordenam repertórios, ajustam o código ao interlocutor, negociam sentidos e ativam estratégias conversacionais – que efetivamente gera mudança linguística e transforma os sistemas em contato.

Também se faz necessário entender que o multilinguismo é um conceito que opera em diferentes escalas, Riehl (2014, p. 15) mostra que compreendê-lo exige distinguir seus níveis de manifestação. No plano individual, trata-se do repertório de um falante que utiliza duas ou mais línguas de forma dinâmica, assimétrica e contextual, variando em proficiência, domínio funcional e trajetória de aquisição ao longo da vida. Já o multilinguismo societal refere-se à organização das línguas dentro de uma comunidade: é a ecologia que distribui funções entre diferentes variedades, estabelece hierarquias de prestígio, define domínios de uso e molda processos de manutenção ou substituição linguística. Nesse plano, atua também o *monolingualer Habitus* – o “habitus monolíngue” – que naturaliza a ideia de que a homogeneidade linguística é a norma e relega as demais línguas ao silêncio ou à marginalidade. Há ainda o multilinguismo institucional, que emerge quando escolas, órgãos públicos e políticas educacionais regulam, reconhecem ou silenciam determinadas línguas, seja por meio de modelos bilíngues e iniciativas de promoção, seja pela reprodução de práticas que privilegiam apenas a língua majoritária e reforçam esse mesmo *habitus* monolíngue. A articulação entre esses três níveis – o indivíduo, a sociedade e as instituições – é o que determina as condições reais de vitalidade, transmissão e uso das línguas em contextos contemporâneos, abrindo ou restringindo caminhos para a diversidade linguística.

Outro fenômeno observado pelo contato linguístico fica evidente em estudos da toponímia paralela na região de imigração alemã do Rio Grande do Sul. Este fenômeno caracteriza-se pela coexistência de denominações em ambas as línguas para um mesmo lugar, sendo comum que a versão em português tenha caráter oficial, enquanto a versão em alemão permaneça ativa na oralidade e na memória cultural local. Por exemplo, o nome “Novo Hamburgo”, em português, deriva de “*Neuer Hamburg*”, em alemão, evidenciando um processo de tradução e adaptação cultural (Barros; Löff-Machado, 2022, p. 148). Esse fenômeno pode ser relacionado aos conceitos de Labov (1994, 2008) sobre a manutenção e mudança linguística, uma vez que a presença de nomes paralelos reflete tanto o prestígio do português como língua dominante quanto a resiliência das tradições linguísticas alemãs locais.

Além disso, o contato entre línguas na região resultou em uma série de adaptações lexicais e semânticas que se refletem nos topônimos. A presença de elementos híbridos, como “Picada Café” (*Kaffeschneis*) e “Arroio do Tigre” (*Tigerbach*), revela um processo contínuo de integração lexical e cultural (Barros; Löff-Machado, 2022, p. 152). Segundo Fishman (1989, 2001), esse tipo de adaptação é comum em comunidades bilíngues que buscam preservar elementos de sua identidade cultural enquanto se ajustam ao ambiente linguístico dominante. Esses exemplos reforçam a ideia de que as línguas em contato não apenas influenciam umas às outras, mas também moldam o espaço geossocial e simbólico das comunidades que as utilizam.

Essas interações, quando sistemáticas, podem resultar em transformações linguísticas tanto na língua dominante quanto na língua minoritária. O bilinguismo, portanto, é um motor de mudança linguística, sendo responsável pela introdução de novos padrões, estruturas e usos dentro das línguas em contato e, comprova-se, ainda, no caso do Hunsrückisch, com empréstimos ao português, como nas palavras *cuca* (de “*Kuche*”) e *chimia* (de “*Schmier*”), evidenciando que o contato linguístico atua numa via de mão dupla (Spinassé, 2017, p. 103). Weinreich, Labov e Herzog (2006, p. 114) corroboram ao demonstrar que os fatores sociais moldam as línguas, como a relação de prestígio entre as línguas e os contextos nos quais cada uma é usada e, dessa forma, como o bilinguismo flui entre os diferentes contextos de uso para atender a essas expectativas linguísticas é utilizado por seus falantes como um recurso.

O bilinguismo reflete essas dinâmicas de prestígio e funcionalidade, adaptando-se aos diferentes domínios sociais e servindo como recurso comunicativo estratégico para os falantes. Em áreas onde as duas línguas são igualmente valorizadas e utilizadas em diferentes domínios (família, trabalho, educação), o bilinguismo tende a preservar as línguas em contato. Por outro lado, em contextos de subordinação social, onde uma língua é considerada mais prestigiosa que a outra, o bilinguismo pode levar à substituição ou marginalização da língua menos prestigiada, o que pode justificar a língua portuguesa tendo emprestado mais palavras ao Hunsrückisch do que o contrário, já que a língua “alta” é o português.

Ferguson (1972, p. 237) argumenta que a manutenção de uma língua minoritária em um contexto diglótico depende, em grande parte, do valor cultural e prestígio atribuído a essa língua e de sua transmissão intergeracional. Fishman (2003) ampliou o conceito introduzido por Ferguson (1972), associando-o a contextos multilíngues e à manutenção de línguas minoritárias. No caso do Hunsrückisch, a preservação é sustentada por práticas familiares, mas enfrenta desafios significativos, como a diminuição do número de falantes fluentes e a ausência de um ensino formal, ainda que desde o início dos anos 2000 a educação bilíngue português-Hunsrückisch esteja aumentando com base nos projetos de leis de cooficialização. A abordagem de Ferguson (1972) corrobora com a necessidade de políticas linguísticas que promovam o uso da língua minoritária em contextos variados, ampliando seu domínio funcional e assegurando sua vitalidade. Contudo, as formas intermediárias descritas na literatura – isto é, estruturas híbridas, itens lexicalmente misturados e zonas de transição entre o português e o Hunsrückisch – apontam justamente para a quebra do modelo diglótico clássico formulado por Ferguson (1972).

Além de descrever a coexistência de línguas ou variedades linguísticas com funções sociais distintas, Fishman (1964, 1965a, 1965b, 2003) aprofundou o estudo do fenômeno diglótico ao analisar sua manutenção e ruptura em níveis nacionais. Ele examinou como fatores sociais, culturais e psicológicos, como o bilinguismo composto e coordenado, influenciam a estabilidade ou a transição entre códigos linguísticos. Fishman (2003, p. 361) destacou que a diglossia pode ser encontrada tanto em sociedades multilíngues quanto em contextos em que há variações dialetais ou registros linguisticamente diferenciados. Ao integrar questões sociológicas e psicológicas, temos novas perspectivas para compreender a dinâmica diglótica, incluindo como padrões sociais governam o uso de uma variedade linguística em detrimento de outra. Seu trabalho complementa a abordagem de Ferguson (1972) ao enfatizar as interações entre as variedades em contextos maiores, oferecendo ferramentas conceituais para investigar a interação entre línguas ou registros em diferentes escalas sociais.

A influência do português sobre o Hunsrückisch já não pode ser explicada apenas pela antiga imagem de uma diglossia rígida entre uma variedade “alta” e outra “baixa”. O que se observa hoje é um ambiente de contato muito mais diversificado, no qual o Hunsrückisch interage simultaneamente com distintos repertórios locais: variedades regionais do português, o italiano teuto-serrano em áreas de colonização mista, o espanhol nas zonas de fronteira e, em algumas comunidades, formas de *Hochdeutsch* introduzidas por iniciativas religiosas ou educacionais. Nessa ecologia ampliada, o português segue exercendo forte pressão simbólica e lexical, mas essa pressão não implica uma simples hierarquia funcional. Ela se inscreve em processos de acomodação, mixing (mistura), nivelamento e inovação que caracterizam as práticas contemporâneas dos falantes. Assim, longe de se limitar a uma oposição estável entre “língua alta” e “língua baixa”, o Hunsrückisch mostra vitalidade ao reorganizar seus próprios domínios de uso e ao dialogar com uma multiplicidade de repertórios que moldam, simultaneamente, sua transformação e sua continuidade dentro do multilinguismo brasileiro. Isso é algo que viemos destacando em diferentes momentos do texto justamente para demonstrar que o Hunsrückisch possui uma força linguística e resiliência que dependem, em grande parte, apenas de políticas públicas adequadas para contribuir com sua vitalidade.

Além das transformações linguísticas, o contato entre o português e o Hunsrückisch gerou percepções variadas sobre a língua de imigração. As atitudes linguísticas, frequentemente influenciadas por fatores históricos e sociais, desempenham um papel central na manutenção ou no abandono da língua. Conforme Fishman (1989, p. 211), a valorização ou o estigma associado a uma língua está diretamente relacionado ao seu prestígio em contextos sociais específicos, evidenciando a importância das políticas linguísticas e do reconhecimento institucional.

4 Conclusão

A análise desenvolvida neste artigo permite retomar a pergunta de pesquisa e confirmar, ainda que sob a forma de um recorte, a hipótese de que o *code-mixing* e os empréstimos lexicais não constituem apenas indícios de “erosão” do Hunsrückisch frente ao português, mas operam como mecanismos de vitalidade em contextos de contato prolongado com essa língua entre outras línguas que compõem o mosaico multilíngue brasileiro. Ao reconstruir o percurso histórico da língua – da formação das colônias e do bi/multilinguismo funcional à repressão do Estado Novo e às políticas recentes de inventário e cooficialização –, mostrou-se que o repertório hunsriqueano contemporâneo não se organiza numa oposição rígida entre “alemão” e “português”, mas num contínuo em que alternância de códigos, integrações lexicais, *Lehnprägungen*, híbridos morfológicos e processos de *koineização* e formação de variedades intermediárias se entrelaçam. Os exemplos lexicais discutidos, bem como os dados grafemáticos e psicocognitivos disponíveis na literatura, sugerem que o Hunsrückisch metaboliza o português como recurso, incorporando formas, adaptando-as à sua fonologia e morfologia, redistribuindo significados e, com isso, mantendo-se funcional em domínios marcados por forte pressão do monolinguismo institucional e por ecologias multilíngues complexas.

Do ponto de vista teórico, os resultados dialogam com os modelos de contato e processamento bi/multilíngue de Weinreich, Labov e Herzog (2006), Riehl (2009, 2014) e Matras (2020), ao evidenciar que *code-switching*, *code-mixing* e empréstimos se distribuem num con-

tínuo de graus de integração, mais do que em categorias estanques. No caso do Hunsrückisch, isso se traduz na coexistência de alternâncias claramente segmentáveis, inserções pontuais ainda não plenamente lexicalizadas (*nonce borrowings*) e itens já integrados ao léxico, bem como em formas híbridas que se apoiam em padrões estruturais germânicos para acomodar material português. Em termos sociolinguísticos, o quadro descrito reforça a ideia de que a vitalidade das línguas de imigração não se mede apenas pelo número de falantes, mas pela capacidade de negociar domínios de uso, redefinir fronteiras estruturais e se reancorar em práticas *in vivo*, inclusive na escrita digital e em repertórios intermédios que escapam ao modelo diglótico clássico.

Ao mesmo tempo, a análise apresenta limites importantes. O corpus mobilizado é predominantemente lexical, documental e descritivo, ancorado em dicionários, estudos geolinguísticos, relatos históricos, registros de escrita digital e experimentos psicocognitivos já publicados, o que restringe o acesso a dados interacionais sistemáticos de fala espontânea em diferentes comunidades e gerações. Além disso, o recorte geográfico e temático privilegia certas regiões e tradições de pesquisa, sem abarcar toda a extensão territorial e toda a diversidade interna do Hunsrückisch na América do Sul, nem explorar em profundidade as dimensões etnográficas e atitudinais apenas afloradas em algumas passagens. Uma alternativa desejável seria articular, em pesquisas futuras, corpora de fala bi/multilíngue gravada em distintos contextos de uso, análises variacionistas de padrões de *code-switching* e *code-mixing*, investigações psicocognitivas comparativas com outras línguas de imigração (como o pomerano) e estudos etnográfico-discursivos que incorporem as metalinguagens dos próprios falantes sobre seus repertórios híbridos.

Apesar dessas limitações, o percurso proposto contribui para reposicionar o Hunsrückisch no debate sobre línguas de imigração e multilinguismo, deslocando o foco de uma narrativa de perda linear para uma perspectiva que enfatiza a agência dos falantes e a criatividade estrutural do contato. Ao evidenciar que *code-mixing*, *code-switching* e empréstimos lexicais podem funcionar como estratégias de preservação e renovação, o artigo sugere que políticas de documentação, patrimonialização e educação bilíngue que pretendam fortalecer o Hunsrückisch precisam reconhecer e incorporar essas práticas *in vivo*, em vez de tratá-las como “desvios” a serem corrigidos por normas externas. Mais amplamente, os resultados indicam que análises de repertórios de línguas de imigração em ecologias multilíngues contemporâneas ganham em precisão quando abandonam dicotomias rígidas entre pureza e mistura e passam a descrever, com maior fineza, os modos como as línguas se mantêm justamente ao se deixar atravessar pelo contato.

Agradecimentos

Agradeço aos professores da disciplina de Sociolinguística e Dialectologia (2024/2) da Universidade Federal de Santa Catarina (UFSC) – Izete Lehmkuhl Coelho, Valter Pereira Romano, Marco Antônio Rocha Martins e Monica Tavares Orsini. Este trabalho resulta das discussões da disciplina e do interesse pessoal do autor pela língua falada por seu avô e bisavós maternos, bem como por sua avó paterna. Agradeço também aos amigos Kaan Akbaba e Alina Reul pela contribuição na validação de palavras apresentadas na Tabela 2. Por fim, agra-

deço ao meu orientador de mestrado, Gilvan Müller de Oliveira, por tornar o Hunsrückisch um ponto central nas reflexões sobre políticas linguísticas.

Referências

- ALTENHOFEN, C. V. Bases para uma política linguística das línguas minoritárias no Brasil. In: NICOLAIDES, Christine; SILVA, Kleber Aparecido da; TILIO, Rogério; ROCHA, Claudia Hilsdorf (orgs.). *Política e Políticas Linguísticas*. Campinas: Pontes Editores, 2013. p. 93-116.
- ALTENHOFEN, C. V.; FREY, J.; KÄFER, M. L.; KLASSMANN, M. S.; NEUMANN, G. R.; SPINASSÉ, K. P. Fundamentos para uma escrita do Hunsrückisch falado no Brasil. *Revista Contingentia*, Porto Alegre, v. 2, n. 2, p. 73-87, nov. 2007. Disponível em: <https://seer.ufrgs.br/index.php/contingentia/article/view/3867>. Acesso em: 16 abr. 2026.
- ALTENHOFEN, C. V.; MORELLO, R. (orgs.). *Hunsrückisch: Inventário de uma Língua do Brasil*. Florianópolis: Editora Garapuvu, 2018.
- BULLOCK, B; TORIBIO, A. J. Themes in the study of code-switching. In: BULLOCK, B; TORIBIO, A. J. (orgs.). *The Cambridge Handbook of Linguistic Code-Switching*. Cambridge, UK, 2009. p. 01-17.
- BARROS, F. H. T.; LÖFF-MACHADO, L. O léxico toponímico da região de imigração alemã do Rio Grande do Sul: notas sobre a sua toponímia paralela. In: DORES, M.; CORDEIRO, M. (orgs.). *Estudos do Léxico: diferentes olhares e perspectivas*. Rio de Janeiro: Letra Capital, 2022. p. 134-156.
- BOLL, P. K. *Dicionário Hunsriqueano Riograndense – Português*. [s. l.], 2021. Disponível em: <https://hunsriqueanoriograndense.wordpress.com/wp-content/uploads/2021/04/dicionario_hrx-por_v3.0.pdf>. Acesso em: 28 out. 2025.
- CALVET, L.-J. *Sociolinguística: uma introdução crítica*. São Paulo: Parábola Editorial, 2002.
- CALVET, L.-J. *As Políticas Linguísticas*. São Paulo: Parábola Editorial, 2007.
- DIÁRIO DA TARDE. Florianópolis, ano 5, n. 1221, 31 ago. 1939.
- EDWARDS, J. V. Foundations of bilingualism. In: BHATIA, T. K.; RITCHIE, W. C. *The Handbook of Bilingualism*. Malden, MA: Blackwell Publishing, 2004. p. 7-31.
- FERGUSON, C. A. Diglossia. In: GIGLIOLI, P. P. (Ed.). *Language and Social Context*. Harmondsworth: Penguin Books, 1972. p. 232-251.
- FISHMAN, J. A. Language maintenance and language shift as a field of inquiry. *Linguistics*, Haia, v. 2, n. 9, 1964. p.32-70.
- FISHMAN, J. A. A Language Maintenance and Language Shift: the American Immigrant Case within a General Theoretical Perspective. *Sociologus*, Berlim, v. 16, n.1, p. 19-39, 1965a.
- FISHMAN, J. A. Who Speaks What Language to Whom and When?. *La Linguistique*, Paris, v. 2, n.1, p. 67-88, 1965b.
- FISHMAN, J. A. *Language and Ethnicity in Minority Sociolinguistic Perspective*. Clevedon: Multilingual Matters, 1989.

- FISHMAN, J. A. *Can Threatened Languages Be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective*. Clevedon: Multilingual Matters, 2001.
- FISHMAN, J. A. Bilingualism with and without diglossia; diglossia with and without bilingualism. In: PAULSTON, Christina Bratt; TUCKER, G. Richard (Eds.). *Sociolinguistics: The Essential Readings*. Malden, MA: Blackwell Publishing, 2003. p. 359-366.
- GÄLZER, F.; MAGALHÃES, M. L. “Mas para que proibir falar em alemão?”: Restrições linguísticas no Vale do Rio dos Sinos/RS durante o Estado Novo (1939-1945). *Estudos Ibero-Americanos*, Porto Alegre, v. 50, n. 1, p. 1-18, e45832, 2024. DOI: <https://doi.org/10.15448/1980-864X.2024.1.45832>
- GREGORY, V. Imigração alemã no Brasil. *Cadernos Adenauer*, [s. l.], v. XIV, edição especial, 2013, p. 9-27. Disponível em: < <http://www.kas.de/wf/doc/10985-1442-5-30.pdf> >. Acesso em: 29 out. 2025.
- GUMPERZ, J. J. *Discourse Strategies*. Cambridge: Cambridge University Press, 1982.
- HABEL, J. M. Os nomes do Hunsrückisch: aspectos linguísticos e extralinguísticos da denominação de línguas de imigração. *Entrepalavras*, Fortaleza, v. 7, n. 2, p. 314-330, ago./dez. 2017. DOI: <https://doi.org/10.22168/2237-6321.7.7.2.314-330>
- HAMMES, E. L. [Fotografia sem título]. [s. l.]: Acervo particular, 2018.
- HEYE, J. *Brasildeutsch, or Diglossia Revisited*. Paper presented to the Sixth International Congress of Applied Linguistics, Lund, 9-15 August 1981. 1981.
- LABOV, W. *Principles of Linguistic Change: Internal factors*. Cambridge: B. Blackwell, 1994.
- LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola Editorial, 2008.
- LANGENSCHIEDT. *Langenscheidt Universal-Wörterbuch Brasilianisches Portugiesisch: Portugiesisch–Deutsch / Deutsch–Portugiesisch*. Berlin; München: Langenscheidt, 2010.
- LIMBERGER, B. K. Leitura de palavras em língua minoritária: a construção do léxico ortográfico em hunsriqueano. *D.E.L.T.A.*, São Paulo, v. 37, n. 2, p. 1-26, 2021. DOI: <https://doi.org/10.1590/1678-460X202148072>
- LIMBERGER, B. K. Dialetoлогия e Psicolinguística em diálogo: estudos sobre processamento do Hunsrückisch e bilinguismo. In: STEFFEN, J.; KRUG, M. J.; LÖFF-MACHADO, L.; PAVAN, C. W.; RADÜNZ, W.; BARROS, F. H. T. (orgs.). *Sprachkontakt, Variation und Mehrsprachigkeit / Contato Linguístico, Variação e Plurilinguismo: perspectivas linguísticas sobre o trabalho acadêmico de Cléo Vilson Altenhofen*. Berlin; Bruxelles; Chennai; Lausanne; New York; Oxford: Peter Lang, 2025. p. 485-497.
- LIMBERGER, B. K.; VON MÜHLEN, F. Variação linguística e alemão como língua estrangeira: contribições a partir da análise de dois livros didáticos. *Pandaemonium Germanicum*, São Paulo, v. 22, n. 37, p. 331-356, maio/ago. 2019. DOI: <https://doi.org/10.11606/1982-88372237331>
- LINDENFELSER, S. „Wolle Ameise Pisye Hunsrik Xraywe?": zur Alltagsverschriftung des brasilianischen Hunsrückisch. *Zeitschrift für Dialektologie und Linguistik*, Stuttgart, v. 91, n. 1, p. 22-77, 2024. DOI: <https://doi.org/10.25162/zdl-2024-0002>
- MACKEDANZ, D.; SPINASSÉ, K. P. Influência da fase etária no uso do léxico em pomerano. *Organon*, Porto Alegre, v. 37, n. 73, p. 245-267, jan./jun. 2022. DOI: <https://doi.org/10.22456/2238-8915.122815>

MASELKO, M. Hunsrik xraywe: a new way in lexicography of the German language island in southern Brazil. *Dialectologia*, Barcelona, Special issue, IV, p. 147-180, nov. 2013.. Disponível em: <http://www.edicions.ub.edu/revistes/dialectologiasp2013/>. Acesso em: 23 out. 2025.

MATRAS, Y. *Language Contact*. 2. ed. Cambridge: Cambridge University Press, 2020.

MUYSKEN, P. Mixed Codes. In: AUER, P.; WIE, L. (eds). *Handbook of multilingualism and multilingual communication*. Berlin: New York: Mouton de Gruyter, 2007. p. 315-339.

NASSIF, L. É PROIBIDO FALAR italiano, alemão e japonês. *Jornal de Todos os Brasis*. 20 mar. 2018. Disponível em: <https://jornalggn.com.br/historia/e-proibido-falar-italianoalemao-e-japones/>. Acesso em: 17 abr. 2026.

NÜBLING, D.; DAMMEL, A.; DUKE, J.; SZCZEPANIAK, R. *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*. 5. Aufl. Tübingen: Narr Francke Attempto, 2017.

OLIVEIRA, G. M. Plurilinguismo no Brasil: repressão e resistência linguística. *Synergies Brésil*, São Paulo, n. 7, p. 19-26, 2009. Disponível em: <https://gerflint.fr/Base/Bresil7/gilvan.pdf>. Acesso em: 11 nov. 2025.

OLIVEIRA, G. M. From Foreign Languages to Brazilian Languages, from One-Language-One-Nation Ideology to Inclusive Co-officialization Policy: the case of Hunsrückisch and Pommersch. In: CAVALCANTI, M. C.; MAHER, T. M. (orgs.). *Multilingual Brazil: Language Resources, Identities and Ideologies in a Globalized World*. New York; London: Routledge, 2018. p. 57-68.

OLIVEIRA, G. M.; ALTENHOFEN, C. V. O *in vitro* e o *in vivo* na política da diversidade linguística do Brasil: inserção e exclusão do plurilinguismo na educação e na sociedade. In: MELLO, H.; ALTENHOFEN, C. V.; RASO, T. (orgs.). *Os contatos linguísticos no Brasil*. Belo Horizonte: Editora UFMG, 2011. p. 187-216.

PREDIGER, A.; HABEL, J. M. Die Weiterentwicklung einer intermediate variety des Deutschen bei Nachfahren nordböhmischer Herkunft in Rio Grande do Sul, Brasilien. *Linguagem & Ensino*, Pelotas, v. 27, n. 1, p. 154-170, jan./abr. 2024. DOI: <https://doi.org/10.15210/rle.v27i1.26620>

RIEHL, C. M. *Sprachkontaktforschung: Eine Einführung*. 2. überarbeitete Auflage. Tübingen: Gunter Narr Verlag, 2009.

RIEHL, C. M. *Mehrsprachigkeit: Eine Einführung*. Darmstadt: WBG – Wissenschaftliche Buchgesellschaft, 2014.

RIO GRANDE DO SUL. Gabinete Deputado Estadual João Fischer (Fixinha). *Hunsrik: língua integrante do patrimônio histórico e cultural do Estado do Rio Grande do Sul*. [s.d.]. Disponível em: <https://aguasmornas.sc.gov.br/uploads/sites/360/2022/05/1649565_Hunsrik_lingua_integrante.pdf>. Acesso em: 6 jan. 2025.

ROMANO, V. P. Desdobramentos, Desafios e Perspectivas da Geolinguística Pluridimensional no Brasil. In: MOTA, J. A.; OLIVEIRA, J. M.; PAIM, M. M. T.; RIBEIRO, S. S. C. (org.). *Contribuições de Estudos Geolinguísticos para o Português Brasileiro: uma homenagem a Suzana Cardoso*. Salvador: EDUFBA, 2020. p. 11-32.

SEVERO, C. G. Transmissão intergeracional e revitalização linguística. *Gragoatá*, Niterói, v. 30, n. 66, jan.-abr. 2025. DOI: <https://doi.org/10.22409/gragoata.v30i66.64442.pt>

SPINASSÉ, K. P. Os imigrantes alemães e seus descendentes no Brasil: a língua como fator identitário e inclusivo. *Revista Conexão Letras*, Porto Alegre, v. 3, n. 3, 2015. DOI: <https://doi.org/10.22456/2594-8962.55637>

SPINASSÉ, K. P. Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung. In: LENZ, A. N. German Abroad. *Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung*. Wien: Vienna University Press, 2016. p. 81-102.

SPINASSÉ, K. P. Contribuição do Português para a Constituição Lexical do Hunsrückisch em Situação de Contato Linguístico. *Revista Virtual de Estudos da Linguagem - ReVEL*, v. 13, n. 3, dez. 2017. DOI: <http://dx.doi.org/10.31513/linguistica.2017.v13n3a16385>

STEFFEN, J.; ALTENHOFEN, C. V. Spracharchipele des Deutschen in Lateinamerika: Dynamik der Sprachvernetzungen im mehrsprachigen Raum. *Zeitschrift für Dialektologie und Linguistik*, Stuttgart, v. 81, n. 1, p. 34-60, 2014. Disponível em: <http://www.jstor.org/stable/24548019>. Acesso em: 13 nov. 2025.

THOMASON, S. G. *Language Contact*. Edinburgh: Edinburgh University Press, 2001

VILELA, S. O alemão lusitano do Sul do Brasil. *Deutsche Welle*, 20 abr. 2004. Disponível em: <https://www.dw.com/pt-br/o-alem%C3%A3o-lusitano-do-sul-do-brasil/a-1174391>. Acesso em: 9 jan. 2025.

VON MÜHLEN, F.; KERSCH, D. F. Usos sociais da escrita do hunsriqueano no Sul do Brasil. *Revista Humanidades & Inovação*, Palmas, v. 8, n. 36, p. 326-338, 28 jun. 2021. Disponível em: <https://revista.unitins.br/index.php/humanidadesinovacao/article/view/4497>. Acesso em: 13 nov. 2025.

WEBER, A. F. O combate à imprensa em língua estrangeira no Brasil: políticas e ideias linguísticas na legislação da era Vargas. In: CAVALHEIRO, A. C. D.; MARCHESAN, A. C.; STÜBE, A. D.; HORST, C.; PAULA, L. M.; LUZ, M. N. S. (org.). *Entre as fronteiras do ensino, da pesquisa e da extensão: estudos na área de Letras* [livro eletrônico]. Chapecó: Editora UFFS, 2020. p. 25-40. DOI: <https://doi.org/10.7476/9786586545258.0002>

WEINREICH, U.; LABOV, W.; HERZOG, M. *Fundamentos Empíricos para uma Teoria da Mudança Linguística*. São Paulo: Parábola Editorial, 2006.

WIESEMANN, U. *Contribuição ao desenvolvimento de uma ortografia da língua Hunsrik falada na América do Sul*. Cuiabá, MT: Associação Internacional de Linguística – SIL Brasil, 2008. Disponível em: https://www.sil.org/system/files/reapdata/72/23/15/72231580797137648352532967849426719009/Hunsrik_49727.pdf. Acesso em: 16 nov. 2025.

ZIEGLER, A. *Deutsche Sprache in Brasilien. Untersuchungen zum Sprachwandel und zum Sprachgebrauch der deutschstämmigen Brazilianer in Rio Grande do Sul*. Essen: Ed. Die Blaue Eule, 1996.

A iconicidade no Português e sua relação com idade de aquisição e classes de palavras

Iconicity in Portuguese and its Relationship with Age of Acquisition and Parts of Speech

Thayná Cristina Ananias

Universidade Federal do Rio Grande do Norte (UFRN) | Natal | RN | BR
thayna.ananias@ufrn.br
<https://orcid.org/0000-0001-6646-501X>

Mahayana Godoy

Universidade Federal do Rio Grande do Norte (UFRN) | Natal | RN | BR
mahayana.godoy@ufrn.br
<https://orcid.org/0000-0002-7499-3290>

Resumo: A iconicidade pode ser caracterizada como uma semelhança entre certos aspectos de uma forma linguística e certos aspectos de seu sentido. Com o objetivo de investigar o papel dessa propriedade na aquisição de palavras do português, replicamos o experimento de Perry, Perlman e Lupyan (2015) e coletamos a percepção de iconicidade de 504 participantes para uma amostra de 521 palavras do português brasileiro (PB). A partir das avaliações coletadas, examinamos a relação da iconicidade e a idade de aquisição de palavras, cuja relação é perpassada por métricas como frequência e concretude. Utilizamos o corpus de palavras e a idade de aquisição para o português disponíveis no MCDI (MacArthur-Bates Communicative Developmental Inventory); recorremos à frequência de palavras disponível no site O Corpus do Português e realizamos um experimento de percepção de concretude. Por meio de análise de regressão linear, nossos resultados indicam que palavras aprendidas mais cedo também apresentam maior índice de percepção de iconicidade no PB. Ao comparar nossos resultados com os de inglês (Winter et al., 2023) e de espanhol (Hinojosa et al., 2021), encontramos que, assim como nas duas línguas, as categorias com maior nota de iconicidade no PB foram onomatopeias e interjeições. Entretanto, divergindo dos indicativos do inglês e do espanhol, os verbos, no PB, obtiveram pontuações mais altas que os substantivos, enquanto não houve diferença significativa no espanhol e, no inglês, os verbos obtiveram notas mais altas.

Palavras-chave: iconicidade; idade de aquisição; classes de palavras; português; comparação interlinguística.



Abstract: Iconicity can be defined as a resemblance between certain aspects of a linguistic form and certain aspects of its meaning. To investigate the role of this property in the acquisition of Portuguese words, we replicated the experiment conducted by Perry, Perlman, and Lupyan (2015) and collected iconicity ratings from 504 participants for a sample of 521 words. Based on these ratings, we examined the relationship between iconicity and age of acquisition, a relationship mediated by factors such as frequency and concreteness. We used the word corpus and age-of-acquisition norms for Portuguese available in the MCDI (MacArthur–Bates Communicative Development Inventory); we relied on word frequency data from O Corpus do Português, and we conducted an independent concreteness-rating experiment. Using linear regression analyses, our results indicate that words acquired earlier tend to show higher iconicity ratings in Brazilian Portuguese. When comparing our findings with those reported for English (Winter et al., 2023) and Spanish (Hinojosa et al., 2021), we found that, as in both languages, onomatopoeias and interjections received the highest iconicity scores in Portuguese. However, diverging from the patterns reported for English and Spanish, verbs in Brazilian Portuguese received higher scores than nouns; in Spanish, no significant difference was observed, whereas in English, verbs received higher ratings.

Keywords: iconicity; age of acquisition; parts of speech; Portuguese; interlinguistic comparison.

1 Introdução

Há mais de um século, encontramos pesquisas sobre a relação de fonemas e certos conceitos. Jespersen (1922), por exemplo, é uma das obras mais antigas e referenciadas sobre a associação de /i/ com pequenas coisas, crianças e animais jovens. Há também trabalhos, em diferentes línguas, sobre a associação de vogais anteriores com palavras relativas a cores claras (como o amarelo) e de vogais posteriores com cores escuras (como o azul), como Jakobson (1962 [1941]), que faz esse levantamento em alemão e tcheco, Reichard, Jakobson e Werth (1949), com sérvio e húngaro, e Masson (1951), com o inglês. Alguns anos depois, Chastaing (1964) indicou que, em francês, 90% das palavras que se referem a pequenas coisas contêm vogais anteriores (como /i/ e /e/ em *pièce* ‘moeda’, por exemplo) e 80% das palavras que se referem a grandes coisas contêm vogais posteriores (como /o/ e /u/ em *voiture* ‘carro’).

Ao longo dos anos, pesquisas voltadas para esse tipo de investigação foram sendo agrupadas sob a alçada do *simbolismo sonoro*, fenômeno cuja definição cobre uma associação motivada entre som e sentido (Imai e Akita, 2024), ou associação entre determinados fonemas e categorias perceptuais, como tamanho, forma, cheiro, luminosidade etc. (Hinton et al. 2006; Lockwood e Dingemanse, 2015). Para exemplificar este fenômeno, podemos nos valer de um clássico experimento da psicolinguística: o efeito *bouba-kiki* (Ramachandran e Hubbard, 2001), inicialmente denominado de efeito takete-maluma (Köhler, 1929). O experimento que nomeia este efeito consiste em apresentar duas imagens contrastantes lado a lado, sendo uma marcada por formas pontudas e a outra, arredondadas, e solicitar aos participantes a escolha de qual nome deve ser associado a cada uma delas, entre as opções *bouba* ou *kiki*. Nas mais diversas línguas, o resultado se repete: a forma geométrica pontuda é majoritariamente associada a *kiki*, e a forma arredondada, a *bouba*. Entre os resultados ratificados, é possível citar pesquisas em inglês (Nielsen e Rendall, 2011), francês (Fort et al., 2013), japonês (Sakamoto e Watanabe, 2018) e português (Godoy et al., 2018), além de outros idiomas (cf. Styles e Gawne, 2017 para uma revisão).

As justificativas para essa associação são diversas, como similaridade entre letras e formatos geométricos, associação entre produção acústica e forma, associação entre articulação sonora e forma (como o arredondamento de /o/ e /u/ e a forma redonda), dentre outras. Sendo tais justificativas aceitas ou rebatidas, o ponto é que a investigação do simbolismo sonoro levantou diversas outras questões a serem discutidas atualmente, como o papel desse fenômeno para além de pseudopalavras, no léxico das línguas propriamente dito.

Winter e Perlman (2021), por exemplo, argumentam que a estrutura de adjetivos referentes a tamanho do inglês é altamente preditiva do seu significado, pois vogais altas anteriores (e.g. /i/) são mais comuns em palavras referentes a dimensões menores (ex. *bitty*, *little*, *mini*), enquanto vogais posteriores baixas (e.g. /a/) são usadas preferencialmente em palavras que denotam dimensões maiores (ex. *giant*, *large*, *vast*). Os autores defendem, então, que seus resultados são a primeira demonstração que o simbolismo sonoro de tamanho é uma propriedade estatística do léxico do inglês.

Ainda quanto a categorias perceptuais, especificamente quanto a textura, Winter et al. (2022) mostram que palavras relacionadas ao conceito de aspereza, em vários idiomas, são muito propensas a conter um segmento vibrante (/r/). Em suas análises, os autores argumentam que o padrão é extremamente robusto, sendo o primeiro padrão generalizado de *iconicidade* documentado não apenas em uma amostra grande e diversificada das línguas faladas em todo o mundo (n = 331), mas também em inúmeras palavras sensoriais dentro de cada língua especificamente.

Neste artigo, ao explorarmos questões relacionadas a esse tema no português, entendemos iconicidade como um “senso de semelhança entre pelo menos algum aspecto de forma e algum aspecto do sentido”¹ (Winter et al., 2023). Tal definição de iconicidade nos leva a uma outra percepção que partiu das investigações sobre simbolismo sonoro, e que se tornou foco de pesquisa: algumas palavras parecem ser melhores representações de seus sen-

¹ A denominação “simbolismo sonoro” é utilizada, muitas vezes, como sinônimo de “iconicidade”, porém seu estudo parece enfatizar a modalidade auditiva e, corriqueiramente, limitar-se ao uso de pseudopalavras. Trabalhos recentes tendem a optar pela padronização de “iconicidade” por abarcar também a modalidade visuoespacial (Winter, Woodin e Perlman, 2023).

tidos que outras. Por exemplo, em inglês, as palavras *hard*, *rip* e *splash* são percebidas como melhores representantes de seus referentes do que *apple* e *find* na percepção de falantes nativos; assim como, em espanhol, *garganta* e *suave* são melhores representantes em comparação com *hormiga* e *doler* (Perry, Perlman e Lupyan, 2015). Isso quer dizer que falantes nativos do inglês estadunidense e do espanhol peninsular, respectivamente, consideram que, por algum aspecto de sua forma, algum aspecto do seu referente é melhor representado; ou, em outras palavras, que há um maior grau de iconicidade em algumas palavras. Essa percepção pode ser justificada por similaridade sonora (Winter, Woodin e Perlman, 2023), por associação metafórica (Taub, 2001; McLean, 2019) ou por outro fator que ainda não foi possivelmente levantado.

No âmbito da Psicolinguística e das Ciências Cognitivas, a iconicidade é tomada como um dos muitos fatores que podem influenciar no processamento e na produção de linguagem. Como exemplo, podemos citar trabalhos que utilizam onomatopeias para investigar o processamento de palavras motivadas. Meteyard et al. (2015), por exemplo, indicam que uma ligação mais forte entre forma e sentido faz com que certas palavras sejam mais resistentes a perdas decorrentes de afasia, enquanto Peeters (2016) reporta uma amplitude de N400 menor para palavras de maior iconicidade em tarefa de decisão lexical auditiva. Acredita-se que a iconicidade pode ter desempenhado um papel na origem das línguas (Perlman e Lupyan, 2018), o que reitera a ideia de sua influência em aspectos de produção e aquisição da linguagem. Há trabalhos, inclusive, que trazem a iconicidade como um fenômeno capaz de facilitar o aprendizado de palavras (Imai e Kita, 2014; Lockwood, Hagoort e Dingemanse, 2016).

Recentemente, o estudo da iconicidade, como semelhança entre certos aspectos de uma forma linguística e certos aspectos do sentido, tem ganhado espaço no campo dos estudos linguísticos das mais diferentes línguas. Um dos estudos iniciais nesse campo foi o de Perry, Perlman e Lupyan (2015), que buscam relacionar a percepção de iconicidade com a idade de aquisição de palavras em inglês e em espanhol e suas respectivas classes de palavras (cf. 2.1 e 2.2 para maiores detalhes). Os resultados dos autores, como detalharemos nas seções subsequentes, indicam que alguns aspectos da iconicidade resultam em padrões comuns na comparação entre línguas, enquanto outros parecem depender das diferenças gramaticais entre os idiomas. Por esse motivo, torna-se relevante pesquisar o papel da iconicidade no léxico do português, e comparar os resultados obtidos para o Português Brasileiro (PB) ao que já foi encontrado em outras línguas. Este artigo², portanto, tem o propósito de investigar as relações da iconicidade com outras métricas em português, além de prover um estudo normativo de iconicidade para o idioma que poderá ser usado por estudos futuros.

2 Estado da arte

O trabalho de Perry et al. (2015), que replicamos em português, investiga a relação da iconicidade com as classes de palavras do inglês e do espanhol e também com a idade de aquisição de cada palavra. Trataremos de cada uma dessas variáveis separadamente para melhor justificar nosso trabalho.

² Os resultados reportados correspondem aos achados da dissertação de mestrado da primeira autora. Este artigo também amplia, portanto, a divulgação dos achados sobre um fenômeno pouco explorado no português brasileiro.

2.1 Iconicidade e idade de aquisição

Por representar uma semelhança potencial entre forma linguística e sentido, o fenômeno da iconicidade levanta um questionamento quanto a sua possível relação com a aquisição da linguagem. Isto é, apesar de parecer atenuar durante nosso crescimento devido ao maior contato com formas não motivadas, somos seres que, naturalmente, buscamos semelhanças entre nomes e referentes (Imai e Kita, 2014). Por essa razão, desde muito cedo, as crianças parecem ser sensíveis aos efeitos da iconicidade. Há, nos mais diversos trabalhos sobre aquisição, a investigação de fatores que podem influenciar nesse processo, inclusive a iconicidade.

Entre os trabalhos selecionados para apresentar esta relação, Maurer *et al.* (2006) realizaram um experimento com dois grupos falantes de inglês: um de adultos e outro de crianças com 2.5 anos de idade. Em uma atividade de escolha forçada, os autores apresentaram duas pseudopalavras (*kaykee* e *boobaa*) e duas figuras, sendo uma com formas arredondadas e outra com formas pontudas. Os resultados demonstraram que as crianças fizeram a associação esperada (*kaykee* à forma pontuda e *boobaa* à forma arredondada) de forma estatisticamente significativa, tanto quanto os adultos.

A sensibilidade de crianças ao reconhecimento da iconicidade sonora também se mostrou presente em crianças japonesas com idade em torno de 2 anos. Todavia, faz-se necessário destacar que a sensibilidade aqui identificada se refere a modo de movimento e a pseudopalavras, baseadas em ideofones³ (Imai et al., 2008). Nesse estudo, as crianças eram apresentadas a palavras criadas (base nos ideofones do japonês) e a dois vídeos, sendo cada um deles apresentando uma pessoa andando de diferentes modos, como correndo com passos pesados ou andando rápido com pequenos passos, por exemplo. Apresentados os vídeos, os pesquisadores solicitaram que as crianças apontassem para o vídeo que mostrava a ação a que a pseudopalavra verbal se referia. Os resultados indicaram que as crianças foram capazes de detectar o simbolismo sonoro ligando tais elementos a vídeos diferentes, o que demonstra a capacidade de aprendizado e reconhecimento de ação em diferentes cenários.

Outro experimento realizado foi o de Kantartzis, Imai e Kita (2011), cujo foco foi demonstrar a facilidade de aprendizado de verbos quando há relação som-referente, independentemente da língua que as crianças estão aprendendo. Essa informação pôde ser observada quando os pesquisadores citados investigaram a aprendizagem de crianças em idade pré-escolar. A um grupo dos participantes, falantes de inglês, foram ensinados novos verbos a partir da estrutura icônica de ideofones do japonês, enquanto o outro grupo foi ensinado que determinadas ações eram denominadas por pseudopalavras baseadas na fonotaxe padrão do inglês. Nos resultados, foi possível perceber que as crianças do grupo dos ideofones conseguiram generalizar a ação ensinada para outra situação com uma facilidade bem maior do que aquelas que foram apresentadas a verbos não-icônicos.

Imai e Kita (2014) partem desses e de outros achados para proporem a *sound symbolism bootstrapping hypothesis*, de acordo com a qual a iconicidade atuaria como facilitadora do pro-

³ Ideofones podem ser definidos como “uma classe de palavras que evocam vividamente impressões sensoriais, por exemplo, sons, movimentos, texturas, padrões visuais e ações.” (Dingemanse et al., 2015, tradução nossa). No Japonês, por exemplo, há uma gama de palavras dessa classe, como é o caso de *goro*, associado a um objeto pesado rolando, e *gorogoro*, a vários objetos pesados rolando; ou *koro*, que em comparação a *goro*, representa um objeto leve rolando.

cesso de aquisição. Segundo os autores, a iconicidade oferta a possibilidade de relacionar som e sentido, estruturando a aprendizagem lexical. A partir dessa tese, algumas considerações são levantadas, como: (i) as crianças pequenas são sensíveis a uma gama mais ampla de possíveis correspondências simbólicas sonoras do que os adultos, mas essa sensibilidade parece ser diminuída e reorganizada à medida que aprendem mais palavras em sua língua nativa; (ii) o simbolismo sonoro ajuda os bebês que acabaram de começar a aprender palavras a entender que os sons da fala se referem a entidades no mundo e (iii) o simbolismo sonoro auxilia as crianças a identificar, mais facilmente, os referentes incorporados em uma cena complexa.

Outro ponto a ser levado em consideração na aquisição é o de que adultos, ao se direcionarem a crianças, tendem a usar palavras mais icônicas. Assim, qualquer que seja o idioma, as crianças pequenas tendem a aprender os itens icônicos mais cedo e tendem a usá-los com mais frequência – o que pode ser reflexo, também, da influência de uma comunicação icônica direcionada às crianças por parte dos pais e/ou cuidadores (Winter e Perlman, 2021).

Ao lado desses achados, faz-se necessário destacar também que, quando se trata crianças mais novas, os resultados para experimentos que envolvem esse fenômeno são diversos. Ozturk *et al.* (2012), por exemplo, ao investigar o efeito *bouba-kiki* com crianças de quatro meses, indicaram que as crianças olharam mais para a forma contrária ao esperado após ouvirem o estímulo auditivo. Para fazer tal medida, os autores utilizaram equipamentos de rastreamento ocular. Ademais, os experimentos realizados por Fort *et al.* (2013), cujo grupo de participantes compunha-se de crianças de cinco e seis meses, não obtiveram respostas significativamente distintas, impossibilitando a formação de uma tendência por parte dessas crianças. Imai e Kita (2014, p. 5) apresentam uma hipótese para essa disparidade de resultados baseada em dois tipos de simbolismo: o universalmente compartilhado e o específico da língua nativa. De acordo com os autores, o primeiro simbolismo a ser desenvolvido é o *universal* nos primeiros anos de vida, e, com mais idade, adquire-se o simbolismo específico de cada língua. Essa hipótese abre espaço para argumentar que, talvez, as crianças participantes de Ozturk *et al.* (2012) e Fort *et al.* (2013) ainda estivessem no processo de desenvolvimento do simbolismo universal, mas que, em algum momento, estariam mais sensíveis ao simbolismo universal e, quando mais velhas, esse simbolismo seria moldado ao conhecimento de um inventário fonológico próprio.

Com o objetivo de investigar essas hipóteses, surgiram outros trabalhos cujos objetivos eram avaliar o comportamento da iconicidade em línguas específicas. Perry, Perlman e Lupyan (2015), por exemplo, buscaram relacionar percepção de iconicidade, categoria lexical e idade de aquisição de palavras em inglês e em espanhol. Para tal, os autores realizaram experimentos com palavras do inglês e do espanhol.

O primeiro experimento dos autores, que serviu como base principal para a investigação empírica apresenta neste trabalho, contou com 442 falantes nativos de inglês, que deram notas de iconicidade para 592 palavras do *MacArthur-Bates Communicative Developmental Inventories* (MCDI), distribuídas em seis categorias: onomatopeia, interjeição, adjetivo, verbo, substantivo e palavra funcional. Os dados obtidos permitem a observação da trajetória de aprendizado de palavras individuais, relacionando a idade em meses (entre 8 e 30 meses) e proporção de crianças que produzem determinadas palavras. Os participantes visualizaram as palavras e eram instruídos a dar uma nota entre -5 e 5, em que -5 indica uma palavra oposta ao seu significado, 0 indica uma palavra arbitrária e 5 indica altamente icônica. Em um segundo momento, os autores repetiram o experimento anterior com a diferença de apre-

sentar as palavras de forma falada para os participantes, mas os resultados não divergiram. O terceiro experimento com palavras do inglês foi realizado com o mesmo *corpus*, mas com um direcionamento diferente: agora, os participantes tinham de dar nota entre 1 e 100 para a possibilidade de um alienígena entender o que aquela palavra significa. Ao comparar, então, as notas de iconicidade com o tempo de aquisição de palavras do MCDI tanto em inglês quanto em espanhol, os autores indicaram que aquelas com as notas mais altas também eram aprendidas mais cedo nas duas línguas.

Considerando, então, os achados de tais estudos e dada a capacidade, mesmo que não consensual, de crianças consideravelmente novas apontarem algum reconhecimento icônico, destaca-se a importância do estudo desse efeito. No Brasil, entretanto, este conceito de iconicidade ainda tem sido pouco explorado, uma vez os estudos feitos até então voltam-se para a associação entre som-referente a partir de pseudopalavras (Godoy et al., 2018; Godoy et al., 2020; Godoy et al., 2021; Ananias e Godoy, 2023). A principal contribuição deste trabalho, portanto, é correlacionar, pela primeira vez, a iconicidade no léxico do português e a idade de aquisição de palavras, de forma a tratar também do efeito quanto às classes de palavras e a relação com outras línguas.

2.2 Iconicidade e classes de palavras

A relação entre categoria gramatical e iconicidade é uma das questões frequentemente exploradas em estudos normativos (Perry, Perlman e Lupyan, 2015; Winter et al., 2023; Hinojosa et al., 2021). Estudos indicam que, independentemente da língua analisada, onomatopeias e interjeições costumam receber as maiores notas de iconicidade. No entanto, a análise de medidas subjetivas de iconicidade para verbos e substantivos apresenta diferenças entre línguas que tendem a ser explicadas a partir de suas características tipológicas. Em Perry, Perlman e Lupyan (2015), a média de iconicidade de onomatopeias e interjeições foi a mais alta em comparação às outras classes gramaticais. Em inglês, verbos foram percebidos como mais icônicos que substantivos; já em espanhol, não houve diferença entre verbos e substantivos considerando sua percepção de iconicidade.

A tendência de, em comparação ao espanhol, verbos do inglês serem considerados mais icônicos do que substantivos é explicada, de acordo com Perry, Perlman e Lupyan (2015), pela tipologia de eventos de movimento proposta por Talmy (2000). De acordo com essa tipologia, há dois⁴ grupos de nosso interesse nos quais as línguas naturais podem ser divididas: línguas centradas no verbo (*verb-framed languages*), que tendem a lexicalizar informações sobre trajetória e movimento no verbo, e línguas centradas em satélites (*satellite-framed languages*), que transmitiriam essas mesmas informações a partir de partículas satélites (como preposições) ligadas ao verbo. Uma das características das línguas centradas em satélites é de que o verbo teria tendência a lexicalizar modo de movimento. O espanhol, por sua vez, tem maior tendência a codificar trajetória no verbo, o que o caracterizaria como uma língua

⁴ Talmy (2000) ainda apresenta um terceiro grupo de caracterização de línguas naturais, mas não entra em nosso foco de pesquisa. Este grupo é responsável por abarcar línguas em que a raiz verbal incorpora elementos semânticos de movimento e de figura. Como exemplo representativo desse padrão, o autor cita o atsugewi, idioma indígena da Califórnia, em que prefixos têm a função de marcar o tipo de evento que desencadeia o movimento, enquanto os sufixos especificam a trajetória percorrida.

centrada no verbo. Para Perry, Perlman e Lupyan (2015), como modo relaciona-se à noção de movimento, e esta é uma das categorias perceptuais com maior potencial de geração de itens lexicais icônicos (Dingemanse, 2012), seria previsível que os verbos em inglês gerassem notas maiores de iconicidade em comparação a outras classes gramaticais.

Recentemente, Hinojosa et al. (2021) reforçam tal hipótese a partir de dados que indicam que as notas de iconicidade para substantivos e verbos no espanhol são iguais para um corpus de mais de 10.000 palavras. Igualmente, análise de 14.000 palavras do inglês também reiteraram que, nessa língua, verbos são percebidos como mais icônicos que substantivos (Winter et al. 2023). Nos dois artigos, a hipótese tipológica de Perry, Perlman e Lupyan (2015) foi citada para explicar os padrões observados, o que nos leva a formular a hipótese de que, em português, os resultados devem seguir o mesmo padrão do espanhol, por serem línguas tipologicamente próximas.

Reitera-se, então, a relevância desta pesquisa, uma vez que nos propomos a testar a hipótese de Perry, Perlman e Lupyan (2015) sobre a relação entre iconicidade e tipologia, e, consequentemente, sistematizar o primeiro estudo normativo com métricas de iconicidade para palavras do PB.

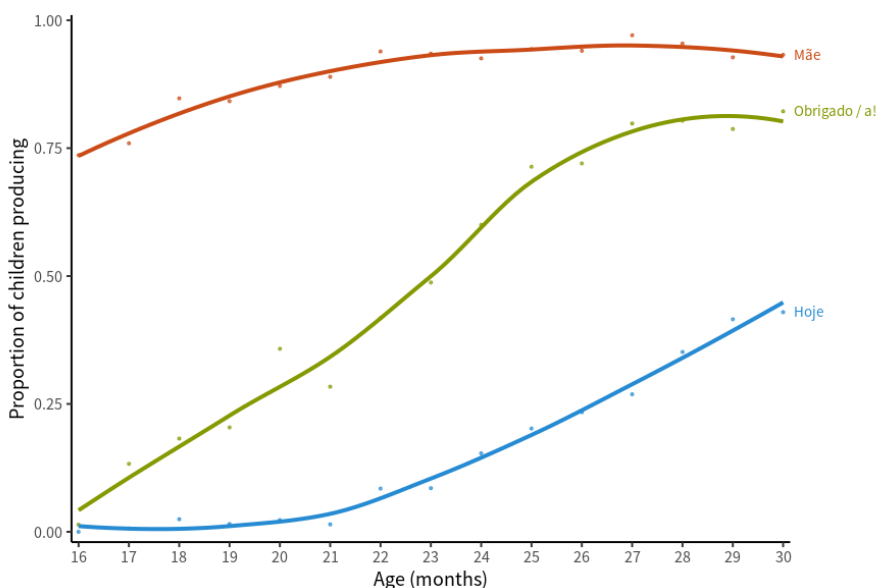
3 Experimento

As hipóteses principais deste trabalho são duas: (i) palavras mais icônicas são aprendidas mais cedo que palavras menos icônicas e (ii) a iconicidade das classes de palavras do português segue o mesmo padrão reportado para o espanhol (não haverá diferença significativa de iconicidade entre substantivos e verbos, por exemplo). Para explorar essas hipóteses, realizamos um experimento de coleta *on-line* para notas de percepção de iconicidade seguindo os procedimentos metodológicos de Perry et al. (2015). Nas subseções a seguir, detalhamos os materiais, procedimentos e caracterizamos os participantes da pesquisa.

3.1 Materiais

O corpus de palavras utilizado para a elaboração do experimento foi selecionado do MCDI, um banco de dados que disponibiliza a relação entre idade (entre 16 e 30 meses) e a proporção de crianças que produzem determinadas palavras e sentenças em diversas línguas, inclusive em português europeu (PE). Na Figura 1, a seguir, trouxemos uma representação dessa relação a partir de três itens: *mãe*, *obrigado/a* e *hoje*.

Figura 1 – Relação entre idade e proporção de crianças que produzem determinadas palavras.



Fonte: Banco de palavras do MCDI, 2025.

Como é de se esperar, a partir dos 16 meses, já há uma alta proporção (quase 75%) de crianças produzindo o termo *mãe*, curva que se mantém basicamente estável, com uma pequena crescente (até em torno dos 22 meses). Em comparação, temos *obrigado/a*, item que tem uma baixíssima proporção de crianças produzindo aos 16 meses, mas que apresenta um crescimento considerável nos meses seguintes e, aos 30 meses, apresenta uma proporção de produção próxima ao item *mãe*. Por fim, o item *hoje*, apesar de apresentar uma crescente a partir dos 22 meses, apresenta uma proporção de produção menor que 50% aos 30 meses. Apesar de termos cotejado tal relação entre apenas três itens, este corpus contempla substantivos, adjetivos, verbos, numerais, onomatopeias, interjeições, preposições, pronomes, conjunções e advérbios. Em prol de investigar a hipótese levantada, e seguindo o procedimento de Perry et al. (2015), operacionalizamos a idade de aquisição de um item como a proporção de crianças que o produzem determinado termo aos 30 meses. Essa medida foi usada para correlacionar idade de aquisição e iconicidade nos itens lexicais analisados.

Com relação à divisão dos itens experimentais em classes de palavras, Perry, Perlman e Lupyan (2015) dividiram os dados do MCDI para espanhol e inglês entre onomatopeias e interjeições, adjetivos, verbos, substantivos e palavras funcionais. Adotamos a mesma classificação ao agruparmos os itens do português que obtivemos a partir do MCDI para palavras do português. Isso implicou, por exemplo, combinar pronomes e advérbios dentro de uma mesma classe de palavras funcionais (cf. Tabela 1). Os numerais não fizeram parte deste agrupamento pois não constam nas análises de Perry, Perlman e Lupyan (2015), mas, uma vez que estavam disponíveis para o português no MCDI, optamos por utilizá-los para composição do *corpus*.

Como os dados para o português do MCDI se referem ao PE, foram feitas adaptações e exclusões no corpus original para realizarmos o experimento com falantes do PB. Quando, por exemplo, era possível perceber uma semelhança considerável entre a forma do PE disponível e a forma escrita do PB, adaptamos a forma (e.g. de 'coco' para 'cocoricó'). Além disso, percebemos que o corpus varia a marcação de número de palavras, ora apresenta palavras no

singular, ora no plural, sem um critério específico. É o caso de ‘ameixas’, que aparece no plural, e ‘banana’, no singular, ainda que as duas palavras façam referência a frutas. Por esse motivo, quando não se tratava de palavras referentes a objetos que usualmente aparecem em pares (e.g. ‘meias’ e ‘olhos’), adaptamos a fim de apresentar palavras no singular (de ‘ameixas’ para ‘ameixa’). Além disso, os itens do corpus que apresentam flexão de gênero (e.g. ‘este’/ ‘esta’/ ‘estes’/ ‘estas’) também foram padronizados em apenas um item masculino singular (‘este’).

Alguns critérios foram considerados para exclusão de palavras em nosso experimento. Primeiramente, as palavras foram excluídas por não serem evidentemente semelhantes o suficiente para serem adaptadas, como ‘tinoni’ (onomatopeia para ambulância) e ‘comboio’. O segundo critério para exclusão considerou palavras de origem estrangeira, como ‘puzzle’, do inglês. Além disso, foram excluídas expressões formadas por mais de uma palavra e palavras compostas, uma vez que o foco do trabalho era, especificamente, palavras simples. Assim, caso o item fosse formado por mais de uma, como *onde está?* e *guarda-chuva*, ele foi excluído. Por ser constituído para identificar idade de aquisição, o corpus dispõe de alguns sintagmas específicos utilizados por crianças, como “palminhas!”, que, apesar de ser apenas uma palavra, é uma expressão infantil específica. Por fim, o sexto e último critério determina o descarte de itens com verbos conjugados, uma vez que optamos por apresentar apenas verbos no infinitivo, de forma a padronizar a classe de palavras.

Ao final da exclusão das palavras coerentes com os critérios elencados, restaram 521 palavras para coleta de notas de iconicidade. A quantidade de palavras por categoria gramatical pode ser vista na Tabela 1:

Tabela 1 – Quantitativo de palavras após junção de categorias

Categoria	Quantidade
Substantivo	308
Verbo	83
Adjetivo	45
Palavra funcional + pronome + advérbio	66
Onomatopeia + interjeição	17
Numeral	10
Total	521

Fonte: Elaboração própria, 2025

Elaboramos o *script* com o auxílio dos pacotes de *software* gratuitos do *Jspych* (De Leeuw, 2015) e hospedamos o experimento no *Mindprobe*, servidor para experimentos on-line. Todas as análises e os dados estão disponíveis na plataforma OSF no seguinte link <https://osf.io/v7ytq/>.

3.2 Procedimentos e participantes

Os participantes foram recrutados via sistemas SINFO/UFRN e a partir de *link* em redes sociais. Como critério de seleção, indicamos a necessidade de ser falante nativo de Português Brasileiro e maior de 18 anos. O número de participantes, de palavras, a escala e as instruções apresentadas a cada um deles segue o protocolo do experimento de Perry, Perlman e Lupyan (2015).

Ao clicar no link para participar do experimento, os participantes eram apresentados ao Termo de Consentimento⁵. Caso aceitasse os termos e quisesse iniciar o experimento, o participante iria para a tela de instruções, cujo comando (traduzido e adaptado de Perry, Perlman e Lupyan, 2015) era dar uma nota entre muito icônica (5), nada icônica (0) e oposto de icônica (-5), para uma lista de 21 ou 22 palavras. Outros estudos (e.g Perry et al., 2015; Hinojosa et al., 2021) já fizeram a comparação entre notas de iconicidade dadas para estímulos visuais e auditivos, e não encontraram diferenças, então optamos por apresentá-las apenas visualmente. De acordo com as instruções, as palavras próximas a nota 5 deveriam refletir, de alguma forma, seu significado. Por exemplo, *roncar* soa como o barulho feito quando algumas pessoas dormem, *mini* soa como algo muito pequeno. Em oposição, àquelas mais próximas à nota -5 representariam algo oposto ao que significam, por exemplo, *microorganismo* é uma palavra grande que significa algo muito pequeno. E, por fim, aquelas alocadas mais próximas ao intermédio (0) provavelmente não apresentariam nem algo próximo ao significado, e tão pouco oposto, como, a princípio, não haver nada canino ou felino nas palavras *cachorro* e *gato*.

A coleta de dados das notas de iconicidade foi realizada entre 14 e 31 de março de 2023. De 512 participantes que finalizaram a tarefa de julgamento, 2 foram descartados por terem menos de 18 anos, 3 por não serem nativos e 3 por responderem que não conheciam mais de 10% das palavras do *corpus*, totalizando, então, 504 participantes. A média de idade dos participantes restantes foi de 31.63 (amplitude de 18 a 69, desvio-padrão de 10.39).

4 Resultados

A seguir, esmiuçamos os resultados de análise referentes às relações em investigação: iconicidade e idade de aquisição e, em seguida, iconicidade e classes de palavras.

4.1 Iconicidade e idade de aquisição

No que se refere às variáveis que podem influenciar a nota de iconicidade de uma palavra, consideram-se as mais diversas. O modelo inicial de Perry, Perlman e Lupyan (2015) indica a variável resposta como sendo a *nota de iconicidade*; as variáveis preditoras sendo *concretude*, *frequência*, *número de fonemas*, *número de morfemas* e *palavras associadas a bebês*⁶; e *participante* como efeito aleatório.

Feita essa análise, os autores extraem os resíduos do modelo para correlacioná-lo com a idade de aquisição. Isto é, eles consideram que aquilo que não foi explicado pelas variáveis preditoras quanto à iconicidade (resíduos) poderia ser explicado pela porcentagem de crianças que produzem a palavra aos 30 meses (referente, neste trabalho à *idade de aquisição*). Após aplicar esse modelo aos resultados do inglês e do espanhol, os autores apontam que os resíduos podem ser explicados pela porcentagem de crianças que falam aquelas palavras aos 30 meses.

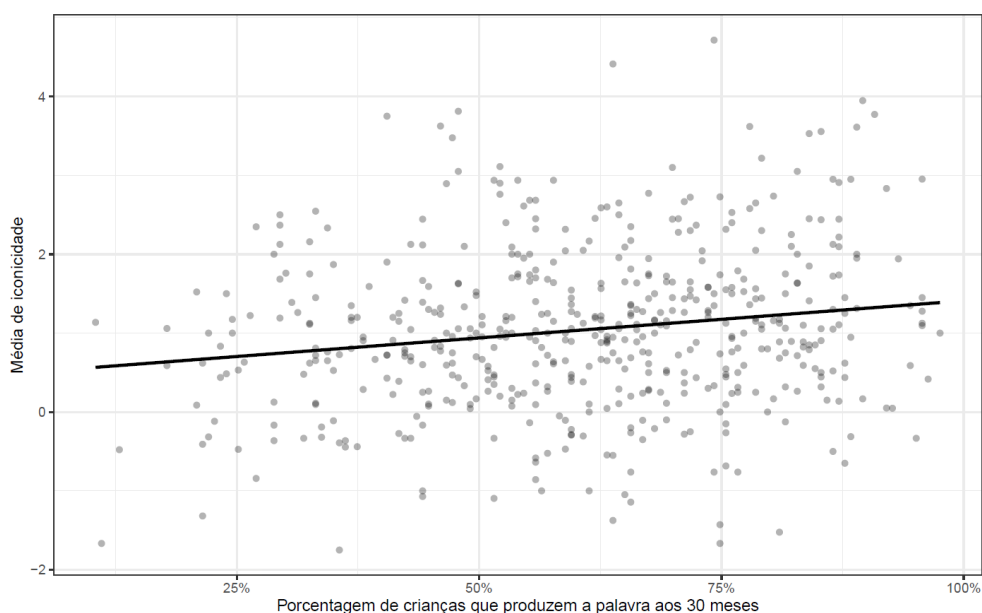
⁵ O experimento foi revisado e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Norte, sob o número de protocolo 63342122.0.0000.5537.

⁶ Métrica utilizada por Perry, Perlman e Lupyan (2015) referente à percepção de adultos quanto a palavras comumente faladas por bebês ou direcionadas a bebês.

Em nosso trabalho, realizamos uma replicação ajustando um modelo linear misto com a nota de iconicidade atribuída por cada participante como variável resposta. As variáveis preditoras foram o logaritmo da frequência (aproveitada dos corpora disponíveis no site *Corpus do Português*⁷), concretude⁸ e número de fonemas. A única diferença com relação ao trabalho de Perry, Perlman e Lupyan (2015) foi a não-inclusão da métrica de palavras associadas a bebês, que não está disponível para o português. Incluímos também interceptos aleatórios por participante, a exemplo dos autores. Foram extraídos os resíduos desse modelo, que foram então correlacionados com a idade de aquisição por meio de um modelo linear. Esse modelo tinha os resíduos como variável resposta e idade de aquisição como variável preditora. Uma comparação com um modelo nulo demonstrou que os resíduos podem ser explicados pela idade de aquisição ($\chi^2(1) = 62.875$, $p < 0.0001$). Em outros termos, esse resultado se mostra compatível com a ideia de que palavras com notas mais altas de iconicidade são, de fato, aprendidas mais cedo. Tal efeito positivo encontrado pode, a princípio, ser associado à quantidade de onomatopeias e interjeições, uma vez que o *corpus* utilizado é focado na produção de crianças, cujas classes são, de fato, mais frequentes no vocabulário infantil (Laing, 2019). Por esse motivo, realizamos uma análise sem as onomatopeias e interjeições, e o resultado foi o mesmo: ainda assim, identifica-se o efeito positivo entre média de iconicidade e idade de aquisição ($F = 31.256$, $p < 0.0001$).

Na Figura 2, é possível visualizar que, apesar de pequeno, há um efeito positivo entre a média de iconicidade das palavras e a idade de aquisição:

Figura 2 – Relação entre idade de aquisição e média de iconicidade por palavra do *corpus*



Fonte: Elaboração própria, 2025.

⁷ Disponível em <https://www.corpusdoportugues.org/xp.asp>

⁸ As notas de concretude foram coletadas em um experimento à parte para este trabalho. Seus métodos de coleta e análise podem ser conferidos em Ananias e Godoy (2024).

Assim como Perry, Perlman e Lupyan (2015) encontraram para o inglês e espanhol, também encontramos, para o português, que palavras percebidas como mais icônicas também tendem a ser produzidas por uma maior porcentagem de crianças.

4.2 Iconicidade e classes de palavras

Como esperado, os resultados da estatística descritiva demonstram que o grupo das onomatopeias e interjeições apresenta a maior média de iconicidade. Em seguida, aparece a classe dos numerais e, após esta, a dos adjetivos; em último, os substantivos têm a menor média⁹. Na tabela 2, é possível observar a média de iconicidade e o desvio padrão por categoria.

Tabela 2 – Média de iconicidade por categoria

Classe	Média de iconicidade	Desvio padrão
Onomatopeias e interjeições	3.25	2.99
Numerais	1.83	2.73
Adjetivos	1.34	2.99
Verbos	1.15	2.95
Palavras funcionais	1.08	3.13
Substantivos	0.815	3.11

Fonte: Elaboração própria, 2025

De forma a especificar mais os resultados e apresentar, de fato, exemplos das palavras vistas e avaliadas pelos participantes, vejamos a Tabela 3, com as palavras percebidas como mais icônicas em nosso experimento:

Tabela 3 – Amostra de palavras mais icônicas do experimento

Palavra	Média de iconicidade	Classe
1 Cocoricó	4.71	Onomatopeia
2 Vrum vrum	4.41	Onomatopeia
3 Au au	3.95	Onomatopeia
4 Uau	3.81	Interjeição
5 Miau	3.77	Onomatopeia
6 Fofo	3.75	Adjetivo
7 Chiu	3.62	Interjeição
8 Triciclo	3.62	Substantivo
9 Não	3.61	Advérbio
10 Chuva	3.56	Substantivo

Fonte: Elaboração própria, 2025

⁹ Destacamos que a classe dos numerais conta com um número baixo de itens, e os substantivos, com o maior número de itens, o que pode refletir na média identificada.

Seguindo a expectativa de maior média de iconicidade das onomatopeias e interjeições, temos representantes majoritários dessas classes. Como as duas representam classes de palavras canonicamente percebidas como mais icônicas nas línguas (Perry, Perlman e Lupyan, 2015; Winter *et al.*, 2023), apresentamos na Tabela 4 as 10 palavras mais icônicas sem considerar essas duas classes.

Tabela 4 – Amostra de palavras mais icônicas do experimento (sem onomatopeias e interjeições)

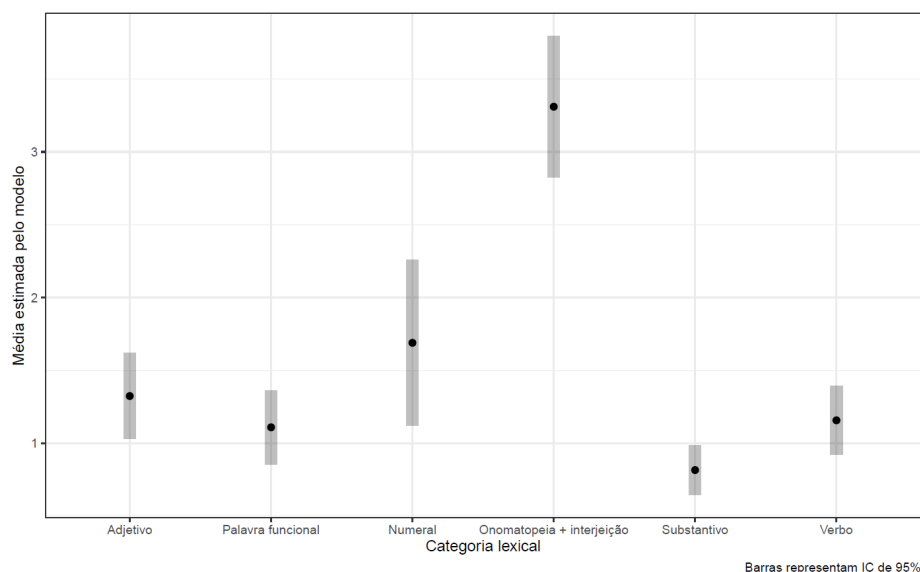
	Palavra	Média de iconicidade	Classe
1	Fofo	3.75	Adjetivo
2	Triciclo	3.62	Substantivo
3	Não	3.61	Advérbio
4	Chuva	3.56	Substantivo
5	Cobertor	3.11	Substantivo
6	Olhar	3.10	Verbo
7	Trator	3.05	Substantivo
8	Balão	2.95	Substantivo
9	Pantufas	2.94	Substantivo
10	Puxar	2.94	Verbo

Fonte: Elaboração própria, 2025

O modelo construído por Perry, Perlman e Lupyan (2015) – replicado neste trabalho – apresenta a nota de iconicidade como variável resposta e a categoria lexical como variável preditora, cujos níveis são duas categorias por vez. Por exemplo, para conferir se havia uma diferença significativa entre substantivos e verbos, foi criado um subconjunto apenas com os itens destas categorias e, com este subconjunto, seria ajustado um modelo linear. Seguimos este mesmo arranjo para a nossa análise inicial, sendo os efeitos aleatórios por item e por participante foram ajustados a partir de modelos lineares mistos com a função lmer do pacote do R lme4 (Bates *et al.*, 2015).

Os nossos resultados indicaram que o grupo formado pelas onomatopeias e interjeições apresenta maiores notas de iconicidade do que todas as outras categorias, sejam adjetivos ($\chi^2(1) = 33.758$, $p < 0.0001$), verbos ($\chi^2(1) = 52.165$, $p < 0.0001$), substantivos ($\chi^2(1) = 92.959$, $p < 0.0001$), numerais ($\chi^2(1) = 37.679$, $p = 0.0002$) ou palavras funcionais ($\chi^2(1) = 40.501$, $p < 0.0001$). Quando comparamos as outras classes entre si, adjetivos tiveram notas mais altas que os substantivos ($\chi^2(1) = 12.657$, $p = 0.00037$), mas não houve diferença significativa entre adjetivos e palavras funcionais ($\chi^2(1) = 1.4945$, $p = 0.2215$); os verbos, assim como os adjetivos, obtiveram notas mais altas que os substantivos ($\chi^2(1) = 10.438$, $p = 0.001235$) e não apresentaram diferença significativa quando comparados a palavras funcionais ($\chi^2(1) = 0.0937$, $p = 0.7595$). Com exceção da comparação com onomatopeias e interjeições, os numerais não demonstraram diferença estatisticamente significativa com nenhuma das outras classes. Tais resultados podem ser visualizados na Figura 3.

Figura 3 – Comparação entre média de nota de iconicidade e categorias gramaticais



Fonte: Elaboração própria, 2025.

Como acréscimo e forma de confirmação, realizamos também uma análise considerando todas as categorias de uma vez. A proposição desta nova análise se dá porque, ao comparar cada categoria entre si, estamos repetindo o mesmo teste, de forma a, possivelmente, aproximar-nos de um falso positivo (Krzywinski e Altman, 2014). Isto é, a probabilidade de encontrarmos uma correlação aumenta cada vez que o teste é repetido. A fim de contornar essa possibilidade, a nova análise foi realizada com o intuito de considerar a correção de p-valor para as múltiplas comparações. Para tanto, consideramos um modelo cuja variável resposta é a nota de iconicidade, a variável preditora é a categoria lexical e os efeitos aleatórios são os itens e os participantes. Ao comparar esse modelo a outro sem a variável preditora, podemos, então, confirmar o resultado da replicação: a categoria lexical da palavra é uma variável considerável para explicar a nota de iconicidade ($\chi^2(6) = 123.6$, $p < 0.0001$). Após perceber que é significativa, realizamos uma análise *post-hoc* para comparar as categorias lexicais entre si, cujos resultados observados inicialmente se mantiveram.

Um fato que nos chamou atenção entre os resultados descritos foi o das palavras funcionais, reconhecidas como mais icônicas do que o esperado, que seria menor que os substantivos. Como é possível verificar na Figura 3, a média das palavras funcionais foi maior do que a dos substantivos e quase idêntica a dos verbos. Com o objetivo de investigar essa singularidade, pormenorizamos a porcentagem de cada uma das classes específicas nas palavras funcionais nos nossos dados e comparamos com a do inglês e do espanhol de Perry et al. (2015). Em português, tivemos uma porcentagem de 40.91% de advérbios no grupo das palavras funcionais. Em inglês, essa porcentagem é de 25.26%, em espanhol, de 24%, como é possível observar na Tabela 5.

Tabela 5 – Divisão das categorias lexicais dentro das palavras funcionais

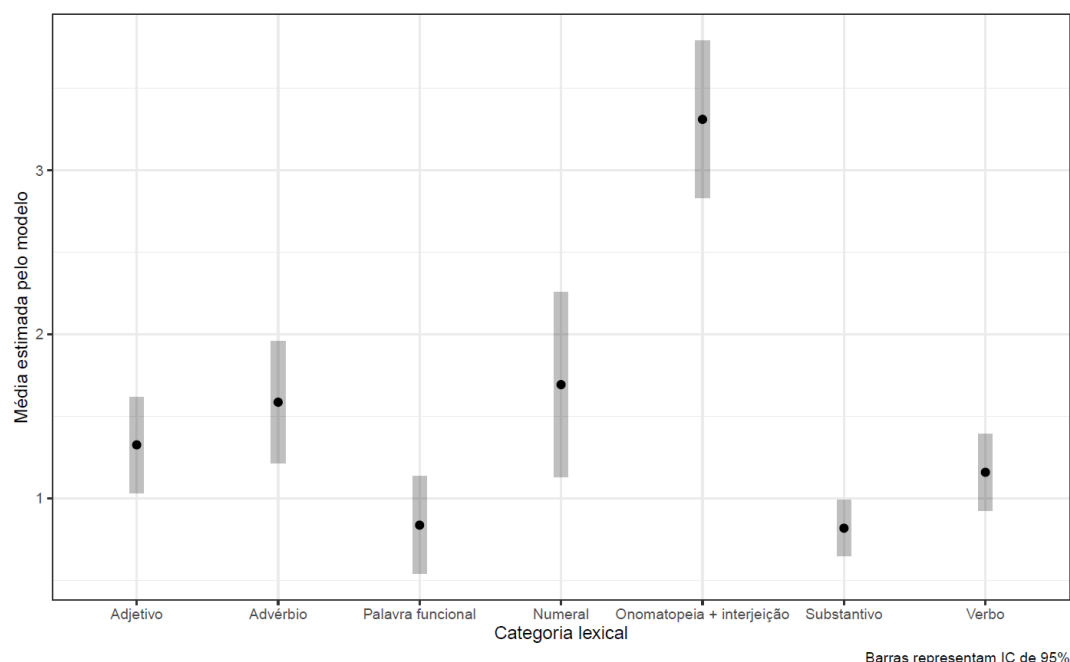
Categoria lexical	Inglês		Espanhol		Português	
	Número absoluto	%	Número absoluto	%	Número absoluto	%
Artigo	3	3,16	8	8,00	-	0,00
Advérbio	24	25,26	24	24,00	27	40,91
Conjunção	2	2,10	4	4,00	3	4,54
Preposição	11	11,58	12	12,00	7	10,61
Pronome	35	36,84	49	49,00	29	43,94
Verbo auxiliar	20	21,05	3	3,00	-	0,00
Total	95	100	100	100	66	100

Fonte: Elaboração própria, 2025

Advérbios, de certa forma, podem fazer referência a características perceptuais do seu referente, como forma, tamanho, velocidade, e, portanto, são uma classe com potencial de se observar padrões de simbolismo sonoro tal qual os reportados em outros estudos. O alto número de advérbios em português comparativamente ao que ocorre nas outras duas línguas pode ter feito com que as palavras funcionais desse idioma tivessem uma nota de iconicidade acima do esperado. Por esse motivo, realizamos uma análise *post-hoc* considerando as palavras funcionais sem os advérbios. Ao realizá-la, encontramos resultados similares quanto às onomatopeias e interjeições, que continuam sendo mais icônicas ($\chi^2 (1) = 40.264$, $p < 0.0001$).

Outro resultado díspar que obtivemos foi em relação aos adjetivos que, na primeira análise, não apresentaram diferença significativa quando comparados às palavras funcionais, e agora passaram a ser considerados mais icônicos ($\chi^2 (1) = 5.8578$, $p = 0.01551$), assim como no inglês e no espanhol. Ao comparar, também, diretamente palavras funcionais e advérbios, tivemos o resultado que advérbios são, de fato, percebidos como mais icônicos que palavras funcionais ($\chi^2 (1) = 10.053$, $p = 0.001521$). Podemos ver esses resultados no gráfico da Figura 4.

Figura 4 – Comparação post-hoc entre média de nota de iconicidade e categorias gramaticais, com palavras funcionais sem presença de advérbios.



Fonte: Elaboração própria, 2025.

Em comparação ao gráfico representado na Figura 3, o gráfico da figura 4 nos permite notar nitidamente como os advérbios elevaram a nota das palavras funcionais. Ainda assim, essas palavras continuam tendo uma nota de iconicidade igual a de substantivos ($z = 2.174$; $p = 0.3099$) e significativamente distinta dos verbos ($z = 4.237$; $p = 0.0005$). Isso difere do reportado no estudo de Perry, Perlman e Lupyan (2015), que apontou, tanto em espanhol quanto em inglês, os verbos com notas significativamente maiores que palavras funcionais.

5 Discussão

Seguindo os procedimentos metodológicos de Perry, Perlman e Lupyan (2015), coletamos e analisamos notas subjetivas de iconicidade atribuídas para um corpus do português de 521 palavras. De modo geral, partindo da premissa de replicação, pudemos observar algumas semelhanças do português quanto aos resultados do inglês e do espanhol, como a correlação positiva entre iconicidade e idade de aquisição. Todavia, a investigação das classes de palavras trouxe resultados que questionam a adequação da hipótese tipológica relacionada a eventos de movimento de Talmy (2000), amplamente utilizada como justificativa em trabalhos normativos prévios, como o de Perry, Perlman e Lupyan (2015).

5.1 Iconicidade e idade de aquisição

Nos dados aqui reportados, foi investigada a percepção de iconicidade por parte de falantes nativos de português brasileiro com mais de 18 anos e a porcentagem de crianças portuguesas que produzem as palavras até os 30 meses. Apesar de dados derivados de duas variações do português, acreditamos que estes podem ser comparados e correlacionados porque realizamos uma seleção criteriosa de palavras a partir do corpus do MCDI (cf. seção 3.1), além de uma análise atenta e consciente da possibilidade de imbróglio entre variações. Reconhecemos, portanto, esta limitação e incentivamos a realização de estudos futuros que objetivem a coleta de idade de aquisição para crianças brasileiras.

A correlação encontrada entre iconicidade e a idade de aquisição pode resgatar algumas etapas da aquisição lexical levantada por Akita e Imai (2022). Isto é, primeiro, podemos considerar a *iconicidade primária*, fase em que são estabelecidas similaridades reconhecidas universalmente, de forma sensorial, entre som-forma. De acordo com os autores, um léxico formado nesta etapa costuma ser mais fácil de aprender e de lembrar, bem como de compartilhar com outros falantes. Dado que estamos lidando com palavras do próprio léxico, é possível ainda trazer à tona, a etapa de *iconicidade emergente*, em que motivação entre som-referente é moldada pela estrutura linguística e, conseqüentemente, é adquirida por meio da experiência com a língua.

Nosso resultado reforça, também, a interpretação de que há uma tendência de adultos, a partir da comparação entre grupos etários distintos, a usar palavras mais icônicas quando se direcionam a crianças (Perry et al., 2018). Nesse raciocínio, é coerente termos indicativos de que as crianças pequenas, de fato, aprendem os itens percebidos icônicos mais cedo (Winter e Perlman, 2021), seja pelo direcionamento dos pares, seja pela pressuposta iconicidade primária/sensorial.

A iconicidade parece se mostrar, então, como princípio da organização da língua e um auxílio para a aquisição, que, até então, não tem sido profundamente explicado e estudado quanto poderia no que se refere às línguas orais.

5.2 Iconicidade e classes de palavras

Como levantado nas seções anteriores, Perry, Perlman e Lupyan (2015) defendem que a diferença sobre a percepção de iconicidade dos verbos no inglês e no espanhol se dão pela tipologia de eventos de movimento (Talmy, 2000). Lembremos que, de acordo com essa tipologia, o espanhol é caracterizado como uma língua centrada no verbo, pois apresenta maior tendência a codificar trajetória nessa categoria; e o inglês, por sua vez, uma língua centrada no satélite, dada a tendências de trajetória em partículas satélites ligadas ao verbo. Tal hipótese tipológica foi, inclusive, reforçada, em trabalhos posteriores, por dados do espanhol (Hinojosa et al., 2021) e do inglês (Winter et al., 2023).

Com base na defesa dos autores, levantamos inicialmente a hipótese de que as classes gramaticais mais icônicas no português seguiriam o padrão identificado no espanhol, dada a proximidade tipológica entre as línguas. Porém, ao investigar a relação entre classe gramatical e iconicidade no português, encontramos um padrão semelhante ao inglês, sendo os verbos percebidos como mais icônicos do que substantivos. Inicialmente, esse resultado pareceu

inesperado. Imaginávamos que o português deveria se comportar como o espanhol pelo fato de ser tradicionalmente descrito como uma língua centrada no verbo, como são as línguas românicas de modo geral, mas o resultado obtido não se mostra compatível com a hipótese de Perry, Perlman e Lupyan (2015).

Nossos resultados não ratificam a explicação dos autores, o que nos leva a crer que a hipótese tipológica pode não estar correta. Dada a natureza classificatória da explicação, endossamos a necessidade de realização de novos estudos que lidem com mais do que duas línguas para que, de fato, possa ser caracterizado como um estudo tipológico. É possível ainda que os resultados reportados pelos autores caracterizem aspectos próprios de cada uma das línguas (no caso, inglês e espanhol) no que se refere ao movimento dos verbos, mas não algo que possa ser delineado por uma tipologia de movimento dos verbos.

Os dados reportados podem também ser explicados por uma revisão do status do PB no âmbito da tipologia talmyiana. Como língua românica, o PB é tradicionalmente classificado como uma língua centrada no verbo. Uma das evidências para essa definição provém de estudos que mostram que construções resultativas, comuns em línguas centradas no satélite, não são típicas ou produtivas no português (Marcelino, 2014; Moreira, 2023). Contudo, a expressão de trajetória e modo no PB sugere padrões mais próximos às línguas de enquadramento satelital.

Estudos recentes têm sugerido que construções verbo-partícula, fenômeno típico das línguas germânicas de frame no satélite, estão presentes no PB coloquial. Baseando-se na tipologia de eventos de movimento de Talmy, o estudo demonstra que o PB informal atende quatro dos cinco critérios típicos das línguas de enquadramento satelital: expressa trajetória por meio de partículas, emprega verbos de modo ou causa, apresenta tendência à telicidade e exibe extensões metafóricas (Moura, 2025). Batoréo e Ferrari (2016) argumentam que certas construções do português (brasileiro e europeu) impedem sua classificação exclusiva como língua centrada no verbo. Santos-Filho (2018), ao catalogar centenas de verbos de modo no PB, propõe que a língua apresenta padrão misto de lexicalização. Outros estudos também questionam a classificação do PB como língua exclusivamente centrada no verbo (Silva Júnior, 2015; Meirelles, 2019).

Uma análise mais detalhada do português brasileiro ainda se faz necessária para avaliar o comportamento da língua à luz da tipologia talmyana. Contudo, se adotarmos a perspectiva de que as línguas são mais bem compreendidas como posicionadas ao longo de um continuum entre enquadramento verbal e satelital (Slobin, 2006), seria razoável situar o PB como uma língua menos centrada no verbo que o espanhol. Isso poderia explicar por que os padrões de iconicidade para verbos nas duas línguas serem diferentes.

Além disso, ainda percebemos uma diferença quanto aos resultados dos advérbios (classe considerada como parte das *palavras funcionais*). Isto é, após as análises iniciais, notamos uma limitação correspondente à disparidade do número de itens (cf. Tabela 1), o que pode ter impactado os resultados entre classes encontrados. Além disso, o grupo das palavras funcionais ser composto pela junção de diferentes classes de palavras (preposições, conjunções, advérbios e pronomes) – também realizada por Perry, Perlman e Lupyan (2015) – representa uma possível influência para o resultado. Ao fazer tal junção, lidamos com a representação de 100 palavras funcionais em espanhol, 95 em inglês e 66 em português. Além dessa comparação numérica, percebemos que as palavras funcionais do PB atingiram uma média de iconicidade maior do que os adjetivos, que, ao compararmos com os adjetivos do inglês e do espanhol, são a classe com maior nota após as onomatopeias e interjeições.

A partir dessa divergência, levantou-se a hipótese de que poderia haver algo, no grupo das palavras funcionais, além da quantidade de itens, que possa ter influenciado no resultado obtido na comparação entre classes, uma vez que as outras duas línguas também reúnem diferentes classes sob a denominação de “palavras funcionais”. Reexaminando os itens, visualizamos uma possível sobre-representação dos advérbios nas palavras funcionais tenha modificado a média de iconicidade dessa categoria no português.

Notamos que as palavras funcionais do inglês e do espanhol, ao serem revisitados após as análises iniciais, são majoritariamente constituídas por pronomes, de modo que há apresentação dos mais diversos pronomes flexionados (por exemplo, *un, una, unos e unas*), enquanto, em nosso corpus, só há uma versão desses pronomes (como exemplo, *um*). Assim, destacamos que a sobre-representação dos advérbios em português, em comparação com as outras línguas, pode ser responsável por aumentar a nota da classe das palavras funcionais.

6 Considerações finais

O objetivo desta pesquisa foi investigar a iconicidade em uma amostra de palavras do português brasileiro. Para tanto, examinamos, com os dados disponíveis, a relação da percepção da iconicidade com métricas diversas, como idade de aquisição, categoria lexical, concretude e frequência. Como arremate, retomamos às hipóteses levantadas e os resultados encontrados:

- ◆ Esperávamos que palavras mais icônicas fossem aprendidas mais cedo que palavras menos icônicas.

De acordo com nossos resultados no português, foi observada uma correlação entre o grau de iconicidade atribuído a palavra e o momento de sua aquisição. Palavras com alta nota de iconicidade também são reconhecidamente aprendidas por uma maior quantidade de crianças portuguesas com até 30 meses (2.5 anos).

- ◆ Esperávamos que a percepção de iconicidade, em português, seguisse o padrão do espanhol quanto aos verbos e substantivos.

Como resultado, não houve validação de tal hipótese tipológica, uma vez que, no português, os verbos obtiveram notas mais altas que os substantivos; enquanto, no espanhol, não houve uma diferença significativa entre as duas categorias. Quanto às outras classes, todavia, uma semelhança geral foi identificada: as onomatopeias e interjeições são os itens lexicais percebidos como mais icônicos em português, inglês e espanhol. Agora, de forma específica para o português, temos que, depois das onomatopeias e interjeições, a maior percepção de iconicidade foi em advérbios, numerais, seguidos dos adjetivos e dos verbos, sendo as palavras funcionais e substantivos igualmente percebidos como pouco icônicos.

Lidamos com a lacuna de trabalhar com dados de crianças portuguesas para idade de aquisição; implicando que, talvez, a porcentagem de crianças brasileiras que produzam as palavras selecionadas possa ser diferente. Apesar de não ter sido uma brecha contemplada para essa pesquisa – especialmente por motivos de tempo de execução –, é uma que pode ser destacada e esperançosamente preenchida por trabalhos futuros. Nossa contribuição, especificamente, recai em ser o primeiro estudo normativo com notas de iconicidade em português

brasileiro, que pode ser usado por outros pesquisadores, em estudos futuros, que possuam interesse em investigar essa métrica.

Declaração de autoria

Declaramo-nos igualmente responsáveis pela escrita do artigo. A primeira autora, Thayná Cristina Ananias, ficou responsável pela coleta de dados, confecção de experimento, análise dos dados, redação e revisão. A segunda autora, Mahayana Godoy, também, pela análise de dados, redação e revisão, além da supervisão do trabalho.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (primeira autora) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processos 304393/2024-O e 201419/2024-6 (segunda autora).

Referências

- AKITA, K., & IMAI, M. The iconicity ring model for sound symbolism. In: LENNINGER, S.; TABAKOWSKA, E.; LJUNGBERG, C.; FISHER, O. (orgs.). *Iconicity in cognition and across semiotic systems*, Amsterdam e Philadelphia: John Benjamins, 2022, p.27-45.
- ANANIAS, T. GODOY, M. Dando nomes: o efeito bouba-kiki em experimento de nomeação livre. *Linguística*, Montevideo, v. 39, n. 2, p.57-74, 2023. DOI: <https://doi.org/10.5935/2079-312X.20230017>.
- ANANIAS, T. GODOY, M. Normas de concretude para 521 palavras do português e sua relação com aquisição da linguagem. *Revista da ABRALIN*, Campinas, v.23, n.1, p.1-18, 2024. DOI: <https://doi.org/10.25189/rabralin.v23i1.2268>.
- BATORÉO, H.; FERRARI, L. Events of motion and Talmyan typology: Verb-framed and satellite-framed patterns in Portuguese. *Cognitive Semantics*, Leiden, v. 1, n. 2, p. 59-79, 2016. DOI: <https://doi.org/10.1163/23526416-00201004>.
- CHASTAING, M. Nouvelles recherches sur le symbolisme des voyelles. *Journal de Psychologie*, v.61, n 1, p. 75-88, 1964. Disponível em: https://www.persee.fr/doc/bupsy_0007-4403_1967_num_20_253_7303_t1_0224_0000_5. Acesso em: 6 jan. 2025.
- DE LEEUW, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, v. 47, n. 1, p. 1–12, 2015. DOI: <https://doi.org/10.3758/s13428-014-0458-y>.
- DINGEMANSE, M. Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass*, v.6, n.1, p. 654-672, 2012. DOI: <https://doi.org/10.1002/lnc3.361>.

DINGEMANSE, M., BLASI, D. E., LUPYAN, G., CHRISTIANSEN, M. H., & MONAGHAN, P. Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, v. 19, n. 10, p. 603-615, 2015. DOI: <https://doi/10.1016/j.tics.2015.07.013>.

FORT, M. WEISS, A. MARTIN, A. PEPERKAMP, S. Looking for the bouba–kiki effect in prelexical infants. In: Proceedings of the 12th International Conference on Auditory-Visual Speech Processing, Annecy, France, 2013, p. 71-76, 2012. Disponível em: https://avsp2013.loria.fr/proceedings/papers/paper_41.pdf. Acesso em: 30 dez. 2024.

GODOY, M.; GOMES, A.; KUMAGAI, G.; KAWAHARA, S. Sound symbolism in Brazilian Portuguese Pokémon names: Evidence for cross-linguistic similarities and differences. *Journal of Portuguese Linguistics*, v. 20, n. 1, p. 1–23, 2021. DOI: <https://doi.org/10.5334/jpl.257>.

GODOY, M. C.; SOUZA FILHO, N. S.; SOUZA, J. G. M.; ALVES, H.; KAWAHARA, S. Gotta name'em all: an experimental study on the sound symbolism of Pokémon names in Brazilian Portuguese. *Journal of Psycholinguistic Research*, v. 49, n. 5, p. 717–740, 2020. DOI: <https://doi.org/10.1007/s10936-019-09679-2>.

GODOY, M. C.; DUARTE, A. C. V.; SILVA, F. L. F.; ALBANO, G. F.; SOUZA, R. J. P.; SILVA, Y. U. A. P. M. Replicando o efeito Takete Maluma em Português Brasileiro. *Revista do GELNE*, Natal, v. 20, n.5, p. 87–100, 2018. DOI: <https://doi.org/10.21680/1517-7874.2018v20n1ID13331>.

HINTON, L.; NICHOLS, J.; OHALA, J. J. (eds.). *Sound symbolism*. Cambridge: Cambridge University Press, 2006.

HINOJOSA, J. A.; HARO, J.; MAGALLARES, S.; DUÑABEITIA, J. A.; FERRÉ, P. Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*, v. 53, n. 3, p. 1262-1275, 2021. DOI: <https://psycnet.apa.org/doi/10.3758/s13428-020-01496-z>

IMAI, M.; KITA, S. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, Londres, v. 369, n. 1651, p. 20130298, 2014. DOI: <https://doi.org/10.1098/rstb.2013.0298>

IMAI, M., KITA, S., NAGUMO, M., & OKADA, H. Sound symbolism facilitates early verb learning. *Cognition*, v. 109, n. 1, p. 54-65, 2008. DOI: <https://doi.org/10.1016/j.cognition.2008.07.015>

JAKOBSON, R. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell, 1962 [1941].

JESPERSEN, O. Sound Symbolism. In: JESPERSEN, O. *Language its nature development and origin*. Londres: George allen e Unwin LTD., 1922, p. 396-411. Disponível em: <https://www.gutenberg.org/files/53038/53038-h/53038-h.htm>. Acesso em: 28 dez. 2024.

KANTARTZIS, K.; IMAI, M.; KITA, S. Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive science*, v. 35, n. 3, p. 575-586, 2011. DOI: <https://doi.org/10.1111/j.1551-6709.2010.01169.x>.

KÖHLER, W. *Gestalt psychology*. New York, Liveright Publishing, 1929.

KRZYWINSKI, M.; ALTMAN, N. Visualizing samples with box plots: use box plots to illustrate the spread and differences of samples. *Nature Methods*, v. 11, n. 2, p. 119-121, 2014. Disponível em: <https://go.gale.com/ps/i.do?id=GALE%7CA361242515&sid=googleScholar&v=2.1&it=r&link-access=abs&issn=15487091&sw=w&p=HRCA&userGroupName=anon%7Ea284e-8cb&aty=open-web-entry>. Acesso em: 05 jan. 2025.

LAINING, A. Exploring what parents of children under 5 years old say could help them engage in a programme about child development. 2019. 148 f. Tese (Doutorado) - University of East London, 2019. Disponível em: https://uel-repository.worktribe.com/OutputFile/484130?__cf_chl_tk=xsyfq8R6ot.eSETxtboF1rs.VrULbl1Qu9LZVHxPmuQ-1775135127-1.0.1.1-BtEckLj9N8oUc7JLyrenxv95gkILYzVGakVbG2TjK18. Acesso em: 10 dez. 2024.

LOCKWOOD, G. DINGEMANSE, M. Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in psychology*, Lausanne, Switzerland, v. 6, n.1, p. 1-14, 2015. DOI: <https://doi.org/10.3389/fpsyg.2015.01246>.

LOCKWOOD, G., HAGOORT, P., & DINGEMANSE, M. How iconicity helps people learn new words: neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra*, Oakland, EUA, v.2, n.1, p.1-15, 2016. DOI: <https://doi.org/10.1525/collabra.42>.

MARCELINO, M. Resultativas em português brasileiro. *Revista Veredas*, v.18, n.1, p.121-137, 2014. Disponível em: <https://periodicos.uff.br/index.php/veredas/article/view/24976>. Acesso em: 05 jan. 2025.

MASSON, D. I. Synesthesia and sound spectra. *Word*, v. 8, n. 1, p. 39-41, 1952. DOI: <https://doi.org/10.1080/00437956.1952.11659418>.

MAURER, D.; PATHMAN, T.; e MONDLOCH, C. J. The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, v. 9, n. 3, p. 316-322, 2006. DOI: <https://doi.org/10.1111/j.1467-7687.2006.00495.x>.

MEIRELLES, L. L. Verbos de movimento do português brasileiro: evidências contra uma tipologia binária. *Revista de Estudos da Linguagem*, Campinas, v. 27, n. 2, p. 1101, 2019. DOI: <https://doi.org/10.17851/2237-2083.27.2.1101-1124>.

METEYARD, L.; STOPPARD, E.; SNUDDEN, D.; CAPP, S. F.; VIGLIOCCO, G. When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, v.76, n.1, p. 264-275. DOI: <https://doi.org/10.1016/j.neuropsychologia.2015.01.042>.

MOREIRA, B. E. da C. A new look at the unproductivity of resultative constructions in Brazilian Portuguese. *DELTA: Documentação de Estudos Em Lingüística Teórica e Aplicada*, São Paulo, v.39, n.4, p.1-27, 2023. DOI: <https://doi.org/10.1590/1678-460X202339457357>.

MOURA, H. Telicidade, trajetória e phrasal verbs no português brasileiro. *Cadernos de Estudos Linguísticos*, Campinas, v.67, n.00, p.1-13, 2025. DOI: <https://doi.org/10.20396/cel.v67i00.8679884>.

NIELSEN, A.; RENDALL, D. The sound of round: evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology*, Ottawa, v. 65, n. 2, p. 115-124, 2011. DOI: <https://doi.org/10.1037/a0022268>.

OZTURK, O.; KREHM, M.; VOULOUMANOS, A. Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of experimental child psychology*, v. 114, n. 2, p. 173-186, 2013. DOI: <https://doi.org/10.1016/j.jecp.2012.05.004>.

PEETERS, D. Processing consequences of onomatopoeic iconicity in spoken language comprehension. In: *Cognitive Science Society*, 2016, University of California, Proceedings of the Annual Meeting of the Cognitive Science Society, California, 2016, p. 1637-1642. Disponível em: <https://www.mpi.nl/publications/item2304565/processing-consequences-onomatopoeic-iconicity-spoken-language>. Acesso em: 18 dez. 2024.

- PERLMAN, M.; LUPYAN, G. People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific reports*, v. 8, n. 2634, p. 1-14, 2018. DOI: <https://doi.org/10.1038/s41598-018-20961-6>.
- PERRY, L.; PERLMAN, M.; LUPYAN, G. Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one*, v. 10, n. 9, p. 137147, 2015. DOI: <https://doi.org/10.1371/journal.pone.0137147>.
- PERRY, L. K., PERLMAN, M., WINTER, B., MASSARO, D. W., & LUPYAN, G. Iconicity in the speech of children and adults. *Developmental Science*, v. 21, n. 3, 2018. DOI: <https://doi.org/10.1111/desc.12572>.
- RAMACHANDRAN, V. S.; HUBBARD, E. M. Synaesthesia-a window into perception, thought and language. *Journal of consciousness studies*, v. 8, n. 12, p. 3-34, 2001. Disponível em: <https://www.ingenta-connect.com/content/imp/jcs/2001/00000008/00000012/1244>. Acesso em: 10 dez. 2024.
- REICHARD, G. A.; JAKOBSON, R.; WERTH, E. Language and synesthesia. *Word*, v. 5, p. 223-233, 1949. DOI: <https://doi.org/10.1080/00437956.1949.11659507>.
- SAKAMOTO, M.; WATANABE, J. Bouba/Kiki in touch: associations between tactile perceptual qualities and Japanese phonemes. *Frontiers in Psychology*, v. 9, p. 295, 2018. DOI: <https://doi.org/10.3389/fpsyg.2018.00295>.
- SANTOS-FILHO, D. A expressão do modo de movimento no português brasileiro. 2018. 340f. Tese (Doutorado em Linguística) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, 2018. Disponível em: <https://repositorio.ufsc.br/handle/123456789/198776>. Acesso em: 14 dez. 2024.
- SILVA JÚNIOR, I. R. Verbos de movimento e sua representação na sua estrutura léxico conceptual. 2015. 176f. Tese (Doutorado em Linguística) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2015. Disponível em: <https://repositorio.ufsc.br/handle/123456789/158387>. Acesso em: 14 dez. 2024.
- SLOBIN, D. I. What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In: HICKMANN, M.; ROBERT, S. (org.). *Space in languages: Linguistic systems and cognitive categories* John Benjamins Publishing Company, 2006. p. 59-81. DOI: <https://doi.org/10.1075/tsl.66.05slo>.
- STYLES, S. J.; GAWNE, L. When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*, v. 8, n. 4, p. 1-17, 2017. DOI: <https://doi.org/10.1177/2041669517724807>.
- TALMY, L. Toward a cognitive semantics, Vol. II: Typology and process in concept structuring Cambridge: *The MIT Press*, 2000.
- TAUB, Sarah F. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press, 2001.
- WINTER, B. Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, v. 31, n. 8, p. 975-988, 2016. DOI: <https://doi.org/10.1080/23273798.2016.1193619>.
- WINTER, B.; PERLMAN, M. Size sound symbolism in the English lexicon. *Glossa: a journal of general linguistics*, v. 6, n. 1, p.1-13, 2021. DOI: <https://doi.org/10.5334/gjgl.1646>.

WINTER, B., SÓSKUTHY, M., PERLMAN, M., & DINGEMANSE, M. Trilled/r/is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, v. 12, n. 1035, p. 1-11, 2022. DOI: <https://doi.org/10.1038/s41598-021-04311-7>.

WINTER, B.; LUPYAN, G.; PERRY, L. K.; DINGEMANSE, M.; PERLMAN, M. Iconicity ratings for 14,000+ english words. *Behavior Research Methods*, v.56, n.1, p. 1640-1655, 2023. DOI: <https://doi.org/10.3758/s13428-023-02112-6>.

WINTER, B.; WOODIN, G.; PERLMAN, M. Defining iconicity for the cognitive sciences. In: FISHER, O., AKITA, K., PERNISS, P. *The Oxford Handbook of Iconicity in Language*. New York, Oxford, 2026, p. 138-150.

Sobre as propagandas dos governos Dilma e Bolsonaro: uma análise dos sentidos atribuídos à escola (2013-2019)

On the Government Campaigns of Dilma and Bolsonaro: An Analysis of the Meanings Attributed to School (2013-2019)

Nilton César Ferreira

Universidade Estadual de Campinas
(UNICAMP) | Campinas | SP | BR
nc.ferreirah@hotmail.com
<https://orcid.org/0000-0003-2694-2154>

Francisco Renato Lima

Universidade Estadual de Campinas
(UNICAMP) | Campinas | SP | BR
fcorenatolima@hotmail.com
<https://orcid.org/0000-0002-1372-5444>

Resumo: Este artigo analisa os sentidos produzidos nas campanhas publicitárias do Ministério da Educação (MEC), veiculadas entre 2013 e 2019, comparando os discursos educacionais de governos progressistas e conservadores. A análise da materialidade linguística permite compreender o funcionamento ideológico da linguagem, evidenciando concepções distintas de educação e disputas simbólicas sobre o papel da escola pública no Brasil. O estudo fundamenta-se na Análise do Discurso (AD) de linha francesa, elaborada por Michel Pêcheux na década de 1960, que concebe a linguagem como mediação entre o homem e a realidade, possibilitando tanto a permanência quanto a transformação do sujeito e do mundo em que ele vive. Nesse processo, as discursividades políticas que atravessam a educação e a escola pública orientam práticas pedagógicas e produzem sentidos sobre os lugares sociais dos sujeitos envolvidos no processo de ensino e aprendizagem – estudantes, professores e comunidade escolar. O *corpus* é composto por duas propagandas produzidas em contextos distintos: uma, durante o governo de Dilma Rousseff, alinhada a uma escola democrática voltada à inclusão e à cidadania; e outra, durante o governo de Jair Bolsonaro, vinculada à militarização das escolas públicas, enfatizando disciplina, ordem e controle moral. O estudo adota um viés comparativo, ao examinar as diferentes posições discursivas que atravessam as campanhas publicitárias,



evidenciando como cada governo materializa projetos distintos de educação e de escola. Conclui-se, assim, que os discursos analisados revelam a tensão entre um projeto de escola democrática e iniciativas voltadas à subserviência e ao controle social.

Palavras-chave: campanhas governamentais; escola pública; análise do discurso; ideologia.

Abstract: This article analyzes the meanings produced in advertising campaigns by the Brazilian Ministry of Education (MEC), aired between 2013 and 2019, comparing the educational discourses of progressive and conservative governments. The analysis of linguistic materiality allows for an understanding of the ideological functioning of language, highlighting distinct conceptions of education and symbolic disputes over the role of public schools in Brazil. The study is based on French-line Discourse Analysis (DA), developed by Michel Pêcheux in the 1960s, which conceives language as a mediation between humans and reality, enabling both the persistence and the transformation of the subject and the world in which they live. In this process, the political discursivities that traverse education and public schooling guide pedagogical practices and produce meanings regarding the social positions of the subjects involved in the teaching and learning process—students, teachers, and the school community. The corpus consists of two advertisements produced in distinct contexts: one during Dilma Rousseff’s government, aligned with a democratic school oriented toward inclusion and citizenship; and another during Jair Bolsonaro’s government, linked to the militarization of public schools, emphasizing discipline, order, and moral control. The study adopts a comparative approach, examining the different discursive positions embedded in the advertising campaigns, revealing how each government materializes distinct educational and school projects. It is concluded that the analyzed discourses reveal the tension between a democratic school project and initiatives aimed at subservience and social control.

Keywords: government campaigns; public school; discourse analysis; ideology.

1 Introdução

Toda imagem é produto, toda imagem é um dado da cultura, contribuindo para a significação ligada à ideologia da época, mesmo quando a contesta. Antes e atrás de cada imagem há, pelo menos, um homem, um olhar e a mão que cria. Esse homem, que vive entre homens, num tempo-lugar, entre natureza e máquinas, com lembranças familiares, desejos, crenças e inquietações, não pode despir-se totalmente desses vínculos. Dessa forma, por trás de grandes cadeias de produção –jornal, publicidade, propaganda, computação, revista, cinema, televisão e satélites –estão sempre os interesses pessoais, políticos e econômicos.

(Silveira, 2001, p. 49)

Na história da humanidade, a escola, enquanto espaço cultural de legitimação do saber letrado e institucionalizado, assumiu diferentes configurações e representações, conforme os contextos políticos, culturais e ideológicos de cada governo. Sempre orientada por uma base ideológica, a imagem da educação pública construída nas propagandas governamentais reflete os ideais e os interesses de cada gestão, definindo o que se acredita que a educação – e, por extensão, a escola – deve ou pode ser. Como observa Michel Pêcheux (2011), a propaganda governamental, antes restrita ao âmbito militar, passou a atuar internamente para legitimar as ações do Estado, moldando a percepção da população sobre a pertinência de seus projetos. Nesse sentido, as propagandas educacionais do governo de Jair Bolsonaro (2019-2022) projetaram uma lógica de autoritarismo e militarização, incorporando valores de ordem, disciplina e controle, em contraste com os discursos de governos anteriores, como o de Dilma Rousseff (2011-2016), que enfatizavam inclusão e cidadania (Goulart; Amaral, 2023).

No Brasil, a militarização das escolas públicas disputa espaço com o projeto de escola democrática, configurando um movimento de desdemocratização e contrariando o princípio constitucional de gestão democrática (Goulart; Amaral, 2023). Esse processo não é recente: em Goiás, desde o final da década de 1990, durante o governo de Marconi Perillo (PSDB), foi criado o Colégio da Polícia Militar de Goiás (CPMG), inspirado em escolas estaduais, sob gestão compartilhada entre Secretaria de Educação e Polícia Militar. A experiência goiana consolidou-se e, a partir de 2019, com o Programa Nacional das Escolas Cívico-Militares (PECIM), no governo de Jair Messias Bolsonaro, foi replicada em outros estados, como Amazonas, Bahia, Distrito Federal, Rondônia, Tocantins, Ceará e Paraná. Mais recentemente, o governo Tarcísio de Freitas, em São Paulo, alinhado ao projeto bolsonarista, também tentou implementar escolas militares, mas o Programa está atualmente suspenso por decisão da Justiça, diante das irregularidades apontadas por movimentos contrários (Dávila, 2025).¹

Sob uma perspectiva histórica e sociológica, a escola pode ser compreendida como um *Aparelho Ideológico do Estado* (Althusser, 2010), em meio a projetos em disputa – como a escola democrática e a escola militarizada – que refletem concepções distintas de sociedade,

¹ O governo de São Paulo, sob a gestão de Tarcísio de Freitas, instituiu em 2024 o programa das “escolas cívico-militares”, com previsão de implementação em até cem unidades da rede estadual. Contudo, o programa foi suspenso pela Justiça de São Paulo em agosto de 2024, após liminar que questionou sua constitucionalidade, bem como pela decisão do Tribunal de Contas do Estado de São Paulo (TCESP) em setembro de 2025, que apontou irregularidades como uso de verbas da educação para pagamento de policiais, falta de estudos orçamentários e ausência de metas claras, conforme matéria na Folha de São Paulo (Dávila, 2025).

cidadania e formação. As propagandas educacionais evidenciam como o Estado produz sentidos sobre a educação, projetando imagens positivas de cada governo e legitimando diferentes modelos de escola. Recentemente, essas campanhas estiveram associadas ao governo de Jair Bolsonaro, reforçando a continuidade de agendas autoritárias e neoliberais iniciadas no governo de Michel Temer (Freitas; Silva, 2018), em contraste com o projeto mais democrático do governo de Dilma Rousseff. Nesse contexto, como observa Pêcheux (2011), a propaganda deixa de ser apenas um instrumento institucional, tornando-se uma preocupação central dos governos na formação social capitalista, influenciando diretamente a construção de sentidos sobre o papel da escola e da educação.

Neste artigo, objetiva-se analisar como o funcionamento da discursividade veiculada nas propagandas do Ministério da Educação (MEC), durante os governos de Dilma Rousseff e de Jair Messias Bolsonaro, comparando o discurso em torno de um projeto educativo voltado para a inclusão, a diversidade e a formação cidadã, sustentado pelo ideal de uma educação democrática, no primeiro governo; e as propostas de militarização das escolas públicas brasileiras, associada à disciplina e ao controle moral, no segundo governo. Ademais, busca-se compreender de que modo essa instituição (o MEC), opera como *Aparelho Ideológico do Estado* (Althusser, 2010), produzindo uma discursividade que associa a escola militarizada a um modelo considerado mais eficaz – tanto no combate à violência e à indisciplina quanto na melhoria dos índices internacionais de educação.

A pesquisa fundamenta-se na Análise do Discurso (AD), de linha francesa, formulada por Michel Pêcheux (2009), que concebe o discurso como prática social atravessada por ideologias e relações de poder, refletindo as condições históricas de sua produção e as posições ocupadas pelos sujeitos. Como observa Orlandi (2000, p. 17), “não há discurso sem sujeito e não há sujeito sem ideologia”, evidenciando que todo enunciado é atravessado por processos ideológicos. Nesse contexto, a análise investiga como as campanhas publicitárias do MEC constroem representações de professores, estudantes e da própria instituição escolar, influenciando percepções sobre o papel da educação pública. Para Pêcheux (2009), o sentido não é pré-estabelecido, mas historicamente construído em meio a disputas ideológicas, emergindo das tensões entre diferentes posições ocupadas pelos sujeitos. Assim, as campanhas operam como instrumentos de interpelamento ideológico, produzindo sentidos que definem o que *pode* e *deve* ser a educação e a escola pública, ao mesmo tempo em que colocam em disputa projetos políticos e sociais, ao tempo que refletem dinâmicas contemporâneas de autoritarismo e militarização no campo educacional.

O *corpus* de análise concentra-se nas campanhas do MEC veiculadas entre 2013 e 2019, nos governos de Dilma Rousseff e de Jair Bolsonaro, entendidas como espaços de produção e de circulação de sentidos sobre a educação pública e a escola. Nesses discursos institucionais, emergem representações de professores, estudantes e da própria instituição escolar, evidenciando projetos sociais distintos: o modelo de educação democrática, voltado à inclusão e à formação cidadã, e o modelo militarizado, centrado na disciplina e no controle moral. Conforme Pêcheux (2011, p. 74), “desde Napoleão, a propaganda é um negócio estratégico, um elemento indispensável no prosseguimento das operações militares contra o inimigo exterior”, destacando a dimensão política da comunicação estatal. Assim, a AD oferece o arcabouço teórico-metodológico para compreender como essas campanhas produzem sentidos e consolidam diferentes visões sobre a educação e a escola.

Assim, o artigo organiza-se em três grandes partes: na primeira, apresenta-se uma breve abordagem sociológica sobre educação pública e escola; na segunda, são examinados os pressupostos teóricos e metodológicos da AD, em diálogo com o referencial adotado; e, por fim, desenvolve-se a análise comparativa das campanhas publicitárias do MEC, veiculadas nos governos de Dilma Rousseff e de Jair Bolsonaro, evidenciando, por meio de um trabalho *da/na/pela* língua(gem), os sentidos atribuídos tanto à educação pública quanto à escola – em cada contexto político-ideológico e as concepções de escola que cada governo buscou legitimar.

2 Do ponto de vista sociológico, a educação e a escola

Segundo Luckesi (1994), a educação – e, por extensão, a escola – historicamente se articulou a partir de três tendências político-filosóficas que refletem diferentes concepções sobre seu papel social: educação como redenção, como reprodução e como transformação da sociedade. Cada tendência revela tensões ideológicas no campo educacional.

A concepção redentora, segundo Luckesi (1994, p. 38), entende a sociedade como “um conjunto de seres humanos que vivem e sobrevivem num todo orgânico e harmonioso”, em que desvios de grupos ou indivíduos resultam de marginalidade ou inserção recente. O papel da educação, nessa perspectiva, é reintegrar tais elementos à estrutura social, promovendo a coesão e a manutenção da ordem. João Amós Comênio, educador do século XVI, exemplifica essa visão ao propor a educação como meio de redimir a sociedade, investindo nas novas gerações e orientando suas ações pelos ensinamentos bíblicos.

Saviani (2012, p. 15) caracteriza essa perspectiva como uma “teoria não crítica da educação”, por compreender a educação apenas como um instrumento capaz de corrigir desvios e transformar a sociedade, sem considerar as determinações sociais e históricas que influenciam o processo educativo:

Como já assinalai, o primeiro grupo de teorias concebe a marginalidade como um desvio, tendo a educação por função a correção desse desvio. A marginalidade é vista como um problema social e a educação, que dispõe de autonomia em relação à sociedade, estaria capacitada a intervir eficazmente, transformando a sociedade, tornando-a melhor, corrigindo as injustiças; em suma, promovendo a equalização social. Essas teorias consideram, pois, apenas a ação da educação sobre a sociedade. Porque desconhecem as determinações sociais do fenômeno educativo, eu as denominei de “teorias não críticas” (Saviani, 2012, p. 15, grifos nossos).

Na perspectiva reprodutivista, a escola integra a sociedade e cumpre a função de reproduzir o modelo vigente, formando sujeitos para ocupar posições específicas na divisão do trabalho e perpetuar a estrutura de classes (Saviani, 2012). Diferentemente da concepção redentora, não busca corrigir desvios, mas assegurar a manutenção da ordem social. Nessa linha, Luckesi (1994) retoma Althusser (2010), define a escola como *Aparelho Ideológico do Estado*, responsável por reproduzir os meios materiais e humanos necessários à continuidade da produção e da sociedade, pois uma formação social que não reproduzir as condições de produção ao mesmo tempo em que as produz, não conseguirá sobreviver por muito tempo.

Inserida nessa lógica, a escola produtivista desempenha papel estratégico na preparação de sujeitos para funções específicas na divisão social do trabalho, naturalizando desigualdades e consolidando a estrutura de classes. Althusser (2010, p. 79) ressalta que “uma última parcela chega ao final do percurso, seja para cair num semidesemprego intelectual, seja para fornecer [...] os agentes da exploração, os agentes da repressão e os profissionais da ideologia”, evidenciando como a educação contribui diretamente para a reprodução social e das hierarquias existentes.

A tendência crítico-transformadora compreende a educação como prática social voltada à transformação da realidade, e surge em oposição às concepções redentora e reprodutivista. Nessa perspectiva, a escola atua na democratização do ensino e na emancipação dos sujeitos do processo de ensino-aprendizagem, reconhecendo os condicionantes históricos, ideológicos e culturais que a atravessam. Segundo Luckesi (1994, p. 49), “se o projeto for transformador, medeia a transformação”, o que evidencia que a função da educação depende do projeto que orienta a prática pedagógica. Embora frequentemente subordinada à lógica reprodutivista da sociedade capitalista, a educação pode também configurar-se como instrumento crítico, capaz de intervir na realidade e de promover mudanças sociais, desde que se reconheçam suas contradições internas e se aja politicamente no âmbito pedagógico (Luckesi, 1994).

Os debates filosóficos sobre educação mostram que a escola não possui uma função única ou neutra. Inserida em uma sociedade de classes e atravessada por interesses sociais, culturais e políticos antagônicos, ela configura-se como espaço de disputa simbólica e ideológica, no qual diferentes grupos sociais projetam seus valores e finalidades. Como observa Tozoni-Reis (2010, p. 56), a escola “não é uma instituição social neutra”, sendo o processo educativo influenciado pelos valores e intenções dos grupos que disputam seu lugar na hierarquia social. Assim, a educação escolarizada pode ter funções redentoras, reprodutivas ou transformadoras, dependendo do projeto que orienta sua prática (Luckesi, 1994).

É sob essa perspectiva de disputa ideológica e simbólica (Althusser, 2010) que este estudo analisa a construção da imagem de educação e de escola pública em campanhas do MEC nos governos de Dilma Rousseff e Jair Bolsonaro. Busca-se compreender como tais discursos produzem sentidos sobre a função social da escola, articulando modelos distintos – da escola democrática, voltada à inclusão e cidadania, e da escola militarizada, associada à disciplina e ao controle moral – e evidenciando de que forma a comunicação institucional contribui para legitimar determinados projetos políticos e sociais.

3 Do ponto de vista da AD, o sujeito e a ideologia

Para a AD, não há falante, locutor, muito menos emissor. Há sujeito (alternativamente, enunciador). O que é, evidentemente, na esteira das rupturas com a Pragmática e com as teorias linguísticas dominantes, outra ruptura, provavelmente, a mais importante para a teoria. Segundo Possenti (2011, p. 388), “a AD rompe com a concepção de sujeito uno, livre, caracterizado pela consciência (isto é, sem inconsciente, sem ideologia) e tomado como origem”. O sujeito não é o indivíduo biológico, sujeito empírico, mas o sujeito do discurso, que, segundo Grigoletto (2005, p. 01), “carrega consigo marcas do social, do ideológico, do histórico e tem a ilusão de ser a fonte do sentido. A teoria do discurso trabalha a ilusão do sujeito como origem, através dos processos discursivos, mostrando que linguagem e sentido não são transparentes”.

Althusser (1973, p. 67) está na origem das formulações essenciais da AD, também no que diz à questão do sujeito e do sentido. É possível ler em seus textos, passagens como as seguintes: “os homens (no plural) concretos são necessariamente sujeitos (plural) *na* história, pois atuam *na* história enquanto sujeitos (no plural). Mas não há Sujeito (no singular) *da* história. Iria até mais longe: “os homens” não são “os sujeitos” da história”. Segundo Possenti (2011, p. 387), o texto, como se vê, “está repleto de distinções sutis, com itálicos, parênteses e aspas, que ficam mais claras em outras passagens”. Assim, Althusser (1973, p. 67) diz que:

[...] os indivíduos humanos, ou seja, sociais, são *ativos* na história – como *agentes* das diferentes práticas sociais do processo histórico de produção e de reprodução – é um fato. Mas [...] não são sujeitos “livres” e “constituintes” no sentido filosófico desses termos” (Althusser, 1973, p. 67, grifos nossos).

Posto isso, é possível resumir o breve texto de Althusser (1973, p. 67) da forma seguinte: “não há Sujeito (no singular)”, há sujeitos; não há sujeitos da história, há sujeitos na história. Isto equivale a dizer que a história não se reduz a um Sujeito que seja Único e Absoluto, que satisfaça os predicados atribuídos, por exemplo, a Deus, um ser que seja Origem da história, Aquele que a faz. De outro modo, a história é feita por sujeitos, mas não o sujeito em sua concretude, ou seja, não é o sujeito empírico, mas o sujeito discursivo, que tem a ilusão de ser um alguém plenamente consciente e capaz de controlar o seu dizer, mas que, na verdade, se constitui através da contradição, pois, ele é ao mesmo tempo livre e submisso: “pode tudo dizer, contanto que se submeta à língua para sabê-la” (Orlandi, 2000, p. 50).

Ainda, Althusser (2010, p. 102) mostra a necessidade do desdobramento do Sujeito em sujeitos, e do próprio Sujeito em sujeito-Sujeito (o dogma da Trindade). E, então, diz que a estrutura de toda ideologia é especular e duplamente especular, pois, “submete os sujeitos ao Sujeito, dando-lhes no Sujeito [...] a garantia de que certamente trata-se deles e Dele”. Por sua vez, a estrutura duplicada da ideologia, conforme Althusser (2010, p. 102-103), assegura: 1. a interpelação dos “indivíduos” como sujeitos; 2. sua submissão ao Sujeito; 3. o reconhecimento mútuo entre os sujeitos e o Sujeito, e entre os próprios sujeitos e, finalmente, o reconhecimento do sujeito por ele próprio; 4. a garantia absoluta de que está tudo bem, assim, e sob a condição de os sujeitos reconhecerem o que são e se conduzirem de acordo, tudo irá bem: “assim seja”.

A partir dessas reflexões em Althusser (2010, p. 104), entende-se que a definição de *sujeito*: a de ser sujeito e a de assujeitar-se. No sujeito, há uma “*subjetividade livre*” (um centro de iniciativa, responsável pelos seus atos) e um *ser submetido* (sujeito a uma autoridade maior, sem liberdade, salvo a de aceitar livremente a sua submissão). É assim que se expressa o “livre arbítrio”, em sua duplicidade: sujeitos *submetidos* ao Sujeito/sujeitos *reconhecidos* pelo Sujeito. Segundo Orlandi (2009, p. 242), aplicado, por exemplo, ao conteúdo da ideologia religiosa, esta “se constitui de uma contradição, uma vez que a noção de livre arbítrio traz, em si, a de coerção”. Em se tratando de coerção, é preciso dizer que não se trata de coerção física: “o funcionamento da ideologia transforma a força em direito e a obediência em dever”.

Gerth e Mills (1973), por sua vez, colocam a noção de “autoridade” como fundamental para o conceito de instituição. Enquanto um poder legitimado, a autoridade (do professor etc.) garante a instituição (no caso, educacional). O comando só pode ser exercido, se tiver como contraponto a obediência, a qual só se mantém pela coerção. Sem a legitimação, o poder não se exerce e, para se legitimar, é preciso que o poder se recubra de atributos como,

por exemplo, a escola, como um saber letrado, e outros valores culturais que definem “finalidades” aceitáveis, assim como as “responsabilidades” daqueles que exercem o poder. O mais forte, como diz Rousseau (1983, p. 25), “nunca é suficientemente forte para se sempre o senhor, senão transformando sua força em direito e a obediência em dever”.

Ainda, a obediência se baseia, segundo Gerth e Mills (1973), na ideia que aquele que obedece tem daquele que exerce uma posição de poder. Retomando Pêcheux (2010), Lagazzi (1988, p. 38-39) diz que é preciso falar essa relação em termos de formações imaginárias: “aquele que obedece faz uma imagem de si mesmo, do lugar que ocupa na relação com o outro, desse outro que exerce o poder, do lugar ocupado por este, e assim sucessivamente”. Por sua vez, “aquele que exerce o poder também constrói uma imagem de si próprio, do lugar que ocupa, do outro que está sob sua autoridade, do lugar deste. É um movimento que se prolonga numa sucessão de imagens, como num jogo de espelhos”. Na relação discursiva, são as formações imaginárias que constituem as diferentes posições discursivas.

Do ponto de vista da AD, falar de ideologia é deslocar esse conceito para a “formação ideológica” (doravante FI), que serve para designar um elemento suscetível de intervir como:

[...] uma força em confronto com outras forças na conjuntura ideológica característica de uma formação social em dado momento; desse modo, cada formação ideológica constitui um conjunto complexo de atitudes e de representações que não são nem “individuais” nem “universais” mas se relaciona mais ou menos diretamente a posições de classes em conflitos umas com as outras (Pêcheux; Fuchs, 2010, p. 163, grifos nossos).

Visto nesta perspectiva, uma FI comporta necessariamente mais de uma posição, capaz de confrontar uma com a outra. Não obstante, em uma FI, as forças não precisam estar necessariamente em confronto, elas podem entreter entre si relações e aliança ou de dominação (Pêcheux; Fuchs, 2010). Por sua vez, é preciso esclarecer que o conceito de FD é utilizado pela AD para designar o lugar onde se articulam o discurso e a ideologia:

[...] as formações ideológicas de que acabamos de falar “comportam necessariamente”, como um de seus componentes, uma ou várias *formações discursivas* interligadas que determinam o que pode e deve ser dito (articulado sob a forma de uma arenga, um sermão, um panfleto, uma exposição, um programa etc.) a partir de uma posição dada numa conjuntura, isto é, numa certa relação de lugares no interior de um aparelho ideológico, e inscrita numa relação de classes (Pêcheux; Fuchs, 2010, p. 163-164, grifos nossos).

Nesse sentido, uma FD é marcada por “regras de formação”, ou mecanismos de controle, que determinam o interno (o que pertence) e o externo (o que não pertence) de uma FD. Ao se definir em relação a um externo, a FD não se trata de um espaço estrutural fechado. Assim, ao se considerar, por exemplo, uma FD capitalista, sua constituição ocorrerá a partir de um sistema de paráfrases, pois os enunciados que buscam promover uma educação puramente reprodutivista são retomados e reformulados “num esforço constante de fechamento de suas fronteiras em busca da preservação de sua identidade” (Brandão, 2012, p. 48), na qual “a escola é a sede da reprodução cultural” (Orlandi, 2009, p. 22).

4 Sobre a metodologia

A presente pesquisa analisa de que maneira a educação pública e a escola são significadas em campanhas publicitárias do MEC, a partir de dois vídeos produzidos em contextos distintos: durante o governo de Dilma Rousseff (2013-2016) e durante o governo de Jair Bolsonaro (2019-2022). O primeiro vídeo aborda políticas de Educação Integral, Programa Universidade para Todos (ProUni), Fundo de Financiamento ao Estudante do Ensino Superior (Fies) e Ciência sem Fronteiras² (Brasil, 2013)³, enquanto o segundo apresenta o lançamento do Programa Nacional das Escolas Cívico-Militares (Brasil, 2019a)⁴. O propósito é compreender, à luz da AD, de filiação francesa, como se configuram as formações discursivas e imaginárias inscritas nas diferentes propostas de governo, tomando como referência as categorias analíticas já estabelecidas e que orientam a leitura crítica dos enunciados.

Os vídeos foram publicados pelo MEC no *YouTube*, plataforma que alcança estudantes, familiares, profissionais da educação e a sociedade em geral. Como suporte digital que redefine práticas de circulação e recepção de sentidos, constituem manifestações de gêneros emergentes. Segundo Marcuschi (2010), o sucesso da tecnologia decorre de reunir texto, som e imagem em um só meio, permitindo a incorporação simultânea de múltiplas semioses, o que reforça a pertinência de considerar as campanhas governamentais sob uma abordagem discursiva.

O *corpus* se constitui a partir desses vídeos, que foram transcritos, tabulados e organizados em categorias analíticas. A análise seguiu os princípios teóricos e metodológicos da AD (Orlandi, 2000), considerando o contexto sócio-histórico como parte constitutiva do discurso. Para tanto, foi utilizada a noção de sequência discursiva (SD), que permite depreender o *corpus* de análise em sequências linguísticas nucleares, “cuja realização representam, no fio do discurso (ou intradiscurso), o retorno da memória (a repetibilidade que sustenta o interdiscurso)” (Mariani, 1998, p. 53).

A escolha da AD justifica-se por sua capacidade de analisar a ordem dos enunciados e evidenciar a disputa ideológica nas campanhas governamentais, mostrando como a ideologia produz efeitos na língua(gem) (Orlandi, 2000). Essa abordagem permite comparar discursos de governos progressistas e conservadores e revelar como políticas e discursos institucionais atribuem sentidos distintos sobre a educação: inclusão e escola democrática no governo Dilma, e hierarquia e disciplina no governo Bolsonaro. Assim, a metodologia integra a análise das SD, das condições de produção e das articulações ideológicas, oferecendo compreensão de como a educação foi discursivizada entre 2013 e 2019.

² É preciso esclarecer que os programas de Educação Integral, ProUni e Fies foram iniciados no governo de Luiz Inácio Lula da Silva (2003-2010) e mantidos por Dilma Rousseff, com o objetivo de ampliar o tempo escolar, oferecer bolsas de estudo e financiar cursos universitários. Já o Ciência sem Fronteiras, lançado por Dilma, promovia intercâmbios acadêmicos no exterior.

³ Disponível em: <https://www.youtube.com/watch?v=zN7cqOj-4A8>. Acesso em: 01 abr. 2026.

⁴ Disponível em: <https://www.youtube.com/watch?v=Jz5Q5PAFSL0>. Acesso em: 01 abr. 2026.

5 Sobre as propagandas produzidas pelo Ministério da Educação

A AD considera constitutivo do sentido o contexto histórico-social e as condições de produção do discurso (Brandão, 2012), neste caso, as propagandas do MEC veiculadas entre 2013 e 2019, nos governos Dilma Rousseff e Jair Bolsonaro. A partir de 2014, intensificaram-se movimentos conservadores que impulsionaram o clamor por um “novo” modelo de escola pautado na disciplina e na obediência, favorecendo a militarização da educação. Nesse cenário, o movimento *Escola Sem Partido*⁵ e o discurso de campanha do governo de Jair Bolsonaro convergiram em torno da moralização e do silenciamento do pensamento crítico, configurando o que Orlandi (2010) denomina “política do silêncio”, em que ao afirmar determinados sentidos, outros possíveis são apagados. Sobre a “política do silêncio”, Orlandi (2010, p. 75) aponta que: “[...] se define pelo fato de que ao dizer algo apagamos necessariamente outros sentidos possíveis, mas indesejáveis, em uma situação discursiva dada”.

O processo de militarização das escolas públicas brasileiras ocorreu em duas etapas. A primeira, em Goiás (1999), com a criação das Escolas Públicas Militarizadas (EPMs), transferindo parte da gestão civil à Polícia Militar e ao Corpo de Bombeiros, prometendo disciplina e eficiência administrativa e funcionando como laboratório de controle sobre a educação pública (Goulart; Amaral, 2023). A segunda etapa, no governo Bolsonaro, quando instituiu o Programa Nacional das Escolas Cívico-Militares (PECIM), pelo Decreto n.º 10.004/2019, que padronizou o modelo híbrido de gestão, associando hierarquia, disciplina e financiamento federal, mantendo práticas autoritárias sem avanços pedagógicos significativos (Dias; Ribeiro, 2021).

Goulart e Amaral (2023), inspirando-se em Tilly (2013), entendem a desdemocratização como o enfraquecimento da participação social e a concentração de poder. Nesse contexto, a militarização das escolas públicas expressa esse fenômeno, impondo hierarquia e disciplina, restringindo a gestão pedagógica e a participação da comunidade escolar. A indicação de diretores por autoridades militares e a limitação da participação estudantil a alunos maiores de idade e com “comportamento exemplar” reforçam esse processo, enfraquecendo princípios da gestão democrática⁶ (Goulart; Amaral, 2023, p. 11). Nesse contexto, esse movimento revela uma disputa ideológica sobre os sentidos da educação pública, ancorando-se em uma escola historicamente marcada por uma tendência produtivista.

Diante disso, a proposta de escola militarizada surge na contramão das políticas educacionais implementadas por governos anteriores, especialmente os de Luiz Inácio Lula da Silva (2003-2010) e Dilma Rousseff (2011-2016). Enquanto esses governos priorizaram a ampliação do acesso e da permanência na educação, por meio de programas como a Educação Integral, o ProUni e o Fies, voltados, respectivamente, à ampliação do tempo escolar, à oferta de bolsas de estudo e ao financiamento de cursos universitários, também investiram em iniciativas como o Ciência sem Fronteiras, que promovia intercâmbios acadêmicos e o fortale-

⁵ O *Escola Sem Partido* é um movimento iniciado em 2004, que busca “neutralizar” a educação, limitando debates sobre gênero, sexualidade e política (Silva, 2020). Ganhou visibilidade a partir de 2014 e, no governo de Jair Bolsonaro, suas ideias se articulam com a proposta de escolas militarizadas, promovendo disciplina e valores conservadores, sendo divulgadas por Carlos Bolsonaro como forma de combater a chamada “doutrinação” escolar.

⁶ O princípio da gestão democrática, previsto no artigo 206, inciso VI, da Constituição Federal de 1988 (Brasil, 1988), assegura a participação de diferentes segmentos da comunidade escolar – como professores, estudantes, pais e funcionários – nos processos de decisão e na organização das instituições de ensino, orientando.

cimento da pesquisa científica. Essas ações expressavam um projeto de ressignificação da educação pública, entendida como instrumento de transformação social, inclusão e participação democrática (Brasil, 2013). Contudo, essa perspectiva foi interrompida com o governo de Michel Temer (2016-2018), cujo plano “Ponte para o Futuro” marcou uma guinada neoliberal na política educacional, ao promover cortes orçamentários, flexibilização curricular e priorização de interesses de mercado, resultando na supressão das políticas de democratização e justiça social que vinham sendo consolidadas (Tenório Neto; Ericson, 2020).

Freitas e Silva (2018), por sua vez, destacam que as propagandas governamentais não se limitam a técnicas ou efeitos psicológicos sobre os sujeitos envolvidos no processo do ensino-aprendizagem, mas funcionam como práticas discursivas historicamente situadas, capazes de projetar uma imagem positiva do governo e de seus feitos. Nesse sentido, seguindo Pêcheux (2010, p. 85), essas campanhas mobilizam formações imaginárias, ou seja, “representações imaginárias das diferentes instâncias do processo discursivo”, que derivam de processos discursivos anteriores e se inscrevem em contextos sócio-históricos específicos. No caso das propagandas sobre educação, tais práticas visam produzir sentidos favoráveis às políticas anunciadas, mesmo que não resultem em benefícios concretos, configurando-se como instrumentos de disputa ideológica sobre a significação da escola e da educação pública.

A partir das reflexões sobre as tendências educacionais, inicia-se, neste momento, a análise da primeira campanha publicitária do MEC, veiculada em 2013 durante o governo de Dilma Rousseff, a qual aborda políticas de Educação Integral, ProUni, Fies e Ciência sem Fronteiras (Brasil, 2013), para, posteriormente, compará-la com as campanhas do governo de Jair Bolsonaro:

SD₁ – “As ações do governo federal estão dando um novo impulso a Pernambuco. Quer ver um exemplo?” “Hoje 2400 escolas daqui adotam educação de tempo integral.” “Eles estudam num turno” “No outro horário, reforço escolar, com aulas de português e matemática.” “Hip-hop, capoeira, futebol.” “Tem contato com a cultura local.” “Cinema, teatro.” “Isso traz uma melhoria no seu desempenho escolar.” “A gente fica feliz, né? A gente termina aprendendo mais.” “A gente se diverte e aprende, tudo ao mesmo tempo” (Brasil, 2013, grifos nossos).

Em 2013, durante o governo Dilma Rousseff, o MEC lançou uma campanha publicitária que apresentava Pernambuco como símbolo dos avanços das políticas públicas em educação, exemplificado na SD₁: “As ações do governo federal estão dando um novo impulso a Pernambuco”. Nesse enunciado, a nominalização⁷ “ações” transforma decisões políticas complexas em um substantivo abstrato, conferindo-lhes aparência de objetividade, enquanto a metáfora econômica “impulso” sugere que a educação funciona como um investimento que gera resultados mensuráveis, deslocando o foco do direito à educação para o retorno sobre o investimento. Esse jogo de linguagem, na SD₁, conduz o interlocutor a se deparar com uma discursividade que apresenta a política educacional no governo Dilma como eficiente e progressista, reforçando a ideia de que essa agenda governamental promove melhorias concretas, contrapondo a imagem de um estado historicamente marcado pela precariedade educacional à de um

⁷ Termo referido a partir de Basilio (2017) que, sob uma perspectiva funcional, refere que a nominalização constitui-se de um processo de associação lexical sistemática entre os verbos e os nomes, apresentando, com isso, funções sintáticas e textual-discursivas, seja de natureza predadora, designadora ou anafórica.

território em ascensão, sustentada pelos dados do Índice de Desenvolvimento da Educação Básica (IDEB), que em 2013 registrou o maior crescimento no Ensino Médio, subindo 12 posições no *rancking* nacional (Brasil, 2014).

O adjetivo “*novo*” – como evidenciado, na SD₁, na formulação “*um novo impulso a Pernambuco*” – opera como marcador de economia e configura sentidos típicos de uma ideologia neoliberal, ao apresentar a educação como investimento que deve gerar resultados mensuráveis e ao sugerir que é possível produzir mais com menos recursos. A metáfora econômica desloca o sentido de educação como um direito social para um bem de consumo, regulado por parâmetros de produtividade e de rentabilidade (Foucault, 2008). Pernambuco, situado historicamente entre os piores índices educacionais, é mobilizado como exemplo eficaz para evidenciar o “*retorno*” desse investimento estatal, funcionando como prova do sucesso da política do governo Dilma. Embora o discurso se marque como pertencente a uma agenda democrática e crítica de escola (Luckesi, 1994; Saviani, 2012), inscreve-se também em uma FD de orientação capitalista, subordinando a educação à racionalidade econômica e aos valores de desempenho e de eficiência.

Essa discursividade articula sentidos alinhados à tendência crítico-transformadora – sustentada em enunciados como “*com aulas de português e matemática*”, “*hip-hop, capoeira, futebol*” –, que entende a educação como prática social voltada à emancipação e à transformação da realidade (Luckesi, 1994; Tozoni-Reis, 2010). Entretanto, embora se anuncie transformadora e inclusiva, opera segundo uma lógica produtivista. A expressão “*um novo impulso*” mobiliza a ideia de fazer funcionar com economia e eficiência – como em “*a gente termina aprendendo mais*” –, ecoando a “*racionalidade de mercado*” descrita por Foucault (2008), segundo a qual é necessário produzir maiores resultados com recursos limitados. Assim, o discurso da eficiência vincula-se à promessa de transformação social, como em “*melhoria no seu desempenho escolar*”, em que a nominalização transforma o ato de desempenhar em algo concreto e mensurável – um resultado escolar –, legitimando a atuação do Estado como promotor do progresso educacional, ainda que atravessado por valores neoliberais. Esse funcionamento produz um efeito de objetividade e naturalização, como se o desempenho fosse um dado técnico e quantificável, e não o resultado de condições sociais, afetivas e pedagógicas.

Na SD₁, o interlocutor é levado a perceber a educação como resultado direto do investimento estatal, como se os avanços decorressem exclusivamente das “*ações do governo federal*”, e não do protagonismo da comunidade escolar. A nominalização “*ações*” opera como recurso de legitimação (Gerth; Mills, 1973), reforçando a imagem de um Estado ativo e comprometido, enquanto apaga outros sujeitos do processo educativo. O protagonismo da comunidade escolar é deslocado para o Estado e para o capital, como se observa em “*um novo impulso*”, em que o adjetivo “*novo*” produz efeito de renovação e de aceleração das políticas governamentais. Essa metáfora econômica associa a educação à lógica da produtividade e do investimento. Nesse funcionamento, “*impulso*” opera como nominalização que materializa o sentido de investimento, sustentando a ideia de que o progresso educacional depende da eficiência das políticas estatais. Ao articular educação e transformação social, o discurso mobiliza sentidos de eficiência e de produtividade, próprios de uma *governamentalidade neoliberal* que converte o investimento público em capital e resultados mensuráveis (Foucault, 2008). Assim, coexistem o discurso da inclusão e da valorização do ensino público e aquele da produtividade e da economia, revelando a inscrição de uma ideologia neoliberal em um projeto político que se apresenta como democrático e transformador.

Pêcheux (2009) afirma, assim, que o lugar do sujeito é preenchido por uma *forma-sujeito*, isto é, pelo modo como o indivíduo se inscreve em uma FD e se reconhece nos sentidos que ela (re)produz. Nesse funcionamento, o enunciador que afirma “*A gente fica feliz*” inscreve-se na *forma-sujeito capitalista*, própria de uma FD neoliberal, em que a felicidade é mensurada em termos de produtividade e de eficiência. O adjetivo “*feliz*” remete à satisfação pessoal e ao reconhecimento das identidades – como em “*hip-hop, capoeira, futebol*” –, mas dissimula a lógica produtivista, associando a “*alegria do aprender*” à ideia de resultado. A formulação “*A gente termina aprendendo mais*” reforça tal racionalidade: o verbo “*termina*” funciona como metáfora de produtividade, enquanto o advérbio “*mais*” intensifica a noção de progresso contínuo. Assim, a *forma-sujeito* realiza um movimento de incorporação-dissimulação dos sentidos do interdiscurso⁸, reinscrevendo sentidos ligados à concepção reprodutivista de educação (Luckesi, 2014; Tozoni-Reis, 2010), na qual o prazer e o sucesso dependem da adaptação às exigências do sistema capitalista.

Ainda na SD₁, é a *forma-sujeito capitalista* que sustenta o efeito de unidade-evidência do sujeito da discursividade produzida na campanha publicitária. Como observa Grigoletto (2005), esse efeito é apenas aparente, pois constrói-se a imagem de uma educação “*transformadora*” – sugerida por enunciados que evocam práticas culturais, como “*hip-hop, capoeira, futebol*” ou “*tem contato com a cultura local*”. Sob essa aparência de inclusão, produz-se uma unidade imaginária: a de uma escola que, embora se enuncie democrática, permanece atravessada por uma *governamentalidade capitalista*, submetendo a ideia de transformação aos valores da produtividade, do investimento e da adaptação do sujeito às exigências do mercado. Essa racionalidade se materializa em enunciados como “*eles estudam num turno*” e “*no outro horário, reforço escolar*”, que operam uma lógica de organização do tempo escolar análoga à lógica fabril – marcada por turnos, metas e rendimento. A escola assume o lugar de uma linha de produção, em que o aprender se converte em tarefa e o aluno em operário do próprio desempenho (Foucault, 2008). Nesse sentido, a educação funciona como instrumento de formação de subjetividades ajustadas a essa lógica (Veiga-Neto, 2011).

Os compromissos da campanha governamental acabam subordinados a uma agenda mercadológica que redefine a finalidade do ensino público: formar sujeitos “*competentes*” e “*empreendedores*”, ajustados ao sistema produtivo (Veiga-Neto, 2011). Esse funcionamento discursivo se manifesta no enunciado “*A gente termina aprendendo mais*”, em que o verbo “*terminar*” opera metaforicamente como marcador de uma lógica de produtividade e resultado, deslocando o aprender do campo formativo para o do desempenho. Mesmo ao mobilizar expressões como “*hip-hop, capoeira, futebol*” ou “*contato com a cultura local*”, que evocam inclusão e diversidade, a campanha reinscreve essas práticas em uma racionalidade instrumental voltada à eficiência e à empregabilidade. Dessa forma, a *forma-sujeito capitalista* reinscreve sentidos que concebem a escola como reprodutora da ordem social (Luckesi, 1994), sustentando uma ideologia orientada pela racionalidade neoliberal. O “*saber-fazer*” promovido pela política educacional, como observa Althusser (2010), implica também um “*saber-interpretar*” que, sob o discurso de uma educação emancipadora, assegura a reprodução das ideologias dominantes e a naturalização do sujeito ajustado ao capital.

⁸ O conceito de interdiscurso, aquilo que fala antes e alhures de outro lugar, possibilita ao analista relacionar cada dizer a uma rede de outros dizeres, inscritos em uma memória coletiva. Dessa forma, torna-se possível reconhecê-lo em sua historicidade e em sua significância, evidenciando os compromissos políticos e ideológicos que o sustentam (Orlandi, 2000).

Nesse movimento, Pêcheux (2011, p. 80) observa que pensar a campanha publicitária na formação social capitalista requer “*fazer história*”, isto é, relacionar os sentidos às condições materiais de sua produção. A campanha do MEC, veiculada em 2013, produz o efeito de uma educação inclusiva e transformadora ao mobilizar enunciados como “*hip-hop, capoeira, futebol*” e “*a gente se diverte e aprende, tudo ao mesmo tempo*”. Entretanto, sob essa aparência democrática, o discurso é atravessado por uma racionalidade produtivista que redefine o aprender em termos de desempenho e de eficiência, convertendo a escola em instrumento de preparação para o trabalho (Luckesi, 1994). A discursividade produzida na SD₁, assim, reafirma seus compromissos com a *governamentalidade capitalista* que naturaliza o sujeito ajustado às demandas do capital (Foucault, 2008). O enunciado “*a gente*” mobiliza ideias de coletividade e totalidade: funciona como um “*todo*”, dando impressão de que todos compartilham as mesmas experiências e aprendizagens. Esse “*todo*” cria um efeito de completude que aparenta harmonia, ao mesmo tempo em que apaga as contradições sociais, de gênero, sexual e racial, entre outras, presentes na realidade escolar e nas relações de inclusão, diversidade e desigualdade.

O enunciado “*tudo ao mesmo tempo*”, presente na SD₁, sugere que é possível conciliar o currículo escolar com atividades como “*hip-hop, capoeira, futebol*”, “*cinema, teatro*” e o reforço escolar, com aulas de português e matemática, criando a aparência de uma infância lúdica e diversificada. A pronominalização “*a gente*” reforça essa ideia ao produzir sentidos de coletividade e de universalidade, reduzindo a pluralidade de sujeitos a um “*todo*” homogêneo, como se *todos* compartilhassem das mesmas experiências, condições e oportunidades. Essa ilusão de completude apaga contradições sociais presentes na realidade escolar e reforça a crença meritocrática de que o sucesso depende apenas do esforço individual (Foucault, 2008). Desse modo, o efeito de harmonia produzido por enunciados como “*A gente fica feliz*” reforça a imagem de uma educação emancipadora e prazerosa, quando, na verdade, submete a infância a uma lógica produtivista, na qual horários, metas e resultados mensuráveis passam a organizar sua rotina, convertendo o direito de aprender em um dever de desempenho.

Ainda analisando a campanha do MEC durante o governo Dilma, a atenção recai sobre políticas como Educação Integral, ProUni, Fies e Ciência sem Fronteiras (Brasil, 2013):

SD₂ – “Também tem muito apoio federal para os universitários. Olha só!” “O PROUNI garante bolsas para jovens de baixa renda.” “O FIES financia o estudo de milhares de jovens e o Ciência sem Fronteiras premia bons estudantes, como o Filipe, Marcelo e Alan com bolsas no exterior.” “Estudar nos Estados Unidos, numa universidade de ponta.” “Poder estudar com pessoas do mundo inteiro.” “Os laboratórios, as clínicas.” “E tudo isso com investimento do Brasil nos jovens.” “Até a próxima.” Brasil, um país rico é um país sem pobreza (Brasil, 2013, grifos nossos).

A democracia participativa implica a criação “de novas formas de representação social, na qual os cidadãos devem estar dispostos a pressionar e contribuir para as transformações necessárias” (Macedo; Lamosa, 2015, p. 135), ou seja, requer engajamento ativo na construção das políticas públicas, e não apenas a presença formal do Estado. Essa perspectiva é mobilizada nas propagandas do governo de Dilma Rousseff, que apresentam a educação como via de realização dessa democracia. Na SD₂, o enunciado “*Também tem muito apoio federal para os universitários*” lexicaliza o termo “*apoio*”, transformando-o em substantivo por meio de nominalização, o que confere estabilidade e legitimidade à ação estatal. O advérbio “*muito*” intensifica o apoio, sugerindo abrangência e alcance da intervenção pública. Esse *apoio* desliza para

sentidos de garantismo de direitos e efetivação de políticas voltadas a “*jovens de baixa renda*” e “*minorias sociais*”, mas sem explicitar o papel ativo dos cidadãos. Nesse mesmo movimento, o efeito de realidade produzido pela campanha mobiliza sentidos de engajamento e inclusão, mas não convoca efetivamente os cidadãos à ação, limitando-se a criar uma ideia de participação e evidenciando a distância entre a promessa de democracia participativa e o real do discurso publicitário (Orlandi, 2000), que se apresenta como resultado de investimentos.

Ao se observar mais detidamente a materialidade discursiva na SD₂ – como em “*Tem muito apoio federal para os universitários*” –, percebe-se que o funcionamento da discursividade produzida na campanha publicitária se constitui por uma contradição fundamental. O advérbio “*muito*” opera como intensificador, produzindo o efeito de abundância, eficiência e produtividade da ação estatal, enquanto a formulação “*apoio federal*” desliza para sentidos de capital financeiro, de investimento monetário que o governo direciona à educação. Em outros termos, o governo Dilma se representa como aquele que aplica recursos mensuráveis para alcançar resultados imensuráveis, convertendo a educação em um campo de cálculo e de rendimento. Nesse movimento, o enunciado faz operar uma racionalidade econômica no interior do discurso político, em que o Estado assume a posição de gestor eficiente dos recursos públicos – um modo de governar afinado à *governamentalidade capitalista*, que submete as políticas educacionais à lógica da produtividade e do investimento.

Essa discursividade ganha relevo quando situada no contexto de emergência de uma onda conservadora que, nos anos seguintes, desestabilizaria o governo Dilma Rousseff (Demier, 2016). A campanha publicitária, ao enfatizar a eficiência e o investimento estatal, projeta a imagem de um Estado gestor, capaz de administrar bem os recursos públicos. Essa representação, contudo, foi tensionada quando a presidenta passou a ser acusada por setores conservadores ligados a grupos religiosos, militares e ruralistas de má gestão e de “pedaladas fiscais”, associadas à manutenção de políticas sociais e educacionais. O discurso de eficiência e abundância de investimentos – como na SD₂ em “*investimento do Brasil*” – contrasta, assim, com as acusações que sustentaram o *impeachment*, revelando o embate entre duas formas de gestão: o Estado social, que busca legitimar-se pelo investimento público, e a racionalidade gerencial, própria da governamentalidade neoliberal, centrada na austeridade e no controle fiscal. Nesse sentido, como observa Foucault (2008, p. 114), na governamentalidade neoliberal, “a economia produz legitimidade para o Estado, que é seu avalista”, de modo que o desempenho econômico passa a funcionar como critério de soberania e de validação da ação governamental.

Assim, a política educacional é discursivamente apresentada não como direito, mas como investimento: o “*apoio*” é representado como meio de retorno social e econômico. O discurso publicitário da SD₂ inscreve-se em uma FD orientada à racionalidade neoliberal, na qual a educação é significada como capital humano e o estudante, como recurso produtivo. Embora anuncie democratização e inclusão, o acesso aos benefícios é condicionado ao desempenho individual, como em “*o Ciência sem Fronteiras premia bons estudantes*”. A modalização do verbo “*premiar*” reforça a ideia de seleção e de exclusividade, deslocando a responsabilidade pela ascensão social do Estado para o sujeito e naturalizando a lógica meritocrática que sustenta essa racionalidade. Nessa perspectiva, o discurso educacional opera sob a lógica do mercado, em que a valorização do desempenho e da competição individual substitui a noção de direito coletivo. Como observa Foucault (2008, p. 116), “o mercado livre, economicamente livre, liga politicamente e manifesta vínculos políticos”, indicando que a liberdade econômica se converte em princípio político de governo, legitimando práticas de gestão e hierarquização social.

Segundo Althusser (2010), a escola, como *Aparelho Ideológico de Estado*, reproduz as relações de produção capitalistas, ao formar sujeitos que internalizam os valores de mérito, esforço e produtividade. O enunciado “*bons estudantes*” representa esse sujeito ideal, que se legitima pelo desempenho individual e dissimula as desigualdades de classe, sustentando a crença de que “*não estuda quem não quer*”. Essa lógica se aproxima da “*escola reprodutivista*” criticada por Luckesi (1994, p. 41), orientada pela eficiência e pela adaptação do indivíduo às exigências do mercado, uma vez que, nesse modelo, “a educação faz, integralmente, parte da sociedade e a reproduz”. Nesse movimento, o verbo “*estudar*” aparece reiteradamente em enunciados como “*Estudar nos Estados Unidos*” e “*Poder estudar com pessoas do mundo inteiro*”, operando como metáfora de investimento que associa o saber ao capital. Assim, ao afirmar, na SD₂, “*E tudo isso com investimento do Brasil nos jovens*”, a campanha materializa essa metáfora econômica, convertendo a educação em rendimento e o estudante em ativo, reafirmando a racionalidade neoliberal que transforma o jovem em empreendedor de si mesmo.

Em diálogo com Luckesi (1994), percebe-se que a propaganda mobiliza sentidos de escola crítico-transformadora, ao enfatizar inclusão, pensamento crítico e até o lúdico, como nos trechos que destacam a experiência internacional e o contato com “*pessoas do mundo inteiro*” e “*laboratórios, clínicas*”. Contudo, há traços evidentes da tendência redentora, que pretende salvar o aluno pela via dos “*investimentos*” estatais, atribuindo à escola o poder de emancipar não apenas pelo saber, mas pelo aporte econômico que garante oportunidades. Essa promessa aparece, por exemplo, no enunciado da SD₂ “*O Ciência sem Fronteiras premia bons estudantes*”, em que o mérito é apresentado como caminho de ascensão e libertação, sustentado pelo financiamento público. Ao mesmo tempo, essa mesma formulação revela traços da tendência reprodutivista, pois naturaliza desigualdades ao transformá-las em mérito, como se apenas o esforço individual fosse suficiente para garantir acesso e mobilidade. O discurso, assim, tensiona entre a tendência redentora – promessa de reparação histórica, embora pela via econômica – e a da “*educação reprodutora da sociedade*” (Luckesi, 1994, p. 41) – reafirmação da ordem social vigente, exigindo (re)produtividade, competitividade e mérito como critérios de inclusão.

O slogan “*Brasil, um país rico é um país sem pobreza*” reinscreve sentidos de cidadania, democracia e economia, produzindo um efeito de evidência que naturaliza a relação entre riqueza nacional e superação da pobreza. A discursividade da SD₂ constrói uma verdade que se apresenta como “*óbvia*”, interpelando o sujeito a reconhecer-se nesse imaginário de progresso em que a economia aparece como meio de superação das desigualdades. A nominalização em “*país rico*” e “*país sem pobreza*” transforma processos históricos – gerar riqueza e eliminar a pobreza – em realidades “*simplificadas*”, apagando as condições políticas que os determinam. Já o efeito metafórico, entendido como “o fenômeno semântico produzido por uma substituição contextual [...] esse deslizamento de sentido entre x e y é constitutivo do sentido designado por x e y” (Pêcheux, 2010, p. 96), manifesta-se através do verbo *ser*, estabelecendo equivalência entre riqueza e ausência de pobreza. Essa operação desloca o discurso para o campo econômico, em que a educação é significada como “*investimento*” capaz de promover reparação histórica. Assim, o discurso publicitário apresenta a escola, na SD₂, como espaço de oportunidades, mas submetido à lógica meritocrática, mobilizando sentidos de que o crescimento econômico bastaria para eliminar as desigualdades históricas que o próprio sistema capitalista sustenta.

A seguir, dando continuidade à comparação entre as campanhas do MEC produzidas durante os governos Dilma e Bolsonaro, analisa-se a campanha de 2019, do *Programa Nacional das Escolas Cívico-Militares* (Brasil, 2019a, 2019b), que evidencia a discursividade oficial do governo Bolsonaro e os sentidos produzidos em torno de suas políticas educacionais. Nesse momento, quem se pronuncia é o então Ministro da Educação, Abraham Weintraub, apresentando o modelo de escola cívico-militar adotado pelo governo.

SD₃ – É um modelo que está tendo uma taxa de sucesso muito boa. Existe um respeito muito grande, uma autoestima de pertencer a uma escola cívico-militar. Não tem pichação, não tem bagunça, não tem vandalismo. As crianças estão lá para aprender e o desempenho que a gente vê, é muito acima da média brasileira. Os recursos têm que ser destinados a resgatar aquelas famílias, aquelas crianças que ficaram para trás (Brasil, 2019a, grifos nossos).

A noção de autoridade, segundo Gerth e Mills (1973), é central para compreender a discursividade produzida por Weintraub, que estabelece uma relação direta entre educar e obedecer. Na SD₃, ao afirmar “um respeito muito grande”, o termo respeito não se configura como valorização da diversidade ou da crítica, mas como marcador de disciplina e conformidade. Os intensificadores “muito” e “grande” produzem efeito de evidência e de magnitude, reforçando a obediência esperada e conferindo estabilidade ao sentido de autoridade. Essa lógica se intensifica quando Weintraub enuncia “modelo que está tendo uma taxa de sucesso muito boa” e “o desempenho que a gente vê é muito acima da média”. Aqui, “desempenho” funciona como nominalização, deslocando o sentido de aprendizagem para disciplina e obediência, enquanto “modelo” aponta para um verdadeiro laboratório neoliberal, sustentado por uma educação militarizada orientada pelo saber-obedecer. Os intensificadores “muito boa” e “muito acima da média” operam como modalizadores e marcadores econômicos, legitimando o modelo como mais eficiente e econômico, apagando contradições e silenciando outras formas de pensar a educação.

Na SD₃, a formulação “Os recursos têm que ser destinados a resgatar aquelas famílias, aquelas crianças que ficaram para trás” inscreve a escola cívico-militar em uma discursividade redentora que combina moral cristã, obediência e dever estatal. O “resgate” é lexicalizado como ato de salvação, mas, sob essa aparência benevolente, o que se efetiva é o controle e a disciplinarização dos corpos. A modalização em “têm que ser” institui a obrigatoriedade como princípio moral e pedagógico, legitimando a intervenção sobre os sujeitos considerados “desviantes”. Já o uso do pronome “aquelas” distancia e categoriza os alunos e famílias pobres como “outros” a serem corrigidos, reiterando a desigualdade como diferença natural. Nesse discurso, o humano é substituído pela norma, e a educação se converte em mecanismo de enquadramento social. Assim, o “resgate” não emancipa, mas captura, e a escola cívico-militar se consolida como instrumento de exclusão, que, sob o pretexto de promover ordem e sucesso, reafirma hierarquias e silencia os efeitos estruturais da desigualdade.

Nesse funcionamento discursivo, a negação reiterada em enunciados como “não tem pichação, não tem bagunça, não tem vandalismo” evidencia como a lexicalização atua na produção de sentidos sobre a escola cívico-militar. O verbo “ter”, utilizado sucessivamente em sua forma negativa, organiza o discurso pelo que a escola não tem, pela exclusão de tudo o que remete à expressão, à espontaneidade ou à diferença. Assim, o sujeito discursivo constrói a imagem de uma escola “limpa”, “organizada” e “eficiente”, fundada na ausência, pois não há ruído, não há conflito, não há corpo. Nesse processo, Weintraub (de)nega tudo aquilo que possa ser humano

ou expressão do corpo, suprimindo a dimensão afetiva e participativa do espaço escolar. A negação funciona, então, como forma de denegação, pois, conforme Indursky (1990, p. 118), “através da denegação, o sujeito pode mascarar aquilo que, por ter sido censurado pelo superego e recalcado no inconsciente, não lhe é facultado dizer”. Assim, o discurso enuncia sem dizer, ao afirmar o que a escola cívico-militar não tem, projeta sobre a escola pública tudo o que ela é – humana, diversa, contraditória –, reinscrevendo, sob a forma do apagamento, a diferença que pretende eliminar. A lexicalização negativa opera, portanto, como estratégia discursiva de contraste e mascaramento, legitimando o modelo militarizado como sinônimo de ordem e moralidade, ao mesmo tempo em que silencia a humanidade da escola pública.

O que se nega – a desordem, o vandalismo, a indisciplina – é o que se projeta sobre a escola pública, produzindo o efeito de que ali impera o caos. Assim, expressões como “*taxa de sucesso muito boa*” e “*desempenho muito acima da média*” consolidam a escola cívico-militar como modelo de eficiência e de moralidade, silenciando a precarização e a falta de investimentos estruturais. O discurso, ao converter desigualdades históricas em falhas morais, mobiliza memórias sociais que associam a escola pública ao fracasso (Orlandi, 2000) e reafirma os valores da classe dominante, naturalizando a subordinação de alunos e professores da rede pública. Nesse sentido, o que está em jogo é o controle dos corpos que, na maioria das vezes, são pessoas consideradas desajustadas socialmente, pelo critério socioeconômico, *aqueles que*, historicamente, o capitalismo excluiu.

Desse modo, o silenciamento não se limita à exclusão de vozes dissidentes, mas se inscreve como forma de governar, sustentando a racionalidade econômica que transforma a escola em instrumento de controle e produtividade. Segundo Orlandi (2010, p. 105), impor o silêncio “não é calar o interlocutor, mas impedi-lo de sustentar outro discurso”. Sob a lógica da governamentalidade econômica (Foucault, 2008), a escola cívico-militar se apresenta como “*mais eficiente*”, por alinhar-se à reprodução da ordem vigente e ao apagamento dessas contradições estruturais. Indicadores como “*taxa de sucesso*” e “*desempenho*” traduzem a educação em produtividade e custo-benefício, funcionando como índices de mercado que legitimam o poder estatal e mascaram a precarização da escola pública. Essa racionalidade se articula com a tendência reprodutivista, descrita por Luckesi (1994), que privilegia a obediência em detrimento da transformação social. Ao reduzir o processo educativo a métricas de eficiência disciplinar, o discurso atua como aparelho ideológico, naturalizando desigualdades e disciplinando alunos considerados “desajustados” – aqueles historicamente excluídos pelo capitalismo. O silenciamento, nesse quadro, impede o desenvolvimento do pensamento crítico e neutraliza qualquer possibilidade de questionar o sistema que produziu a própria marginalidade desses sujeitos.

Nesse contexto, a escola militarizada funciona como *laboratório da governamentalidade capitalista*, onde se testam formas de controle, disciplina e docilização dos sujeitos. A modalização em “*é um modelo que está tendo uma taxa de sucesso muito boa*” produz efeito de evidência e inevitabilidade, atualizando o saudosismo da ditadura e o imaginário da “ordem” como valor supremo. A educação, assim, deixa de ser espaço crítico para tornar-se instrumento de produtividade e contenção social. Essa racionalidade neoliberal articula-se à tendência reprodutivista, destacada por Luckesi (1994), que reduz o ensino à mensuração de resultados e à obediência, em vez de promover reflexão e transformação. O Estado, ao operar pela economia da obediência, legitima-se pela disciplina e pelo controle dos conflitos, enquanto naturaliza a precarização da escola pública e silencia os sujeitos considerados “desajustados”. Nesse movi-

mento, “a economia produz legitimidade para o Estado, que é seu avalista” (Foucault, 2008, p. 114), e a escola cívico-militar se consolida como o espaço onde a ordem e o silenciamento se apresentam como “sucesso” educacional.

Dando continuidade à análise da campanha do MEC no governo Bolsonaro, apresenta-se a divulgação do modelo de escola cívico-militar no Programa Nacional das Escolas Cívico-Militares (Brasil, 2019a, 2019b), com pronunciamento do próprio presidente Jair Bolsonaro:

SD₄ – O que nos tira da miséria e da pobreza, da ignorância, obviamente, é o conhecimento. Nós queremos integrar, botar na cabeça de toda essa garotada, a importância dos valores cívico-militares (Brasil, 2019a, grifos nossos).

A SD₄ evidencia como a educação é mobilizada para construir e controlar sujeitos, segundo uma discursividade ultraliberal e conservadora. Termos como “*miséria*” e “*pobreza*” não se limitam a descrever condições socioeconômicas; funcionam também como marcadores sociais carregados de estigma, historicamente associados a grupos excluídos e entendidos como sinônimos de “*ignorância*” ou inferioridade. Dessa forma, a pobreza deixa de ser apenas uma condição material e se transforma em um rótulo que legitima preconceitos e reforça desigualdades, produzindo uma patologização do pobre, construído como alguém que carece de intervenção e precisa ser resgatado. A escola é apresentada como a instituição capaz de transformar esses indivíduos, reforçando que a exclusão histórica e a desigualdade social devem ser corrigidas por meio de uma mediação moral e educativa centrada na obediência e na assimilação de valores impostos (Dias; Ribeiro, 2021). Como evidencia na SD₄, “*O que nos tira da miséria e da pobreza, da ignorância, obviamente, é o conhecimento*”, em que a modalização em “*obviamente*” reforça o efeito de evidência e naturalização, como se a ascensão social e o saber fossem consequências inevitáveis do acesso à educação, quando a educação cívico-militar prioriza a disciplina e a obediência em vez do ensino-aprendizagem crítico.

O caminho proposto para a “redenção” do aluno é indicado por verbos como “*integrar*” e “*botar na cabeça*”, funcionando como metáforas de submissão e imposição. A expressão “*botar na cabeça*” sugere inculcar ideias conservadoras, orientando a educação menos para a reflexão crítica e mais para a internalização de normas e comportamentos. Na prática, a integração funciona como exclusão da autonomia, sem espaço para participação social ou questionamento do sistema que produz a marginalidade dos sujeitos. Na SD₄, a formulação “*Nós queremos integrar, botar na cabeça*” evidencia a imposição do saber, naturalizando a obediência como única via de assimilação do conhecimento (Althusser, 2010). A pronominalização em “*nós queremos*” reforça a posição da governamentalidade ultraliberal como sujeito de um saber autorizado e legitimado, enquanto o aluno é colocado como *agente passivo*. Esse funcionamento também se articula com uma racionalidade econômica (Foucault, 2008), em que o Estado se apresenta como gestor eficiente, disciplinando e silenciando em vez de enfrentar as desigualdades sociais e os problemas estruturais da educação pública.

A referência aos “*valores cívico-militares*” remete à tradição do regime militar brasileiro, marcada pelo civismo coercitivo e pela disciplina hierárquica. Como apontam Dias e Ribeiro (2021), nas ECIMs o civismo historicamente formou o indivíduo segundo pilares de Educação, Trabalho e Pátria, dentro de uma lógica de segurança nacional, ordem, progresso e desenvolvimento econômico, mediante medidas ideologicamente coercitivas. Na SD₄, a expressão “*a importância dos valores cívico-militares*” associa o conhecimento à integração/exclusão moral e

cívica, exigindo obediência, respeito e submissão às regras que definem o comportamento adequado. A padronização de comportamentos, a desvalorização da identidade individual e a inibição ao questionamento reproduzem relações de poder que limitam a agência do aluno e naturalizam a hierarquia e o controle social. Assim, evidencia-se a governamentalidade capitalista autoritária (Foucault, 2008), na qual a educação funciona como instrumento de gestão de recursos e comportamentos, disciplinando e silenciando em consonância com o ultraliberalismo e conservadorismo do governo Bolsonaro.

Nesse movimento, a SD₄ combina tendência redentora e reprodutiva (Luckesi, 1994), problematizando o aluno como sujeito marcado pela pobreza, miséria e ignorância, que precisa ser “resgatado”, e propondo uma solução via educação autoritária, por meio da internalização de valores cívico-militares e da submissão às normas, apagando qualquer possibilidade de crítica ou questionamento do sistema que produziu essa exclusão. O conhecimento não é apresentado como instrumento de emancipação crítica, mas como ferramenta de integração moral e disciplinar (Dias; Ribeiro, 2021), assegurando que a escola cívico-militar funcione como laboratório da governamentalidade capitalista (Foucault, 2008), regulando corpos, comportamentos e subjetividades em consonância com uma lógica de dominação histórica e manutenção da ordem social. Enunciados da SD₄ – “*O que nos tira da miséria e da pobreza, da ignorância, obviamente, é o conhecimento*” e “*Nós queremos integrar, botar na cabeça de toda essa garotada, a importância dos valores cívico-militares*” – evidenciam a patologização do sujeito, a imposição de normas e valores e a constituição da *forma-sujeito*, reforçando o efeito de evidência e naturalização característico da discursividade ultraliberal e conservadora do governo Bolsonaro.

5 Conclusão

A análise das campanhas educacionais do MEC, veiculadas nos governos de Dilma Rousseff (2011-2016) e de Jair Messias Bolsonaro (2019-2022), evidenciou como o Estado, por meio do discurso publicitário, produz e legitima diferentes projetos de educação. Em ambos os casos, a escola é tratada como espaço estratégico de formação de sujeitos e manutenção da ordem social, projetando também, modelos distintos de sociedade e revelando a estreita relação entre educação, escola, ideologia e poder. No governo Dilma, a educação é representada como investimento e possibilidade de mobilidade social (Brasil, 2013); enquanto no governo Bolsonaro, como instrumento de controle e disciplinamento moral (Brasil, 2019a).

Na campanha do governo Dilma, a educação é significada como investimento e oportunidade, articulando o sucesso do aluno à meritocracia, à lógica do capital e a um modelo de educação mercantilista. O sucesso do aluno parece depender tanto dos recursos investidos quanto de sua própria disposição, evidenciando uma tensão produtivista (Luckesi, 1994). Embora subordinada à lógica do desempenho e da eficiência, a educação mantém espaço para a diversidade e o diálogo com um modelo de escola redentora, ao conceber o aluno como sujeito que precisa ser “salvo” por reparação histórica e investimentos. Pernambuco, historicamente com baixos índices educacionais, é mobilizado como exemplo do retorno desse investimento, reforçando a imagem de um Estado eficiente e progressista. Mesmo ao valorizar práticas culturais e esportivas – “*hip-hop, capoeira, futebol*” –, a educação se organiza sob uma lógica de *saber-fazer* e *saber-interpretar* (Althusser, 2010), que naturaliza o sujeito ajustado às

demandas do capital, subordinando a inclusão e o prazer de aprender à produtividade e à eficiência (Foucault, 2008).

Já na campanha do governo Bolsonaro, a noção de *desempenho* mobiliza obediência, disciplina e cumprimento de normas, como nos enunciados “*Existe um respeito muito grande*” e “*Não tem pichação, não tem bagunça, não tem vandalismo*”. Ao definir pelo que se acredita que a escola não deve/pode ter – vandalismo, bagunça, entre outros –, ela se inscreve pelo que lhe falta: o fator humano, a autonomia, a participação e a capacidade de gerar reflexão crítica. A escola também mobiliza a ideia de “*resgate*” dos alunos considerados desajustados – os *aqueles que* –, esquecidos pelo capitalismo, pobres ou marginalizados. Contudo, a solução apresentada é exclusivamente reprodutivista, baseada na força e na coerção moral, silenciando questionamentos críticos sobre o sistema que produziu a exclusão desses sujeitos. Nesse sentido, a escola militarizada se configura como uma política orientada para a desdemocratização, impondo hierarquia e disciplina, restringindo a gestão pedagógica e a participação da comunidade escolar (Goulart; Amaral, 2023). Observa-se, assim, o silenciamento das políticas públicas de inclusão, cidadania e diversidade, substituídas por um discurso que exalta ordem, moral e submissão.

Como observa Han (2017, p. 23), não vivemos mais em uma sociedade disciplinar baseada em instituições de controle direto, mas em uma “sociedade de desempenho”, na qual o capitalismo moderno se organiza por investimentos, recompensas e competição. Sob essa perspectiva, tanto o modelo de Dilma quanto o de Bolsonaro se inscrevem na lógica da governamentalidade capitalista, orientada pelo *saber-fazer* e pelo *saber-interpretar* (Althusser, 2010). No entanto, o modelo de Dilma manifesta uma forma mais atual dessa racionalidade, ao conceber a educação como investimento e destinação de recursos, abrindo espaço para práticas inclusivas e para uma tendência crítico-transformadora. Já o modelo de Bolsonaro revela um capitalismo disciplinar obsoleto, excludente e autoritário, sustentado na coerção moral e na submissão hierárquica, reduzindo a educação a instrumento de gerenciamento, controle e silenciamento.

Nesse cenário, a escola cívico-militar funciona como laboratório da governamentalidade capitalista, que reproduz valores da classe dominante e interpela os sujeitos à obediência, apagando contradições sociais e naturalizando a precarização da escola pública. A escola militarizada, marcadamente produtivista, promove um *saber-fazer* e um *saber-interpretar* que, no contexto das escolas cívico-militares, se configuram como saber “aceitar” a ordem de comando militar, transformando o aprendizado em reprodução da disciplina e da própria submissão. A modalização em “*é um modelo que está tendo uma taxa de sucesso muito boa*” produz efeito de evidência e de legitimação, como se não houvesse alternativa possível. Ao operar pela economia da obediência, o Estado se legitima pela disciplina e pela contenção dos conflitos sociais, convertendo a educação em instrumento de gestão da desigualdade (Foucault, 2008). Trata-se, portanto, de uma escola que, como aponta Luckesi (1994), realiza-se sob a tendência (re)produtivista, ou seja, obedece à lógica do rendimento e da submissão, reafirmando a desigualdade estrutural, transformando ensino em eficiência e aprendizagem em comportamento adequado.

Assim, a campanha educacional revela-se mais que um instrumento de publicidade institucional, sendo uma prática discursiva de produção de subjetividades e de legitimação do poder. Ao comparar os governos Dilma e Bolsonaro, percebe-se que ambos se sustentam na racionalidade econômica da governamentalidade capitalista, mas o primeiro a atualiza por

meio de discurso de investimento e inclusão – constituindo uma FD neoliberal –, enquanto o segundo a radicaliza em sua face autoritária e disciplinar – inscrita em uma FD marcadamente ultraliberal, na qual a educação surge como projeto de obediência e de subserviência. Essa virada discursiva evidencia o papel da propaganda na sustentação do poder e no apagamento do caráter político da escola, transformada em dispositivo de gestão da vida e dos corpos.

Por esse viés, evidencia-se que educação e sociedade se constituem como dimensões indissociáveis de um mesmo tecido social, nas quais cada enunciado sobre a escola opera como tentativa de interpelar sujeitos, configurando o tipo de cidadão e projetando o futuro social que se pretende instituir. Entre as tensões do ideal democrático e das práticas autoritárias, a escola se apresenta como espaço de disputa discursiva (Althusser, 2010), no qual se articulam diferentes modos de governar. Assim, falar de educação implica, simultaneamente, falar de poder, na medida em que o Estado, por meio da linguagem, atua na produção de sentidos sobre como os sujeitos devem ser, pensar, agir e aprender na sociedade contemporânea.

Declaração de autoria

Nilton César Ferreira foi responsável pela análise do *corpus*, com enfoque na Análise do Discurso de linha francesa, além de ter contribuído diretamente para a redação do artigo. Francisco Renato Lima colaborou com o embasamento teórico, na sugestão de leituras voltadas para o campo pedagógico e linguístico, bem como participou das análises e realizou a revisão final do texto.

Referências

ALTHUSSER, L. Observações sobre uma categoria: processo sem sujeito e sem fim(s). In: ALTHUSSER, L. *Posições-1*. Rio de Janeiro: Graal, 1973. p. 66-71.

ALTHUSSER, L. *Aparelhos Ideológicos de Estado*: nota sobre os Aparelhos Ideológicos de Estado. 7. ed. Rio de Janeiro: Graal, 2010.

BASILIO, M. *Formação e classes de palavras no português do Brasil*. 3. ed. São Paulo: Contexto, 2017.

BRANDÃO, H. H. N. *Introdução à análise do discurso*. 3. ed. Campinas: Editora da Unicamp, 2012.

BRASIL. *Constituição da República Federativa do Brasil*. Brasília: Congresso Nacional, 1988. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf. Acesso em: 01 abr. 2026.

BRASIL. Governo Federal. *Propaganda do Governo Federal: Escola integral - ProUni - Fies - Ciência sem Fronteiras* CsF. Brasília: Ministério da Educação, 2013. Disponível em: <https://www.youtube.com/watch?v=zN7cqOj-4A8>. Acesso em: 01 abr. 2026.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Resultados do IDEB 2013 - Ensino Médio por Unidade da Federação*. Brasília: INEP/MEC, 2014.

- BRASIL. Governo Federal. *Presidente lança Programa Nacional das Escolas Cívico-Militares*. Brasília: Ministério da Educação, 2019a. Disponível em: <https://www.youtube.com/watch?v=Jz5Q5PAFSLo>. Acesso em: 01 abr. 2026.
- BRASIL. Decreto nº 10.004, de 5 de setembro de 2019. Institui o Programa Nacional das Escolas Cívico-Militares. *Diário Oficial da União*, Brasília, DF, 6 set. 2019b.
- DÁVILA, S. TCESP suspende contratação de PMs para escolas cívicomilitares de Tarcísio de Freitas. Edição de Educação. *Folha de S. Paulo*, 04 set. 2025. Disponível em: <https://www1.folha.uol.com.br/educacao/2025/09/tribunal-de-contas-suspende-contratacao-de-pms-para-escolas-civico-militares-de-tarcisio.shtml>. Acesso em: 01 abr. 2026.
- DEMIER, F. O barulho dos inocentes: a revolta dos “homens de bem”. In: DEMIER, F.; HOEVELER, R. (Orgs.). *A onda conservadora: ensaios sobre os atuais tempos sombrios no Brasil*. Rio de Janeiro: Mauad, 2016. p. 09-24.
- DIAS, Z. R.; RIBEIRO, A. C. Escolas cívicomilitares: conservadorismo e retrocesso na educação brasileira. *Revista Teias*, Rio de Janeiro, v. 22, p. 406426, out./dez. 2021. DOI: <https://doi.org/10.12957/teias.2021.59634>
- FOUCAULT, M. *Nascimento da biopolítica: curso dado no Collège de France (1978-1979)*. São Paulo: Martins Fontes, 2008.
- FREITAS, R. A. de; SILVA, S. D. “Se a base da educação é a mesma, as oportunidades também serão”? A propaganda sobre educação no governo Temer. *Revista de Informação e Notícias*, Recife, v. 24, n. 1, p. 0117, 2018. DOI: <https://doi.org/10.51359/2175-294x.2018.237922>.
- GERTH, H.; MILLS, W. *Caráter e estrutura social: a psicologia das instituições sociais*. Rio de Janeiro: Civilização Brasileira, 1973.
- GOULART, J. M. de O.; AMARAL, D. P. do. Militarização de escolas e a gestão democrática: os sentidos da desdemocratização do ensino público. *Revista Brasileira de Política e Administração da Educação*, Brasília, v. 39, n. 1, p. 0121, 2023. DOI: <https://doi.org/10.21573/vol39n12023.128742>.
- GRIGOLETTO, E. Do lugar social ao lugar discursivo: o imbricamento de diferentes posições-sujeito. In: SEMINÁRIO DE ESTUDOS EM ANÁLISE DO DISCURSO (SEAD), II., 2005, Porto Alegre. *Anais...* Porto Alegre: UFRGS, 2005. p. 01-11. Disponível em: <https://www.ufrgs.br/analisedodiscurso/anaisdo-sead/2SEAD/SIMPOSIOS/EvandraGrigoletto.pdf>. Acesso em: 01 abr. 2026.
- HAN, B.-C. *Sociedade do cansaço*. Trad. Enio Paulo Giachini. 2. ed. Petrópolis: Vozes, 2017.
- INDURSKY, F. Polêmica e denegação: dois funcionamentos discursivos da negação. *Cadernos de Estudos Linguísticos*, Campinas, v. 19, p. 117-122, jul./dez. 1990. DOI: <https://doi.org/10.20396/cel.v19i0.8636829>.
- LAGAZZI, S. *O desafio de dizer não*. Campinas: Pontes, 1988.
- LUCKESI, C. C. *Filosofia da Educação*. São Paulo: Cortez, 1994.
- MACEDO, J. M. de; LAMOSA, R. A. A regulação do trabalho docente no contexto da reforma gerencial da educação. *Revista Contemporânea de Educação*, Rio de Janeiro, v. 10, n. 20, p. 133-152, jul./dez. 2015. DOI: <https://doi.org/10.20500/rce.v10i20.2288>.

- MARCUSCHI, L. A. Gêneros textuais emergentes no contexto da tecnologia digital. In: MARCUSCHI, L. A.; XAVIER, A. C. (Orgs.). *Hipertexto e gêneros digitais: novas formas de construção de sentido*. 3. ed. São Paulo: Cortez, 2010. p. 15-80.
- MARIANI, B. S. C. *O PCB e a imprensa: os comunistas no imaginário dos jornais (1922-1989)*. Campinas: Editora da Unicamp, 1998.
- ORLANDI, E. P. *Análise de discurso: princípios e procedimentos*. Campinas: Pontes, 2000.
- ORLANDI, E. P. *A linguagem e seu funcionamento: as formas do discurso*. 5. ed. Campinas: Pontes, 2009.
- ORLANDI, E. P. *As formas do silêncio: no movimento dos sentidos*. Campinas: Editora da Unicamp, 2010.
- PÊCHEUX, M. Análise Automática do Discurso (AAD-1969). In: GADET, F.; HAK, T. (Orgs.). *Por uma Análise Automática do Discurso: uma introdução à obra de Michel Pêcheux*. Trad. Eni Puccinelli Orlandi. Campinas: Unicamp, 2010. p. 59-158.
- PÊCHEUX, M.; FUCHS, C. A propósito da Análise Automática do Discurso: atualização e perspectivas [1975]. In: GADET, F.; HAK, T. (Orgs.). *Por uma Análise Automática do Discurso: uma introdução à obra de Michel Pêcheux*. Trad. Eni Puccinelli Orlandi. Campinas: Unicamp, 2010. p. 159-250.
- PÊCHEUX, M. Foi “propaganda” mesmo que você disse? [1979]. Trad. Eni Orlandi. In: ORLANDI, E. P. (Org.). *Análise de Discurso: Michel Pêcheux*. 3. ed. Campinas: Pontes, 2011. p. 73-92.
- PÊCHEUX, M. *Semântica e discurso: uma crítica à afirmação do óbvio*. Trad. Eni Puccinelli Orlandi et al. 4. ed. Campinas: Editora da Unicamp, 2009.
- POSSENTI, S. Teoria do discurso: um caso de múltiplas rupturas. In: MUSSALIM, F.; BENTES, A. C. (Orgs.). *Introdução à Linguística: fundamentos epistemológicos*, v. 3. 5. ed. São Paulo: Cortez, 2011. p. 353-391.
- ROUSSEAU, J.-J. *Do contrato social*. São Paulo: Abril Cultural, 1983.
- SAVIANI, D. *Escola e democracia*. 42. ed. Campinas: Autores Associados, 2012.
- SILVA, C. R. C. Linguagem, poder e ideologia: uma análise crítica discursiva do Escola Sem Partido. *Educação em Foco*, Belo Horizonte, v. 23, n. 40, p. 174-190, 2020. DOI: <https://doi.org/10.24934/eeef.v23i40.4602>.
- SILVEIRA, M. H. Concepções e funções de TV e vídeo na comunicação educativa. In: FIORENTINI, L. M. R.; CARNEIRO, V. L. Q. (coords.). *TV na escola e os desafios de hoje: curso de extensão para professores do Ensino Fundamental e Médio da rede pública*. Módulo 1. 2. ed. Brasília: Editora da Universidade de Brasília, 2001.
- TENÓRIO NETO, J. F.; ERICSON, S. “Uma ponte para o futuro”: efeitos de sentido do discurso neoliberal no Brasil. *Revista da ABRALIN*, [S. l.], v. 19, n. 3, p. 409-428, 2020. DOI: <https://doi.org/10.25189/rabralin.v19i3.1739>.
- TILLY, C. *Democracia*. Trad. Raquel Weiss. Petrópolis: Vozes, 2013.
- TOZONI-REIS, M. F. de C. A contribuição da Sociologia da Educação para a compreensão da educação escolar. In: PINHO, S. Z. de (Org.). *Cadernos de Formação: Formação de Professores*. Educação, Cultura e Desenvolvimento. v. 3. São Paulo: Cultura Acadêmica, 2010. p. 53-67. Disponível em: <https://acervodigital.unesp.br/bitstream/123456789/169/3/01do9to3.pdf>. Acesso em: 01 abr. 2026.
- VEIGA-NETO, A. Governamentalidades, Neoliberalismo e Educação. In: BRANCO, G. C.; VEIGA-NETO, A. (Orgs.). *Foucault: filosofia & política*. Belo Horizonte: Autêntica, 2011. p. 37-52.

Building the Terminology of an Audiovisual Saga: The Example of Star Wars¹

Elaborando a terminologia de uma saga audiovisual: o exemplo de Star Wars

Guilherme Fromm

Universidade Federal de Uberlândia
(UFU) | Uberlândia | MG | BR
guifromm@ufu.br
<https://orcid.org/0000-0001-5654-0135>

Guilherme Rodrigues Ferreira

Universidade Federal de Uberlândia
(UFU) | Uberlândia | MG | BR
guilherme.ferreira1@ufu.br
<https://orcid.org/0000-0003-2030-2594>

Abstract: The objective of this article is to show how a multimedia franchise can be studied through terminological lenses. The Star Wars saga was chosen for the study due to its huge amount of material available on the internet. The theoretical framework comprises the Terminology area (more specifically, the Communicative Theory of Terminology and Ethnoterminology), the Terminography, a lexical analysis suite (WordSmith Tools) used for the work with concordance lines from the saga and the development of the Terminology in Fiction platform. The methodology presents many steps, from the choice of the material, the selection of term candidates from the WordSmith Tools suite up to the exploration of the concordance lines of these candidates to create entries in the Terminology in Fiction platform. The platform is also analysed, from the creation of the project up to the entries development. As a result, we present one example (of many available at the platform) of a Star Wars monolingual entries in contrast (available in Portuguese and English), useful for translators, fans and for those curious about how terminology permeates even fictional works.

Keywords: terminology; terminography; Star Wars; corpus linguistics; terminology in fiction.

Resumo: O objetivo deste artigo é mostrar como uma franquia multimídia pode ser estudada por meio de lentes terminológicas. A saga Star Wars foi escolhida para o estudo devido à sua enorme quantidade de material

¹ The description of trademarks owned by Disney were used in this paper only for scientific purposes under Brazilian law.



disponível na internet. O arcabouço teórico compreende a área de Terminologia (mais especificamente, a Teoria Comunicativa da Terminologia e a Etnoterminologia), a Terminografia, uma suíte de análise lexical (WordSmith Tools) usada para o trabalho com linhas de concordância da saga e o desenvolvimento da plataforma Terminologia em Ficção. A metodologia apresenta muitas etapas, desde a escolha do material, a seleção de candidatos a termos através da suíte WordSmith Tools até a exploração das linhas de concordância desses candidatos para criar verbetes na plataforma Terminologia em Ficção. A plataforma também é analisada, desde a criação do projeto até o desenvolvimento dos verbetes. Como resultado, apresentamos um exemplo (de muitos disponíveis na plataforma) de verbetes monolíngues de Star Wars em contraste (português e inglês), útil para tradutores, fãs e para aqueles curiosos sobre como a terminologia permeia até mesmo obras de ficção.

Palavras-chave: terminologia; terminografia; Star Wars; linguística de *corpus*; terminologia em ficção.

1 Introduction

Star Wars is one of the most important and influential franchises in the history of cinema, television, books, video games and comics. Since the release of the first film in 1977, the saga created by George Lucas has become a phenomenon and a valuable brand, now owned by The Walt Disney Company. Among the many reasons that motivated the completion of this corpus-based research on Star Wars, the influence of the saga was one of the most significant. It is a large multimedia franchise that has been continually expanding since the 1970s, requiring daily teams of audiovisual translators who must tackle the challenging standardization required for such a vast universe. Currently, the high demands of dubbing and subtitling companies have created a truly frenetic pace, with extremely tight deadlines and increasingly larger projects. Among movies, animations, and TV series, the Star Wars franchise currently has dozens of titles in various stages of production, which will eventually be completed and sent to these companies for their respective translations and adaptations. Considering the work of dubbing and subtitling translators, who deal with complex deadlines and may not necessarily have a deep understanding of fictional universes like that one of Star Wars, terminology assistance is naturally welcomed and motivates the development of this research.

Every aspect of Star Wars presents a vast terminological universe that can be explored. The geography of the saga specifies various planets, terrains, and regions, each with their own names and characteristics. Politics, in turn, involves factions and organizations such as the

Republic, the Rebel Alliance, the Empire, and the First Order, with each institution having a range of affiliations, including military ranks, technologies used, and cultural and economic elements. The franchise also delves into different races, beliefs, philosophies, institutions, and technologies, with each of these divisions carrying plenty of terms (fictional or not) that can be studied in literature or applied in the professional and amateur world of audiovisual translation, given the constant releases in cinema, TV, video games, and literature.

1.1 Goals

The general objective of the project developed by one of the authors TCC² (Ferreira, 2022) was to create a vocabulary proposal about the Star Wars franchise on the Terminology in Fiction platform (version 2; query page available at: <http://ic.votec.ileel.ufu.br/?lang=en>; choose an Area > Sciences and Fiction > Sagas > Star Wars; click the  icon). As for the platform, the intention of this work is to help translation apprentices and professional translators with their jobs. It is expected that, with the implementation of this proposal, the work of translators who deal with the franchise might be improved (eliminating or decreasing the possibility of multiple translations of the same term) and facilitated (from the terminology search interface of the Terminology in Fiction platform).

The TCC specific objectives were to: a. Find candidate terms in English used in subtitles and books; b. Find candidate terms in Portuguese used by translators; c. Analyze the most used vocabulary in English and its Portuguese counterpart in order to reflect on issues of translation; d. Make the online vocabulary proposal available for free through the Terminology in Fiction platform (TF, from now on; Fromm, 2024).

Therefore, this paper assumes that there is an expressive community of translators who can benefit from including Star Wars vocabulary in TF project. In the next section, some of the elements of Star Wars will be addressed, as well as translation decisions that the franchise has gone through around the world over the years.

2 About Star Wars

When released in theaters in 1977, George Lucas's Star Wars was an instantaneous phenomenon. At the time, two sequels were released (The Empire Strikes Back, 1980, and The Return of the Jedi, 1983) forming what is now known as the classic trilogy. One second trilogy was released between 1999 and 2005 and a third one between 2015 and 2019. These films make up the Skywalker Saga, which tells the story of a family of people sensitive to The Force, fighting for the fate of a very distant galaxy (as they point out at the beginning of each film), from the rise, fall and redemption of Anakin Skywalker, the adventures of his children (Luke and Leia), to the young Rey's journey to find her place in the universe. The franchise also has movies independent to the saga, among animations (Star Wars: The Clone Wars, 2008), films produced for television (The Star Wars Holiday Special, 1978; Caravan of Courage: An Ewok Adventure, 1984 and Ewoks - The Battle of Endor, 1985) and spin-offs (Rogue One: A Star Wars Story, 2016

² Trabalho de Conclusão de Curso – undergraduate thesis.

and Han Solo: A Star Wars Story, 2018). Released simply as Star Wars, it was only with the success of VHS that the films were finally renamed along the lines of the original vision of author George Lucas. The first film then becomes “Star Wars: Episode IV – A New Hope”, indicating that the future would bring a new trilogy about the past of this universe.

According to Lucas (2011), the film was not expected to become such a success, since several elements of the project were exactly opposite to the conventions of science fiction cinema and space operas at the time. Inspired by Akira Kurosawa, the first act of A New Hope establishes who are some of the plot’s most powerful and relevant characters, but the story is defined from the point of view of the two figures of minor importance. In Episode 4, these characters are the droids C-3PO and R2-D2, while Kurosawa makes the same dynamic with Tahei and Matashichi, in The Hidden Fortress.

The mythology of Star Wars has been built, since the first film, around real-life elements as politics, religion, and the polarization of good versus evil. Religion, for example, is one of the central themes in the saga, being represented by the Force, a kind of spiritual energy that connects all living elements in the universe. It is then seen as a religion for the Jedi, a group that seeks balance in the galaxy by following a strict moral code linked to ethical and philosophical principles. Politics, on the other hand, plays an important role in the saga, whose narrative is developed in a galaxy with political and ideological fractures, in which characters constantly compete for dominance of power. Topics as tyranny, democracy, corruption, and political oppression, for instance, are addressed throughout of the trilogies involving complex political issues, such as corruption in the Senate and the tyranny of the Galactic Empire.

From the grandeur of the universe established in the classic trilogy, a huge amount of additional material has been produced since the release of the first film in 1977. Among novels, comics, encyclopedias, games and, more recently, several television series, the Star Wars universe gained an established mythology in hundreds of copies in various media, constituting a vast universe that encompasses a significant number of planets, vehicles, groups, technologies, weapons, and droids, for example.

With the constant expansion of the Star Wars universe, the franchise very soon became a brand, present worldwide by becoming a symbol of pop culture. An example of the efforts, so far by George Lucas and the former 20th Century Fox (now a division of the Disney Group), to strengthen the marketing character of the franchise, is evident in the way in which it becomes recognized worldwide with the end of the first trilogy. The films, which until then had their titles translated around the world (*La guerre des étoiles*, in France; *Guerre stelari*, in Italy and *Guerra nas Estrelas*, in Brazil, for example), became known worldwide simply as Star Wars with the release of the second trilogy, in 1999.

Since the release of the first film, it is safe to say that the franchise has secured an important space in popular culture. In 2015, with the release of Star Wars: Episode VII – The Force Awakens, the franchise was officially revived and currently lives its most active moment, with several series, animated and in live-action, in activity and being translated, through dubbing or subtitling, around the world.

Therefore, it is important to highlight the complexity of these translation works in the more diverse media. Due to the short deadlines required by agencies, studios and companies requesting translation and localization work, the professional may encounter difficulties in locating the specifics of the technical vocabulary of franchises like Star Wars. Such difficulties facilitate the incidence of errors, since a shorter research time might result in terminological mistakes.

3 Theoretical framework

In this section, the theoretical references that support the research will be analyzed. We will start from the studies of Terminology (more specifically from the Communicative Theory of Terminology, by Cabré, 1999), Terminography, the studies of Barbosa's Ethnoterminology (2006, extending its principles from literature to audiovisual), and the resources used – WordSmith Tools 7, and the TF project.

3.1 Terminology

It is called “Terminology” (with the initial capital letter) the study about the terms and “terminology” (with the initial lowercase letter) referring to the specific terms and vocabularies of a certain area, being an interdisciplinary field that offers support to several other subjects. According to Lara (2004), “Terminology” has a polysemic meaning, encompassing both theoretical and methodological approaches as well as concrete terminology. A theoretical perspective offers methodologies that describe, organize and transfer knowledge, establishing principles that govern the compilation, formation of terms, structuring of conceptual fields, and the use and administration of terminologies. The meaning of concrete terminology, however, refers to a set of terms related to a specialized language (Lara, p. 234).

Still according to Lara (2005), the Terminological Unit is the object of Terminology. Unlike a word, which encompasses various possibilities of language use, a term refers to words used in specific communication situations, meaning language in specialized contexts. She also establishes that Terminology has both theoretical and practical objectives. The theoretical approach emphasizes the position of terminologies in the knowledge and practice of experts. The goal is to analyze, through specific methods, the use of terminologies in different contexts. The practical approach, on the other hand, occurs, for example, in areas such as “translation, documentation, linguistic standardization, and linguistic planning” (Lara, 2005, p. 6). It is safe to say, then, that it plays a fundamental role in the development of linguistic resources as vocabularies and glossaries, for instance.

Terminological work has seen an increasing use of technological tools developed within the scope of Computational Linguistics and has been increasingly reliant on Corpus Linguistics methodology. The goal, according to Zamora (2015), is to provide a precise description of terminology in a specific field of knowledge, establishing linguistic or terminological equivalences more accurately for research focused on translation (Zamora, 2015). In the realm of literary translation, video game translation, and audiovisual content, such studies represent a potential advancement for the quality of various translated materials. When dealing with expanded universes and well-established sagas, it is quite common to encounter an extensive terminological field which, without proper attention, might result in different translations of the same term as new products are released and translated by different individuals and teams. Terminological analysis through a corpus can assess historically established linguistic patterns by frequency, thereby avoiding or eliminating the possibility of multiple translations for the same term.

Thus, it can be concluded that Terminology plays a fundamental role in the study of specific vocabulary both in real or fictional languages/universes, and by understanding the importance of the field, it is possible to explore its application in many contexts. The study of a particular terminology ensures an understanding of the nature of terms and their uses, which aids in standardization and optimization of work across various fields. In summary, Terminology is, according to Dias (2000, p. 90), in practice, “a set of methods and activities aimed at collecting, describing, processing, and presenting terms; as a product, it is a set of terms, or vocabulary, of a specific specialty”.³

3.2 Communicative Theory of Terminology

Upon analyzing the foundations of Maria Teresa Cabré works, Fromm (2020) observes that the Communicative Theory of Terminology presents itself in an antagonistic manner to the General Theory of Terminology of Wüster. The latter, rooted in a considerably rigid notion regarding term analysis, showed little flexibility in the study of contexts. Thus, the distinctive aspect of the Communicative Theory of Terminology lies in its emphasis on the dimension of how a term can function, varying between common language and specialized areas; this means that any word can be a term in a specific specialized context (in a vocabulary or glossary), while being merely a lexeme (in a dictionary) in other more general linguistic contexts.

Consequently, Cabré (1999) establishes the Communicative Theory of Terminology to recognize the interdisciplinary nature of Terminology, the relationship between general and specialized knowledge, and the interdisciplinarity and plurality of terms. Additionally, the theory values the polysemy of lexical units, the existence of synonyms, and the diversification of discourse concerning the topic.

3.3 Ethnoterminology

According to Latorre, Ethnoterminology consists of the “intersection zone between the studies of literary discourses and those of specialized languages or terminologies” (2013, p. 73). This work uses the concept of Ethnoterminology, widely applied in the literary field, to analyze technical discourse in a multimedia universe of scientific fantasy.

The Star Wars universe is brimming with elements existing in real life but reimagined in its mythology to establish a rich world composition anchored in bases familiar to the audience. An example is the concept of the “Force” itself, which gains religious values when applied in the narrative (as can be seen in the definition present in TF). This movement is reflected in terminological analysis when terms are encountered that, within the established fictional universe, take on new meanings and/or broader or specific definitions. Regarding the connections between common language and specialized languages, Barbosa warns that “[...] lexical units are multifunctional. The precise establishment of their function depends on their insertion into a discursive norm, which then determines the status of word or term” (2006).

³ Originally, in Portuguese: “[...] um conjunto de métodos e atividades voltado para coleta, descrição, processamento e apresentação de termos; como produto, é um conjunto de termos, ou vocabulário, de uma determinada especialidade.”

According to Fromm (2011), the pursuit of verisimilitude in narratives results in the presence of specific terminology in fictional works. To reach this conclusion, one of Fromm's advisees analyzed *Star Trek* (Peixoto, 2014), a series relatively similar to our studies (also a space saga) and observed the massive presence of unique terminology derived from real fields such as Physics, Astronautics, and Astronomy. Similar trends can be found in *Star Wars*, in the fields of Physics (Hyperspace, Lightspeed) and Militarism (Star Destroyer), for example.

Therefore, this paper is based on the premise that the same approach can be applied to analyzing the *Star Wars* universe, considering the rich world-building established by various authors since the 1970s through comics, books, series, animations, and the acclaimed film series.

3.4 Terminography

Terminography is a practical discipline closely linked to Terminology. Finatto (2014) examines the objective of this field of study as follows:

Description of the linguistic, conceptual, and pragmatic properties of terminological units in one or more languages in order to produce reference works such as dictionaries, glossaries, vocabularies in print or electronic format, terminological databases, and specialized knowledge bases (Finatto, 2014, p. 439).⁴

Therefore, Terminography can involve the production of technical vocabularies or glossaries for a specific field, organizing terms and their definitions systematically. It is an area that encompasses the search and systematization of relevant terms within a specific domain, so that other professionals can benefit from reference works that may be enriched by this effort.

3.5 Wordsmith Tools

Wordsmith Tools is a software developed by linguist Mike Scott in 1996 at the University of Liverpool. Even after so many years, it continues to be updated and, at the time of this writing, is in its version 9, released in January 2024. Published by Oxford University Press and Lexical Analysis Software, Wordsmith Tools integrates tools for textual analysis, allowing, for example, the study of word frequency, keyword listing, and identification of collocations. In its interface, Wordsmith Tools presents three main tools (as can be seen at the top of Figure 1): Concord, Keywords, and WordList.

⁴ Descrição das propriedades linguísticas, conceituais e pragmáticas das unidades terminológicas de uma ou mais línguas, a fim de produzir obras de referência, tais como dicionários, glossários, vocabulários em formato papel ou eletrônico, bases de dados terminológicos e bases de conhecimento especializado.”

[illegible]

Berber Sardinha summarizes them well in the following paragraph:

The software also provides secondary tools for corpus processing, such as the Text Converter and the Corpus Checker. According to Ribeiro (2004), the use of tools like Wordsmith Tools can still have certain limitations for translators due to the complexity of corpus compilation. Therefore, the existence of a resource like the Online Technical Vocabulary, or VoTec, by Fromm (2007) and its derivative projects (Fromm; Lisboa, 2024), proves to be important for this purpose.

Rev. Est. Ling., Belo Horizonte, v.34, n.1, p.130-153, 2026

3.6 Terminology in Fiction

Terminology in Fiction, a derived version of the VoTec (Online Technical Vocabulary; available at: <http://votec2.ileel.ufu.br>) project (Fromm; Lisboa, 2024), emerges as a web-based terminological management environment aimed at assisting novice and professional translators in projects that demand heavy terminological translations (Fromm, 2007). The work is based on the premise that there is a significant need for integration between the fields of Translation and Terminology since, in the author's words, "increasingly, the translator also plays the role of terminologist and occasionally terminographer" (Fromm, 2007, p. 13).

The TF platform is a data bank supplied by researchers with terms previously gathered through corpora, processed by software such as Wordsmith Tools, for example. Along with the term candidates (enumerated from the WST's Keywords tool), the platform allows the researcher to input all contexts in which the word under analysis was used (from the WST's Concordance tool), enabling the registration of the term, the contexts that best explain its definition, and the final definition created by the researcher solely based on the selected contexts. The terms registered on the platform go through an approval process by the administrator, who must assess the quality of the registered contexts and their sources, distinctive features, encyclopedic information, the final concept, and the definition of each term.

For the completion of this work, the version of the TF environment shown in the figure 2 below was used.

Figure 2 - Astromech term in TF's databank (not published).

Online Technical Vocabulary

Full Screen | Português

Term: Astromech

Previous step

Save

Quit without save

Contexts

Final Concept / Definition

Contexts

	Example	Concept	Source
1	I believe this C1 astromech is the droid infiltrator	Droid	RebelsEN 06/15/2023
2	"What's worse?" Mill squinted, as if that might make it easier to understand the beeps of astromech droid language.	Beep	SWBrotherhood 06/15/2023
3	Obi-Wan came up alongside them, soon followed by R2-D2. "Scan the area for guards or threats," he said to the astromech, though he gave her space to breathe.	Scan	SWBrotherhood 06/15/2023
4	Poe's X-wing was almost ready. He watched as	There's an astromech compartment in ships	SWSkywalkerLIVRO

Data

Distinctive Traces

Semantics

Equivalent Term

Cross-references

Encyclopedic Information

Data

Part of Speech: Noun

Number: Singular

Gender: Neutral

Acronym:

Complete Word:

Morphosyntactic Variants:

Meaning N°:

Corpus

Frequency order 1667

Term number of 179

02/05/2024 11:58 © 2007 FFLCH - ICMC Jr.

Source: TF databank

In general terms, the Corpus Linguistics approach used in this project is mainly based, as can be seen, on issues of Terminology (a word in a specific context of use, such as in a technical area), Terminography (development of products from terminological studies), the Communicative Theory of Terminology (interdisciplinarity and plurality in the treatment of what is considered a term), Ethnoterminology (the idea that any word can have the status of a term within a fictional universe), the lexical analysis software (WordSmith Tools) necessary to process the material from the chosen saga (Star Wars), and the platform (Terminology in Fiction) that we used to create bilingual entries. In the next section, we will analyze the processes of compilation, cleaning, systematization, and exploration of the corpus.

4 Methodology

The corpus for the project was compiled over the course of three months, with the English files being collected first, followed by the translated version of each one. In both languages, the corpus was divided into five parts: Animated Series/Animações, Books/Livros, Films/Filmes, TV Series/Séries, and Everything/Tudo, with the latter being the compilation of all the others – important for the creation of the WordList later.

In total, 760 files were downloaded, with 380 in the original language and 380 translated into Portuguese. For the corpus composition, only content considered canonical up to the time of compilation was selected, and it needed to be easily accessible. This excluded, for example, comic books, which despite being canonical, presented considerable difficulty in converting the issues into text documents.

Books constitute the most substantial part of the corpus, being nearly three times larger than the second-largest category, which is animations. This material corresponds to a total of 38 canonical novels that have been officially or unofficially translated into Portuguese. The selected works belong, within the Star Wars timeline, to the High Republic Era, the Republic Era, the Imperial Era, and the New Republic Era.

The entire Legends series, formerly known as the Expanded Universe, was disregarded, as were short stories, children's series, art and behind-the-scenes books, dictionaries, and encyclopedias. Clearly, books represent the richest portion of the Star Wars saga's world-building, with the first work being published even before the launch of the cinematic universe. Therefore, if all this data were considered for the study, the result would easily exceed a thousand files, significantly increasing the complexity of the compilation process.

In English, the books were found in online repositories. All of them were downloaded in EPUB format and later converted to TXT. This conversion presented some difficulties initially, as it couldn't preserve the character encoding of the original text correctly, which is a common issue when dealing with the diacritic accents of the Portuguese language.

Finding translations for the books into Portuguese was also a time-consuming process. The solution found to complete the research was to rely on the ongoing work of the non-professional group "Tradutores dos Whills," which translates all Star Wars literature not officially published in Brazil into Brazilian Portuguese. It was particularly interesting to analyze how the team maintained the original visual identity and graphic design, making it more accessible to read works that might never make it to Brazilian bookstores. Therefore, even though many works were officially translated and used in the corpus compilation, it is important to

mention the presence of non-professional translations in this project, which encompass not only books but all other categories as well.

The books that form the corpus are: *A New Dawn*, *Aftermath*, *Aftermath: Empire's End*, *Aftermath: Life Debt*, *Ahsoka*, *Alphabet Squadron*, *Battlefront - Twilight Company*, *Battlefront II - Inferno Squad*, *Bloodline*, *Brotherhood*, *Catalyst*, *Dark Disciple*, *Heir to the Jedi*, *Leia, Princess of Alderaan*, *Lords of the Sith*, *Lost Stars*, *Master & Apprentice*, *Phasma*, *Queen's Peril*, *Queen's Shadow*, *Rebel Rising*, *Resistance Reborn*, *Rogue One*, *Tarkin*, *The Force Awakens*, *The High Republic - Into the Dark*, *The High Republic - Light of the Jedi*, *The High Republic - Midnight Horizon*, *The High Republic - Out of the Shadows*, *The High Republic - The Rising Storm*, *The High Republic - The Fallen Star*, *The Last Jedi*, *The Rise of Skywalker*, *Thrawn - Alliances*, *Thrawn - Treason*, *Thrawn Ascendancy - Chaos Rising*, *Thrawn Ascendancy - Greater Good*, *Thrawn Ascendancy - Lesser Evil*.

After collecting the books, the process of searching for the constituent elements of the second most significant category in the corpus, which comprises animations, was initiated.

The animations that form the corpus are: one season of *The Bad Batch*, seven seasons of *The Clone Wars*, one season of *Droids*, two seasons of *Ewoks*, four seasons of *Star Wars Rebels*, two seasons of *Star Wars Resistance*, one season of *Visions*.

The expressiveness of this corpus arises from the longevity of some of these series, such as *The Clone Wars*, which reached a significant number of 133 episodes spread across seven seasons. *Rebels*, in turn, concluded after 75 episodes in four seasons. In terms of the number of files, the corpus is consequently the most substantial, with 304 subtitle files for each language.

In addition to animations, subtitles were also used as items for the composition of the film and TV series corpora.

Forming the film corpus are: *Star Wars: Episode IV - A New Hope* (1977), *The Star Wars Holiday Special* (1978), *Star Wars: Episode V - The Empire Strikes Back* (1980), *Star Wars: Episode VI - Return of the Jedi* (1983), *Caravan of Courage: An Ewok Adventure* (1984), *Ewoks: The Battle for Endor* (1985), *Star Wars: Episode I - The Phantom Menace* (1999), *Star Wars: Episode II - Attack of the Clones* (2002), *Star Wars: Episode III - Revenge of the Sith* (2005), *Star Wars: The Clone Wars* (2008), *Star Wars: Episode VII - The Force Awakens* (2015), *Rogue One: A Star Wars Story* (2016), *Star Wars: Episode VIII - The Last Jedi* (2017), *Solo: A Star Wars Story* (2018), *Star Wars: Episode IX - The Rise of Skywalker* (2019).

And the series: one season of *The Book of Boba Fett*, two seasons of *The Mandalorian*, one season of *Obi-Wan Kenobi*.

For live-action and animated series, all episodes released up to the time of corpus compilation were considered. *The Mandalorian* was in its second season, while *Obi-Wan Kenobi* and *The Book of Boba Fett* were in their first seasons. Likewise, regarding animations, *The Bad Batch* and *Visions* were in their first seasons. *The Clone Wars* was in its seventh season, *Rebels* in its fourth, and *Resistance* in its second. *Ewoks* and *Droids* were concluded in the 1980s. New releases from the *Star Wars* franchise for the Disney+ streaming platform, such as *Ahsoka*, *Star Wars: Young Jedi Adventures*, *Tales of the Jedi*, *The Acolyte*, and projects subsequent to these, were not included in the corpus as they had not yet been released during the compilation period.

4.1 Corpus Management

The corpora in both languages were divided into five parts. In Portuguese: “Desenhos”, “Filmes”, “Livros”, “Séries”, and “Tudo”. In English: “Animated Series”, “Films”, “Books”, “TV Series”, and “Everything.” All files from each corpus and their respective Wordlists were saved, separated by languages in distinct folders.

For corpus processing, the first step was to convert all subtitles and books to TXT format. For subtitles, SubtitleEdit was used, which is capable of converting and cleaning files in bulk. For books, online file conversion tools like Convertio were employed. A second conversion, this time for encoding, was necessary to use the corpus in WordSmith Tools. To achieve this, the software itself was used, specifically the Text Converter tool, which enables files to be converted into Unicode encoding.

Failure in the corpus encoding can lead to processing errors, as illustrated below in Figure 3 (with what should be the word “didn’t”).

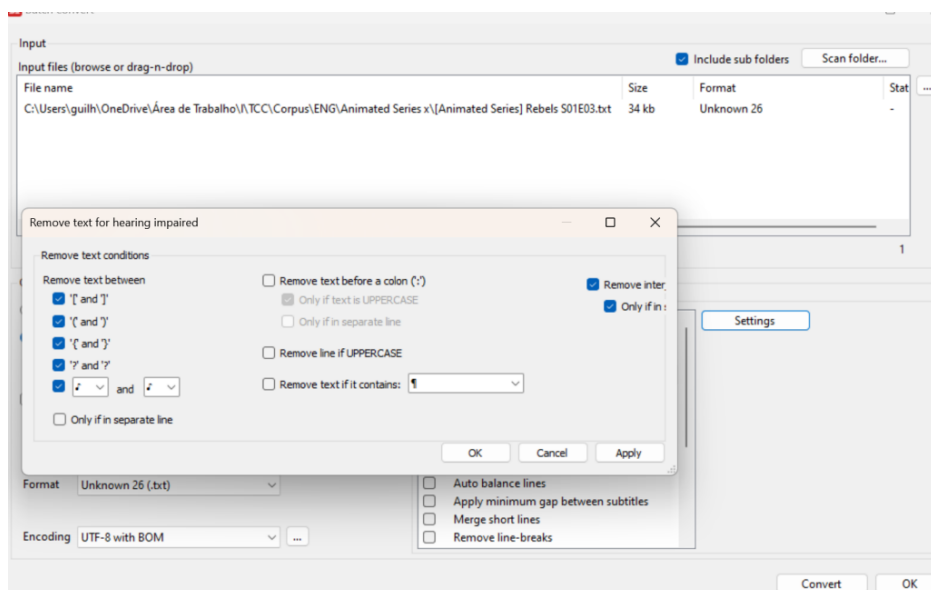
Figure 3 - Example of encoding error in WordSmith wordlist.

67	EVEN
68	KNOW
69	DIDNÂ€™T
70	DOWN
71	THROUGH

Source: extracted from Wordsmith Tools

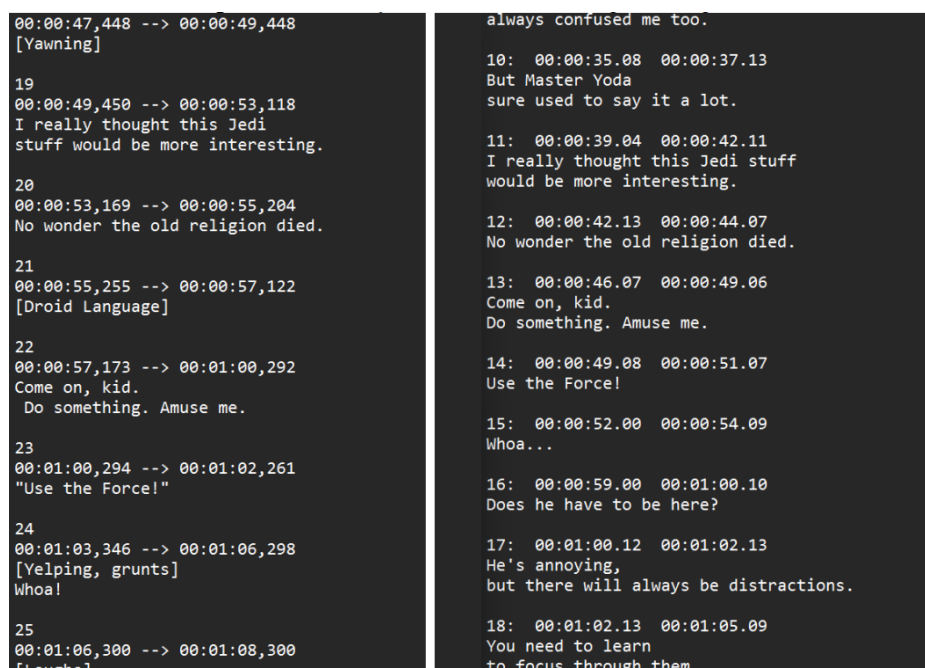
The corpus of animations, as well as series and films, consists of subtitles that were freely available on the internet. The English subtitles found are often typically labeled as “HI,” meaning Hearing Impaired (for those with total or significant hearing loss). This means that these subtitles include audio descriptions that do not correspond to the dialogue, such as music, screams, cries, laughter, sounds, and different intonations. Since such descriptions are not the focus of our analyses and would therefore influence the data obtained, the English subtitles underwent an additional process to remove all this information. Initially, this was done manually, and later with the assistance of the SubtitleEdit software, which performs this process automatically (as seen in Figures 4 and 5).

Figure 4 - Subtitle Edit program configuration.



Source: Subtitle Edit

Figure 5 - Subtitle Edit results.

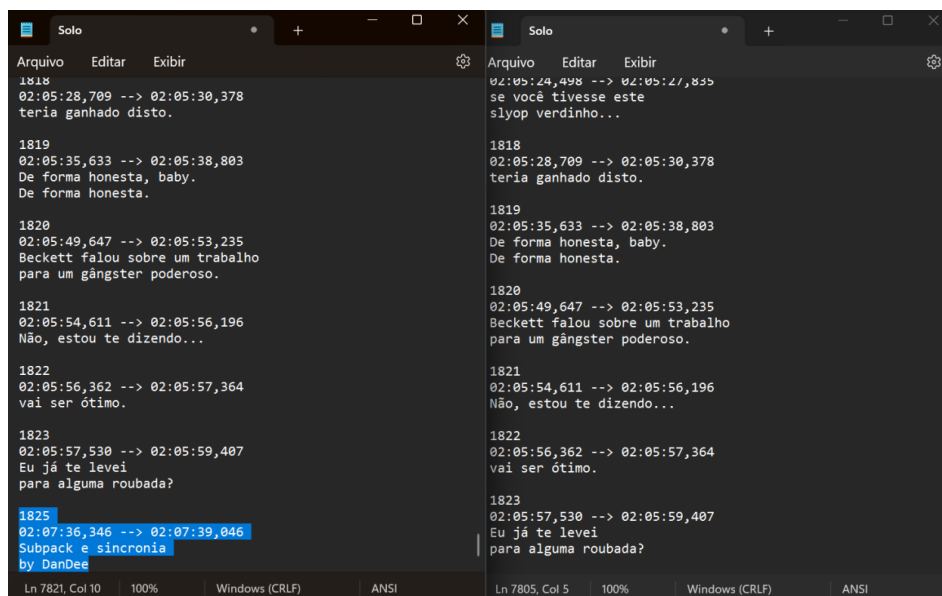


Source: screenshot by the author

In the case of translated subtitles, although HI elements are relatively rare, the items needed to undergo a cleaning process for a different reason. These subtitles, especially in the field of older animations and series, were produced by non-professional groups that credited their work in every episode two or three times. While understanding that the work of these teams was crucial for the completion of this research, the credits had to be removed to prevent the corpus from being polluted. Unlike HI elements, which are systematically enclosed

in square brackets within the subtitle, these credits do not follow a standard pattern, making it impossible for software like SubtitleEdit to perform automatic removal. Therefore, the cleaning of translated subtitles was a manual task (as shown below in Figure 6).

Figure 6 - Subtitle translator group advertised in the subtitle.



Source: screenshot by the author

4.2 Selection of term candidates

When conducting the corpus analysis with Wordsmith Tools, keywords are identified based on their frequency statistically compared to a reference corpus. These keywords stand out due to their significant occurrence in relation to the overall corpus, providing relevant insights into the key terms used in the analyzed texts – in this research, the canonical universe of Star Wars.

The terms chosen for registration in the TF platform were selected from the first 500 words in the KeyWords list generated by WordSmith Tools. To generate this list of keywords, a version of the COCA (Corpus of Contemporary American English) was used as a reference corpus, limited to its first 100,000 words. The result can be seen in Figure 7 below.

Figure 7 - first 33 keywords from the Star Wars corpus.

N	Key word	Freq.	%	Texts	RC. Freq.	%	BIC	Log_L	Log_R	P
1	BACK	9.445	0,20	360	6.366		64.181,05	64.200,92	7,07	0,0000000000
2	TIME	7.762	0,17	364	4.065		55.012,21	55.032,07	7,44	0,0000000000
3	DIDN	5.533	0,12	31	0		49.994,90	50.014,76	139,30	0,0000000000
4	THRAWN	5.392	0,12	22	0		48.720,35	48.740,21	139,26	0,0000000000
5	JEDI	5.215	0,11	242	0		47.120,38	47.140,25	139,21	0,0000000000
6	GOOD	4.646	0,10	362	113		40.910,84	40.930,70	11,87	0,0000000000
7	LL	4.171	0,09	31	0		37.683,29	37.703,16	138,89	0,0000000000
8	VE	3.843	0,08	33	0		34.718,39	34.738,25	138,77	0,0000000000
9	EYES	3.796	0,08	123	461		31.383,93	31.403,79	9,55	0,0000000000
10	LEIA	3.260	0,07	33	0		29.448,44	29.468,30	138,53	0,0000000000
11	DON	4.104	0,09	31	2.176		29.020,97	29.040,83	7,42	0,0000000000
12	WASN	3.189	0,07	31	0		28.806,65	28.826,51	138,50	0,0000000000
13	FIND	3.167	0,07	347	975		24.108,46	24.128,32	8,20	0,0000000000
14	CAPTAIN	2.625	0,06	178	18		23.493,35	23.513,21	13,69	0,0000000000
15	SHIP	5.839	0,13	285	26.683		22.729,52	22.749,38	4,31	0,0000000000
16	VOICE	3.371	0,07	71	2.580		22.363,94	22.383,80	6,89	0,0000000000
17	VADER	2.297	0,05	39	0		20.743,54	20.763,40	138,03	0,0000000000
18	SIDE	2.475	0,05	182	253		20.673,08	20.692,94	9,79	0,0000000000
19	DROID	2.452	0,05	201	256		20.455,55	20.475,41	9,76	0,0000000000
20	PADME	2.255	0,05	22	0		20.363,89	20.383,75	138,00	0,0000000000
21	ANAKIN	2.237	0,05	115	0		20.201,18	20.221,04	137,99	0,0000000000
22	FELT	2.531	0,05	64	848		19.069,94	19.089,80	8,08	0,0000000000
23	THINGS	2.148	0,05	260	45		18.958,82	18.978,68	12,08	0,0000000000
24	COULDN	2.091	0,05	31	0		18.881,43	18.901,29	137,89	0,0000000000
25	SHIPS	2.237	0,05	142	397		17.976,46	17.996,33	9,00	0,0000000000
26	OBI	1.988	0,04	92	0		17.950,38	17.970,24	137,82	0,0000000000
27	ROOM	1.872	0,04	113	157		15.800,18	15.820,05	10,08	0,0000000000
28	MAN	1.927	0,04	160	288		15.693,40	15.713,27	9,25	0,0000000000
29	HADN	1.711	0,04	31	0		15.446,48	15.466,34	137,60	0,0000000000
30	DROIDS	1.735	0,04	178	78		15.021,75	15.041,62	10,98	0,0000000000
31	POE	1.583	0,03	24	0		14.289,44	14.309,30	137,49	0,0000000000
32	POWER	1.813	0,04	220	459		14.092,02	14.111,88	8,49	0,0000000000
33	TARKIN	1.506	0,03	22	0		13.593,41	13.613,27	137,42	0,0000000000

Source: KeyWords tool, WordSmith.

The terms extracted from the first 500 keywords were classified into nine specific semantic categories, which later became sub-areas in the TF database. The categories and pre-selected terms are indicated in table 1.

Table 1 – Gathered terms.

Vehicles	Species	Places	Groups	Weapons	Droids	Techno-logy	Profes-sions	Reli-gions
Starfigh-ter	Wookie	Coruscant	Rebels	Lightsaber	Astromech	Droid	Storm-trooper	Force
Star Destroyer	Ewok	Alderaan	Clones	Blaster	Protocol Droid	Hyperspace	Twilight Soldier	Sith
Spaceport	Hutts	Naboo	Separatists	Laser Cannons	Medical Droid	Hyperdrive	First Senator	Jedi
Imperial Star Destroyer		Manda-lore	Youngling		Server Droid	Datapad		
		Endor	Padawan		Pill Droid	HoloNet		
					Battle Droid	Lightspeed		
					Tactical Droid			
					Droid Tri Fighter			
					Moderator Droid			

Source: KeyWords tool from WordSmith.

In table one, we can see some of the terms found within the selected range of the first keywords for each category. For subsequent registration in TF, two representative examples from each category were chosen to best illustrate the use of the platform and the usefulness of the proposal.

4.3 Exploration of concordance lines

The exploration of terms was made possible through the Concord tool in WordSmith Tools, which allows researchers to identify the contexts in which a specific word or expression occurs in the corpus. For example, if a researcher is working with a Star Wars corpus in English, they can find all the contexts in which the word “Force” was used, whether they are cultural, scientific, religious, etc. Furthermore, it is also possible to analyze the words that surround the term, its cotext, which can even lead to the discovery of new technical phraseology⁶ – as was the case with “Imperial Star Destroyer,” which demonstrated relevance during the exploration of Star Destroyer contexts.

For registration in TF's databank, it is necessary to gather contextual examples that demonstrate different concepts related to the studied term, allowing the creation of a terminological definition based on the word in use.

In figure 8 we can find examples of contexts to define the term “Ewok”:

Figure 8 - Concordance lines for the term EWOK in Portuguese.

N	Concordance	Set	Tag	Word	#Sent	Sent	Para	H	H	Sect	Sect	File	Date	%
					# Pos.	# Pos.	# Pos.	# Pos.	# Pos.	# Pos.	# Pos.			
49	Ewok. 99: 00:07:43.04 00:07:45.15 Quer dizer que o Ewok que vimos... 100: 00:07:45.17 00:07:49.03				817	94	9	0	816	0	816	[Desenho] Ewoks S01E13.txt	2022/nov/14 00:	35%
50	242: 00:18:21.06 00:18:23.22 Esse não pode ser o Ewok que me derrotará. 243: 00:18:26.22 00:18:				1,821	252	10	01..0	01..0	01..0	01..0	[Desenho] Ewoks S02E06.txt	2022/nov/14 00:	82%
51	gorjeta. 319: 00:22:20.13 00:22:24.11 Acho que os Ewok fazem mal para nossos negócios. 320: 00:22:				2,564	324	8	02..3	02..3	02..3	02..3	[Desenho] Ewoks S01E05.txt	2022/nov/14 00:	95%
52	02:48:20 00:02:52.09 Pronto. Kneesaa, você é outra Ewok. 41: 00:02:52.11 00:02:54.14 Sim. Mas qual?				368	30	6	0	367	0	367	[Desenho] Ewoks S02E13.txt	2022/nov/14 00:	15%
53	Chirpa. 43: 00:04:13.08 00:04:15.18 Nem pensar, Ewok. 44: 00:04:16.11 00:04:19.19 Venha e se junte				353	33	7	0	352	0	352	[Desenho] Ewoks S02E04.txt	2022/nov/14 00:	19%
54	26.05 00:17:28.07 O que faremos com a pequena Ewok? 216: 00:17:28.20 00:17:30.00 E a Kneesaa!				1,669	221	11	01..8	01..8	01..8	01..8	[Desenho] Ewoks S02E08.txt	2022/nov/14 00:	74%
55	: 00:08:41.09 00:08:46.16 Então, minha pequena Ewok, não é divertido ser um Jinda? 137: 00:08:				1,146	116	8	01..5	01..5	01..5	01..5	[Desenho] Ewoks S01E05.txt	2022/nov/14 00:	42%
56	- uma pequena bolota dada e ela pelo pequeno Ewok conhecido como Wicket. Ela cultivou a planta				60,2705..	3	19	02..9	02..9	02..9	02..9	[Livro] Aftermath - Divida de Honra.txt	2022/out/14 00:	91%
57	47: 00:04:33.13 00:04:37.07 Agora, pequeno Ewok, junte-se ao meu totem 48: 00:04:39.19 00:04:				379	37	7	0	378	0	378	[Desenho] Ewoks S02E04.txt	2022/nov/14 00:	17%
58	69: 00:05:51.18 00:05:56.07 - Teebo? - Pequeno Ewok, junte-se ao meu totem. 70: 00:05:57.20 00:05:				553	65	4	0	552	0	552	[Desenho] Ewoks S02E04.txt	2022/nov/14 00:	24%
59	134: 00:10:07.19 00:10:12.08 Nada mau, pequeno Ewok, mas tenho mais azar para você. 135: 00:10:				1,139	139	8	01..8	01..8	01..8	01..8	[Desenho] Ewoks S02E11.txt	2022/nov/14 00:	44%
60	a árvore-santuário que lhe fora dada pelo pequeno Ewok Wicket. Ela nunca foi capaz de sentir a árvore				77,8177..	6	14	07..6	07..6	07..6	07..6	[Livro] Aftermath - Fin do Império.txt	2022/out/14 00:	117%
61	14:05 05 Eles devem ter lhe dado aquele pirralho Ewok 174: 00:14:05.07 00:14:08.03 para escalá-la.				1,439	166	12	01..8	01..8	01..8	01..8	[Desenho] Ewoks S01E12.txt	2022/nov/14 00:	64%
62	217: 00:17:57.08 00:18:02.03 Que linda princesa Ewok. 218: 00:18:02.05 00:18:03.10 Não. 219: 00:				1,974	263	8	01..3	01..3	01..3	01..3	[Desenho] Ewoks S02E02.txt	2022/nov/14 00:	79%
63	214: 00:17:43.13 00:17:46.13 Temos uma princesa Ewok para a Rainha Lesmona. 215: 00:17:47.23 00:				1,955	260	8	01..4	01..4	01..4	01..4	[Desenho] Ewoks S02E02.txt	2022/nov/14 00:	78%
64	200: 00:16:37.01 00:16:40.21 Nossa, uma princesa Ewok para a Rainha Lesmona. 201: 00:16:45.04 00:				1,836	243	8	01..5	01..5	01..5	01..5	[Desenho] Ewoks S02E02.txt	2022/nov/14 00:	73%
65	25:03 00:13:30.00 - Vão, tragam-me uma princesa Ewok. - Sim, Rainha Lesmona. 163: 00:13:35.16 00:				1,504	198	10	01..3	01..3	01..3	01..3	[Desenho] Ewoks S02E02.txt	2022/nov/14 00:	60%
66	101: 00:06:58.09 00:07:01.06 uma figura de proa Ewok. 102: 00:07:01.13 00:07:04.22 - Ewok? - Da				762	93	14	0	761	0	761	[Desenho] Ewoks S02E09.txt	2022/nov/14 00:	34%
67	202: 00:13:49.04 00:13:51.12 Levanta se pudesse, Ewok. 203: 00:13:51.14 00:13:53.09 Mas,				1,709	194	8	01..8	01..8	01..8	01..8	[Desenho] Ewoks S01E05.txt	2022/nov/14 00:	62%
68	214: 00:14:02.20 00:14:04.15 Você disse qualquer Ewok. 215: 00:14:05.02 00:14:07.01 Tudo bem, se				1,692	231	8	01..1	01..1	01..1	01..1	[Desenho] Ewoks S02E07.txt	2022/nov/14 00:	66%
69	58:02 00:13:01.23 Aposto que transformo qualquer Ewok em um guerreiro em um dia. 197: 00:13:02.13				1,555	211	9	01..4	01..4	01..4	01..4	[Desenho] Ewoks S02E07.txt	2022/nov/14 00:	61%
70	00:10:54.15 00:10:56.09 - Sabão Ewok! - Sabão Ewok! 150: 00:10:56.19 00:10:59.10 Vamos, rápido.				1,218	145	4	01..7	01..7	01..7	01..7	[Desenho] Ewoks S01E02.txt	2022/nov/14 00:	53%
71	Pequeno! 252: 00:19:16.06 00:19:18.05 O sabão, Ewok! 253: 00:19:20.17 00:19:22.17 Entregue.				2,046	258	7	02..5	02..5	02..5	02..5	[Desenho] Ewoks S01E02.txt	2022/nov/14 00:	89%
72	se limpando. 95: 00:06:47.02 00:06:48.09 Sabão Ewok. 96: 00:06:48.22 00:06:50.19 Não há nada				800	86	6	0	799	0	799	[Desenho] Ewoks S01E02.txt	2022/nov/14 00:	39%
73	temos aqui? 149: 00:10:54.15 00:10:56.09 - Sabão Ewok! - Sabão Ewok! 150: 00:10:56.19 00:10:59.10				1,215	144	7	01..4	01..4	01..4	01..4	[Desenho] Ewoks S01E02.txt	2022/nov/14 00:	53%
74	51:28 08 O cristal pertence a Kaink, a sacerdotisa Ewok. 361: 00:51:31.18 00:51:34.12 Mas antes que				2,805	462	12	02..4	02..4	02..4	02..4	[Filme] Caravana da Coragem.txt	2022/nov/14 00:	65%
75	59: 00:05:12.05 00:05:17.11 Um monstro selvagem Ewok com seis braços 60: 00:05:17.18 00:05:21.02				481	51	8	0	480	0	480	[Desenho] Ewoks S01E13.txt	2022/nov/14 00:	21%
76	frito. 45: 00:03:25.07 00:03:26.17 Eu também, Ewok. 46: 00:03:31.02 00:03:34.00 Mas alguns				370	33	7	0	369	0	369	[Desenho] Ewoks S01E08.txt	2022/nov/14 00:	17%
77	seja. Anasid sorri. - Eu poderia colocá-lo na terapia Ewok, em vez disso. Algumas das criaturas nativas				99,3833..	2	8	09..2	09..2	09..2	09..2	[Livro] Aftermath - Divida de Honra.txt	2022/out/14 00:	150%
78	35:09 00:16:38.07 É melhor fazer um bom trabalho, Ewok. 247: 00:16:38.09 00:16:40.01 Senão. 248: 00:				2,049	240	11	02..8	02..8	02..8	02..8	[Desenho] Ewoks S01E05.txt	2022/nov/14 00:	75%
79	se reunem para uma cerimônia tradicional Ewok. 301: 00:42:19.05 00:42:22.22 Antes de				2,330	398	13	02..9	02..9	02..9	02..9	[Filme] Caravana da Coragem.txt	2022/nov/14 00:	54%
80	: 00:12:05.21 00:12:05.05 É melhor olhar para trás, Ewok. 191: 00:12:17.16 00:12:20.16 Kwark. Corram.				1,591	211	10	01..0	01..0	01..0	01..0	[Desenho] Ewoks S01E07.txt	2022/nov/14 00:	57%
81	27: 00:01:59.21 00:02:04.03 Saudações da tribo Ewok, Rainha das Wisties. 28: 00:02:16.15 00:02:				244	13	8	0	243	0	243	[Desenho] Ewoks S01E01.txt	2022/nov/14 00:	11%
82	: 00:18:18.16 00:18:23.01 Lutaremos até o último ewok para proteger nossas árvores da alma. 275: 00:				2,177	262	9	02..6	02..6	02..6	02..6	[Desenho] Ewoks S01E10.txt	2022/nov/14 00:	84%

Source: Concordance, WordSmith Tools

For this term, five contexts were selected, each one supplying different concepts. The sources vary; in the case of “Ewok,” the excerpts are from the movie “Caravan of Courage: An Ewok Adventure” from 1984 and the animated series “Ewoks” from 1985. From the chosen con-

⁶ “There are various notions of what a phraseologism is, but in this text, I consider it to be any grouping of two or more words whose combined frequency is highlighted by lexical analysis programs.” (Fromm, 2020) Original: Existem várias noções do que seja um fraseologismo, mas considero o mesmo, neste texto, como qualquer agrupamento de duas ou mais palavras cuja frequência combinada seja destacada pelos programas de análise lexical.”

texts, it was possible to discover that Ewoks are a tribe of friends living in trees in the forests of Endor, with their own ceremonies, warriors, and wizards.

In the figure 9, we can analyze the selected contexts in order to create a definition.

Figure 9 - Selected contexts for the term EWOK.

Registered contexts

Example	Concept	Source	Actions
"Gather for a traditional Ewok ceremony."	Ceremonies	CaravanEwok 04/13/2023	edit - delete
"Ewoks! We're the Ewoks, raggedy Ewoks Living in the tall trees Living in the spiral Dancing in the forest On the moon of Endor Ewoks all together And we're having fun Friends together Friends forever Ewoks We're careless, little Ewoks We like adventure Helping friends in danger Out in the forest Sharing in the magic On the moon of Endor Ewoks all together And we're having fun Friends together Friends forever Ewoks We're the Ewoks, yeah"	Friends living in trees in a forest of Endor	Ewoks 04/13/2023	edit - delete
"It's Queen Izrina!" "Greetings from the Ewok tribe, Queen of all the Wisties."	Tribe	Ewoks 04/13/2023	edit - delete
"And the Ewok warriors will stop you good if you even show your face near there."	Warriors	Ewoks 04/13/2023	edit - delete
"But, Latara, no Ewok wizard can do that. Not since the Crystal Cloak was stolen."	Wizards	Ewoks 04/13/2023	edit - delete

registered contexts: 5

Source: TF databank

The use of Concord is not only useful for the elaboration of term definitions but also, among various uses, for exploring translation aspects – one of the themes of this work. Let's return to the example of "Star Destroyer". 385 different contexts of "Star Destroyer" were found in the English corpus, compared to only 21 in the Portuguese corpus. The clear discrepancy demonstrates that there were issues about its translation, and it was up to the researcher to verify one of the English contexts and locate the same passage in the Portuguese version to find the translated term. As a result, it was discovered that the term historically was more frequently translated as "Destróier Estelar" in Brazilian materials. Furthermore, 21 contexts were found where a term was either not mentioned in Brazilian materials or simply not translated, emphasizing the need for a terminological research source like TF, since the original title, in a foreign language, was retained - in contrast to the established standard by the Star Wars Brazilian translation.

After all these steps, the terms were ready for registration in TF. In the next section, we analyze how this stage was carried out.

5 The development of Star Wars entries in the vocabulary

As previously explained, the general objective of the project is to create a proposal for a Star Wars vocabulary in the TF platform. The query page of the TF project can be accessed for free from any computer or mobile device, making it a simplified environment for translators to consult.

5.1 Creation of the Star Wars project in TF

The project was created within the field of Science and Fiction Terminology, where there are already studies on works like “Farscape”, “Doctor Who”, and “Star Trek”, among others. However, this project was the first to include sub-areas, which further refines the research process. This means that when searching for terms, users can navigate through the categories as shown in Table 1.

5.2 Building the entries

Researchers who are TF users have access to the data bank page where they can register new terms and contribute to the platform. On the home page, as you can see in figure 10, researchers can access a list of all terms that have been registered but not yet approved or awaiting review. They also have the options to create a term and register a new source.

Figure 10 - Creating a term in TF.

Change password

New Term

Source register

Quit

Not yet approved card

Term	Language	Collected on	Collected by	Actions
Mandalore		11/05/2023	Guilherme Rodrigues Ferreira	editar
Mandalore		11/05/2023	Guilherme Rodrigues Ferreira	editar
Wookiee		11/05/2023	Guilherme Rodrigues Ferreira	editar
Wookiee		11/05/2023	Guilherme Rodrigues Ferreira	editar
Ewoks		21/04/2023	Guilherme Rodrigues Ferreira	editar
Ewoks		13/04/2023	Guilherme Rodrigues Ferreira	editar
Destróier Estelar		13/04/2023	Guilherme Rodrigues Ferreira	editar
Star Destroyer		12/04/2023	Guilherme Rodrigues Ferreira	editar
Caça Estelar		11/04/2023	Guilherme Rodrigues Ferreira	editar
Starfighter		11/04/2023	Guilherme Rodrigues Ferreira	editar

Source: TF databank

The first step in creating a term is to name it. Some words have the possibility of being in singular or plural form, as is the case with “Ewoks” in figure 10, which was not registered as “Ewok” in the singular form. For this work, we considered the most frequent form in the corpus for registration in TF – which means that in the entire canonical universe of Star Wars, the usage of the term “Ewoks” is more frequent than “Ewok.” However, both forms (singular and plural) were subsequently lemmatized, meaning they were grouped in the corpus as a single term for analysis. This explains the possibility of finding contextual examples in the singular form of a term that was registered in the plural. After creating the term, it is necessary to choose the language, the major area, subarea 1, and subarea 2. The first belongs to the material being worked on, in our case, Star Wars. The second consists of the categories that were previously registered. Figure 11 shows how this process is done, starting with the registration of the term “Wookiee”, for example.

Figure 11 - Creating the term WOOKIE

New Term

[Back to control panel](#)

Step 1

Wookiee

Language
Choose a language: English

Ontology
High Area: Sciences and Fiction
Sub-area 1: Sagas
Sub-area 2: Star Wars
Sub-area 3: Species

- Choose a sub-area
- Droids
- Organizations
- Places
- Professions
- Religion
- Species**
- Technology
- Vehicles
- Weapons

Source: TF databank.

The second step is to register the contexts. The definition of the term will be created solely from data extracted from real contextual fragments, as preconized by the VoTec general project (Fromm, 2007). As previously explained, with the help of the Concord tool, the researcher can search for a specific term in the corpus and analyze all the situations in which it was used. The goal is to capture a variety of concepts.

The context registration screen in TF can be seen in figure 12.

Figure 12 - Registration of the examples and contexts for the WOOKIE term

New Context

Step 2

Context data

Example*:

Concept*:

Source*: [Register new](#)

source

Collected on*: (month/day/year e.g.: 03/18/2007)

Registered contexts

Example	Concept	Source	Actions
"Let him have it. It's not wise to upset a Wookiee."	Dangerous	ANewHope 05/10/2023	edit - delete
General Hux led the way. They passed officers and stormtroopers, droids and maintenance crew, and although a giant, hairy Wookiee occasionally made someone do a double take, Hux's presence gave them unhindered passage through the ship's corridors.	Giant and hairy	RiseSkywalkerBook 05/10/2023	edit - delete
"Thanks for not breaking my neck." The Wookiee replied with a guttural, modulated rumble. Finn chose to interpret it as an apology of sorts.	Guttural voice	TheForceAwakens 05/10/2023	edit - delete
"Even a wookiee can't crush First Order armor."	Strong	TheForceAwakens 05/10/2023	edit - delete

registered contexts: 4

Source: TF databank.

As soon as a passage in the corpus that adequately explains the term is selected, that excerpt is inserted into the "Example" field. The second context registered in Figure 12, for example, serves to showcase the physical characteristics of a Wookiee – this is a kind of concept, and the more concepts we have, the richer the definition will be. Additionally, the source and collection data are also entered.

The next step involves providing details about the term and creating a concept and definition. We enter the grammatical category, number (singular or plural), gender (neutral in English, masculine/feminine in Portuguese), position in the frequency order, and the number of occurrences of the term. We also include the equivalent term in another language, if it has already been registered in TF, along with encyclopedic information.

Based on the selected excerpts, we create a final concept that effectively explains what the term is. The definition, in turn, must provide a single sentence that offers a polished and concise explanation developed from the final concept. The difference between the final concept and the definition can be seen in the figure 13, within the same example.

Figure 13 - Creating the final concept and the definition of the term WOOKIEE

Contexts

Final Concept / Definition

Final Concept / Definition

Final Concept:

Giant, hairy creature that is capable of physically overpowering and potentially harming others if provoked or upset. Their communication is primarily through guttural, modulated rumbles, and they have a reputation for being strong and difficult to defeat in combat.

Definition:

Giant, hairy and strong creature that communicates through guttural rumbles, known for their ability to physically overpower others and their reputation as skilled combatants

Source: FT databank.

With this entire process completed, the term is ready to be submitted to the administrator, who must then evaluate and approve (or not) the work. For this project, this process was carried out 36 times, as two terms were selected and registered twice (in Portuguese and English) for each of the nine categories. In Figure 14, you can see the interface of TF for the end user.

Figure 14 - Term WOOKIEE, as shown in TF



Terminology in Fiction

Project | How to cite | ILEEL | PPGEI | Help

Search term

Q Normal Extended Modular

Sciences and Fiction Sagas Star Wars Species

English

[Go back to search results](#)

Wookiee. *Species. n.m/f.s.* Giant, hairy and strong creature that communicates through guttural rumbles, known for their ability to physically overpower others and their reputation as skilled combatants. Ex.: "Let him have it. It's not wise to upset a Wookiee.". **Corpus:** *Frequency order position:* (1478); *Term number of occurrences:* (204). **Encyclopedic Information:** Wookiees were a species of tall, hairy humanoids that were native to the planet Kashyyyk. em: [Wookiee - Wookieepedia](#)

How to cite: FERREIRA, Guilherme Rodrigues. Wookiee. In: FROMM, Guilherme. **Terminologia em Ficção.** Versão 2. Uberlândia: PPGEI/ILEEL-UFU, 2024. 1 verbete. Disponível em: <http://ic.votec.ileel.ufu.br?trm=671>. Acessado em: 15 out. 2024.

Português

[Voltar ao resultado da busca](#)

Wookiee. *Espécies. s.m.s.* Criatura gigante, peluda e forte que se comunica por meio de estrondos guturais, conhecida por sua capacidade de dominar fisicamente os outros e sua reputação como combatentes habilidosos. Ex.: "Não é bom contrariar um Wookiee.". **Corpus:** *Posição na Ordem de Frequência:* (1608); *Nº de Ocorrências do termo:* (286). **Informações Enciclopédicas:** Os Wookiees eram uma espécie alta e peluda que eram nativos do planeta Kashyyyk Em: [Wookiee - Star Wars Wiki em Português](#)

Como citar: FERREIRA, Guilherme Rodrigues. Wookiee. In: FROMM, Guilherme. **Terminologia em Ficção.** Versão 2. Uberlândia: PPGEI/ILEEL-UFU, 2024. 1 verbete. Disponível em: <http://ic.votec.ileel.ufu.br?trm=672>. Acessado em: 15 out. 2024.

Termo elaborado por [FERREIRA, Guilherme Rodrigues](#)

16/10/2024 01:59 © 2024 Guilherme Fromm - Update: Samuel Victor Silveira de Lima - Development (VoTec v1.0 - 2007): ICMC Jr

Source: TF query page.

The final result can be found on the TF query page (<http://ic.votec.ileel.ufu.br>), where all approved terms are made available for consultation by the community. To the end user, as

Rev. Est. Ling., Belo Horizonte, v.34, n.1, p. 130-153, 2026

150

shown in Figure 14, it is their role to input the areas they would like to explore, search for a term, and analyze the results in the provided languages (Portuguese and English). It is important to notice that the arrangement of the terms shown in Figure 14 are monolingual in contrast.

6 Final Remarks

The entire project presented here is based on our shared assumption (within the Ethnoterminology theory) that many fictional universes can contain a large number of terms, created or re-created (from words already existing in any language) for the narrative purposes of authors in their works (as demonstrated by Oliveira, 2016, in relation to the Harry Potter books). Especially in long sagas (literary or audiovisual ones), the use or creation of terms (even if they do not exist in the real world) is important to differentiate that story from others already told and to maintain verisimilitude in the work as a whole. In an era of mass streaming and a flood of audiovisual products, documenting how terminological elaboration is created in large sagas, such as Star Wars, and the consequent development of a bilingual reference work, can help translators and screenwriters in their work and also serve as a more elaborate database for fans of these sagas.

The studies from Corpus Linguistics have provided a more facilitated analysis of linguistic patterns through the examination of corpora, which can result in the development of tools and applications or even sociolinguistic studies. The relationship between Corpus Linguistics and Translation is longstanding, as highlighted by Kruger, Wallmach, and Munday (2011), and the ability to compare corpora in different languages has become commonplace for researchers and translators. Building upon this principle, two corpora from the Star Wars canonical universe in Portuguese and English were compiled, with the aim of assisting translators and translation students with their works or studies. We also considered the cultural impact and relevance of the franchise, which not only involves the translation of its elements in various releases each year but also in all other media that in some way reference or will reference this material, such as Star Wars episodes in *Family Guy*, *The Simpsons*, and *South Park*.

As a prototype of a bilingual English/Portuguese Star Wars corpus-based vocabulary, the research objectives were fulfilled since the proposal is available on TF query page in a detailed manner, enabling translators who are dealing with the Star Wars universe in their work and research to consult it. However, this study can be further extended by continuing the corpus analysis and its entries in TF. To work with the entire Star Wars terminology, more time would be required to enter all the terms and update the corpora, as the franchise is active and new content is continuously being released. This project also excluded canonical media that are difficult to transpose into corpora, such as comics and video games, which would be relevant in a more in-depth study of this same material.

In conclusion, the Star Wars franchise offers a considerably rich universe for corpus exploration, allowing analyses in the numerous fields of knowledge as literature, technology, etc., which have been created since the late 1970s. In this paper, we followed a more technical path, focusing on the construction of a prototype vocabulary with the intention of assisting the professional and amateur activities of translators who can benefit from TF.

Authorship declaration

Both authors worked on the writing of this article. The first author developed the original studies on his TCC under the supervision of the second author and using his platform. The revision of the article was also performed by both authors.

References

- BARBOSA, M. A. Para uma etno-terminologia: recortes epistemológicos. *Ciência & Cultura*, v. 58, n. abr./maio/ju 2006, p. 48-51, 2006.
- CABRÉ, M. T. *La Terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, 1999.
- CARNEIRO, R. M. O. *Discurso literário de fantasia infantojuvenil: proposta de descrição terminológica direcionada por corpus*. 2016. 281 f. Dissertação (Mestrado em Estudos Linguísticos) - Programa de Pós-graduação em Estudos Linguísticos, Universidade Federal de Uberlândia, Uberlândia, 2016. DOI <http://doi.org/10.14393/ufu.di.2016.445>
- DIAS, C. A. Terminologia: conceitos e aplicações. *Ciência da Informação*, Brasília, v. 29, n. 1, p. 90-92, abr. 2000. IBICT. DOI <http://dx.doi.org/10.1590/S0100-19652000000100009>.
- FERREIRA, G. *Protótipo de um vocabulário bilíngue inglês/português de Star Wars baseado em Corpus*. 2022. 39 f. Trabalho de Conclusão de Curso (Graduação em Letras – Inglês e Literaturas de Língua Inglesa) – Universidade Federal de Uberlândia, Uberlândia, 2023. Disponível em: <https://repositorio.ufu.br/handle/123456789/38772>. Acesso em: 09 set. 2024.
- FINATTO, M. J. B. Orientações para a terminografia: das teorias às práticas em busca de amplitude da informação terminológica. In: ISQUERDO, A. N.; CORNO, G. M. dal. (org.). *As Ciências do Léxico: lexicologia, lexicografia e terminologia*, Campo Grande. MS: Ed.UFMS, vol. VII, 2014, p. 439-459.
- FROMM, G. *Terminologia em Ficção*. Versão 2. Uberlândia: PPGEL/ILEEL-UFU, 2024. 1 plataforma. Disponível em: <http://ic.votec.ileel.ufu.br/>. Acesso em: 09 set. 2024.
- FROMM, G.; LISBOA, J. VoTec terminographic environment over the years: brief overview. *Acta Scientiarum. Language and Culture*, v. 45, n. 2, p. e67669, 23 fev. 2024. DOI: <https://doi.org/10.4025/actascilangcult.v45i2.67669>
- FROMM, G. *VoTec: a construção de vocabulários técnicos eletrônicos para aprendizes de tradução*. 2007. 215f. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês). Departamento de Letras Modernas, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007. DOI: <https://doi.org/10.11606/T.8.2008.tde-08072008-150855>
- KRUGER, A.; WALLMACH, K.; MUNDAY, J. *Corpus-Based Translation Studies: research and applications*. Londres: Continuum, 2011.
- LARA, M, L, G, de. Linguagem documentária e terminologia. *Transinformação*, Campinas, v. 16, n. 3, p. 231-240, dez. 2004. DOI <http://dx.doi.org/10.1590/S0103-37862004000300003>.
- LARA, M. L. G. de. *Elementos da Terminologia*: (apostila para uso didático). São Paulo, 2005.

- LATORRE, V. R. D.. A dialética entre os extremos: da terminologia à etnoterminologia. *Caderno Seminal*, [S.L.], v. 19, n. 19, p. 70-94, 29 ago. 2013. DOI <http://dx.doi.org/10.12957/cadsem.2013.12062>.
- PEIXOTO, L. M. Identificação de unidades fraseológicas no vocabulário de Star Trek: abordagens corpus-driven e corpus-based. *Domínios de Linguagem*, Uberlândia, v. 8, n. 2, p. 139-163, 2014. DOI: 10.14393/DL16-v8n2a2014-8.
- RIBEIRO, G. C. B. Tradução técnica, terminologia e lingüística de corpus: ferramenta Wordsmith Tools. *Cadernos de Tradução*, Rio de Janeiro, v. 2, n. 14, p. 160-174, jan. 2004.
- SARDINHA, T. B. *Pesquisa em Lingüística de Corpus com Wordsmith Tools*. Campinas: Mercado de Letras, 2009.
- SCOTT, M. *WordSmith Tools version 9 (64 bit version)* Stroud: Lexical Analysis Software, 2024.
- Star Wars: Episódio IV – Uma Nova Esperança* (Comentários em áudio de George Lucas, Carrie Fisher, Ben Burt e Dennis Muren). Direção de George Lucas. Estados Unidos: Twentieth Century-Fox Film Corporation, 2011. Blu-ray (121 min)
- ZAMORA, R. E. M. Sobre tradução e terminologia das ciências sociais e humanas: quando a cultura encontra a “cultura”. *Mutatis Mutandis: Revista Latinoamericana de Traducción*, [S. L.], v. 8, n. 2, p. 548-566, 2015.

Tipos de entidades semânticas dos referentes que são tópicos discursivos em descrições, narrativas de experiência e relatos de opinião

Types of Semantic Entities of Referents that are Discourse Topics in Descriptions, Experience Narratives and Opinion Reports

Aline Gomes Garcia

Universidade Estadual Paulista (UNESP)

Araraquara | SP | BR

gomes.garcia@unesp.br

<https://orcid.org/0000-0002-6348-9179>

Resumo: Neste artigo, fundamentamo-nos na Gramática Textual-Interativa objetivando descrever o tipo de entidade semântica dos tópicos discursivos em descrições, narrativas de experiência e relatos de opinião, do Banco de Dados IBORUNA. Considerando que os textos de nosso corpus são representativos de gêneros textuais descritivos, narrativos e argumentativos, compõe ainda nosso propósito discutir a influência de traços de gêneros em que predomina a descrição, a narração ou a argumentação sobre os tipos semânticos dos tópicos na construção de textos. Partindo do método de análise tópica, que permite a identificação dos tópicos discursivos instaurados nos textos, o procedimento utilizado para precisar a categoria semântica de cada um dos tópicos identificados foi a distinção do tipo de entidade semântica denotada pelo termo que expressa o tópico discursivo e por aqueles que mais evidenciam a contração tópica. Assim, demonstramos que, em descrições, há predominância de tópicos que são da classe semântica lugar e, em narrativas de experiência e em relatos de opinião, a ocorrência maior é de tópicos que são da categoria dos estados-de-coisas, com quase exclusividade desse tipo semântico em narrativas de experiência e frequência também expressiva de tópicos que são proposições em relatos de opinião. Argumentamos, então, que os tipos de entidades semânticas dos tópicos em descrições, narrativas de experiência e relatos de opinião são influenciados por características de gêneros textuais descritivos, narrativos



e argumentativos e propomos que os gêneros textuais se diferenciam uns dos outros e podem ser caracterizados pelos tipos de entidades semânticas dos tópicos discursivos em cada um deles.

Palavras-chave: tópico discursivo; ontologia; descrição; narração; argumentação.

Abstract: This paper, based on Textual-Interactive Grammar, aims to describe the type of semantic entity in discourse topics in descriptions, narratives about experiences and opinion reports of the IBORUNA Database. Considering the texts of our *corpus* are representative of descriptive, narrative and argumentative textual genres, our purpose is also to discuss the influence of the genre features in which description, narration or argumentation predominates over the semantic types of topics in the construction of texts. Based on the topic analysis method, which allows the identification of the discourse topics established in the texts, the procedure used to specify the semantic category of each of the identified topics was the distinction of the type of semantic entity denoted by the term that expresses the discourse topic and by those that make the topic centering most evident. We demonstrate that, in descriptions, there is a predominance of topics from the semantic class of place and, in narratives that describe experiences and opinion reports, topics from the category of states-of-things occur the most, with almost exclusivity of this semantic type in narratives about experience and also an expressive frequency of topics that are propositions in opinion reports. We argue, then, that the types of semantic entities of the topics in descriptions, narratives about experiences and opinion reports are influenced by descriptive, narrative and argumentative textual genres characteristics. We further propose that textual genres differ from each other and can be characterized by the types of semantic entities of discourse topics in each of them.

Keywords: discourse topic; ontology; description; narration; argumentation.

1 Introdução

A Gramática Textual-Interativa, também conhecida como Perspectiva Textual-Interativa, é uma vertente da Linguística Textual (LT) que surgiu no interior do Projeto de Gramática do Português Falado (Castilho, 1990) e que tem em sua base teórica princípios de três diferentes áreas da Linguística, sendo uma delas a Pragmática.¹ De acordo com Jubran (2015a), um olhar pragmático respalda a primazia da teoria pela dimensão comunicativo-interacional, orientando a descrição de dados linguístico-textuais essencialmente a partir de seu funcionamento em situações concretas de uso da linguagem, dentro do quadro de suas condições de efetivação. Também o alinhamento da Gramática Textual-Interativa (GTI, daqui em diante) à perspectiva pragmática da LT demonstra a presença da ótica pragmática nesse quadro teórico-metodológico. Conforme Koch (2018 [2004]), com a virada pragmática, a LT passa a pesquisar o funcionamento da língua nos processos comunicativos de uma sociedade concreta, deixando os textos de serem vistos como estruturas acabadas e passando a ser considerados no processo de sua constituição, verbalização e tratamento pelos parceiros da comunicação.²

Tendo a GTI a preocupação com a língua em contextos de uso e, consequentemente, com a inserção dos textos na sua instância de produção, o que, como se pode notar, também está presente na fase pragmática da LT (Koch, 2018), o foco de investigação dessa vertente de base pragmática dos estudos do texto concentra-se, segundo Jubran (2015a), na construção textual integrada aos fatores enunciativos que lhe garantem existência e que se mostram nos próprios procedimentos de constituição do texto. Essa visão que congrega estruturas e processamento de estruturas textuais sustenta a descrição não apenas de regularidades estritamente estruturais dos textos, mas também princípios que governam a atividade verbal, conforme as restrições impostas em cada situação de comunicação verbal e as escolhas facultadas ao falante (Jubran, 2007), o que já sinaliza um pressuposto assumido pela GTI – o de que a construção textual está relacionada às características dos gêneros textuais a que os textos se filiam.

Um trabalho que representa um desenvolvimento recente de estudos textuais de base fundamentalmente pragmática, que descreve fenômenos textuais a partir de seu contexto de uso, buscando compreender como a construção dos textos associa-se a particularidades dos gêneros textuais em que se realiza, é o livro sobre diacronia de processos de construção textual (Penhavel; Cintra, 2022), todos com base na GTI. Como explicam estes autores, o percurso diacrônico de cada processo de constituição do texto descrito na obra situa-se no contexto da evolução histórica do gênero textual em que o processo é estudado. Por exemplo, no capítulo sobre referência em cartas de leitor mineiras, Cintra (2022) analisa cartas de três sincronias – primeira e segunda metades do século XIX e segunda metade do século XX –, discutindo especialmente a referência pronominal na função de sujeito sintático da oração, com pronome expresso ou elíptico, na designação de primeira e segunda pessoas do discurso. Observando traços gerais das cartas de leitor no período diacrônico investigado, o

¹ Os princípios que constituem a base teórica fundadora da teoria também vêm da Linguística Textual e da Análise da Conversação.

² Mesmo mais alinhada a um viés pragmático, a GTI não nega a orientação cognitivista da LT. Jubran (2007), por exemplo, deixa claro que a entrada cognitivista no trajeto da LT permitiu uma visão mais clara do conhecimento sociointeracional, que interessa aos estudos textual-interativos. Por isso, a autora caracteriza a GTI como uma teoria de base pragmática, mas que também tem compatibilidade com a orientação cognitivista da LT.

autor aponta que houve mudança na dinâmica interacional das cartas, tendo ocorrido particularmente mitigação do diálogo simulado entre primeira e segunda pessoas discursivas e consequente inibição da presença dos parceiros presumidos, o que levou à diminuição do caráter interpelativo e emotivo da interação. Assim, Cintra (2022) defende que essa mudança nas atividades interacionais no gênero influenciaram o uso de certas estratégias de referência, haja vista que um dos achados do autor foi a diminuição da utilização do pronome expresso na designação pronominal da primeira e, sobretudo, da segunda pessoa discursiva, indicando, portanto, uma atenuação diacrônica no grau de envolvimento dos participantes na interação, o que foi estampado na materialidade textual pelo processo de referência.

Nesse contexto de desenvolvimento recente de estudos textuais muito alinhados a um viés pragmático, que buscam interpretar os fenômenos textuais assumindo sua relação intrínseca com o contexto de uso e, dessa forma, com as características dos gêneros textuais a que os textos se filiam, neste artigo,³ nos embasamos na GTI com o objetivo de descrever o tipo de entidade semântica dos referentes que adquirem estatuto de tópico discursivo em um conjunto de descrições, narrativas de experiência e relatos de opinião, coletado da Amostra Censo do Banco de Dados IBORUNA (Gonçalves, 2007), o qual reúne amostras linguísticas do português brasileiro contemporâneo falado na região noroeste do estado de São Paulo. Sendo as descrições, as narrativas de experiência e os relatos de opinião de nosso material representativos do que seriam gêneros textuais descritivos, narrativos e argumentativos, compõe ainda nosso objetivo discutir a influência de traços de gêneros textuais em que predomina a descrição, a narração ou a argumentação sobre os tipos de entidades semânticas dos tópicos na construção dos textos aqui em estudo. Essa discussão deixa evidente nossa preocupação em compreender a constituição textual no contexto de suas condições de efetivação, conforme respalda a base pragmática da GTI, e, especialmente, na relação com traços dos gêneros textuais em que se atualiza, como sustenta o referido pressuposto assumido pela GTI.

O artigo está organizado da seguinte forma: depois desta seção de introdução, na seção 2, apresentamos os fundamentos da GTI e a noção de tópico discursivo definida na teoria; depois, na seção 3, sintetizamos os tipos de entidades semânticas dos referentes; na sequência, na seção 4, fazemos nossa discussão de dados, detalhando os tipos de entidades semânticas dos referentes que são tópicos em nosso material; dando continuidade ao artigo, na seção 5, refletimos sobre como características de gêneros textuais descritivos, narrativos e argumentativos motivam as classes semânticas dos textos por nós analisados; na seção 6, apresentamos nossas conclusões, que vêm seguidas de uma seção de agradecimento e de outra dedicada às referências bibliográficas.

2 A Gramática Textual-Interativa e a noção de tópico discursivo

Como uma vertente da LT, a GTI constitui um quadro teórico-metodológico que toma o texto como objeto de estudo, focalizando especificamente a construção textual. Desse modo, os trabalhos desenvolvidos na GTI enfocam os processos de construção de textos, a saber, organização tópica, referência, parentetização, repetição, correção, parafraseamento e

³ O presente artigo origina-se da tese de doutorado intitulada *Uma abordagem textual-interativa da relação entre organização tópica e referência em descrições, narrativas de experiência e relatos de opinião* (Garcia, 2022).

tematização/rematização, bem como os marcadores discursivos, um conjunto de expressões linguísticas de constituição diversificada que colaboram para a efetivação textual-interativa desses processos (Jubran, 2015a; Penhavel; Cintra, 2022).

Segundo Jubran (2007), a GTI assume a concepção de linguagem como interação social, como forma de ação verbal, por meio da qual os interlocutores efetuam tarefas comunicativas de trocas de representações, interesses, metas, sempre no contexto de um espaço discursivo orientado aos parceiros da comunicação. Considerando essa concepção, a linguagem não pode ser estudada sem se levar em conta suas condições de efetivação, como enfatiza Jubran (2007). Alinhada a essa forma de conceber a linguagem está a noção de texto da GTI, entendido como um processo dinâmico, sujeito a pressões de ordem interacional, que se mostram na sua própria materialidade linguística (Jubran, 2007).

Calcada nesses conceitos, a GTI lança seus princípios basilares. Um deles, sinalizado na seção de introdução, é o de que os fenômenos considerados na teoria têm suas propriedades e funções definidas no uso, nas situações concretas de interlocução, coenvolvendo as circunstâncias de enunciação. Outro princípio fundamental da GTI é o de que os fatores interacionais são constitutivos do texto e inerentes à expressão linguística, o que faz com que as condições enunciativas que sustentam a ação verbal estejam inscritas na própria materialidade textual. Alicerçado na assumida imbricação dos procedimentos de formulação textual e de interação, já representada no próprio nome do quadro teórico-metodológico, está um terceiro pressuposto crucial à GTI, o princípio de gradiência. Conforme esse princípio, a análise dos procedimentos de elaboração textual não dicotomiza as funções textual e interativa, mas as conjugam, segundo a dominância de uma ou de outra função (Jubran, 2007, 2015a).

Com base nesse perfil teórico, a GTI define uma categoria de análise de estatuto textual – o tópico discursivo. Analisando a conversação, que envolve a troca de turnos entre, pelo menos, dois interlocutores, Jubran (2015b) explica que a constituição desses textos implica uma construção colaborativa, na qual um turno não é apenas um sucessor temporal de outro, mas é produzido, em alguma medida, por referência ao anterior.⁴ Dessa forma, segundo a autora, a relação de interdependência entre turnos conversacionais pode ser motivada pela preocupação dos falantes de se entrosarem, buscando manter o seu texto em torno de um conjunto de objetos de discurso⁵ compartilhados, foco da interação verbal. Assim, “há como que uma consciência de que se deve *falar sobre algo* e de que o ponto para o qual converge a conversa deve ficar claro para os interlocutores” (Jubran, 2015b, p. 86, destaque nosso). Constatando, então, que a construção textual é orientada pela centração dos falantes em determinados temas, que se estabelecem como foco da interação verbal (Jubran, 2015a), o tópico discursivo é assumido na GTI como um elemento decisivo na constituição dos textos e a topicalidade, ou seja, a estruturação do texto em tópicos discursivos, o fio condutor da organização textual-interativa.

⁴ Como detalha Jubran (2006a), a categoria analítica do tópico discursivo foi definida a partir da análise de dados da modalidade falada da língua, especificamente do estudo da conversação. Mais adiante, o tópico discursivo foi sendo utilizado para a análise de textos de outros gêneros falados e também escritos (Guerra, 2022; Penhavel, 2010, 2020; 2022; Pinheiro, 2005; Souza, 2020; Zanin, 2023), o que demonstra a aplicabilidade da categoria para a *análise textual*, não sendo restrita somente a textos falados, uma vez que o tópico discursivo é um elemento fundamental na construção textual, conforme se poderá notar nesta seção 2.

⁵ Os objetos de discurso não são entidades que têm relação especular com o mundo real, mas entidades interativa e discursivamente produzidas pelos interlocutores ao longo da enunciação, sendo no e pelo discurso que são colocados, delimitados, desenvolvidos e transformados (Mondada, 2001).

Tratando propriamente da noção de tópico discursivo, esta não pode ser confundida com a de tópico sentencial ou da predicação.⁶ Como frisa Jubran (2015b), os segmentos textuais nos quais os tópicos discursivos são instaurados assumem uma extensão que vai além dos limites sentenciais, havendo, via de regra, sequências de sentenças que se mantêm no mesmo tópico discursivo. Sendo, pois, uma categoria de análise pensada conforme propriedades próprias da organização textual-interativa da linguagem, o tópico discursivo pode ser entendido, no sentido geral, como “acerca de’ que se fala” (Jubran, 2006a, p. 35). De forma mais específica, Jubran (2015b) argumenta que o tópico, que é uma categoria analítica abstrata, manifesta-se no texto mediante enunciados formulados pelos interlocutores a respeito de um conjunto de referentes, concernentes entre si e em relevância em determinado ponto do texto. Nesse contexto, a autora define duas propriedades particularizadoras do tópico discursivo – centração e organicidade. A propriedade de centração abrange três traços, que são concernência, relevância e pontualização.

A concernência diz respeito à relação de interdependência entre elementos textuais, firmada por mecanismos coesivos de sequenciação e referenciação, que promovem a integração desses elementos em um conjunto referencial específico de objetos de discurso, explícitos ou inferíveis, instaurado no texto como alvo da interação. O traço da relevância, por sua vez, compreende a proeminência de elementos textuais na constituição desse conjunto de referentes, que são projetados como focais, tendo em vista o processo interativo. Por último, o traço da pontualização corresponde à localização desse conjunto referencial em determinado ponto do texto, fundamentada na concernência e na relevância de seus elementos.

O exemplo em (1), a seguir, ajuda a compreender a noção de tópico discursivo apresentada, bem como os três traços da propriedade tópica de centração.

(1) Tópico discursivo: Atividades profissionais do marido de L1

Doc. – o seu marido sempre exerceu essa profissão que ele tem agora?

L1 – não ele teve escritório no início da carreira ... teve escritório durante ... oito anos:: mais ou menos ... depois ... ainda com escritório ... e como ele tinha liberdade de advogar ele também ... exercia a:: profi/ o a advocacia do Estado né?: ... e depois ... é que ele começou a lecionar quando houve ... a necessidade de regime de dedicação exclusiva [...] ... então:: ele::: começou a lecionar foi convidado e:: [...] e::: e deu-se muito bem no magistério ... ele se realiza sabe? [...] ele se dedica muitíssimo a ... tanto à carreira de procurador como de professor (tá?) [...] [D2 SP 360: 1.160-91]

(Jubran, 2015b, p. 88)

Segundo Jubran (2015b), em (1), a concernência é constatada pela observação da construção de um conjunto de referentes relativo às profissões do marido de L1. A interdependência entre os elementos constitutivos desse conjunto referencial é instaurada, por exemplo, pela coesão lexical estabelecida entre lexemas como *profissão*, *escritório*, *carreira*, *advogar*, *lecionar*, *magistério*, *procurador* e *professor*, que circunscrevem o tema “profissão”. Já a relevância

⁶ De acordo com Camacho (2002), na predicação, isto é, em uma estrutura que envolve um predicado e um conjunto de termos apropriados inseridos nesse predicado (*João escreve a carta*, por exemplo, cf. Dik 1997), um constituinte com função tópica apresenta a entidade sobre a qual se faz uma predicação, como “ele”, em “No Brasil, ele evoluiu muito [o teatro]” (Camacho, 2002, p. 272. Adaptado).

tópica decorre da posição focal assumida por dados referentes na construção do tópico. Assim, quando observamos os temas sentenciais, notamos que estes, muitas vezes, são expressos pelo pronome “ele” e (re)categorizam o elemento “marido”, colocando-o em proeminência. A esses temas, conectam-se remas que expressam a dominância do tópico “profissões”, na medida em que especificam o exercício das profissões, como em “ele começou a lecionar”, e manifestam avaliações sobre a realização das atividades profissionais, como em “ele se realiza sabe?”. A partir da identificação dos traços de concernência e relevância tópica, pode-se pontualizar, no texto, o segmento textual que realiza o tópico discursivo identificado, o que já diz respeito ao terceiro traço da centração, a pontualização.

Como dissemos, o tópico discursivo é uma categoria analítica abstrata e, com base nos traços da centração, identificamos, na materialidade textual, o segmento textual que realiza um tópico discursivo. Esse segmento de texto que materializa um tópico é chamado, na GTI, de “Segmento Tópico” (SegT) e os segmentos textuais nos quais são instaurados os tópicos discursivos mais específicos do texto, no que diz respeito a uma hierarquização temática, são conhecidos como “SegTs mínimos”. Neste artigo, os exemplos apresentados em nossa discussão de dados sempre corresponderão a SegTs mínimos.

Tendo sido sinalizada a possibilidade de haver tópicos mais e menos específicos em um texto, passamos a tratar da segunda propriedade particularizadora do tópico discursivo, a organicidade, a qual se manifesta, como define Jubran (2015b), por relações de interdependência tópica que se estabelecem simultaneamente em dois planos, o hierárquico e o linear. No primeiro, ocorrem as dependências de superordenação e subordinação entre tópicos que se implicam pelo grau de abrangência do assunto, podendo haver em um texto, portanto, níveis de hierarquização na estruturação tópica, indo de um tópico mais amplo até tópicos mais específicos. Por outro lado, no segundo plano, as articulações intertópicas se dão em termos de adjacência ou interposições de tópicos diferentes na linearidade textual, sendo as relações de linearidade tópica basicamente caracterizadas por continuidade e descontinuidade. Neste trabalho, apesar de reconhecermos que a organicidade nos auxilia na identificação dos diversos tópicos de um texto, não daremos um tratamento mais detalhado a essa propriedade tópica porque o foco deste nosso estudo não é discutir as relações de interdependência tópica.

Ainda tratando da noção de tópico discursivo, como nosso propósito neste trabalho é discutir os tipos de entidades semânticas dos referentes que são tópicos em descrições, narrativas de experiência e relatos de opinião, importa argumentar que todo tópico discursivo pode ser admitido como um referente, conforme assumimos em Garcia (2022).

No que toca à concepção de referente, Lyons (1977b) ressalta que é comum dizer que palavras ou expressões, mas não conceitos, referem-se a coisas. Nesse sentido, focalizando a referência, o autor afirma que, quando utilizamos uma sentença descritiva simples, é frequentemente apropriado dizer que o que estamos fazendo envolve dizer algo sobre alguém ou alguma coisa, sobre uma entidade particular (ou entidades), ou grupo(s) de entidade(s) do mundo. Para exemplificar, Lyons (1977b) destaca que, em uma situação em que uma expressão como “Napoleão é um corso” é empregada para fazer uma declaração, diz-se que o falante se referiu a um certo indivíduo (Napoleão) pelo uso de expressão referencial. Sendo a referência bem sucedida, a expressão referencial permitirá que o ouvinte identifique corretamente o indivíduo em questão, que é o referente.⁷ Ainda para o autor, em sentenças com apenas uma

⁷ Na Semântica, área em que o trabalho de Lyons (1977b) se insere, o referente é tido como uma entidade do mundo extralinguístico, um objeto de mundo, concepção que se difere daquela da LT (também adotada neste trabalho), em que o referente é uma entidade textual, como é possível compreender com base na nota 5.

expressão referencial, a expressão usada para referir aquilo sobre o qual estamos falando (o referente) é tipicamente o sujeito da sentença (Lyons, 1977b).⁸ Também reconhecendo que o referente está relacionado àquilo acerca do que se fala, Brown e Yule (1983), partindo da leitura de Lyons (1977b), dizem o seguinte: “O termo *referência* pode [...] ser [...] reservado àquela função pela qual os falantes (escritores) indicam, pelo uso de uma expressão linguística, as entidades sobre as quais estão falando (escrevendo)” (Brown; Yule, 1983, p. 205, tradução nossa).⁹

Baseados nessa concepção de referente, defendemos que todo tópico discursivo pode ser reconhecido como um referente porque o tópico, na condição de tema do texto – “tema” concebido como aquilo sobre o que se fala em um discurso (Brown; Yule, 1983) –, também pode ser visto como algo do mundo, (re)construído no texto,¹⁰ acerca do que se fala no processo textual-interativo. Essa visão referencial do tópico discursivo, pressuposta em trabalhos basilares para a definição da categoria (Jubran *et al.*, 2002; Jubran, 2006a, 2006b, 2015b), pode ser demonstrada a partir do exemplo em (2).

(2) Tópico discursivo: Bronquite do filho do meio da informante

aí nasceu [...] o E. [...] mas aí ele já... ele já tinha problema de bronquite aos dois meses	1
ele começô(u) com problema de bronqui::te aí a gente... eu fiz muita nove::na... levei	2
muito no mé::dico... ele fez muito tratamen::to... até:: simpatia que eu nunca fui	3
MUI::to de crediTÁ(r)... né? mas quando você... éh:: tá naquele momento assim de	4
desesPEro... éh::... se você num sabe mais o que faz/ que cê faz você foi no médico você	5
já fez tratamento... cê tá com/ dan(d)o reMÉdio aí num melhorava num melhorava aí	6
cê (a)caba até... cain(d)o ne/ fazen(d)o simpatia né?... eu acabei fazen(d)o simpatia...	7
pra vê(r) se o E. melhorava da bronquite... e com o passar dos anos realmente graças	8
a Deus ele ficô(u) bom... ¹¹	9

[BDI-AC-102; NE: L. 43-53; 6/9]¹²

⁸ Ao versar sobre a “referência” como tratada em Lyons (1977b), importa explicar a distinção feita pelo autor entre “referência” e “denotação”. De acordo com Lyons (1977b), quando se fala em “referência”, deve-se compreender a relação entre uma expressão e o seu contexto de uso específico. Já quando se fala em “denotação” de um lexema ou expressão, deve-se entender a relação entre esse lexema e as pessoas, os lugares, as propriedades e os processos que não compõem o sistema linguístico. Assim, segundo o autor, a denotação é válida independentemente das ocasiões de enunciação e pode ser compreendida como a classe de entidades a que uma expressão poderia se referir, como a expressão “vaca”, que denota ontologicamente uma classe particular de animais. Portanto, é no uso da língua que uma expressão linguística deixa apenas de ter valor denotativo para se tornar também uma expressão referencial.

⁹ The term reference can [...] be [...] reserved for that function whereby speakers (writers) indicate, via the use of a linguistic expression, the entities they are talking (writing) about (Brown; Yule, 1938, p. 205).

¹⁰ Nesse sentido, todo tópico discursivo, desde o mais abrangente até os mais específicos de um texto, pode ser visto como um objeto de discurso, ou seja, um referente, nos termos da LT (Garcia, 2022).

¹¹ Como um método para facilitar o reconhecimento de um tópico discursivo, nos exemplos que analisamos, temos destacado em negrito trechos do SegT que mais diretamente indicam seu respectivo tópico.

¹² Ao final de cada exemplo discutido, apresentaremos, entre colchetes, a indicação da fonte da qual eles foram extraídos, como recomendado pelo Banco de Dados IBORUNA, além da ordem em que o tópico discursivo aparece na linearidade do texto que compõe e da quantidade de tópicos hierarquicamente mais específicos identificada no texto. Por exemplo: [BDI-AC-102; NE: L. 43-53; 6/9] indica que o dado foi extraído do Banco de Dados IBORUNA (BDI); da entrevista 102 da Amostra Censo (AC-102); de uma narrativa de experiência (NE, que pode variar com DE, para descrição, e RO, para relato de opinião); das linhas 43 a 53 da transcrição da entrevista; e que o tópico materializado no SegT exemplificado é o sexto, na ordem linear da construção tópica, de um total de 9 tópicos discursivos daquela narrativa de experiência.

O SegT em (2) faz parte de uma narrativa de experiência em que a informante narra a gravidez e o início da vida de seus três filhos, centrando-se, em certo ponto do texto, na bronquite do filho do meio logo depois de seu nascimento, como vemos no exemplo em apreço. Observamos a construção de um conjunto referencial relativo a tal problema de saúde do filho pela integração semântica dada principalmente entre as passagens destacadas em negrito, como “ele já tinha problema de bronquite” (L. 1), “aos dois meses ele começô(u) com problema de bronqui::te” (L. 1-2) e “eu acabei fazen(d)o simpatia... pra vê(r) se o E. melhorava da bronquite” (L. 7-8), que promovem mais diretamente um recorte temático acerca do tema “bronquite do filho”, o que já demonstra a concernência, primeiro traço da centração tópica. Também verificamos a construção desse conjunto referencial pela proeminência assumida pelos referentes “filho” e “bronquite”. Já no início do desenvolvimento do tópico, na linha 1, há a menção ao nome do filho do meio – o “E.”. Na continuação do tópico, esse objeto de discurso vai sendo mantido em foco, principalmente pelo pronome “ele” (L. 2, 3 e 9) e pela nova menção ao nome da criança (L. 8). Com relação ao referente “bronquite”, notamos a sua ativação explícita na linha 1, sua reiteração nas linhas 2 e 8 e ainda a relação semântica desse objeto de discurso com outros, como “tratamento”, “remédio” e “médico”, que também podem ser considerados do campo semântico “problema de saúde”, como o objeto “bronquite”. Assim, a saliência dos referentes “filho” e “bronquite” sugere a construção do tópico *Bronquite do filho*, o que já se relaciona com a relevância tópica, segundo traço da centração. Como localizamos, na materialidade textual, um conjunto referencial concernente entre si e relevante em determinado ponto do texto sobre a bronquite do filho do meio da informante, pontualizamos o trecho da narrativa de experiência em que esse conjunto referencial se concretiza, o que mostra o terceiro traço da centração.

Fundamentando-nos nessa análise tópica, assumimos que todo tópico discursivo pode ser visto como um referente porque é sobre a entidade de mundo “Bronquite do filho do meio da informante”, reconstruída no processamento textual-iterativo, que se fala no SegT. Em outros termos, é acerca de uma entidade (isto é, um referente) – “Bronquite do filho do meio da informante” – que se instaura centração tópica (ou seja, que se instala um tópico). É por isso, portanto, que entendemos que todo tópico discursivo é sempre um referente. Sendo um referente, o tópico discursivo pode ser classificado pelos tipos de entidades semânticas dos referentes, segundo tipologia apresentada na seção 3.

3 Tipos de entidades semânticas dos referentes

A tipologia de entidades semânticas denotadas ontologicamente pelas expressões linguísticas é, inicialmente, proposta por Lyons (1977b) e, posteriormente, expandida por outros autores, como Hengeveld (1988, 1992) e Hengeveld e Mackenzie (2008). Assim, Lyons (1977b) define três tipos de entidades semânticas: indivíduo, estado-de-coisas e proposição, sendo que cada um desses tipos de entidades é categorizado segundo uma ordem particular, sendo indivíduo uma entidade de primeira ordem, estado-de-coisas, de segunda ordem, e proposição, de terceira ordem, classificação que vai da ordem mais concreta (primeira) para a mais abstrata (terceira).

Conforme a proposta de Lyons (1977a, 1977b), podemos chamar de entidades de primeira ordem os objetos físicos, ou seja, as pessoas, os animais e as coisas que, sob condições

normais, são relativamente constantes quanto às suas propriedades perceptuais, que são localizáveis em algum ponto no tempo e no espaço, podendo ocupar um espaço tridimensional, e que podem também ser observáveis publicamente. Dessa forma, segundo classificação do autor, indivíduos podem ser reconhecidos como distintos, concretos e identificáveis, o que exclui dessa classe as abstrações e também os estados psicológicos. Ademais, os elementos que compõem essa classe são avaliados em termos de sua existência,¹³ além de terem cor, tamanho e peso, podendo, por isso, ser alvo de atribuições de propriedades. Para Lyons (1977b), “homem”, “gato” e “mesa” são lexemas que denotam entidades de primeira ordem.

Já as entidades de segunda ordem – estado-de-coisas – dizem respeito, de acordo com Lyons (1977b), a eventos, ações, processos, experiências, estados e posições que são localizados em algum ponto ou intervalo do tempo, são observáveis e ocorrem em algum lugar, mas não existem. Assim, estados-de-coisas são avaliados em termos de seu estatuto de realidade.¹⁴ Segundo Lyons (1977b), “chegada”, “morte” e “manutenção” são exemplos de nomes que denotam entidades dessa categoria.

No tocante às entidades de terceira ordem, proposições dizem respeito, conforme Lyons (1977b), a construtos mentais (crenças, expectativas e julgamentos) e não podem ser localizadas nem no espaço nem no tempo, sendo passíveis de serem avaliadas a partir do seu valor de verdade (podem ser verdadeiras ou falsas),¹⁵ mas não pela sua existência ou realidade. Ainda, proposições podem ser asseveradas ou negadas, lembradas ou esquecidas, podem ser razão, mas não causa material. Também podem ser qualificadas em termos de atitudes proposicionais (certeza, dúvida, descrença). Sobre essa categoria de entidades, Hengeveld e Mackenzie (2008), baseados em Lyons (1977b), explicam que proposições podem ser factuais, quando são peças do conhecimento ou crença a respeito do mundo real, ou não-factuais, quando são esperanças ou desejos a respeito de um mundo imaginário. Exemplos de nomes que denotam essa ordem de entidades são: “razão”, “ideia” e “crença”.

Ampliando a tipologia de Lyons (1977b), Hengeveld (1988, 1992) propõe uma quarta ordem de entidades, a qual recobre os atos de fala. Conforme o autor, entidades de quarta ordem caracterizam-se por poderem ser localizadas no espaço e no tempo em que os interlocutores estão envolvidos, por serem pronunciadas, compreendidas e avaliadas em termos de suas condições de felicidade.¹⁶ Tais entidades dizem respeito às diferentes ilocuções que transmitem as intenções comunicativas do falante, caracterizando os atos de fala, como

¹³ Um indivíduo existe se é localizado em algum lugar do espaço.

¹⁴ Um estado-de-coisas é real se ocorre em algum lugar e em algum momento do tempo.

¹⁵ Conforme Oliveira (2001) o valor de verdade de uma sentença diz respeito às condições, no mundo, que lhe permitem ser verdadeira ou falsa. Assim, uma sentença como “está chovendo” é verdadeira se estiver chovendo no momento em que o falante a proferir. Desse modo, determinar se uma sentença é verdadeira ou falsa requer conhecimento da situação particular de enunciação, observando se as suas condições de uso são satisfeitas.

¹⁶ Austin (1962) denomina de condições de felicidade de um proferimento performativo as condições que o contexto deve satisfazer para que o uso de uma determinada expressão possa ser adequado. Assim, uma sentença declarativa como “eu declaro que a seção está aberta” preenche suas condições de felicidade se é dita, por exemplo, por um presidente de uma assembleia, em uma sala de reuniões, mas essas condições não se efetivam se um vigia entra no mesmo local e enuncia a sentença. Portanto, as condições de felicidade podem ser entendidas como as condições contextuais necessárias para que um ato de fala tenha sucesso.

as declarações, as perguntas, as exclamações.¹⁷ Camacho, Dall’Aglio-Hattner e Gonçalves (2014) destacam que alguns substantivos como *pergunta*, *afirmação*, *exclamação*, *ordem* e *pedido* podem denotar esse tipo de entidade semântica, mas não representam, por eles mesmos, o ato ilocutivo que designam.

Também Hengeveld e Mackenzie (2008) expandem a tipologia de entidades semânticas proposta por Lyons (1977b), distinguindo, dentre outros tipos de entidades, as categorias “propriedade”, “lugar” e “tempo”. Segundo os autores, propriedades não podem ser caracterizadas a partir dos parâmetros de espaço e de tempo, não têm existência independente e somente podem ser avaliadas em termos de sua aplicabilidade ou a outros tipos de entidades ou à situação que elas descrevem em geral. Assim, a propriedade “verde” aplica-se a entidades de primeira ordem (indivíduos), a propriedade “recente” a entidades de segunda ordem (estados-de-coisas) e a propriedade “inegável” a entidades de terceira ordem (proposições). Ainda conforme classificação de Hengeveld e Mackenzie (2008), outros nomes que designam essa categoria semântica são “cor”, “tamanho” e “peso”.

Sobre os tipos semânticos “lugar” e “tempo”, Hengeveld e Mackenzie (2008) explicam que os conceitos de espaço e de tempo especificam dimensões das categorias semânticas, constituindo, portanto, classes independentes. Abordando particularmente a distinção da classe lugar em relação à categoria indivíduo, esses autores argumentam que devemos assumir que a conceptualização de indivíduos, como “cobertor”, “pedra” ou “Martin Luther King”, difere-se, em alguma medida, da conceptualização de lugares, tais como “ambiente”, “norte” ou “Geórgia”. Contudo, um mesmo fenômeno da realidade externa pode ser construído mentalmente tanto como um indivíduo quanto como um lugar, dependendo dos objetivos de quem conceptualiza. Consideremos o exemplo do termo “casa”. Do ponto de vista de um comprador, uma casa pode ser entendida como lugar, um espaço para viver. Em contrapartida, do ponto de vista de um corretor de imóveis, uma casa pode ser conceptualizada como indivíduo, um produto a ser vendido. Conforme Hengeveld e Mackenzie (2008), os lexemas “aeroporto”, “cais” e “lugar” são expressões que tipicamente designam a classe lugar.

Quanto à expressão da categoria tempo, Hengeveld e Mackenzie (2008) explicam que algumas expressões que designam essa classe estão ligadas à interpretação contextual do momento de fala, como “hoje”, “próximo ano”, outras marcam um ponto relativo na linha do tempo (“antes de sábado”, “duração”), enquanto outras estão relacionadas ao calendário socialmente estabelecido (“segunda-feira”, “dia de Natal”). Também se pode dizer que algumas expressões temporais identificam um ponto na linha do tempo (“momento”, “12h”), outras, um intervalo temporal (“período”, “abril”).

Seguindo essa tipologia de entidades semânticas dos referentes, que contempla as classes *indivíduo*, *estado-de-coisas*, *proposição*, *ato de fala*, *propriedade*, *lugar* e *tempo*, na seção 4, então, discutiremos os tipos de entidades semânticas dos referentes que são tópicos discursivos em descrições, narrativas de experiência e relatos de opinião.

¹⁷ De acordo com a Teoria dos Atos de Fala (Austin, 1962; Searle, 2002), a linguagem é uma forma de agir. Desse modo, os atos de fala correspondem aos enunciados em situações reais de uso da linguagem, que são formas de ação verbal. Uma das ações realizadas por meio da linguagem é chamada, conforme a Teoria, de ato ilocucional, que diz respeito às ações que realizamos quando falamos (ordenar, perguntar, declarar, afirmar, informar, reclamar, permitir, prometer, parabenizar, agradecer e etc.). São essas ações ilocucionais que fazem referência, então, às entidades semânticas de quarta ordem.

4 Discussão de dados

Como definimos na seção 1, o objetivo deste artigo é descrever o tipo de entidade semântica dos referentes que são tópicos discursivos em um conjunto de descrições, narrativas de experiência e relatos de opinião, discutindo a influência de traços de gêneros textuais descritivos, narrativos e argumentativos sobre a classe semântica dos tópicos na construção desses textos, tendo em vista que as amostras linguísticas em apreço são representativas de gêneros textuais em que predomina a descrição, a narração ou argumentação, nessa ordem.

Conforme também já antecipamos na seção 1, as descrições, as narrativas de experiência e os relatos de opinião que compõem nosso *corpus* provêm da Amostra Censo do Banco de Dados IBORUNA (Gonçalves, 2007),¹⁸ do qual coletamos um total de 30 textos, sendo 10 descrições, 10 narrativas de experiência e 10 relatos de opinião, combinando os diferentes perfis sociais estratificados no IBORUNA. Esse quantitativo de textos foi definido porque suspeitávamos, com base em outros trabalhos que realizaram análise textual a partir da categoria do tópico discursivo (Oliveira, 2016; Penhavel, 2010; Souza, 2020), que o montante de 30 textos já nos daria dados suficientes para a descrição de regularidades quanto à classe semântica dos diferentes referentes que são tópicos discursivos. Ademais, escolhemos descrições, narrativas de experiência e relatos de opinião porque, sendo esse material representativo de gêneros textuais descritivos, narrativos e argumentativos, acreditávamos que essa tipologia variada de textos, conforme tipologia textual descrita por Koch e Fávero (1987), poderia nos oferecer também uma diversidade de tipos de entidades semânticas dos tópicos discursivos.¹⁹ Essas duas expectativas que orientaram a escolha do material foram confirmadas com o desenvolvimento das análises.

No IBORUNA, descrições, narrativas de experiência e relatos de opinião são textos de natureza predominantemente monológica. Contudo, ainda que o informante seja o principal locutor dessas amostras linguísticas, os momentos de interação explícita entre informante e documentador evidenciam que este último assume uma função textual-interativa fundamental na construção desses textos. Isso porque, para a coleta de cada tipo de texto oral da Amostra Censo do IBORUNA, foi definida uma proposta de “pergunta base”, a ser feita pelo documentador no momento da gravação e que consideraria a natureza do texto que o informante deveria produzir na situação em particular. Assim, para a coleta de descrições, a pergunta base foi “descreva, diga como é o local onde você mais gosta de ficar, passear ou brincar [...]” (a casa, a escola o local de trabalho, etc.); para a coleta de narrativas de experiência, a pergunta consistiu basicamente em “conte uma história que tenha acontecido com você [...]” (uma briga, uma viagem, o nascimento de um filho, etc.); e, finalmente, para o levantamento de relatos de opinião, “o que você pensa/acha sobre/de...” (educação, política, vantagens e problemas da própria profissão, etc.) compreendia a pergunta base (Gasparini-Bastos; Gonçalves, 2005). Desse modo, dizemos que o documentador assume uma função textual-interativa fundamental na construção desses textos porque, dentre outras questões, ele tende

¹⁸ O seguinte endereço eletrônico dá acesso ao IBORUNA: <https://alip.ibilce.unesp.br/bancos-de-dados/banco-de-dados-iboruna>.

¹⁹ Esclarecemos que o tipo textual descritivo (em termos de sequência tipológica) não é a mesma coisa que os textos do IBORUNA rotulados, pelo próprio Banco de Dados, de “descrição”, mas as descrições do IBORUNA são textos em que há predominância do tipo descritivo, conforme noção de tipo textual apresentada na seção 5.

a direcionar a categoria semântica dos tópicos discursivos, como chamaremos a atenção na discussão do caso em (4).²⁰

Passando a explicar a metodologia de análise de dados, esta foi realizada em duas etapas: (i) análise tópica de descrições, narrativas de experiência e relatos de opinião e (ii) análise do tipo de entidade semântica de cada um dos referentes que se constitui como tópico discursivo nesses textos. A primeira etapa envolveu a aplicação do método de análise tópica (Jubran, 2015b), o qual, com base nas propriedades tópicas de centralização e organicidade, explicadas na seção 2, permite reconhecer os SegTs e tópicos discursivos de um texto. Esse método, já ilustrado na apresentação dos exemplos em (1) e (2), também na seção 2, levou-nos à identificação de 185 SegTs mínimos e seus respectivos tópicos discursivos em todo nosso material, distribuídos da seguinte forma: 78 em descrições, 61 em narrativas de experiência e 46 em relatos de opinião.²¹

A segunda etapa de análise de dados, por sua parte, embasou-se na tipologia de entidades semânticas dos referentes, apresentada na seção 3, e foi realizada seguindo dois procedimentos. O primeiro deles consistiu em observar a categoria de entidades denotada pelo termo utilizado para descrever, representar o próprio tópico discursivo, como em (1) e em (2), em que “Atividades profissionais do marido de L1” e “Bronquite do filho do meio da informante” denotam, nos dois casos, estados-de-coisas. O segundo procedimento incluiu observar o tipo de entidade designada por trechos do SegT que mais evidenciam a centralização tópica, o que representa uma maneira de distinguir a classe semântica do referente com estatuto tópico com base na materialidade linguística do SegT e nos leva a identificar quais tipos de temáticas são mais prototípicas em cada uma das amostras linguísticas analisadas. Por questões de limites de espaço, durante nossa discussão de dados, apresentaremos apenas a análise semântica dos tópicos discursivos, que é o foco deste artigo, ficando suprimida a análise tópica dos SegTs.

Como resultado das duas etapas de análise, a Tabela 1, a seguir, mostra a frequência e a proporção de ocorrências de cada tipo de entidade semântica dos referentes que se constituem como tópicos discursivos nas três diferentes amostras linguísticas que compõem nossos dados.

²⁰ Como também se poderá notar em (4), em nossos dados, acontece de o documentador direcionar não apenas a categoria semântica do tópico discursivo, mas já sugerir ao informante um ou outro determinado referente como tópico.

²¹ Como sintetizado na seção 3, Lyons (1977b) define três tipos de entidades semânticas – indivíduo, estado-de-coisas e proposições –, definição que vai da classe mais concreta (indivíduo) para a mais abstrata (proposição). Assim, cruzando esse quantitativo de SegTs identificado em cada amostra linguística investigada com os resultados sumarizados na Tabela 1, ainda nesta seção, pensamos que pode haver uma relação entre maior/menor concretude da categoria semântica dos referentes que são tópicos discursivos nos tipos de amostras linguísticas em estudo e a quantidade de tópicos instaurada nessas amostras, sendo a produtividade tópica da amostra linguística menor quanto maior for o grau de abstração da categoria semântica dos referentes que são tópicos discursivos naquela amostra.

Tabela 1 – Frequência e proporção de tipos de entidades semânticas dos referentes que são tópicos discursivos em descrições, narrativas de experiência e relatos de opinião

Tipos de amostras de fala dos tópicos Tipo de entidade semântica	Descrições % (número)	Narrativas de experiência % (número)	Relatos de opinião % (número)	Total % (número)
Indivíduo	24,36% (19/78)	0	0	10,27% (19/185)
Estado-de-coisas	20,51% (16/78)	90,16% (55/61)	63,04% (29/46)	54,06% (100/185)
Proposição	0	8,20% (5/61)	36,96% (17/46)	11,89% (22/185)
Ato de fala	0	0	0	0
Propriedade	3,85% (3/78)	0	0	1,62% (3/185)
Lugar	51,28% (40/78)	0	0	21,62% (40/185)
Tempo	0	1,64% (1/61)	0	0,54% (1/185)
Total de tópicos mínimos	42,16% (78/185)	32,97% (61/185)	24,87% (46/185)	100% (185/185)

Fonte: Elaboração própria.

Observando os resultados expostos na Tabela 1, vemos que, em descrições, um objeto de discurso que atua como tópico discursivo pode ser, além de um indivíduo ou um lugar, um estado-de-coisas ou uma propriedade; enquanto em narrativas de experiência, pode ser estado-de-coisas, proposição ou tempo, com baixa frequência dessas duas últimas categorias; já em relatos de opinião, é possível que um referente com estatuto de tópico seja ou estado-de-coisas ou proposição apenas. Como também se pode verificar, em nenhum dos três tipos de amostras linguísticas encontramos tópicos discursivos que são atos de fala, não sendo característica desses textos a centração em uma ação que realizamos ao falarmos (ordem, pergunta, declaração). Outros dados interessantes que demonstra a Tabela 1 é a maior ocorrência de proposições em relatos de opinião, e não em descrições ou narrativas de experiência, apesar de não ser essa a categoria predominante nos relatos de opinião, e a maior frequência de estados-de-coisas (54,06%, 100 de 185 casos) em meio ao total de todas as categorias identificadas, considerando a soma dos tipos semânticos apurados nos três tipos de amostras linguísticas.

A partir do SegT em (3), recortado de uma descrição, discutimos um tópico discursivo que é uma entidade semântica do tipo lugar, a classe mais recorrente nesse tipo de amostra linguística, com 51,28% das ocorrências.²²

²² Conforme se pode notar observando a pergunta-base que guiou a coleta de descrições do IBORUNA (Gasparini-Bastos; Gonçalves, 2005), o documentador deveria sugerir ao informante que descrevesse um lugar, não priorizando a descrição de pessoas ou roupas, por exemplo. Isso faz com que as descrições do IBORUNA sejam tipicamente caracterizadas como descrições de lugares e é também uma evidência da importância do documentador no processo textual-interativo nesses textos.

(3) Tópico discursivo: Cômodos da casa da informante

e também vô(u) falá(r) um po(u)quinho da **minha casa... minha casa...** éh:: há **qua-** 1
tro... quatro cômodos e o banhe(i)ro... éh::... na sala... tem/ há dois sofá::... éh:: uma 2
 rack né? uma televisão:: um vídeo o/ um rádio... éh:: aí tem o **meu quarto** onde eu 3
 durmo né?... tem a cama de casal::... o guarda-ro(u)pa... a beli::che... uma cômoda 4
 tem uma televisão... também... aí há o **banhe::(i)ro** né?... é um **banhe(i)ro** pe/ é 5
 gran::de assim:: éh do jeito assim éh:: bonito assim éh tem vaso pia coisa que todo 6
banhe(i)ro tem aí no o(u)tro quarto da minha mãe tem... a ca::ma dela a/ a cômoda o 7
 guarda-ro(u)pa... né? aí também tem a:: **cozinha da minha ca::sa** que há uma me::sa 8
 com seis cade::(i)ras... um armá::rio... uma geladei::ra... e há também um fogão e a 9
 pia... ((vozes ao fundo)) 10

[BDI-AC-038: DE: L. 163-172; 4/9]

Na descrição que comporta o SegT em (3), a informante descreve sua cidade, focalizando diferentes lugares do município, como a sua casa e, mais especificamente, os cômodos desse local, tópico em torno do qual se constitui acentração no caso ora analisado. Entendemos que o tópico em (3) é um referente do tipo lugar em razão de que o termo “cômodos”, que expressa o tópico, denota uma entidade dessa categoria, a qual reúne referentes conceptualizados como ambientes concretos, espaços que podem ser ocupados por indivíduos. Também foi possível chegar a essa categorização semântica porque observamos que trechos do SegT que mais diretamente vão indicando o conjunto referencial em torno do qual se constrói acentração, como “cômodos”, “banheiro”, “sala”, “quarto” e “cozinha”, denotam entidades do tipo lugar. Assim, considerando a categoria semântica designada por esses elementos textuais e a integração entre eles, instaura-se um objeto de discurso que é tópico discursivo também da classe lugar.

O caso em (4), na sequência, serve-nos para ilustrar um tópico que é um indivíduo, segunda classe semântica mais recorrente nos tópicos discursivos em descrições (24,36% dos casos).

(4) Tópico discursivo: Pessoa de quem a informante gosta muito

Doc.: agora eu queria que cê... fizesse uma descrição pra mim... descrevesse ou **uma** 1
peessoa::a ou um lugar:: um lugar que cê foi:: achô(u) interessan::te uma cida::de uma 2
 pra::ia... uma descrição de qualquer coisa 3
 Inf.: descrição?... uhm::... uma descrição... eu vô(u) descrevê(r) **uma pessoa que eu** 4
gosto muito... [Doc.: uhum ((concordando))]. precisa falá(r) o nome [ou só fala 5
 quando for precisá(r)?] 6
 Doc.: [não... (só o que cê pod/ acho que sempre)] 7
 Inf.: então éh::... **ela é uma pessoa assim muito amoro::sa...** eu conheci ((a filha da 8
 informante fala com terceiros)) eu conheci **ela...** quando eu tinha dezessete a::nos... 9
ela foi uma pessoa que me ajudô(u) mui::to... é uma pessoa assim que::... **você conta** 10
as coisa pra ela ela consegue te entendê::(r) é uma pessoa que tem empati::a... tam- 11
 bém uma pessoa que sempre me ajudô::(u)... **ela::...uma pessoa::...** como eu posso 12
 dizê(r)? de aparência até num é tão bela mas... é:: intimamente assim é **uma pes-** 13
soa maravilhosa... ela tem a:: paciência de:: escutá::(r)... e também de te aconselhá(r) 14
 conforme você precise... né?... **ela é:: é morena...** tem os olhos verdes... **morena não** 15

loira assim mas o cabelo dela é castanho claro... tem os olhos verdes é branca... mora	16
aqui perto de mim e foi uma pessoa muito importante na minha vida... E... e mais	17
assim é uma pessoa excelente é muito amoro::sa... e:: é o que eu te falei ela conse-	18
gue::... entendê(r) ou quando a gente fala escu::ta... pra depois aconselhá(r)	19
Doc.: uhum ((concordando))...	20
[BDI-AC-064: DE: L. 64-84; 1/4]	

No SegT em (4), o primeiro, na linearidade textual, de um conjunto de 4 SegTs da descrição da qual foi extraído, o documentador pede à informante que faça uma descrição de uma pessoa ou um lugar, como vemos nas linhas 1-3, e com base na ativação desses referentes, a falante opta por descrever uma pessoa de que gosta muito, conforme anunciado nas linhas 4-5. É com base em exemplos como esse que dissemos, então, que, apesar de os textos de nosso material serem de cunho predominantemente monológico, o documentador também assume uma função textual-interativa fundamental na construção textual, haja vista que, ao sugerir um ou outro referente como tópico – “eu queria que cê [...] descrevesse ou *uma pesso::a* ou *um lugar::*” – ele já dá um direcionamento sobre a classe semântica do tópico discursivo a ser instaurado, a qual, nesse caso, é um indivíduo, categoria semântica definida pela informante a partir de uma relação textual-interativa com a fala do documentador.

Como fizemos com o exemplo em (3), uma forma de justificar essa análise semântica é a identificação da categoria semântica denotada pelo termo que expressa o tópico discursivo, no caso, “pessoa”, que designa um indivíduo, referente ontologicamente caracterizado, por exemplo, como concreto, localizável no tempo e no espaço e passível de ser tocado. Também entendemos que tal classificação semântica é possível porque, na materialidade linguística do SegT, expressões como “pessoa”, novamente, e “ela”, que estão presentes em toda a organização textual-interativa do SegT e vão nos mostrando a relevância do tópico discursivo identificado, também denotam entidades do tipo indivíduo.²³

Em (5), exemplificamos um SegT cujo tópico é um estado-de-coisas em descrições, categoria semântica identificada em 20,51% dos dados desse tipo de amostra linguística.

(5) Tópico discursivo: Semelhança do jipe do informante com jipes usados na II Guerra Mundial

é um jipe que foi:: usado na::... na segunda guerra mundial... ele é um po(u)co mais	1
moderno um po(u)co mais novo né?... do que do o usado na segunda guerra mundial	2
mas ele é basicamente a mesma coisa... ele é/ a aparên::cia... e até:: a mecâ::nica dele	3
é praticamente igual ao usado na guerra né? na segunda guerra mundial...	4
[BDI-AC-049: DE: L. 116-120; 2/13]	

²³ Em nossos dados, é admissível ainda assumir uma correlação entre entidade semântica de um tópico discursivo e a categoria da língua (nome ou predicação) que mais diretamente remete ao tópico na materialidade do SegT. Por exemplo, nas ocorrências em que o tópico é ou um lugar ou um indivíduo, são os sintagmas nominais que indicam, de forma mais explícita, o tópico no interior do SegT, como vemos em (3), com o sintagma “cômodos” (L. 2), e também em (4), com o sintagma “uma pessoa que eu gosto muito” (L. 4-5). Por outro lado, quando o tópico é ou um estado-de-coisas ou uma proposição, é comum que sejam as predicações as responsáveis por indicar o tópico de modo mais direto na materialidade do SegT, conforme se poderá notar em (7) e em (8), com

Na descrição que recobre o SegT em (5), o falante descreve diferentes partes do carro e também trata da semelhança do veículo com outros do modelo usados na II Guerra Mundial, o que é foco da interação no exemplo ora em análise. Assumimos que o tópico em (5) é um estado-de-coisas porque o termo “semelhança do jipe do informante com jipes usados na II Guerra Mundial” designa uma entidade de segunda ordem, uma vez que pode ser considerado em si mesmo como um estado (ser semelhante), testemunhável e que pode ter uma duração temporal demarcada. Ainda, essa categorização semântica foi possível porque interpretamos que algumas passagens linguísticas em foco, que mais evidenciam de forma direta a construção do tópico, já denotam um estado-de-coisas, como fazem os excertos “ele é basicamente a mesma coisa” (L. 3) e “a aparên::cia... e até:: a mecâ::nica dele é praticamente igual ao usado na guerra né?” (L. 3-4). Sendo assim, é possível depreender que essas passagens expressam um estado, um modo de ser do jipe (o seu modo semelhante relativamente aos carros usados na II Guerra), o qual pode ser observável, pontualizado no tempo e avaliado em termos de sua realização.

A seguir, utilizamos o SegT em (6) para discutir um caso em que o tópico discursivo é do tipo propriedade, classe semântica menos recorrente em descrições (3,85% das ocorrências).

(6) Tópico discursivo: Tonalidade do acabamento da casa do informante

e:: no nos quarto também TUdo cla::ro nas duas salas já é um po(u)co mais escuro	1
Doc.: ah nos quarto também é cla::ro	2
Inf.: é:: é claro...[Doc.: hum:::] e são todos enverniza::dos portas... made(i)ra aqui tudo	3
de cereje(i)ra clara... [Doc.: hum:::]	4

[BDI-AC-139: DE: L. 258-262; 8/11]

O SegT em (6) compõe uma descrição na qual o documentador pede ao informante que descreva a casa onde mora e, com base nesse referente ativado pelo entrevistador, são construídos tópicos centrados, por exemplo, em diferentes cômodos do lugar e também em propriedades perceptuais da residência, como ocorre em (6). Desse modo, dizemos que no SegT em pauta emerge um referente com estatuto tópico que é do tipo propriedade porque o termo “tonalidade” designa essa classe semântica, expressando uma propriedade perceptual da casa. Além disso, na materialidade do SegT, verificamos que a centração vai sendo mais diretamente evidenciada por termos como “claro” e “escuro”, os quais nos orientam no reconhecimento do recorte temático e também denotam a categoria semântica propriedade.

Iniciamos a análise das narrativas de experiência com o SegT em (7), no qual o tópico discursivo é um estado-de-coisas, categoria semântica bastante predominante (90,16% dos casos) nos tópicos nesse tipo de amostra linguística.

as respectivas predicções “nos tornamos Testemunhas de Jeová” (L. 4-5) e “se ele tivesse ficado aqui eu não teria talVEZ engravidado tão Rápido” (L. 14-15). A nosso ver, essa correlação pode ser explicada pela categoria da língua que codifica as classes de entidades semânticas dos referentes, haja vista que indivíduos e locativos só podem ser codificados por nomes, ao passo que estados-de-coisas e proposições podem ser codificados por nomes ou por predicções.

(7) Tópico discursivo: Conversão da informante e do marido em testemunhas de Jeová	
depois eu/ depois quando a C. nasceu:... depois de:... um a::no um ano e pouquinho	1
nem isso... aí/ nós começamo(s) a estudá(r) a Bíblia também... aí nós começamo(s)...	2
freqüentá(r) o Salão do Rei::no né?... começamo(s) também:... í(r) nas reuniõ::es e aí	3
depois no/ nós (nos) batiza::mo(s) aí tem a prime(i)ra comunhão depois... nos torna-	4
mos Testemunhas de Jeová::... e agora levamo(s) as fi::lhas né? também	5
[BDI-AC-064: NE: L. 14-18; 3/4]	

No SegT em (7), a informante se centra no processo de conversão dela e do marido em testemunhas de Jeová, tópico da interação nesse ponto da narrativa de experiência. Classificamos esse tópico como um estado-de-coisas porque o termo “conversão”, que expressa o tópico, denota entidades dessa categoria semântica, dado que expressa um processo que ocorreu em um certo período do tempo, podendo ser avaliado pela sua realização. Outro fator que nos levou a essa classificação semântica foi a observação da instauração da centração sobremaneira por meio de trechos do SegT que designam um episódio formado por diferentes estados-de-coisas, como “nós começamo(s) a estudá(r) a Bíblia” (L. 2), “nós começamo(s) ... freqüentá(r) o Salão do Rei::no” (L. 2-3), “começamo(s) também:... í(r) nas reuniõ::es” (L. 3), “nós (nos) batiza::mo(s)” (L. 4) e “nos tornamos Testemunhas de Jeová::” (L. 4-5), os quais exprimem uma sequência de eventos que podem ser testemunhados, localizados em algum momento do tempo (no caso, depois do nascimento da filha C.) e também do espaço (podemos interpretar, por exemplo, que o batismo ocorreu no Salão do Reino), sendo passíveis de serem avaliados pelo seu estatuto de realidade.

Assim, considerando a classe semântica designada por trechos que fundamentalmente demarcam o conjunto referencial concernente em foco, instala-se um objeto de discurso tematicamente abrangente no SegT, com estatuto de tópico discursivo, que é uma entidade do tipo estado-de-coisas, uma vez que diz respeito a um processo pelo qual a informante e o marido passaram (a conversão a testemunhas de Jeová), o qual também pode ser avaliado pela sua realização.

Na sequência, exemplificamos um SegT em que o tópico é uma proposição, segunda categoria mais identificada nos tópicos nas narrativas de experiência (8,20% dos casos).

(8) Tópico discursivo: Possibilidade de retardação da gravidez da informante caso o pai não tivesse se separado da mãe	
Doc.: e:: essa separação do seu pai e da sua mãe tem alguma relação de assim de você	1
tê(r) engravidado tão nova?	2
Inf.: eu acho que não... ma/... ah se tem alguma coisa a vê(r) né?... AH TEM tem muito	3
a vê(r) porque aconteceu o seguinte meu PAI... éh foi embo::ra então até que ele tava	4
aqui eu apesar/ eu era uma meni::na... mas o que que acontecia? eu tinha me::do...	5
PAI sempre é mai/ éh num sei alguns são mais rigoroso então eu tinha me::do eu	6
num saía de ca::sa... eu eu era assim só ficava atrás dele aí o que aconteceu? ele foi	7
embo::ra... o tempo passô::(u) eu amadureci:: fiquei moci::nha... aí comecei saí(r)...	8
COIsa que se ele tivesse aqui talvez ele teria me segura::do um po(u)co mais... e eu	9

já a/ obedecê(r) minha mãe eu já não obedeCla eu/ eu ia eu queria saí(r) eu saí::a... 10
e com ele não eu acho que se ele tivesse fica::do aqui::... éh que nem quan/ quando 11
ele so(u)be que eu estava grávida ele éh ele/ minha mãe falô(u) pra ele – “porque que 12
você não ficô(u) aqui pa ajuDÁ(r) eu criá(r) seus filhos até eles cresce/ eles crescesse 13
mais?” – então acho que se ele tivesse ficado aqui eu não teria talVEZ engravidado 14
tão RÁpido como foi porque eu tinha quatorze ano eu ainda era:: uma menina né? 15
num tinha juízo nenhum... e aí:: eu acho eu acho que também por isso porque::... eu 16
acho éh:: se ele tivesse ficado aqui talvez teria evitado... mas... éh:: ele acabô(u) in(d)o 17
embo::ra né?... e eu::/ aí já comecei saí(r) eu acho que a reVOLta também:: né? porque 18
eu fiquei revoltada... e foi onde eu comecei fazê(r) coisas erradas e acabei... ten(d)o 19
a minha/ os meus/ a minha/ o meu filho antes da hora... -- pode continuá(r)? [Doc.: 20
((acena que sim))--... 21
[BDI-AC-038: NE: L. 33-52; 5/8]

Como apontado, no SegT em (8) a centração focaliza a possibilidade de a informante não ter engravidado na adolescência caso a separação dos seus pais não tivesse ocorrido ainda na sua infância. Analisando a classe semântica do tópico discursivo identificado, apuramos que ele é uma proposição em virtude de que o termo “possibilidade”, que o expressa, denota essa categoria, ou seja, diz respeito a um construto mental, relativo a um mundo imaginário, não-factual, tratando a informante de uma realidade que apenas potencialmente poderia ter sido concretizada.

Além da classe semântica designada por “possibilidade”, observamos que segmentos textuais que vão mais diretamente construindo o tópico, como aqueles destacados em negrito no interior do SegT, denotam proposições, expressando um construto mental da informante, o qual não pode ser localizado nem no tempo, nem no espaço e é passível de ser avaliado pelo seu valor de verdade, podendo ser verdadeiro ou falso. Assim, podemos entender a partir, por exemplo, do trecho “então acho que se ele tivesse ficado aqui eu não teria talVEZ engravidado tão Rápido” (L. 14-15) que a possibilidade da gravidez mais tardia da informante é algo que só existe na mente dela, um julgamento que ela mesma faz sobre uma realidade potencial, que poderia ser negada por algo como a realização da gravidez mesmo com a presença do pai em casa durante a adolescência da jovem. A partir dessas passagens linguísticas que designam entidades de terceira ordem e mais evidenciam o tópico discursivo, instaura-se um objeto de discurso, com estatuto tópico, que tem a mesma natureza semântica das entidades designadas pelos trechos que mais demonstram o foco textual-interativo. Ainda, a classificação do tópico em (8) como uma proposição pode ser sustentada pela ocorrência de expressões epistêmicas que marcam a conjectura da informante, como “eu acho que...” (L. 11, 14, 16, 18) e o advérbio “talvez” (L. 9, 14 e 17), indicando uma atitude proposicional²⁴ (no caso, de dúvida) em relação à possibilidade de ocorrência de um estado-de-coisas não verdadeiro (a gravidez mais tardia).

Com o SegT em (9), discutimos o único caso encontrado em nossos dados em que o tópico é da classe semântica tempo, o qual representou 1,64% das ocorrências nas narrativas de experiência.

²⁴ Conforme apresentado na seção 3, proposições podem ser qualificadas em termos de atitude proposicional.

(9) Tópico discursivo: Período de trabalho do informante na C.P.F.L.

ma::s... esse período que eu tive na companhia então eu tive MUita experiência fiz	1
MUita amizade conheci MUita gente	2
Doc.: quantos anos o senhor ficô(u) lá?	3
Inf.: eu fiquei na paulista... vinte e nove anos	4
Doc.: vinte nove [anos?]	5
Inf.: [vinte] nove anos só na C.P.F.L.... só na C.P.F.L. foi vinte no::ve... viajando:: vinte	6
nove acredito que uns vinte e seis anos só de viagem	7

[BDI-AC-139: NE: L. 82-88; 8/10]

Nesse SegT, o informante focaliza o período em que trabalhou na Companhia Paulista de Força e Luz, a C.P.F.L. Classificamos o tópico discursivo em questão como da classe tempo porque a expressão da língua que permite nomear o tópico – “período” – designa essa categoria semântica, identificando um intervalo temporal em que o informante trabalhou na tal empresa. Ainda como comprovação dessa nossa classificação, chamamos a atenção para expressões recorrentes na materialidade do SegT, que vão nos indicando mais diretamente o tópico e que também designam o tipo semântico tempo, a exemplo de “período” (L. 1), uma vez mais, e “anos” (L. 3, 4, 5, 6 e 7).

Segundo apresentamos, em relatos de opinião, os tópicos discursivos são ou estados-de-coisas ou proposições, com predomínio não muito amplo de estados-de-coisas (63,04% dos casos). Vejamos um exemplo no qual o tópico é dessa classe semântica predominante.

(10) Tópico discursivo: Despreocupação do informante com política

aGOra... eu realmente eu num:: num procuro me... PREOcupá(r) muito com a política sabe? porque [Doc.: aham ((concordando))... às vez cê vê dólar so::be dólar (a)	1
ba(i)xa num sei q/ bolsa isso aquilo... e pra mim que eu trabalho com vendas eu eu	2
num procuro vê(r) isso mui::to porque ele num acaba... sendo até (mal) pra mim por-	3
quE... cê pega os que... cedo é – “ah o dólar subiu ah a bolsa (a)ba(i)xô(u)” – num sei...	4
parece que no o(u)tro dia quando cê sai pra trabalhá(r) cê já sai meio que com aquilo	5
na cabeça né? ah:: então... ((conversas ao fundo)) ih/ sei lá viu... (a)caba até meio que	6
prejudican(d)o então eu num... num... vô(u) muito atrás de política não num... num	7
procuro s/ sabê(r) muito não num procuro discutí(r) também muito sobre política não	8
Inf.: não	9
Doc.: não?	10
	11

[BDI-AC-075: RO: L. 311-322; 4/4]

Entendemos que o tópico *Despreocupação do informante com política* é um referente da natureza dos estados-de-coisas porque podemos o reconhecer por meio de um termo – “despreocupação” – que designa essa classe de entidades, expressando uma experiência do informante (não) vivida por suas faculdades sensoriais.

Ao observar a materialidade linguística do SegT, também constatamos que excertos que constroem o tópico mais diretamente, como “eu num:: num procuro me... PREOcupá(r) muito com a política” (L. 1-2), “eu num procuro vê(r) isso mui::to (L. 3-4) “ e “eu num... num... vô(u) muito atrás de política não” (L. 8), designam entidades da ordem dos estados-de-coi-

sas, demonstrando algo que o falante (não) experimenta e que pode ser determinado em um período de tempo. Assim, é possível dizer, por exemplo, que, durante os anos que trabalha como vendedor, o informante não se preocupa com política e não se interessa pelo assunto, o que não acontecia quando exercia, em um intervalo temporal anterior, outro ofício. Considerando, pois, a categoria de entidades denotada por enunciados que mais evidenciam o tópico construído, cria-se um objeto de discurso que é tópico discursivo da mesma ordem semântica designada por esses enunciados.

Por fim, mostramos um SegT que materializa um tópico discursivo que é uma proposição em um relato de opinião, classe semântica que também se mostrou recorrente nos tópicos nesse tipo de amostra linguística, tendo sido identificada em 36,96% das ocorrências.

(11) Tópico discursivo: Necessidade de se ter um curso básico de administração para ser político

você pega aí ... vamo(s) pegá(r) pela nossa cidade né? São José do Rio Preto... na época 1
de política você vê vários candidatos que aparecem na televisão aí... cê vê que... a pes- 2
soa realmente... NÃO... POR ELA assim mas ela num TEM um::... num vô(u) falá(r) 3
um nível cultural mas vô(u) falá(r) assim num tem um nível de estudo né? num tem 4
uma formação:: acadê::mica alguma coisa assim... pra sê(r) político porque...(então) 5
a gente VÊ que/ que que vê? isso é tudo jogada de partido né? pa puxá(r) legenda pra 6
cá puxá(r) legenda pra lá... mas eu sô(u) contra isso porque eu penso assim a pessoa 7
pra sê(r) político eu sempre comentei isso com várias pessoas... eu acho que ela teria 8
que tê(r)... NUM VÔ(U) falá(r) uma faculda::de porque seria aí uma coisa ca::ra mui- 9
tos não teriam condições de pagá(r)... mas eu acho que... a própria prefeit::ra no 10
ca::so o próprio gover::no deveria fazê(r)... tipo d'um curso porque... pra você sê(r) 11
político nada mais é do que você administrá(r)... [Doc.: uhum ((concordando))] eu 12
dei exemplo de um vereador ou um PRÓprio prefeito num tem noção neNHUma de 13
administração... cê fala assim pra mim –“ah u/ o/ vô(u) te dá(r) pra você sê(r) um vere- 14
ador”– TUDO BEM eu vô(u) sê(r) vereador... am/ u/ BOM eu vô(u) lá eu vô(u) ganhá(r) 15
dinhe(i)ro oh num plu/ num sei né? me colocam lá um exemplo assim... mas eu NUM 16
tenho formação neNHUma pa administrá(r) então eu acho que é:: complicado... 17
porque o que eu falei... nada mais é... o preFEito mesmo... é administrá(r) tá admi- 18
nistran(d)o... os bem NOSSO o da ciDAde da comunidade em si tá administrando... 19
então eu acho que deveria tê(r)... uma faculdade? NÃO... mas um curso... BÁ::sico 20
sei lá de administração... certo?... e a pessoa teria que tê(r) um mínimo de estudo 21
pra podê::(r)... sê(r)... um... um político né?... [Doc.: aham ((concordando))]... cê vê... 22
cê pega tantos político aí que num tem noção nenhuma... e tá aí fazen(d)o alguma 23
coisa... às vez tem uns que tem no/ MUIta noção mas... e RO(u)ba (há tempo) e acho 24
que sabia até demais... agora eu acho que pra sê(r) político deveria tê(r) um... um 25
curso básico de administração NO mínimo [Doc.: uhum ((concordando))] é pelo 26
menos o que ele deveria tê(r) assim... ((conversas ao fundo)) essa é a minha idéia pra 27
podê(r) s/ pessoa... éh:: engajá(r) aí na política 28
Doc.: tá [Inf.: hum]... 29
[BDI-AC-075: RO: L. 274-300; 2/4]

Em (11), uma justificativa para a classificação semântica do tópico *Necessidade de se ter um curso básico de administração para ser político* é que ele pode ser entendido, particularmente, como um ponto de vista, uma avaliação do falante sobre o mundo atual, sobre ser necessário, segundo a própria visão dele, ter um curso básico de administração para ser político. Esse entendimento de que o tópico está relacionado com um ponto de vista do informante é sugerido por elementos textuais que integram o SegT, como “eu penso assim” (L. 7), “eu acho que ela teria que ter” (L. 8-9), “eu acho que deveria” (L. 20) e “essa é a minha idéia pra podê(r) s/ pessoa... éh:: engajá(r) aí na política” (L. 27-28).

A relação do tópico com uma avaliação do falante também pode ser apreendida a partir da categoria de entidades designada por “necessidade de se ter um curso básico de administração para ser político”, uma entidade da ordem das proposições, já que manifesta um julgamento acerca do mundo, e também por trechos do interior do SegT que indicam de forma mais direta o tópico instaurado, como “a própria prefeitu::ra no ca::so o próprio gover::no deveria fazê(r)... tipo d’um curso” (L. 10-11), “a pessoa teria que tê(r) um mínimo de estudo pra podê::(r)... sê(r)... um... um político né?” (L. 21-22), “pra sê(r) político deveria tê(r) um... um curso básico de administração” (L. 25-26), os quais designam proposições, isto é, construtos mentais, que, portanto, só estão na mente do falante e que expressam uma avaliação dele a respeito de que deveria ser convencionado socialmente, por meio de obrigação moral ou legal, que quem quiser assumir um cargo político tenha um curso de administração. Nesse contexto, enquanto construto mental, essa avaliação poderia ser negada, por exemplo, por um estudo que mostrasse que não há relação direta entre políticos com curso de administração e a qualidade da gestão pública feita por eles.

Assim, como interpretamos que a própria expressão linguística que expressa o tópico designa uma entidade do tipo proposição, o que também ocorre com passagens que mais evidenciam o tópico instaurado, entendemos que o objeto de discurso com estatuto tópico é uma entidade da ordem das proposições.

5 Influência de características de gêneros textuais descritivos, narrativos e argumentativos sobre os tipos de entidades semânticas dos tópicos discursivos em descrições, narrativas de experiência e relatos de opinião

Conforme apresentado na seção 1, como uma vertente da LT filiada sobretudo a um viés pragmático, o foco de investigação da GTI é a construção textual integrada aos fatores enunciativos que lhe dão existência. Por isso, no âmbito desse quadro teórico-metodológico, assume-se que a construção dos textos está relacionada a características dos gêneros textuais a que os textos se filiam (Penhavel; Cintra, 2022). Seguindo esse pressuposto, nesta seção, discutimos a influência de traços de gêneros textuais descritivos, narrativos e argumentativos sobre os tipos de entidades semânticas dos referentes que ocorrem como tópicos discursivos em nossos dados, os quais são representativos, como reiteramos, de gêneros textuais dessas três tipologias. Para cumprir esse nosso propósito, primeiramente, apresentaremos sinteticamente a concepção de gênero textual por nós adotada, bem como a de tipo textual.

De acordo com Koch (2018), os gêneros textuais constituem formas padrão e relativamente estáveis de enunciados, diretamente relacionadas às diferentes situações sociais, marcadas socio-historicamente e com características temáticas, composicionais e estilísticas próprias. No que toca ao conceito de tipo textual, Marcuschi (2008) compreende tipo textual como uma espécie de construção teórica definida pela natureza linguística de sua composição (aspectos lexicais, sintáticos, tempos verbais, relações lógicas, estilo). Conforme o autor, o tipo caracteriza-se muito mais como sequências linguísticas do que como textos materializados, sendo, a rigor, modos textuais que, em um número limitado e sem tendência a aumentar, recobre as seguintes categorias: narração, argumentação, exposição, descrição e injunção. Ainda para o autor, quando predomina um modo num determinado texto concreto, dizemos que esse texto é narrativo ou argumentativo ou expositivo ou descritivo ou injuntivo.

Feita essa breve explicação, passamos a nos valer de especificidades dos tipos textuais descritivo, narrativo e argumentativo descritas em Koch e Fávero (1987), pensando em como tais especificidades estariam associadas aos resultados apresentados na seção anterior.

Segundo Koch e Fávero (1987), duas das características do tipo descritivo são a caracterização física ou psicológica de personagens ou espaços e a demonstração de qualidades do ser descrito. Assim, interpretamos que a predominância de objetos de discurso com estatuto tópico que são do tipo lugar em nossas descrições e a identificação da classe semântica indivíduo como a segunda mais recorrente nesse tipo de amostra linguística (51,28% e 24,36% das ocorrências, respectivamente) estão relacionadas justamente com a peculiaridade de gêneros descritivos de apresentarem uma caracterização física de personagens ou espaços, posto que, como explicado na seção 3, podemos chamar de indivíduos os objetos físicos, as pessoas, os animais e as coisas que são concretos e identificáveis e classificar como da classe lugar os ambientes, os espaços físicos que podem ser ocupados por indivíduos. Portanto, como a caracterização de personagens ou espaços físicos é tipicamente focalizada no processo textual-interativo no tipo textual descritivo, é comum que se instaure centração em torno desses elementos físicos em descrições, levando-os a adquirir estatuto de tópico discursivo.

Ainda, é possível dizer que a ocorrência de tópicos que são estados-de-coisas nas descrições de nosso material pode ter relação com o traço do tipo descritivo de demonstrar qualidades do ser descrito, já que, segundo também explicamos na seção 3, entidades de segunda ordem podem dizer respeito a estados, o que pode permitir que tópicos que são estados-de-coisas expressem um estado, uma qualidade de um indivíduo ou um lugar, como em (5), em que o tópico é a *Semelhança do jipe do informante com jipes usados na II Guerra Mundial*. Com relação aos tópicos que são propriedades nas descrições de nosso corpus, estes também podem estar relacionados à peculiaridade do tipo descritivo de apresentar qualidades do ser descrito, dado que os casos de tópicos dessa natureza estão centrados em uma propriedade perceptual de um referente, como o tópico *Tonalidade do acabamento da casa do informante*, em (6).

No tocante a particularidades de gêneros narrativos que podem ajudar a compreender as entidades semânticas dos objetos de discurso que são tópicos discursivos nas narrativas de experiência, Koch e Fávero (1987) assinalam que alguns dos traços do tipo narrativo são a asserção de enunciados de ação e a captação de eventos numa sucessão cronológica. Então, a grande predominância de tópicos discursivos que são estados-de-coisas neste tipo de amostra linguística (90,16% dos casos) pode ser explicada pelo traço do tipo narrativo de apresentar enunciados de ação, tendo em vista que essa categoria semântica envolve ações, bem como pela sua particularidade de captar eventos cronologicamente ordenados, sendo que cada

tópico discursivo linearmente ordenado em uma narrativa de experiência vai expressando, via de regra, um evento cronologicamente organizado e proeminente naquela narrativa.

Essa especialidade do tipo narrativo de apresentar eventos conforme uma ordem temporal também pode permitir entender por que a única ocorrência de tópico discursivo do tipo tempo foi apurada em uma narrativa de experiência. Focalizando esse tipo de amostra de fala eventos, acontecimentos situados no tempo, pode ocorrer centração em torno do próprio tempo, o que ajuda a marcar temporalmente eventos sobre os quais se concentra a interação em outros pontos do texto. Ainda, podemos inferir que esse traço do tipo narrativo de focalizar eventos também permite explicar os casos de tópicos que são proposições em narrativas de experiência, dado que, nessas amostras, os tópicos dessa categoria semântica podem expressar, por exemplo, um julgamento, uma expectativa sobre certo acontecimento que era visto como possível ou esperado num mundo hipotético, conforme é o tópico *Possibilidade de retardação da gravidez da informante caso o pai não tivesse se separado da mãe*, discutido a partir de (8).

Quanto a características do tipo textual argumentativo que podem estar associadas às classes semânticas dos objetos de discurso que são tópicos discursivos em relatos de opinião, Koch e Fávero (1987) apontam que o tipo argumentativo é esquematizado conforme categorias como argumentos, contra-argumentos e tese. Sendo assim, tópicos que são da classe predominante nestas amostras linguísticas, ou seja, que são estados-de-coisas, estariam relacionados a fatos, acontecimentos, que funcionariam como argumentos para sustentar uma tese proposta no relato de opinião.

Na mesma direção, a ocorrência consideravelmente constante de tópicos que são proposições nestas amostras de fala (36,96%) pode ser associada à categoria “tese” do tipo argumentativo. Assim, enquanto tópicos que são entidades da ordem dos estados-de-coisas poderiam ser vistos como argumentos que suportariam uma tese do relato de opinião, tópicos que são da classe das proposições estariam mais relacionados a um julgamento, um ponto de vista, uma avaliação explícita do informante sobre o assunto tratado, o que poderia, portanto, ser entendido como uma tese presente no relato de opinião. Dito de forma resumida, tópicos que são estados-de-coisas atuariam como argumentos para uma tese proposta no texto, expressa mais evidentemente por um tópico que é uma proposição.

6 Conclusões

Privilegiando o desenvolvimento recente de estudos do texto de base fundamentalmente pragmática e buscando descrever fenômenos textuais na relação com seu contexto de uso, particularmente na associação com particularidades dos gêneros textuais nos quais esses fenômenos se realizam, neste trabalho, fundamentamo-nos na GTI com o objetivo de descrever os tipos de entidades semânticas dos referentes que são tópicos discursivos em descrições, narrativas de experiência e relatos de opinião, discutindo a influência de traços de gêneros textuais descritivos, narrativos e argumentativos, respectivamente, sobre as categorias semânticas dos tópicos identificados em nosso material.

Como mostramos, em descrições, os referentes que são tópicos discursivos podem ser das categorias lugar, indivíduo, estado-de-coisas ou propriedade, com percentuais maiores de tópicos que são do tipo lugar. Já nas narrativas de experiência, os objetos de discurso com estatuto tópico são da classe dos estados-de-coisas, na quase totalidade dos casos, ou da

categoria das proposições ou, ainda, em um percentual consideravelmente baixo, da classe tempo. Por último, nos relatos de opinião, nossas análises demonstram que a categoria dos estados-de-coisas é a que recobre a maioria dos tópicos nesse tipo de amostra linguística, tendo sido, no entanto, a classe das proposições também recorrente nesse material. A partir desses resultados, argumentamos que os tipos de entidades semânticas dos tópicos discursivos em descrições, narrativas de experiência e relatos de opinião têm motivação em características de gêneros textuais descritivos, narrativos e argumentativos.

Ao argumentarmos que as classes semânticas dos referentes que adquirem estatuto de tópico discursivo são influenciadas por particularidades dos gêneros textuais em que os tópicos são instaurados, reafirmamos nossa filiação à GTI, cujo foco de investigação, como lembramos, é a construção textual integrada aos fatores enunciativos que lhe garantem existência e que se mostram nos próprios procedimentos de constituição do texto (Jubran, 2015a), o que também se demonstrou no trabalho de Cintra (2022).

Como nossas análises revelam, a predominância do tipo de entidade semântica dos referentes que são tópicos discursivos irá variar de acordo com diferentes gêneros textuais, em razão de fatores envolvidos no contexto real de uso da linguagem e, por conseguinte, de particularidades dos gêneros a que os textos se filiam. Essa constatação aponta para uma proposta por nós pensada (Garcia, 2022) – a de que os gêneros textuais se diferenciam uns dos outros e, portanto, podem ser caracterizados pelos tipos de entidades semânticas dos referentes que são tópicos em cada um deles. Tal proposta abre caminhos para o desenvolvimento de trabalhos futuros que envolvam um conjunto de pesquisas voltado para a análise dos tipos semânticos dos tópicos nos mais variados gêneros textuais, visando descrever os diferentes gêneros de texto a partir da categoria tópico discursivo e entender as classes semânticas de tópicos que orientam a construção do texto em diversos gêneros. Investigações dessa natureza continuariam se preocupando em compreender como os textos são construídos na sua relação com o contexto comunicativo, colaborando para o desenvolvimento recente de estudos textuais de base pragmática, como o faz este estudo.

Agradecimento

A pesquisa que deu origem a este artigo foi realizada com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. À CAPES, agradecemos o financiamento.

Referências

- AUSTIN, J. L. Locutionary, illocutionary, perlocutionary. In: HARNISH, R. (org.). *Basic topics in the philosophy of language*. New Jersey: Prentice Hall, 1962. p. 30-39.
- BROWN, G.; YULE, G. *Discourse analysis*. Cambridge: Cambridge University Press, 1983.
- CAMACHO, R. G. O papel da estrutura argumental na variação de perspectiva. In: KOCH, I. V. (org.). *Gramática do Português Falado*. 2. ed. rev. Campinas: Editora da Unicamp, 2002. p. 259-279. v.6.

- CAMACHO, R. G.; DALL'AGLIO-HATTNER, M. M.; GONÇALVES, S. C. O substantivo. In: ILARI, R. (org.). *Gramática do português culto falado no Brasil: palavras de classe aberta*. São Paulo: Contexto, 2014. p. 13-63. v.3.
- CASTILHO, A. T. O português culto falado no Brasil – história do Projeto Nunc no Brasil. In: PRETI, D. URBANO, H. (org.). *A linguagem falada culta na cidade de São Paulo*. v. IV, Estudos. São Paulo: T.A. Queiroz/Fapesp, 1990. p. 141-197.
- CINTRA, M. R. A referência em cartas de leitor mineiras. In: PENHAVAL, E.; CINTRA, M. R. (coord.). *História do português brasileiro: diacronia dos processos de construção de textos*. São Paulo: Contexto, 2022. p. 124-143.
- DIK, S. C. *The Theory of Functional Grammar: The Structure of the Clause*. 2. ed. rev. Berlin: Mouton de Gruyter, 1997.
- GARCIA, A. G. *Uma abordagem textual-interativa da relação entre organização tópica e referência em descrições, narrativas de experiência e relatos de opinião*. 2022. 259 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2022. Disponível em: < <https://repositorio.unesp.br/entities/publication/3f751cfa-115d-4f52-9b51-3093307f601a> >. Acesso em: 4 out. 2025.
- GASPARINI-BASTOS, S.D.; GONÇALVES, S. C. L. Roteiro para coleta de amostras de fala: projeto ALIP (Amostra Linguística do Interior Paulista). 2005. Disponível em: <http://www.alip.ibilce.unesp.br>. Acesso em: 12 abr. 2022.
- GONÇALVES, S.C.L. *Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista*. 2007. Disponível em: <http://www.alip.ibilce.unesp.br/>. Acesso em: 30 out. 2024.
- GUERRA, A. R. A organização tópica em cartas de leitor paulistas. In: PENHAVAL, E.; CINTRA, M. R. (coord.). *História do português brasileiro: diacronia dos processos de construção de textos*. São Paulo: Contexto, 2022. p. 38-67.
- HENGEVELD, K. Illocution, mood and modality in a functional grammar of Spanish. *Journal of Semantics*, v. 6, 1988, p. 227-269.
- HENGEVELD, K. *Non-verbal predication: theory, typology, diachrony*. Berlin: Mouton de Gruyter, 1992.
- HENGEVELD, K.; MACKENZIE, J. L. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press, 2008.
- JUBRAN, C. C. A. S. (Org.). *Gramática do português culto falado no Brasil: A construção do texto falado*. São Paulo: Contexto, 2015a.
- JUBRAN, C. C. A. S. Tópico Discursivo. In: JUBRAN, C. C. A. S. (org.). *Gramática do português culto falado no Brasil: A construção do texto falado*. São Paulo: Contexto, 2015b. p. 85-126.
- JUBRAN, C. C. A. S. Uma Gramática textual de orientação interacional. In: CASTILHO, A. T.; MORAIS, M. A. T.; LOPES, R. V.; CYRINO, S. M. L. (org.). *Descrição, história e aquisição do português brasileiro*. São Paulo: FAPESP; Campinas: Pontes Editores, 2007. p. 313-327.
- JUBRAN, C. C. A. S. Revisitando a noção de tópico discursivo. *Cadernos de Estudos Linguísticos*, Campinas, v. 48, n. 1, p. 33-41, 2006a. DOI: 10.20396/cel.v48i1.8637253.
- JUBRAN, C. C. A. S. Tópico Discursivo. In: JUBRAN, C. C. A. S. J.; KOCH, I. G. V. (org.). *Gramática do português culto falado no Brasil: Construção do texto falado*. Campinas: Editora da Unicamp, 2006b. p. 89-132.

- JUBRAN, C. C. A. S.; RISSO, M.; URBANO, H.; FÁVERO, L.; KOCH, I. G. V.; MARCUSCHI, L. A.; TRAVAGLIA, L. C.; SOUZA E SILVA, M. C. P.; ANDRADE, M. V.; AQUINO, Z.; SANTOS, M. C. Organização tópica da conversação. In: ILARI, R. (org.). *Gramática do português falado: Níveis de análise linguística*. 4. ed. rev. Campinas: Editora da UNICAMP, 2002. p. 341-428.
- KOCH, I. G. V.; FÁVERO, L. L. Contribuição a uma tipologia textual. *Letras & Letras*, Uberlândia, v. 3, n. 1, p. 3-10, 1987.
- KOCH, I. V. *Introdução à linguística textual: trajetória e grandes temas*. 2. ed. São Paulo: Contexto, 2018.
- LYONS, J. *Semântica*. Lisboa: Editorial Presença, 1977a. v. 1.
- LYONS, J. *Semantics*. Cambridge: Cambridge University Press, 1977b. v. 2.
- MARCUSCHI, L. A. *Produção textual, análise de gêneros e compreensão*. São Paulo: Parábola, 2008.
- MONDADA, L. Gestion du topic et organisation de la conversation. *Cadernos de Estudos Linguísticos*, Campinas, v. 41, p. 7-35, 2001. DOI: 10.20396/cel.v41io.8636999.
- OLIVEIRA, G. A. *Estudo do processo de estruturação interna de segmentos tópicos mínimos em cartas de leitores de jornais paulistas do século XXI*. 2016. 194 p. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/144361>. Acesso em: 25 nov. 2024.
- OLIVEIRA, R. P. *Semântica formal: uma breve introdução*. Campinas: Mercado das Letras, 2001.
- PENHAVAL, E. A organização tópica em editoriais paulistas. In: PENHAVAL, E.; CINTRA, M. R. (coord.). *História do português brasileiro: diacronia dos processos de construção de textos*. São Paulo: Contexto, 2022. p. 68-101.
- PENHAVAL, E. *Marcadores Discursivos e Articulação Tópica*. 2010. 168 p. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2010. Disponível em: https://bdtd.ibict.br/vufind/Record/CAMP_ece1coc40655b1d18d29fc578a922483. Acesso em: 13 dez. 2021.
- PENHAVAL, E. O processo de organização intratópica em narrativas de experiência. *Revista Diálogo e Interação*, Cornélio Procopio, v. 14, n. 1, p. 119-145, 2020. Disponível em: <https://www.faccrei.edu.br/revista/index.php/revista-dialogo-e-interacao/issue/view/4>. Acesso em: 9 dez. 2021.
- PENHAVAL, E.; CINTRA, M. R. (coord.). *História do português brasileiro: diacronia dos processos de construção de textos*. São Paulo: Contexto, 2022. v. 11.
- PINHEIRO, C. L. *Estratégias textuais-interativas: a articulação tópica*. Maceió: EDUFAL, 2005.
- SEARLE, J. R. *Expressão e significado: estudos da teoria dos atos de fala*. 2. ed. São Paulo: Martins Fontes, 2002.
- SOUZA, A. D. *Estudo da organização intratópica e das relações retóricas em minissagas*. 2020. 322 p. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2020. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/194468/souza_ad_dr_sjrp.pdf?sequeseq=3. Acesso em: 1 dez. 2021.
- ZANIN, I. C. A. *A diacronia da organização tópica em cartas de redator e editoriais de jornais paulistas*. 2023. 197 p. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2023. Disponível em: <https://repositorio.unesp.br/items/oef61b9a-f214-43ab-8b10-7b28b802cf26>. Acesso em 18 out. 2024.

Variação diacrônica do grau de transparência linguística do português brasileiro: o impacto do aumento percentual do uso de “a gente” em lugar de “nós”

Diachronic Variation of Linguistic Transparency in Brazilian Portuguese: The Role of the Percentage Increase in the Use of “a gente” (“we”) Instead of “nós” (“we”)

Alessandra Regina Guerra

Universidade Estadual Paulista (UNESP)

Araraquara | SP | BR

alessandreguerra@gmail.com

<https://orcid.org/0000-0002-9665-0002>

Resumo: Este artigo aborda a variação diacrônica do grau de transparência linguística do português brasileiro (PB). O objetivo do trabalho é analisar como o grau de transparência do PB é afetado por uma mudança linguística em curso na língua, o aumento da frequência de uso da forma pronominal *a gente* em detrimento do emprego do pronome *nós*. O quadro teórico-metodológico adotado funda-se na Gramática Discursivo-Funcional, complementada por um conjunto de trabalhos que descrevem a mudança *nós/a gente*. São examinados dados do PB distribuídos do século XIX ao século XXI, extraídos de peças de teatro relativas a esse período. O artigo argumenta, em termos teóricos, que a transparência do PB é, de fato, influenciada pela mudança *nós/a gente* (em correlação com outros fatores) e demonstra, então, mediante análise quantitativa, o impacto específico dessa mudança, mostrando que ela atua na direção de diminuição do grau de transparência da língua.

Palavras-chave: transparência linguística; mudança linguística; argumento-sujeito; referência cruzada.

Abstract: This article addresses the diachronic variation in the degree of linguistic transparency of Brazilian Portuguese (BP). The purpose of the paper is to analyze how the transparency of BP is affected by an ongoing linguistic change, the increasing use of the pronoun “a gente” (“we”) instead of the pronoun “nós” (“we”). The theoretical framework is based on Functional Discourse



Grammar, which is complemented by a group of works that describe the change “*nós/a gente*”. The data examined are distributed from the 19th to the 21st century and they are selected from Brazilian theatre plays. The paper argues, on a theoretical basis, that the transparency of BP is in fact influenced by the change “*nós/a gente*” (in interaction with other factors). Then, it demonstrates the specific role of this change based on quantitative analysis, showing that the diachronic variation between “*nós/a gente*” contributes to decrease the transparency of BP.

Keywords: linguistic transparency; linguistic change; subject argument; cross-reference.

1 Introdução

Nos últimos anos, no campo da Gramática Discursivo-Funcional (GDF) (Hengeveld; Mackenzie, 2008), vem sendo produzida uma série de pesquisas sobre a propriedade linguística chamada de *transparência*. Em linhas gerais, a noção de transparência linguística diz respeito à existência, em uma dada língua, de relações biunívocas entre unidades de forma e unidades de significado na estruturação linear dos enunciados, uma noção similar à modalidade de iconicidade normalmente conhecida como *isomorfismo*. Quanto mais uma língua apresenta correspondências biunívocas entre forma e significado, mais transparente a língua é; quanto menos exibe essas correspondências, menos transparente (ou mais opaca) ela é. Assim, cada língua pode ser caracterizada em termos de um grau de transparência, que consiste na medida, maior ou menor, em que a língua prevê tais correspondências forma-conteúdo.

Em um conjunto de trabalhos recentes (Guerra, 2017, 2021, 2022), temos investigado o comportamento *diacrônico* da propriedade da transparência linguística, observando-o particularmente na história do português brasileiro (PB). O propósito geral desses trabalhos tem sido analisar como o grau de transparência de uma língua varia diacronicamente (se esse grau aumenta, diminui, se caminha ora em uma, ora em outra direção) e como diferentes fatores linguísticos influenciam a evolução diacrônica da transparência. O presente artigo vincula-se a esse conjunto de trabalhos, assumindo o objetivo de analisar como o grau de transparência do PB é influenciado por uma mudança particular em curso na língua, a saber, o aumento percentual de uso da forma pronominal *a gente* em detrimento do emprego do pronome *nós*.

Como exemplos iniciais da influência da mudança *nós/a gente* sobre a transparência do PB, podem ser consideradas as sentenças em (1) e (2):

- (1) Nós cantávamos bem.
- (2) A gente cantavaØ bem.

Em (1), o pronome *nós* e a desinência *mos* expressam uma mesma unidade de significado—que, com base no quadro teórico-metodológico da GDF, trataremos como o argumento-sujeito de primeira pessoa do plural. Nesse caso, duas unidades estruturais correspondem a uma mesma unidade de significado, configurando uma construção em que não há biunivocidade entre forma e conteúdo, isto é, uma construção não transparente. Por outro lado, em (2), apenas a forma *a gente* corresponde àquela unidade semântica, já que a desinência número-pessoal não é materialmente expressa (sendo representada por morfema zero). Dessa vez, há biunivocidade forma-conteúdo, configurando uma construção transparente, o que indica a existência, aqui em pauta, de relação entre a alternância *nós/a gente* e a transparência do PB.

Acreditamos que, em linha com o propósito mais geral de nossos estudos acima mencionados, os resultados aqui oferecidos podem, de fato, cooperar para um melhor entendimento sobre o comportamento diacrônico da propriedade da transparência linguística, reflexão relevante para a teoria sobre transparência, bem como para a compreensão do funcionamento de diferentes línguas. As pesquisas já realizadas sobre transparência têm abordado diferentes aspectos do fenômeno, especialmente diferenças de grau de transparência entre dialetos e entre línguas (Grández Ávila, 2011; Hengeveld, 2011a, b; Hengeveld; Leufkens, 2018; Leufkens, 2013, 2015; Nordhoff, 2011; Seguin, 2018; entre outros). Todavia, a questão da *diacronia* da transparência ainda não tem sido igualmente focalizada, lacuna que endossa a discussão aqui proposta. Além disso, a motivação para este artigo encontra-se em nossa expectativa de que os resultados fornecidos possam também contribuir, em particular, para a descrição da história do PB, tanto no que se refere ao percurso diacrônico de seu grau de transparência, quanto no que tange à mudança *nós/a gente*, a qual integra um conjunto de alterações associadas a substancial transformação do quadro pronominal do PB e de seu respectivo paradigma de flexão verbal (Duarte, 1993, 1995; Tarallo, 1993).

O artigo encontra-se organizado do seguinte modo: na seção 2, explicamos o conceito de transparência aqui adotado e argumentamos, teoricamente, que a transparência do PB é, de fato, influenciada diacronicamente pela mudança *nós/a gente*, procurando, inclusive, delimitar qual parte do sistema gramatical da língua é afetada por tal mudança; na seção 3, sintetizamos alguns trabalhos que relatam a ocorrência da variação diacrônica *nós/a gente* no PB e expomos dados quantitativos diacrônicos acerca desse processo, levantados em pesquisa empírica por nós realizada; na seção 4, procedemos, então, a um cálculo quantitativo para mostrar, especificamente, o impacto da mudança em estudo no grau de transparência do PB; na seção 5, apresentamos as considerações finais.

2 Fundamentação e discussão teórica

A GDF, abordagem em que é postulada a noção de transparência aqui adotada, é uma teoria funcionalista da estrutura gramatical das línguas. O modelo abrange quatro níveis de análise linguística: o nível interpessoal (NI), o nível representacional (NR), o nível morfossintático (NM) e o nível fonológico (NF). Os dois primeiros são dedicados à descrição dos significados dos enunciados, e os outros dois níveis descrevem seus aspectos formais.

O NI comporta a descrição de aspectos pragmáticos da significação dos enunciados, incluindo unidades como *atos discursivos* (as menores unidades linguísticas de comportamento comunicativo), *ilocuções* (intenções comunicativas convencionalizadas, como a ilocução declarativa, a interrogativa, a imperativa etc.) *subatos de atribuição* (que fazem a evocação de propriedades) e *subatos de referência* (que evocam referentes). O NR descreve aspectos semânticos dos enunciados, contendo unidades como *conteúdos proposicionais*, *estados de coisas* e *indivíduos*; ainda o NR descreve os enunciados em termos de esquemas de predicação, com predicados e argumentos, sendo também responsável pela marcação de funções semânticas, como agente e paciente. O NM trata da estrutura morfossintática dos enunciados, com unidades como *orações*, *sintagmas*, *palavras*, *radicais* e *afixos*. No NF, é feita a representação da estrutura fonológica dos enunciados, mediante unidades como *sintagmas entoacionais*, *palavras fonológicas*, *pés* e *sílabas*.

A título de exemplificação do modelo, observe-se a sentença em (3), atentando-se para o NI, o NR e o NM, mais relevantes no presente artigo:

(3) Eu comprei um livro.

Na GDF, a análise desse exemplo passa por descrever, no NI, que a sentença, dentre outros aspectos, é formada por um ato discursivo declarativo, composto por um subato atributivo (para indicar que a sentença inclui a evocação de uma propriedade, que será, no caso, a ação de comprar), dois subatos de referência que evocam o próprio falante (um a ser realizado pelo pronome *eu* e outro, pela desinência *i*), e um novo subato de referência que evoca um referente não identificável pelo ouvinte. O NR registra, por exemplo, o esquema de predicação, que, em (3), inclui um predicado (preenchido, então, pelo item lexical *comprar*, que descreve semanticamente a evocação de propriedade feita no subato atributivo do NI), seu primeiro argumento (que funciona como agente e corresponde ao falante) e seu segundo argumento (que funciona como paciente e é preenchido pelo lexema *livro*, que propicia a descrição semântica da evocação de referente não identificável feita no NI). O NM registra, por exemplo, que a sentença é composta por uma oração, constituída por um primeiro sintagma nominal (com a função sintática de sujeito), um sintagma verbal e um segundo sintagma nominal (que atuará como objeto), nessa ordem, identificando ainda a estrutura morfológica das palavras;¹ no NM, entre outros procedimentos, é inserido o artigo *um* (que expressa o significado de referente *não identificável*, indicado no NI) e é estabelecida a ordenação sequencial dos itens *um* e *livro*.

Essa breve ilustração do modelo permite também destacar a categoria do *argumento-sujeito* (crucial para este trabalho), que será o argumento (do esquema de predicação do NR) que, na implementação de uma oração, é selecionado para a função de sujeito (marcada no NM) – no exemplo em (3), que se encontra na voz ativa, será o argumento 1. O argumento-sujeito, no caso do PB, em orações finitas, é expreso (no que tange a informações de número

¹ Na GDF, um sintagma verbal é um sintagma com uma palavra verbal como seu núcleo, e não a combinação de verbo e objeto, conforme é mais comumente concebido (cf. Hengeveld; Mackenzie, 2008).

e pessoa) por desinência verbal (explícita ou representada por morfema zero); além disso, pode ser expresso (ou não) por sintagma nominal na posição de sujeito da oração.²

Na GDF, o grau de transparência de uma língua é aferido examinando-se em que medida a língua possui fenômenos que violam a relação de um-para-um entre os quatro níveis de descrição gramatical (mais precisamente, a relação de um-para-um-para-um-para-um entre esses quatro níveis). São reconhecidos dois tipos de fenômenos linguísticos que violam a transparência. O primeiro inclui fenômenos cuja descrição implica mapeamento nos quatro níveis, porém esse mapeamento não é transparente, envolvendo, por exemplo, uma unidade num nível e duas unidades em cada um dos outros. Um fenômeno do tipo, segundo Hengeveld e Leufkens (2018), é a *referência cruzada* (RC), pela qual uma sentença faz referência a uma mesma entidade duas (ou mais) vezes, mediante elementos lexicais e gramaticais. Trata-se de recurso que acarreta uma correspondência entre uma unidade no NR e duas unidades no NI (assim como duas respectivas unidades no NM e duas no NF). Para os autores, a noção de RC só é aplicável quando cada um dos diferentes mecanismos linguísticos fazendo referência à mesma entidade é capaz de fazer referência por si mesmo, de forma independente dos outros mecanismos. O exemplo em (4) é um caso de RC, no português europeu, segundo Leufkens (2013, p. 340):

- (4) *Eu chegue-i.*
1SG arrive-1SG

Para a autora, a sentença em (4) faz referência a uma mesma entidade duas vezes: gramaticalmente, pela desinência *i*, e lexicalmente, pelo pronome *eu*. A autonomia referencial da desinência é comprovada, segundo a autora, porque o uso do verbo, com essa desinência, é gramatical mesmo sem o pronome, ou seja, o verbo sozinho é capaz de constituir um enunciado completo. Verifica-se uma correspondência não transparente entre os níveis da GDF, no caso, entre um argumento, no NR, e duas unidades em cada um dos demais níveis, sendo dois subatos de referência no NI e suas respectivas expressões formais no NM e no NF.

O segundo tipo de fenômeno violador de transparência reúne procedimentos puramente formais, que, assim, pressupõem descrição apenas no NM e no NF, não sendo analisados nos outros dois níveis. É o caso da *concordância verbal*, que se verifica, nos termos da GDF, no caso em que uma propriedade de uma unidade é copiada para outra unidade dentro da oração. O exemplo em (5), do francês, retirado de Hengeveld e Mackenzie (2008, p. 350), ilustra o processo de concordância verbal:

- (5) *Nous chant-ons.*
We sing-1.pl
“We sing” (nós cantamos).

² A variação entre expressão e não expressão do argumento-sujeito por sintagma nominal corresponde ao que, no Modelo de Princípios e Parâmetros da perspectiva gerativista (Raposo, 1992), é tratado pelo parâmetro do sujeito nulo. Já o sistema de dupla expressão do argumento-sujeito (realizada por desinência verbal e sintagma nominal) corresponde ao que, comumente, é conhecido como *concordância verbal*; na GDF, porém, conforme explicado adiante, o sistema de expressão dupla será tratado como *referência cruzada*, sendo a noção de *concordância verbal* reservada para outro tipo de caso.

Como explicam os autores, a sentença em (5) exibe um pronome de primeira pessoa do plural (*nous*) e uma flexão verbal também de primeira pessoa do plural (*ons*). Nesse caso, diferentemente do que ocorre em (4), a flexão não é capaz de fazer referência por si própria (o que se nota pela agramaticalidade do exemplo em (5), caso o pronome seja omitido), sendo apenas uma cópia da informação contida no pronome. Ou seja, a flexão não é uma codificação morfosintática de significados interpessoais ou semânticos, decorrendo, na verdade, de uma operação que tem origem no próprio NM. Trata-se, assim, de um fenômeno não transparente, ao envolver descrição no NM e no NF, mas não nos outros dois.

Hengeveld e Leufkens (2018) explicam que diversos fenômenos podem ser adotados para se calcular a transparência de uma língua, entre os quais selecionam os seguintes: (i) aposição, (ii) referência cruzada, (iii) relações gramaticais, (iv) alternância de raiz morfo(fono) logicamente condicionada, (v) descontinuidade de constituintes, (vi) ausência de paralelismo entre as estruturas morfosintática e fonológica, (vii) expletivos, (viii) gênero gramatical, (ix) concordância sintagmática, (x) concordância verbal, (xi) adaptações fonológicas. Os seis primeiros representam o primeiro tipo de fatores de violação de transparência, e os demais integram o segundo tipo. Quanto menor o número de fenômenos que uma língua apresenta, entre os listados, mais transparente a língua é. Quanto mais a língua apresentar fenômenos entre esses, menos transparente ela é.

Neste artigo, assumimos que o PB apresenta o fenômeno da RC – similarmente ao que, segundo Leufkens (2013), ocorre no português europeu – e procuramos demonstrar de que forma a transparência da gramática do PB, mais precisamente a transparência de uma parte da gramática que está relacionada à RC, é afetada diacronicamente pela mudança *nós/a gente*.

Note-se que a GDF, tendo em vista interesses mais gerais de investigação tipológica, ao definir a RC como um dos parâmetros para determinação do grau de transparência de uma língua, leva em conta a existência ou não do fenômeno na língua. Sua existência é tida como fator de opacidade e sua inexistência, como fator de transparência. Todavia, quando se focaliza a investigação de uma língua particular, pensamos que um tipo de análise que é viável (e que propicia, inclusive, uma descrição linguística mais detalhada e completa) consiste em estender o exame da transparência para além da verificação de (in)existência de RC na língua e avaliar a transparência no contexto de um domínio gramatical mais amplo, do qual a RC é parte.

Entendemos que, no PB, a RC constitui *uma* possibilidade, na construção de orações finitas, de se expressar o argumento-sujeito, possibilidade que integra um sistema maior de opções, incluindo alternativas transparentes e não transparentes (opacas). Esse sistema pode ser tratado como *o sistema de expressão do argumento-sujeito*. Ele decorre diretamente do fato de o argumento-sujeito (conforme mencionado) ser expresso mediante desinência verbal, podendo a desinência ser explícita ou representada por morfema zero, e do fato de esse argumento também poder, ou não, ser expresso mediante sintagma nominal (SN). Em outras palavras, esse sistema é o conjunto de possibilidades, disponibilizadas pela gramática do PB, de combinação entre expressão/não expressão do argumento-sujeito por desinência verbal número-pessoal explícita e expressão/não expressão do argumento-sujeito por SN.

Assim, esse sistema compreende quatro tipos de expressão do argumento-sujeito, em termos de transparência ou opacidade: (i) expressão por RC, que ocorre quando o argumento-sujeito é expresso tanto por desinência explícita quanto por SN (como na oração *nós cantávamos bem*), o que configura uma construção opaca, já que um único argumento do NR corresponderá a dois subatos referenciais no NI e a duas unidades no NM, o pronome e a

desinência verbal, bem como a duas respectivas unidades no NF; (ii) um tipo de expressão que pode ser denominado de *referência simples* (RS), verificado quando o argumento-sujeito é expresso apenas por desinência verbal número-pessoal explícita (*Ø cantávamos bem*), o que constitui uma construção transparente, pois nela o argumento-sujeito é expresso uma única vez; (iii) um segundo tipo de RS, também transparente, instaurado quando o argumento-sujeito é expresso apenas por SN, com desinência verbal representada por morfema zero (*a gente cantavaØ bem*); (iv) um tipo de expressão que chamamos de *referência zero* (RZ), o qual se manifesta quando a oração não expressa o argumento-sujeito nem por desinência explícita (morfema zero) nem por SN (como em *Ø cantavaØ bem*, no sentido de *a gente cantava bem*), tipo opaco de construção, já que se caracteriza por uma correspondência entre um argumento no NR e nenhuma unidade nos demais níveis.³

É sobre a transparência desse sistema que terá reflexo a mudança *nós/a gente*. Mais precisamente, essa mudança incide sobre um domínio particular desse sistema. Primeiramente, deve-se notar que a mudança *nós/a gente* atuará sobre o domínio circunscrito à expressão do argumento-sujeito de *primeira pessoa do plural*. Ademais, observe-se que a expressão do argumento-sujeito por SN (seja na primeira pessoa do plural, seja nas demais pessoas) inclui expressão mediante construção nominal (como em *meus velhos amigos e eu cantávamos bem*) e mediante pronome (*nós cantávamos bem.*). A mudança *nós/a gente* afetará, naturalmente, apenas a parte do sistema que envolve a expressão pronominal. A forma *a gente* (como veremos adiante) entra no quadro pronominal do PB como variante do pronome *nós*, competindo com *nós* na codificação do significado de primeira pessoa do plural em si, não como forma alternativa a significações lexicais de expressões nominais (como a significação de *meus velhos amigos*). Desse modo, a mudança *nós/a gente* estará relacionada com a parte do sistema de expressão do argumento-sujeito que envolve a expressão *pronominal* (não a expressão *nominal*), combinada, então, com a expressão *desinencial* do argumento-sujeito. Enfim, pode-se dizer que a mudança *nós/a gente* incidirá precisamente sobre o *sistema de expressão pronominal e desinencial do argumento-sujeito de primeira pessoa do plural*. No restante do artigo, é desse sistema que estamos tratando, entendendo-o como uma parte (um subsistema) da gramática do PB. Seguiremos nos referindo a esse sistema, de forma mais sintética, como *sistema EAS-1PP* (como abreviação de *sistema de expressão do argumento-sujeito de primeira pessoa do plural*).

Especificamente, o sistema EAS-1PP é composto pelo conjunto de todas as possíveis combinações entre: expressão/não expressão dos pronomes *nós/a gente* e a forma verbal correspondente a cada um desses pronomes em cada um dos nove tempos verbais do PB (os seis do indicativo e os três do subjuntivo), tendo em vista que, entre essas formas verbais, umas têm desinência número-pessoal explícita, e outras não. O Quadro 1 apresenta esse sistema:

³ Não incluímos, em nosso estudo, a análise do fenômeno da simplificação da morfologia verbal flexional que leva a construções como *nós canta*. Também não incluímos na pesquisa o uso da forma pronominal *a gente* combinada com a desinência verbal de primeira pessoa do plural (*a gente cantamos*). Essas exclusões foram estabelecidas porque, após análise prévia, constatamos que esses fenômenos se verificam em nosso corpus muito raramente (não oferecendo quantidades de dados minimamente suficientes para análise) e apenas em casos muito particulares.

Quadro 1 – Sistema de expressão pronominal e desinencial do argumento-sujeito de 1ª pessoa do plural⁴

Pro-nome	Expres-são pronominal	Ind. pres.		Ind. pret. imp.		Ind. pret. perf.		Ind. pret. m.q.perf.		Ind. fut. pres.		Ind. fut. pret.		Subj. pres.		Subj. pret. imp.		Subj. fut.		Totais
		DE	TR	DE	TR	DE	TR	DE	TR	DE	TR	DE	TR	DE	TR	DE	TR	DE	TR	
nós	+	+	RC	+	RC	+	RC	+	RC	+	RC	+	RC	+	RC	+	RC	+	RC	9
		nós cantamos		nós cantávamos		nós cantamos		nós cantáramos		nós cantaremos		nós cantaríamos		nós cantemos		nós cantássemos		nós cantarmos		
	-	+	RS	+	RS	+	RS	+	RS	+	RS	+	RS	+	RS	+	RS	+	RS	9
		ø cantamos		ø cantávamos		ø cantamos		ø cantáramos		ø cantare-mos		ø cantaríamos		ø cantemos		ø cantássemos		ø cantarmos		
a gente	+	-	RS	-	RS	+	RC	-	RS	-	RS	-	RS	-	RS	-	RS	-	RS	9
		a gente cantaø		a gente cantavaø		a gente cantou		a gente cantaraø		a gente cantaráø		a gente cantariaø		a gente canteø		a gente cantasseø		a gente cantarø		
	-	-	RZ	-	RZ	+	RS	-	RZ	-	RZ	-	RZ	-	RZ	-	RZ	-	RZ	9
		ø cantaø		ø cantavaø		ø cantou		ø cantaraø		ø cantaráø		ø cantariaø		ø canteø		ø cantasseø		ø cantarø		
Totais		4		4		4		4		4		4		4		4		4		36

Fonte: Elaboração própria.

Conforme mostra o quadro, esse sistema é composto por 36 possibilidades de expressão do argumento-sujeito: cada um dos dois pronomes de 1PP expresso, combinado com cada um dos nove tempos verbais, o que resulta em 18 combinações ($2 \times 9 = 18$), somadas às respectivas versões sem expressão pronominal de cada uma dessas 18 combinações, resultando nas 36 combinações possíveis ($18 + 18 = 36$).

Também de acordo com o quadro, as formas verbais relativas ao pronome *nós*, em *todos* os tempos, *contêm* desinência verbal número-pessoal explícita. Já em relação à forma *a gente*, a situação é (quase) completamente inversa: nesse caso, as respectivas formas verbais, em *quase todos* os tempos, *não contêm* desinência número-pessoal explícita (sendo essa flexão representada por morfema zero), com exceção apenas do pretérito perfeito do indicativo.

Essa diferença é crucial para fazer com que a mudança *nós/a gente* acarrete alteração na transparência do sistema EAS-1PP. De acordo com esse sistema, o uso de *nós* (em vez de *a gente*) combinado com a opção por expressão pronominal (em vez de não expressão) implica sempre expressão *opaca* do argumento-sujeito, já que as formas verbais relativas a esse pronome, em todos os tempos, *contêm* desinência número pessoal explícita – a expressão do pronome somada à existência de desinência explícita resulta sempre em construção *opaca*, mediante RC (*nós cantávamos*). Inversamente, no caso do emprego de *a gente* associado à opção por expressão pronominal, o sistema passa a favorecer amplamente a expressão *transparente* do argumento-sujeito, já que as respectivas formas verbais, em quase todos os tempos, *não têm* desinência explícita – a expressão pronominal somada à ausência de desinência explícita resulta em construção *transparente*, mediante RS (*a gente cantavaØ*). Dessa

⁴ No Quadro 1, “DE” significa desinência verbal número-pessoal explícita e “TR”, o tipo de referência realizada pela expressão do argumento-sujeito (RC, RS ou RZ). Os sinais “+” e “-”, na coluna de cada tempo verbal, indicam presença e ausência de desinência número-pessoal explícita nas formas dos verbos regulares do PB padrão, no cruzamento entre o tempo considerado e os pronomes de 1ª pessoa, de acordo com Câmara Júnior (1989) e Bechara (2009). Nesse sentido, o sistema, aqui considerado, de expressão do argumento-sujeito refere-se à porção da gramática do PB que envolve o domínio regular de conjugação verbal da língua, em sua variedade padrão.

forma, pode-se concluir que, no âmbito da expressão pronominal do argumento-sujeito (em oposição ao domínio da não expressão pronominal), o sistema EAS-1PP, a partir de um gradativo aumento diacrônico nos índices de uso de *a gente* em detrimento de *nós*, terá seu grau de transparência aumentado, isto é, o sistema passa a favorecer o emprego de construções transparentes em lugar de opacas.

Por outro lado, no domínio da não-expressão pronominal do argumento-sujeito, o sistema opera de forma oposta. A não expressão do pronome *nós* ocasiona construções *transparentes* em todos os tempos verbais, por meio de RS ($\emptyset$ *cantávamos*), enquanto a não expressão de *a gente* propicia, em quase todos os tempos (com exceção do pretérito perfeito do indicativo), construções *opacas*, mediante RZ ($\emptyset$ *cantava* $\emptyset$, no sentido de *a gente cantava*). Ou seja, conclui-se que, no domínio da não expressão pronominal, o sistema torna-se mais opaco, em razão da mudança *nós/a gente*.

Como indicam essas constatações, o aumento diacrônico do emprego da forma *a gente* em lugar de *nós* acarretará, de fato, alteração na transparência do sistema em questão.⁵ Como demonstraremos na seção 4 adiante, o impacto dessa alteração dependerá da proporção em que o pronome *nós* vai dando lugar ao pronome *a gente*, o que iremos calcular em termos dos percentuais de uso dessas duas formas ao longo do tempo.

Tais constatações evidenciam ainda que a variação de transparência do sistema irá depender também de outra mudança em curso no PB. Observe-se que o efeito da mudança *nós/a gente* sobre a transparência do sistema EAS-1PP depende do cruzamento dessa mudança com a opcionalidade entre expressão e não expressão pronominal do argumento-sujeito: ocorre aumento de transparência, no que tange à associação entre crescimento no índice de *a gente* (em detrimento de *nós*) e expressão pronominal do argumento-sujeito; por outro lado, ocorre diminuição de transparência, na esfera da combinação entre crescimento do índice de *a gente* e não expressão pronominal. Dessa forma, o aumento e a diminuição de transparência dependem também das taxas de expressão e não expressão pronominal e, por isso, serão afetados pela mudança, amplamente atestada no PB, segundo a qual se verifica aumento diacrônico nos índices de expressão pronominal em detrimento da não-expressão (mudança tratada, na gramática gerativa, em termos do percurso do PB na direção de língua de sujeito nulo para língua de sujeito não nulo (ou de sujeito parcialmente nulo).

Em Guerra (2021), focalizamos justamente o impacto particular que a mudança entre expressão e não expressão pronominal acarreta sobre a diacronia da transparência do PB. Já no presente artigo, optamos por não contemplar os efeitos da alteração diacrônica das taxas de expressão/não expressão pronominal, com o objetivo de focalizar, especificamente, o impacto da mudança *nós/a gente* sobre a transparência do PB. Na seção 4, então, procedemos a esse cálculo, sintetizando, antes disso, na seção 3, alguns trabalhos que tratam da mudança *nós/a gente*, bem como dados quantitativos que levantamos a esse respeito.

⁵ Considerando que a mudança *nós/a gente* afeta a transparência do PB e considerando (como mostrado na seção seguinte) que essa mudança tem a ver com o processo de gramaticalização da forma *a gente*, pode-se ver que, em uma língua, processos de mudança linguística, como a gramaticalização, além de promoverem outras alterações na língua, podem ter influência sobre seu grau de transparência.

3 O aumento diacrônico do índice de uso do pronome *a gente* em lugar de *nós*

A alternância entre o uso das formas *nós* e *a gente* para expressão de 1PP é um fenômeno que pode ser tomado como característico do PB. Conforme afirma Castilho (2010), o quadro de pronomes pessoais do PB, na atualidade, inclui as duas formas para expressão de 1PP (embora diversos manuais e materiais didáticos ainda não reconheçam essa realidade). De fato, vários estudos, concentrados principalmente nas áreas de Teoria da Gramaticalização, Variação e Mudança Linguística e Sociolinguística Histórica, discutem diferentes aspectos envolvendo a variação entre as duas formas e a inserção de *a gente* no quadro pronominal do PB.

Omena e Braga (1996) explicam que o substantivo latino *gens, gentis* dá origem, na língua portuguesa, ao termo *gente*, que nomeia, de modo coletivo, indeterminador, mais ou menos geral, um conjunto de seres humanos, identificados entre si por objetivos, ideias, qualidades, nacionalidade ou posição. De acordo com as autoras, no decorrer da história do PB, a palavra *gente*, acompanhada do artigo *a*, passa a incorporar a referência à primeira pessoa do discurso, especialmente à primeira do plural, tornando-se uma forma alternativa ao pronome *nós*.⁶

Mais especificamente, as autoras defendem que a entrada da forma *a gente* no quadro pronominal e sua concorrência com o pronome *nós* decorrem de um processo de gramaticalização daquela forma.⁷ Em termos gerais, pode-se dizer que o processo de gramaticalização constitui um tipo de mudança linguística em que um item lexical torna-se gramatical ou um item já gramatical torna-se ainda mais gramatical. Omena e Braga (1996) procuram evidenciar a ocorrência da gramaticalização de *a gente* pela aplicação dos critérios de gramaticalização propostos por Hopper (1991): *estratificação, divergência, especialização, persistência e descategorização*.

Segundo as autoras, a estratificação significa que, numa língua, num dado domínio funcional (como o da expressão de 1PP), podem coexistir recursos de expressão linguística variantes, isto é, camadas (do inglês *layering*) alternantes, o que pode acontecer porque a emergência de novas camadas não implica a eliminação das já existentes. Como esclarecem as autoras, essas camadas podem corresponder a itens lexicais ou tipos de construção, exercem funções similares ou idênticas e podem ter significados ligeiramente diferentes ou ser reconhecidas como alternativas estilísticas. Para as autoras, o comportamento do item *a gente* é um legítimo caso de *layering*, uma vez que o item origina a forma *a gente*, que, no domínio da expressão de 1PP, passa a competir com a forma *nós*, ambas exercendo funções similares.

A divergência, como sintetizam Omena e Braga (1996), é a preservação da forma lexical que origina um processo de gramaticalização, a qual pode permanecer como item lexical autônomo e sofrer, ou não, as mudanças que atingem a classe a que pertence. Observam as autoras que o item *gente* permanece no PB como item lexical autônomo, sem sofrer mudanças. A forma inovadora é a que passa por alterações. Ela cristaliza a relação determinante-determinado, fixando-se com o artigo definido, sem admitir outras especificações, e passa a ter mobilidade mais restrita, identificando-se com os pronomes.

⁶ Omena e Braga (1996) afirmam que o significado da expressão *a gente* incorpora também referência à 1ª pessoa do singular, sendo alternativa ao pronome *eu*, o que, porém, salientam, já constitui um uso menos frequente e bastante impreciso.

⁷ Lopes (1999, 2004) também considera que a forma *a gente* teria passado por um processo de gramaticalização e afirma que o processo teria se implementado principalmente durante os séculos XVIII e XIX.

Quanto ao critério da especialização, que remete ao estreitamento de escolhas sofridas pelas construções alternantes (a forma em gramaticalização e seu(s) concorrente(s)), Omena e Braga (1996) constatarem que a forma *a gente*, com o andar de sua gramaticalização, predomina sobre o pronome *nós*. Avaliando a distribuição dessas formas pelas funções sintáticas, as autoras detectam que a forma *a gente* é majoritária nas funções de sujeito, complemento e adjunto adverbial, prevalecendo o pronome *nós* na função de adjunto adnominal.

Ainda de acordo com Omena e Braga (1996), o critério da persistência refere-se à manutenção de traços semânticos da forma fonte na forma que sofre gramaticalização, quando esta não se encontra em estágios muito avançados de gramaticalização, traços esses que explicam, muitas vezes, certas restrições que a forma gramaticalizada experimenta em sua distribuição. Segundo as autoras, o traço de coletividade do termo *gente* contribui para uma referência indeterminadora verificada em certas ocorrências da forma pronominal *a gente*, o que gera determinadas restrições em seu uso: enquanto o pronome *nós* pode ser modificado por quantificadores, numerais, especificadores, esse mesmo tipo de procedimento não é possível com a forma *a gente* – por exemplo, *todos, cada um de, nenhum de* podem modificar *nós*, mas não *a gente*.

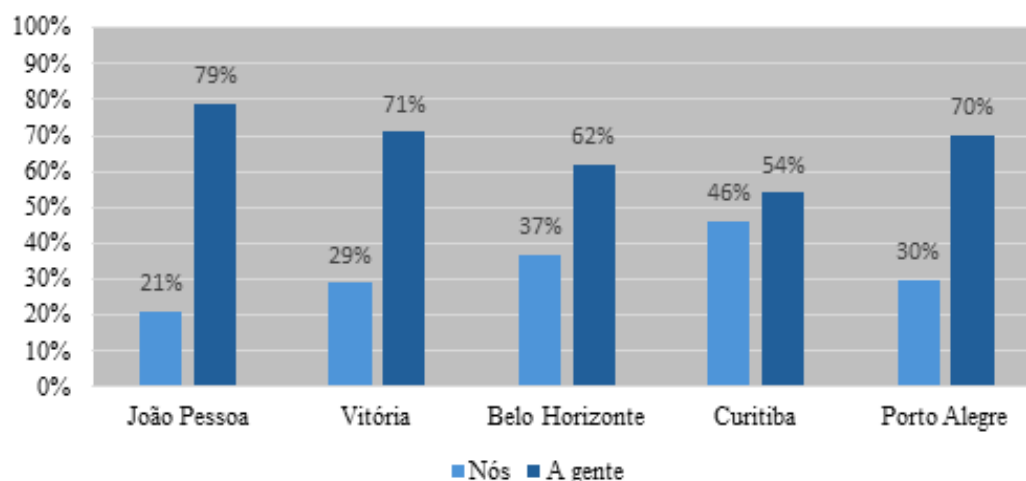
Já o critério da descategorização determina que os itens em gramaticalização normalmente perdem ou neutralizam marcadores morfológicos e traços sintáticos de categorias plenas e assumem aspectos de categorias secundárias. Como apontam as autoras, o termo *gente* conserva a mobilidade de colocação dos substantivos, admite flexão de número, derivação de grau, pode ser modificada por quantificadores, determinantes, possessivos etc., enquanto a forma pronominal *a gente* não é compatível com esses procedimentos; além disso, embora mantenha com o verbo a mesma relação que o substantivo *gente*, concordando com formas verbais de terceira pessoa do singular, a forma *a gente* exibe a variação *a gente vamos*, que revela maior vínculo com o quadro pronominal, por concordar com a primeira pessoa do plural.

Enfim, com base nos critérios de Hopper (1991), Omena e Braga (1996) demonstram que a forma *a gente* gramaticaliza-se, processo responsável pela introdução de *a gente* no quadro pronominal do PB e por sua concorrência com o pronome *nós*. Nesse contexto, passa a se instaurar no PB a mudança diacrônica abordada neste artigo referente ao aumento da frequência de uso de *a gente* em detrimento do emprego de *nós*.

Vários estudos de ordem sincrônica, realizados com dados relativos ao final do século XX e/ou início do século XXI, fornecem resultados que evidenciam a mudança *nós/a gente*. Rubio e Gonçalves (2010) analisam dados do Banco de Dados Iboruna (Gonçalves, 2007), que reúne amostras de língua falada de diferentes gêneros textuais, recolhidas de informantes da região noroeste do estado de São Paulo, entre os anos de 2004 e 2006. Examinando o uso de *nós* e *a gente* na função sintática de sujeito, apuram o emprego dessas formas nos percentuais de 25,4% e 74,6%, respectivamente. Seara (2000), averiguando entrevistas de informantes da cidade de Florianópolis, Santa Catarina, coletadas junto ao Projeto VARSUL (Projeto de Variação Linguística da Região Sul do Brasil) na década de 1990, e também observando a função de sujeito, constata os índices de 28% de *nós* e 72% de *a gente*. Santana (2014), por sua vez, descreve o uso de *nós* e *a gente*, também na função de sujeito, no português popular de Salvador, Bahia, adotando como *corpus* o Programa de Estudos do Português Popular Falado em Salvador, composto por entrevistas do tipo diálogo entre informante e documentador, em situação de fala semiespontânea, levantadas entre 1998 e 2000. O autor encontra as taxas de 24% e 76% para as formas *nós* e *a gente*, respectivamente.

Vianna e Lopes (2012) compilam resultados de diversos trabalhos (que lançam mão de *corpora* variados, representativos do fim do século XX e começo do XXI), a fim de comparar os empregos de *nós* e *a gente* em distintas capitais brasileiras. Os resultados indicam uso predominante de *a gente* em todas as capitais observadas. Os dados reunidos pelas autoras constam no gráfico que elaboramos abaixo:

Gráfico 1 – Percentuais de *nós* e *a gente* em diferentes capitais brasileiras, segundo dados de Vianna e Lopes (2012)



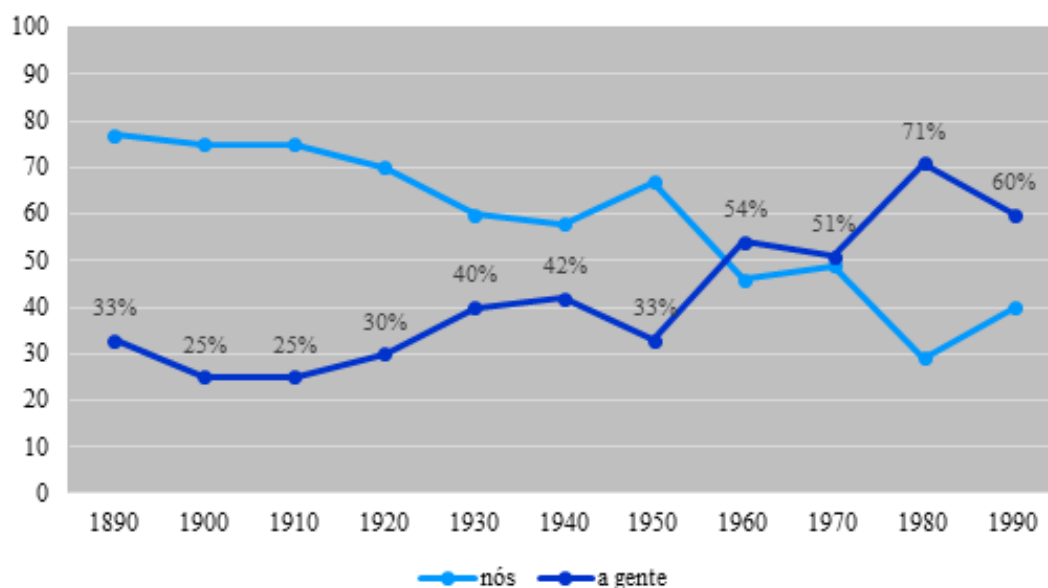
Fonte: Elaboração própria.

Pacheco (2018) procede a um levantamento bibliográfico similar ao de Vianna e Lopes (2012), reunindo quantidade, inclusive, muito extensa de trabalhos variacionistas, realizados desde a década de 1980, que comparam os percentuais de uso de *nós* e *a gente*. A autora levanta estudos que descrevem o PB de várias regiões do Brasil (incluindo tanto diferentes estados quanto diferentes partes de mesmos estados). Segundo ela, referindo-se ao final do século XX e início do XXI, o PB, de uma forma geral, privilegia o uso de *a gente* em detrimento de *nós*. Até em comunidades mais isoladas, a forma *a gente* é majoritária.

Levando em conta que, em séculos passados do PB, não se verifica emprego de *a gente* como forma pronominal, ou esse uso é pouco recorrente, e considerando, conforme apontam os estudos citados, que, atualmente, a frequência de uso da forma pronominal *a gente* supera os índices de *nós*, pode-se depreender que o PB, de fato, passou pela ampliação da frequência de emprego de *a gente* em prejuízo de *nós*.

Com efeito, algumas pesquisas levantam dados diacrônicos que indicam essa mudança. Borges (2004, 2019), por exemplo, analisa a inserção e a propagação da forma pronominal *a gente* em onze peças de teatro escritas por autores gaúchos, no período de 1896 a 1995. Segundo sua investigação, nesse contexto, a partir da década de 1960, a forma *a gente* na função de sujeito passa a competir efetivamente com o pronome-sujeito *nós* e, dessa década em diante, *a gente* apresenta percentual superior a 50%, superando o pronome *nós*. Para efeito ilustrativo, os dados do autor são apresentados no gráfico abaixo:

Gráfico 2 – Evolução diacrônica do uso de *nós* e *a gente*, segundo dados de Borges (2004, 2019)



Fonte: Elaboração própria, mediante adaptação de Borges (2004, 2019).

Lopes (1999) estuda a variação linguística envolvendo as formas *nós* e *a gente*, examinando entrevistas de informantes da cidade do Rio de Janeiro, junto ao Projeto NURC-RJ (Projeto da Norma Urbana Oral Culta do Rio de Janeiro), realizadas na década de 70 e na década de 90 (com os mesmos informantes entrevistados na década de 70 e com informantes diferentes). Como resultado, a autora identifica, na amostra relativa à década de 70, os percentuais de 42% e 58% de *a gente* e de *nós*, respectivamente; na amostra de 1990, com os mesmos informantes recontactados, são apurados os índices de 54% e 46% de *a gente* e de *nós*, e, com informantes diferentes, os percentuais de 75% e de 25% de *a gente* e de *nós*. Segundo a autora, o resultado sinaliza uma aceleração da substituição de *nós* por *a gente*, nesse contexto, no período em estudo.

Também nas pesquisas que temos realizado sobre a diacronia da transparência do PB, levantamos dados diacrônicos quantitativos acerca do emprego das formas *nós* e *a gente* – em Guerra (2017), apresentamos esses dados, que retomamos aqui, para os fins do presente artigo.⁸ Nossos dados sobre as formas *nós* e *a gente* foram coletados em um *corpus* que montamos para essas pesquisas, o qual é composto por peças teatrais brasileiras produzidas do século XIX ao XXI.

Assumindo que a mudança *nós/a gente* (bem como outras mudanças relevantes para estudo da transparência) pode ser mais bem detectada em gêneros discursivos típicos da oralidade e considerando a inviabilidade de reunir material de língua falada de períodos mais antigos, escolhemos *peças de teatro* para formação do *corpus*, visando examinar textos mais próximos da oralidade, tendo em vista que peças teatrais costumam retratar a língua falada de sua época. O espaço temporal recoberto pelo *corpus* – século XIX ao XXI – deveu-se

⁸ De acordo com nossa pesquisa bibliográfica, não é extenso o corpo de estudos que disponibilizam dados quantitativos diacrônicos sobre a mudança *nós/a gente*. Nesse sentido, nosso trabalho, ao fornecer dados diacrônicos sobre o uso dessas formas, procura, entre seus demais objetivos, representar uma contribuição para o registro da mudança *nós/a gente*.

a limitações naturais de acesso a peças produzidas antes do século XIX, decorrendo ainda do entendimento de que o intervalo temporal selecionado contempla, de fato, um período considerável da história do PB. Com base em procedimento metodológico comumente empregado em análises diacrônicas (ver, por exemplo, procedimentos adotados junto ao *Projeto para a História do Português Brasileiro* (Castilho, 2019), compusemos o *corpus* dividindo cada século em primeira e segunda metade, o que proporcionou cinco recortes sincrônicos para serem comparados diacronicamente: primeira e segunda metade do século XIX, primeira e segunda metade do século XX e primeira metade do século XXI, períodos que chamaremos aqui, respectivamente, de 19-1, 19-2, 20-1, 20-2 e 21-1.

O *corpus* foi constituído por 55 peças teatrais, distribuídas entre as cinco sincronias, compreendendo, por exemplo, peças como as seguintes: sincronia 19-1: *Os Irmãos das Almas* (de 1844, escrita por Martins Pena), *Angélica e Firmino* (1845, Araújo Porto Alegre); sincronia 19-2: *A Torre em Concurso* (1863, Joaquim Manuel de Macedo), *Cahio o Ministerio* (1883, França Júnior); sincronia 20-1: *Ministro do Supremo* (1921, Armando Gonzaga), *Álbum de Família* (1946, Nelson Rodrigues); sincronia 20-2: *A Invasão* (1962, Dias Gomes), *Ópera do Malandro* (1978, Chico Buarque); sincronia 21-1: *Síndromes* (2003, Maria Carmem Barbosa e Miguel Falabella), *Ilustríssimo Filho da Mãe* (2008, Leilah Assumpção). A seleção de peças para composição do *corpus* obedeceu a uma série de condições, dentre as quais podem ser destacadas: a seleção somente de obras publicadas oficialmente; a escolha da primeira edição da obra ou de versão próxima da primeira edição; o descarte de obras que retratavam um tempo diferente daquele em que foram produzidas; a exclusão de peças e trechos escritos em verso; a coleta de dados a partir, apenas, das falas dos personagens (com exclusão de outras partes, como descrição de cenário, de personagens, estruturação da obra etc.) por se tratar da parte do gênero que estaria mais próxima da oralidade do período em questão.⁹

A partir desse material, estabelecemos a meta de levantar, para cada uma das sincronias, 100 ocorrências de expressão pronominal de argumento-sujeito de 1PP, para identificarmos, em cada período, a frequência em que essa expressão é realizada por *nós* e por *a gente* (apenas no material da sincronia 19-1, a meta de 100 ocorrências não foi obtida, amostra na qual foi possível encontrar 58 ocorrências). Para levantamento das ocorrências, foram considerados os seguintes critérios:

- i) estabelecemos o limite de 25 ocorrências por peça (para diversificação das fontes dos dados, restrição não aplicada à sincronia 19-1, pela indisponibilidade de *corpus* mais extenso representativo desse período);¹⁰
- ii) coletamos apenas ocorrências em que o argumento-sujeito de 1PP é explicitamente expresso por *nós* ou *a gente* (sujeito pronominal não nulo), excluindo ocorrências de não expressão pronominal (sujeito pronominal nulo);¹¹

⁹ O conjunto total de peças do *corpus* e a lista completa de critérios para seleção das obras podem ser encontrados em Guerra (2017).

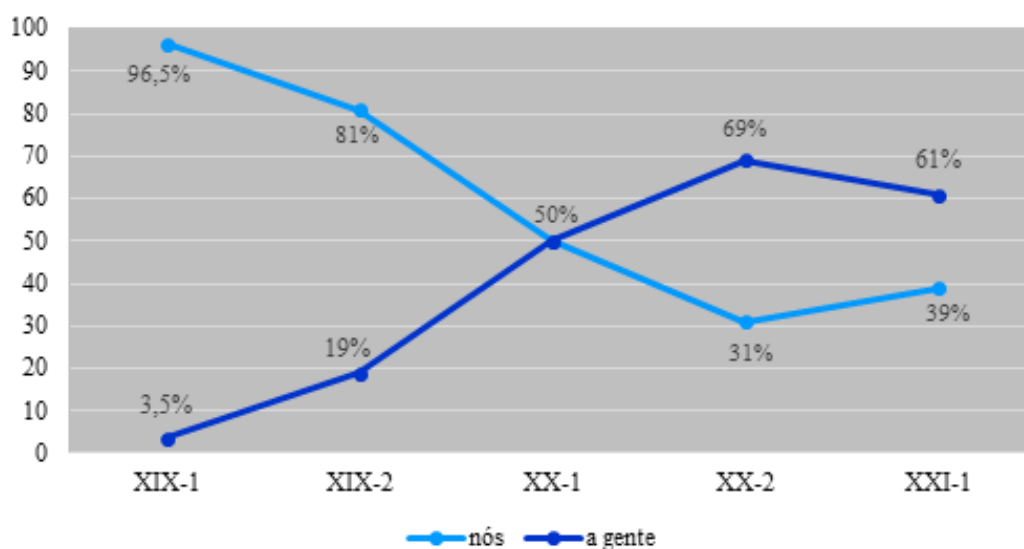
¹⁰ Embora, a princípio, o limite de 25 ocorrências por peça não tenha sido aplicado à sincronia 19-1, nenhuma peça dessa sincronia conteve mais do que 25 dados. Assim, nesse período, houve naturalmente diversificação de peças no levantamento das ocorrências.

¹¹ Na coleta dessas 25 ocorrências, foram selecionadas, de cada peça, as primeiras 25 ocorrências de pronome explícito de 1PP.

- iii) para as ocorrências de *a gente*, recolhemos somente os casos em que a forma atua como pronome de 1PP, desconsiderando casos em que a forma opera como item não-gramaticalizado, composto pelo substantivo *gente*, antecedido do artigo *a*;
- iv) para as ocorrências de *nós*, excluimos casos em que o pronome é modificado por quantificadores, como em *nós todos*, *nós três*, *cada um de nós*, entendendo que tais construções não admitem variação entre *nós* e *a gente*.

O Gráfico 3 mostra, então, os resultados de nosso levantamento de dados:

Gráfico 3 – Evolução diacrônica do uso de *nós* e *a gente*



Fonte: Elaboração própria.

Como se pode ver, nossos resultados corroboram e especificam o reconhecimento do fenômeno do aumento percentual de uso da forma *a gente* em detrimento do pronome *nós*.¹² Apenas no intervalo entre a sincronia 20-2 para a 21-1, há queda (relativamente pequena) do emprego de *a gente*, em favor do aumento do uso de *nós*. Em todo o restante do período, verifica-se aumento significativo do percentual de *a gente*, com destaque para o espaço entre 19-2 e 20-1, em que o índice de *a gente* sobre 31 pontos percentuais, empatando com a taxa de *nós* em 50%. A partir daí, e até o final do tempo observado, a frequência de *a gente* é superior à de *nós*. Na seção seguinte, usaremos os índices dispostos no Gráfico 3 (que retrata apenas pronomes sujeitos explícitos) para cálculo do impacto da mudança *nós/a gente* sobre o grau de transparência do PB.

Finalizando a presente seção e a título de ilustração e registo de dados, os exemplos abaixo fornecem uma ocorrência de *nós* e uma de *a gente* de cada um dos recortes sincrônicos analisados:

¹² Além de corroborar a ocorrência da mudança *nós/a gente*, nosso trabalho complementa outros estudos que oferecem dados diacrônicos sobre a mudança, ao descrevê-la com base em um *corpus* distinto do usado por pesquisas prévias e que contempla período diferente, abrangendo desde a primeira metade do século XIX até início do século XXI – note-se, por exemplo, o estudo de Borges (2004, 2019), que recobre o período de 1890 a 1990.

- (6) LUIZA – Tenho tanta repugnância de servir-me d'este dinheiro.
 JORGE – Porque?
 LUIZA – Dinheiro de esmolas que pedes para as almas!
 JORGE – E então o que tem isso? – é verdade que peço para as almas; mas **nós** também não temos alma? (*Os Irmãos das Almas*; Pena, 1847, p. 3).
- (7) CÂNDIDA – Homem com vontade de ferro, que não muda ainda mesmo doente! Ao menos o mano Gustavo não é assim. Nos momentos de bom humor cede a tudo, principalmente se **a gente** lhe diz que ele é bonito. Faz tudo quanto se lhe pede. (*Angélica e Firmino* [1845]; Cafezeiro; Guerra, 1997, p. 86).
- (8) CRESPIM – [...] Pascoal, **nós** somos dous ingleses, tão ingleses como a própria lama de Londres.
 PASCOAL – Mas o bôllo!... o bôllo!... o bôllo!... (*A Torre em Concurso*; Macedo, 1863, p. 100).
- (9) COELHO – Mas, minha filha, tu pensas que em politica **a gente** sóbe unicamente por seus bellos olhos? Não sou rico, já estou velho, não tenho pai alcaide, se deixar fugir as ocasiões, quando serei ministro?
 D. BARBARA – E para que você quer ser ministro, seu Chico?
 COELHO – Ora, tens ás vezes certas perguntas? Para que? Para governar, para fazer o que os outros fazem (*Cahio o Ministério*; França Júnior, 1883, p. 31).
- (10) MANECO – O Vicente é meu velho camarada. **Nós** brigámos ás vezes, mas eu gosto delle. (*Ministro do Supremo* [1921]; Gonzaga, 1940, p. 71).
- (11) GLÓRIA (como que falando para si mesma) – Ficou sentida – tão sentida! – por que eu contei que...
 GUILHERME (desesperado) – Contou o quê?
 GLÓRIA – ... que tôda vez que **a gente** se beijava, eu fechava os olhos e via direitinho a fisionomia de papai (*Album de Família*; Rodrigues, 1946, p. 86).
- (12) MALÚ – O senhor acha que **nós** não devíamos ter tomado parte na invasão?
 JUSTINO – Acho é que a gente nunca devia ter saído de nossa terra. Nunca. (*A Invasão*; Gomes, 1962, p. 14).
- (13) DURAN – Que debutante? Não, hoje não chegou aqui nenhuma debutante. [...]. (Toca a campainha) Eu vou ter que desligar, Chaves, **a gente** se fala depois. (Toca a campainha; Duran desliga o telefone e berra) Entra, porra! (*Ópera do Malandro*; Buarque, 1978, p. 28).
- (14) PRYSCILLA – Eu quero ir pra casa. Tou com fome, mãe!
 JANDIRA – **Nós** já vamos. (*Síndromes*; Barbosa; Falabella, 2003, p. 287).
- (15) JORGE – E o que tanto fala do Caio? O que é isso? Vocês têm se visto?
 LENA – Desde que ele se separou.
 JORGE – (surpreso) Como assim? Ele não me disse nada...
 LENA – O Caio é uma pessoa muito sensível. **A gente** tem muita afinidade. (*Ilustríssimo Filho da Mãe* [2008]; Assumpção, 2010, p. 593).

4 Cálculo do impacto da mudança nós/a gente sobre a transparência do PB

Como estamos considerando neste trabalho, o sistema EAS-1PP pode ser caracterizado, em cada recorte sincrônico da língua, como tendo um dado *grau de transparência*. Suponha-se que o argumento-sujeito de 1PP fosse expresso pronominalmente apenas pela forma *nós* (sem variação com *a gente*) e que a expressão pronominal fosse obrigatória (sem variação com a não-expressão pronominal). Num cenário assim, o sistema EAS-1PP propiciaria a expressão do argumento-sujeito sempre por RC, pela combinação entre o pronome obrigatório e a desinência número-pessoal, que é explícita em todas as formas verbais relativas ao pronome *nós*. O sistema seria caracterizado por ser completamente opaco. É esse tipo de característica que observamos para identificar o grau de transparência do sistema. No caso hipotético em questão, diríamos que o sistema teria um grau de transparência de 0% (ou um grau de opacidade de 100%).

Diferentemente, porém, desse cenário, o sistema EAS-1PP, como argumentado na seção 2, proporciona, em certa medida, construções opacas e, em certa medida, construções transparentes. Seu grau de transparência será definido, então, pela proporção em que favorece a expressão transparente do argumento-sujeito.

Conforme explicado também na seção 2, a transparência do sistema é determinada por três fatores (tendo em vista todo o recorte temporal aqui em pauta): (i) propensão da língua para tratar o argumento-sujeito de 1PP em termos do pronome *nós* e em termos do pronome *a gente* (variação *nós/a gente*); (ii) propensão da língua para expressar e para não expressar o argumento-sujeito por pronome (variação expressão/não expressão pronominal); (iii) presença e ausência de desinência número-pessoal explícita nas formas verbais relativas a esses pronomes, nos nove tempos verbais do PB.

A interação desses fatores resulta nas 36 possibilidades de expressão do argumento-sujeito que compõem o sistema EAS-1PP, como mostra o Quadro 1 anterior, sintetizado abaixo no Quadro 2:

Quadro 2 – Sistema de expressão pronominal e desinencial do argumento-sujeito de 1ª pessoa do plural

Pronome	Expressão pronominal	Ind. pres.	Ind. pret. imp.	Ind. pret. perf.	Ind. pret. m.q.perf.	Ind. fut. pres.	Ind. fut. pret.	Subj. pres.	Subj. pret. imp.	Subj. fut.	Totais
nós	+	RC	RC	RC	RC	RC	RC	RC	RC	RC	9
	-	RS	RS	RS	RS	RS	RS	RS	RS	RS	9
a gente	+	RS	RS	RC	RS	RS	RS	RS	RS	RS	9
	-	RZ	RZ	RS	RZ	RZ	RZ	RZ	RZ	RZ	9
Totais		4	4	4	4	4	4	4	4	4	36

Fonte: Elaboração própria.

O Quadro 2 destaca que, das 36 possibilidades de expressão do argumento-sujeito, exatamente metade (18 cruzamentos) compreende expressão transparente (casos tachados no quadro), e a outra metade inclui combinações opacas. Isso ocorre porque, em qualquer combinação entre pronome e tempo verbal, se a expressão pronominal implica construção opaca (*nós cantávamos*), a respectiva versão sem expressão pronominal será transparente ($\emptyset$ *cantávamos*); em

sentido oposto, se a expressão pronominal implica transparência (*a gente cantava-Ø*), a respectiva versão sem expressão pronominal será opaca (*Ø cantava-Ø*). Desse modo, uma vez que o sistema, naturalmente, comporta uma metade de combinações com expressão pronominal e outra metade com as respectivas versões sem expressão, há mesmo número de combinações transparentes (18) e opacas (18). Dessa forma, poder-se-ia concluir que, em qualquer sincronia em que houvesse opcionalidade entre expressão e não-expressão pronominal (como em todas as sincronias aqui consideradas), o sistema EAS-1PP apresentaria um grau de transparência de 50%.

No entanto, os 36 cruzamentos de fatores, em qualquer das sincronias, não são igualmente fomentados pelo sistema, que irá promover maior emprego de umas ou outras das combinações – o que se verifica porque o sistema, a cada sincronia, irá propiciar a expressão de 1PP em termos de *nós* e de *a gente* em diferentes proporções, bem como irá propiciar a expressão e a não-expressão pronominal também em diferentes medidas. Assim, o aparente equilíbrio entre transparência e opacidade (pela simples igualdade numérica de combinações desses dois tipos) deixa de existir, e o sistema pende para um ou outro desses lados. Ou seja, a transparência do sistema, conforme aqui analisamos, não depende apenas da quantidade de combinações transparentes e opacas que ele disponibiliza, mas depende também de quanto ele favorece o emprego de cada uma das combinações.

Nesse sentido, para calcular o grau de transparência do sistema em cada sincronia, nosso método consiste em identificar em que proporção o sistema promove cada uma das 36 combinações e, a partir dos índices de promoção das combinações transparentes, determinar o grau de transparência do sistema como um todo. Para isso, partimos das frequências de uso de *nós/a gente* e de expressão/não expressão pronominal levantadas em *corpus* de análise, entendendo que essas frequências refletem justamente as propensões da língua – no caso, as propensões do sistema EAS-1PP – para emprego de *nós/a gente* e para expressão/não expressão pronominal e, assim, as propensões do sistema para uso de umas ou outras das combinações disponibilizadas. O Quadro 3 mostra como o cálculo é feito para a sincronia 19-1¹³.

Quadro 3 – Cálculo do grau de transparência do sistema EAS-1PP na sincronia 19-1¹³

Pro-nome	Expres-são pronominal	Ind. pres.	Ind. pret. imp.	Ind. pret. perf.	Ind. pret. m.q.perf.	Ind. fut. pres.	Ind. fut. pret.	Subj. pres.	Subj. pret. imp.	Subj. fut.
nós 96,5%	+	RC	RC	RC	RC	RC	RC	RC	RC	RC
	18%	17,4%	17,4%	17,4%	17,4%	17,4%	17,4%	17,4%	17,4%	17,4%
	-	RS	RS	RS	RS	RS	RS	RS	RS	RS
	82%	79,1%	79,1%	79,1%	79,1%	79,1%	79,1%	79,1%	79,1%	79,1%
a gente 3,5%	+	RS	RS	RC	RS	RS	RS	RS	RS	RS
	18%	0,6%	0,6%	0,6%	0,6%	0,6%	0,6%	0,6%	0,6%	0,6%
	-	RZ	RZ	RS	RZ	RZ	RZ	RZ	RZ	RZ
	82%	2,9%	2,9%	2,9%	2,9%	2,9%	2,9%	2,9%	2,9%	2,9%
Valor total de transp. por tempo verbal		79,7%	79,7%	82%	79,7%	79,7%	79,7%	79,7%	79,7%	79,7%
Grau médio de transp. do sistema		79,9%								

Fonte: Elaboração própria.

¹³ No cálculo explicado no Quadro 3, os percentuais de *nós* e *a gente* (que aparecem na primeira coluna) são aqueles apresentados no Gráfico 3 acima. Esses percentuais foram levantados a partir de 100 ocorrências de cada sincronia (com exceção apenas da sincronia 1, como acima explicado), considerando-se ocorrências apenas de pronome expreso, como forma de focalizar a opcionalidade entre *nós* e *a gente* (sem incluir, nesse levantamento,

Para elaborar o cálculo, primeiramente levamos em conta as frequências de uso de *nós* e *a gente* na sincronia em questão (19-1), as quais, conforme nossos dados expostos na seção anterior, são, respectivamente, 96,5% e 3,5%, valores registrados na primeira coluna do Quadro 3. Assim, uma primeira indicação a ser notada no Quadro 3 é a de que, na sincronia 19-1, o sistema EAS-1PP fomenta o emprego de *nós* e *a gente* na proporção de 96,5% e 3,5% respectivamente; ou seja, em 19-1, a probabilidade de um falante expressar o argumento-sujeito de 1PP em termos de *nós* é de 96,5%, e a probabilidade de o expressar em termos de *a gente* é de 3,5%.

Em seguida, consideramos os percentuais de uso de expressão e não-expressão pronominal de 1PP. Em Guerra (2017), apresentamos resultados de pesquisa diacrônica sobre o PB, que realizamos com o mesmo *corpus* exposto na seção anterior, e apuramos os percentuais de 18% de expressão e 82% de não-expressão pronominal de 1PP, índices que constam na segunda coluna do Quadro 3. Desse modo, uma segunda indicação a ser focalizada no Quadro é a de que, em 19-1, o sistema EAS-1PP promove o uso de expressão pronominal e de não expressão pronominal na proporção de 18% e 82%.

A partir, então, das taxas de emprego de *nós/a gente* e de expressão/não-expressão pronominal, calculamos a proporção em que o sistema EAS-1PP promove o uso de cada uma das 36 combinações do sistema, para cada tempo verbal.

Em qualquer tempo verbal, a proporção em que o sistema promove a combinação entre *nós* e expressão pronominal (*nós cantamos*) será de 17,4% (como se vê, por exemplo, na coluna 3, linha 2 do Quadro 3), valor que representa 18% (disposição do sistema para a expressão pronominal) da tendência de 96,5% (propensão do sistema para emprego de *nós*, em vez de *a gente*). Similarmente, a proporção em que o sistema favorece o emprego da combinação entre *nós* e não expressão pronominal (*Ø cantamos*) é de 79,1% (coluna 3, linha 3), o que significa 82% (inclinação do sistema para a não expressão pronominal) da tendência de 96,5% (tendência para o uso de *nós*). A proporção em que o sistema patrocina a combinação entre *a gente* e expressão pronominal (*a gente canta*) (coluna 3, linha 4) é de 0,6%, o que equivale a 18% (propensão para expressão pronominal) de 3,5% (propensão para o emprego de *a gente*). Já a proporção em que o sistema leva à combinação entre *a gente* e não expressão pronominal (*Ø canta*) (coluna 3, linha 5) é de 2,9%, isto é, 82% (tendência para não expressão pronominal) de 3,5% (inclinação para o uso de *a gente*).

Por exemplo, considere-se um conjunto de 1.000 ocorrências de argumento-sujeito de 1PP. Em relação a esse total, o sistema prevê, para qualquer tempo verbal, a expressão do argumento-sujeito nas seguintes proporções:

- a) em 96,5% das 1.000 ocorrências (965 ocorrências), o argumento-sujeito tenderá a ser tratado em termos do pronome *nós* e, desse conjunto de 965 ocorrências, 18% delas deverão ser realizadas com expressão pronominal (*nós cantamos*), o que resulta em 174 ocorrências, isto é, 17,4% do total inicial de 1.000 ocorrências (coluna 3, linha 2);

também a opcionalidade entre expressão e não expressão de sujeito). Os percentuais dessa segunda opcionalidade (que aparecem na segunda coluna do Quadro 3) foram levantados em trabalho anterior (Guerra, 2017), que apurou essa opcionalidade considerando *nós* e *a gente* em conjunto, isto é, considerando os percentuais de expressão e não expressão de 1PP (desse modo, o Quadro 3 considera um mesmo percentual de expressão para *nós* e *a gente* e um mesmo percentual de não expressão para os dois pronomes). Ainda assim, é necessário separar os dois pronomes no Quadro 3, porque as formas verbais relativas a cada um diferem em termos da presença/ausência de desinência número-pessoal explícita (com *nós*, todas as formas têm desinência explícita; com *a gente*, têm desinência explícita apenas as formas de pretérito perfeito).

- b) no conjunto de 965 ocorrências tratadas em termos de *nós*, 82% provavelmente serão realizadas sem expressão pronominal ($\emptyset$ *cantamos*), o que resulta em 791 casos, representando 79,1% do total inicial de 1.000 ocorrências (coluna 3, linha 3);
- c) em 3,5% das 1.000 ocorrências (35 ocorrências), o argumento-sujeito tenderá a ser tratado em termos do pronome *a gente* e, nesse conjunto de 35 ocorrências, 18% delas deverão ser realizadas com expressão pronominal (*a gente canta*), o que resulta em 6 ocorrências, isto é, 0,6% do total inicial de 1.000 ocorrências (coluna 3, linha 4);
- d) no conjunto de 35 ocorrências tratadas em termos de *a gente*, 82% provavelmente serão realizadas sem expressão pronominal ($\emptyset$ *canta*), resultando em 29 ocorrências, que representam 2,9% do total inicial de 1.000 ocorrências (coluna 3, linha 5).

Uma vez calculada, a cada tempo verbal, a proporção em que o sistema fomenta o uso de cada uma das quatro combinações entre *nós/a gente* e expressão/não expressão pronominal, nosso método inclui identificar o valor total de transparência para cada tempo. No presente do indicativo, são transparentes (RS) a combinação entre tratamento de 1PP como *nós* e não expressão pronominal ($\emptyset$ *cantamos*) e a combinação entre tratamento de 1PP como *a gente* e expressão pronominal (*a gente canta* $\emptyset$). O sistema promove o emprego dessas duas combinações nas proporções de 79,1% (coluna 3, linha 3) e 0,6% (coluna 3, linha 4), respectivamente, o que soma 79,7% (coluna 3, linha 6), valor total de transparência do sistema no presente do indicativo. Isso significa que, quando o argumento-sujeito de 1PP ocorre em uma oração no presente do indicativo, a probabilidade de que ele seja expresso de forma transparente é de 79,7%.¹⁴

O mesmo raciocínio explica o valor de transparência do sistema em cada um dos demais tempos verbais. Como se pode ver no quadro, o valor de quase todos os tempos, com exceção do pretérito perfeito do indicativo, é o mesmo (79,7%), já que todos eles, com a exceção ressaltada, comportam-se do mesmo modo em termos de existência ou não de desinência número-pessoal explícita: nesses tempos, todas as formas verbais relativas ao pronome *nós* têm desinência explícita e todas as formas relativas a *a gente* não têm. Assim, nesses tempos, as combinações transparentes entre *nós/a gente* e expressão/não expressão pronominal são sempre as mesmas: *nós* e não expressão pronominal ($\emptyset$ *cantamos*); *a gente* e expressão pronominal (*a gente canta* $\emptyset$). No pretérito perfeito, tempo em que há desinência explícita tanto na forma relativa a *nós* quanto na forma relativa a *a gente*, as combinações transparentes são: *nós* e não expressão pronominal ($\emptyset$ *cantamos*) (como nos outros tempos); *a gente* e não expressão pronominal ($\emptyset$ *cantou*) (diferentemente dos demais tempos). Por isso, no pretérito perfeito do indicativo, o valor total de transparência é diferente, 82% (coluna 5, linha 6), que é a soma dos valores das combinações transparentes nesse tempo: *nós* e não expressão pronominal (79,1%, coluna 5, linha 3); *a gente* e não expressão pronominal (2,9%, coluna 5, linha 5).

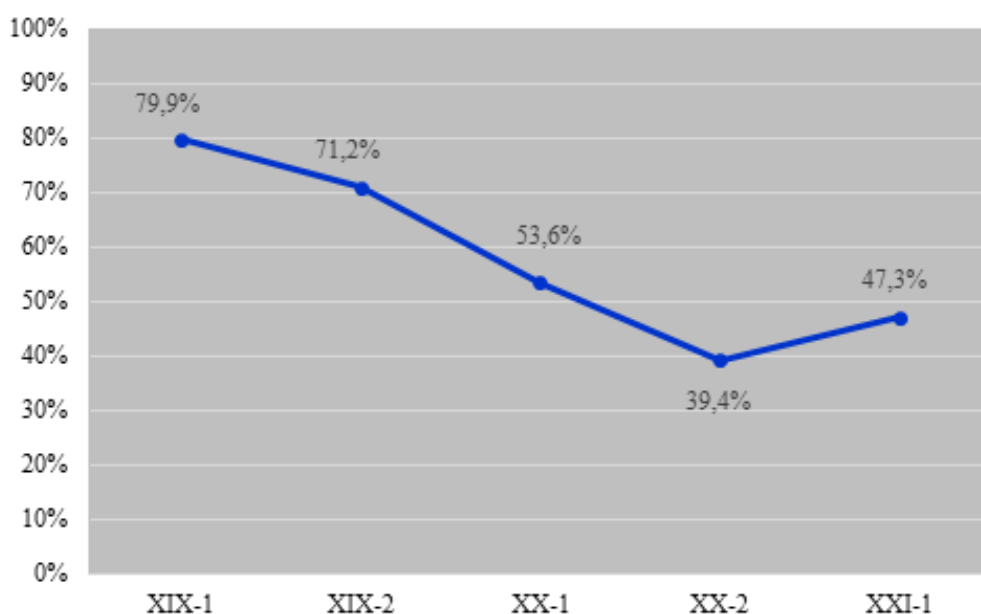
¹⁴ O cálculo expresso no Quadro 3, para cada combinação entre pronome e expressão ou não expressão de sujeito, resulta em percentuais iguais para os diferentes tempos verbais. Mesmo assim, é necessário incluir no cálculo os diferentes tempos separadamente uns dos outros, porque essas diferentes combinações entre pronome e expressão e não expressão nem sempre resultam num mesmo tipo de referência. No domínio do pronome *nós*, em todos os tempos verbais, as referências transparentes (que o artigo focaliza) ocorrem quando o pronome é não expresso. Já no domínio da forma *a gente*, as referências transparentes ocorrem quando o pronome é expresso, em quase todos os tempos verbais, mas não em todos (com exceção do pretérito perfeito). Assim, para se identificar a transparência do sistema, é necessário olhar os tempos separadamente e, ao final, fazer uma média entre os resultados de todos.

Se o valor de transparência de todos os tempos verbais fosse sempre o mesmo, por exemplo 79,7% (como na maioria dos tempos no caso em análise), esse valor já constituiria o próprio grau de transparência do sistema como um todo – em qualquer tempo verbal, a proporção em que o sistema levaria à expressão transparente do argumento-sujeito seria sempre de 79,7%. Todavia, como o valor de transparência não é o mesmo em todos os tempos, nosso procedimento é realizar uma média aritmética simples entre os valores de todos os tempos, o que resulta no valor final de 79,9%, como sendo, então, o grau de transparência do sistema como um todo. Esse valor final significa que, na sincronia 19-1, o sistema EAS-1PP promove o emprego de combinações transparentes de expressão do argumento-sujeito na proporção de 79,9%; isto é, a probabilidade de que o falante expresse esse argumento de forma transparente é de 79,9%.

Para identificar o grau de transparência do sistema nas demais sincronias, aplicamos exatamente o mesmo cálculo explicado acerca da sincronia 19-1. Alteramos apenas os índices de uso dos pronomes *nós* e *a gente* (coluna 1 do Quadro 3), que variam a cada recorte sincrônico, utilizando, nos cálculos, os índices de cada sincronia relatados no Gráfico 3 da seção anterior. No PB, a cada sincronia também variam, como mencionado, as taxas de expressão e não expressão pronominal, porém, nos cálculos dos diferentes períodos, foram usadas sempre as mesmas taxas da sincronia 19-1, justamente para que nossos cálculos captassem o impacto apenas da variação *nós/a gente* sobre a transparência do PB (como é o objetivo deste artigo). Mediante esse procedimento, foram identificados, para as sincronias 19-2, 20-1, 20-2 e 21-1, os seguintes graus de transparência, respectivamente: 71,2%, 53,6%, 39,4%, 47,3%.

A identificação desses valores, juntamente com o relativo à sincronia 19-1, oferece, enfim, o traçado diacrônico buscado neste trabalho. Os graus de transparência de todas as sincronias são expostos também no gráfico seguinte:

Gráfico 4 – Variação diacrônica do grau de transparência do sistema EAS-1PP, considerando apenas o impacto da variação *nós/a gente*



Fonte: Elaboração própria.

Como se pode notar, o resultado obtido confirma nosso entendimento de que a variação *nós/a gente* tem influência, diacronicamente, sobre o grau de transparência do PB. Especificamente, os dados mostram que a alternância *nós/a gente*, no período em pauta, é uma força que, mantidas as demais características da língua, atuaria para a diminuição da transparência do sistema EAS-1PP. Trata-se de um impacto considerável sobre o sistema, envolvendo uma diminuição de transparência que chega a 32,6 pontos percentuais (na escala de 0 a 100%) ao final do período (diferença entre 79,9%, em 19-1, e 47,3% em 21-1). Essa queda pode ainda ser vista como maior, considerando-a em relação ao próprio grau de transparência do sistema no início da contagem (79,9%): em relação a esse valor inicial de 79,9% de transparência, a queda de 32,6% significa, na verdade, uma redução da ordem de 40,8%. Na sincronia 20-2, a transparência do sistema chegaria a uma redução até maior, de 40,5 pontos percentuais (na escala de 0% a 100%, sendo a diferença entre 79,9% e 39,4%), o que equivale a uma redução de 50,7% quando se considera a redução em relação ao valor inicial da transparência do sistema. São quedas, inclusive, que levariam o sistema, nas sincronias 20-2 e 21-1, a se tornar mais opaco do que transparente, já que nessas duas sincronias o índice de transparência ficaria abaixo de 50% (39,4% em 20-2 e 47,3% em 21-1).

A diminuição de transparência se mostra especialmente relevante quando se observa que o recorte diacrônico estudado é pouco extenso em relação à história de uma língua. Hodge (1970 *apud* Slobin, 1977) identifica, no egípcio antigo, uma alteração entre a tendência da língua a priorizar construções linguísticas analíticas *versus* a tendência da língua a preferir construções sintéticas, o que configura uma distinção comparável à diferença entre transparência e opacidade. O autor apura que o egípcio percorre dois ciclos inteiros, indo de uma a outra dessas tendências. No entanto, segundo o autor, esses ciclos completam-se no decorrer de quase 4.000 anos de história da língua. Tendo esse relato em vista, pode-se dizer que é, de fato, significativo o resultado obtido em relação ao PB, considerando o recorte diacrônico observado, que, embora recubra espaço considerável da história da língua, pode ser tomado como relativamente curto.

Os dados calculados de queda na transparência do sistema EAS-1PP, enfim, corroboram e especificam quantitativamente um cenário que pode ser sintetizado como sendo a passagem de uma situação na qual o sistema privilegiava a expressão do argumento-sujeito por RS para uma situação na qual privilegiaria a expressão por RZ. No início do período em análise (19-1), predominavam o tratamento de 1PP em termos *nós* (em vez de *a gente*) e a não expressão pronominal (em lugar da expressão pronominal); o predomínio dessa combinação, somado ao fato de todas as formas verbais relativas a *nós* terem desinência número-pessoal explícita, levava à prevalência de construções de RS como $\emptyset$ *cantamos* (que sobrepujava *nós cantamos*, *a gente canta* $\emptyset$ e $\emptyset$ *canta* $\emptyset$). Já no final do período em apreço (21-1), observando que predomina o tratamento de 1PP em termos *a gente* (em vez de *nós*), considerando inalterada a preferência por não expressão pronominal (para os efeitos do foco deste artigo) e tendo em conta que a maioria das formas verbais relativas a *a gente* não tem desinência número-pessoal explícita, obteríamos um cenário em que o sistema passaria a privilegiar construções de RZ como $\emptyset$ *canta* $\emptyset$ (em detrimento de $\emptyset$ *cantamos*, *nós cantamos* e *a gente canta* $\emptyset$). Em síntese, no período em questão, a mudança *nós/a gente*, em si, levaria à passagem do predomínio de construções de RS como $\emptyset$ *cantamos*, $\emptyset$ *cantávamos*, $\emptyset$ *cantaríamos*, $\emptyset$ *cantaremos*, $\emptyset$ *cantás-*

sems para a predominância de construções de RZ como $\emptyset$ *canta* $\emptyset$, $\emptyset$ *cantava* $\emptyset$, $\emptyset$ *cantaria* $\emptyset$, $\emptyset$ *cantará* $\emptyset$, $\emptyset$ *cantasse* $\emptyset$.¹⁵

Conforme mencionado, este artigo delimita-se a descrever o impacto da mudança *nós/a gente* sobre a transparência do sistema EAS-1PP, justamente como uma forma de identificar o papel específico dessa mudança sobre a língua. Naturalmente, uma compreensão mais completa da diacronia da transparência do sistema em pauta requer uma análise integrativa entre os efeitos aqui apurados com aqueles advindos do aumento diacrônico nos índices de expressão pronominal do argumento-sujeito. Na próxima seção, esboçamos esse tipo de discussão.

5 Considerações finais

Neste artigo, analisamos o fenômeno da variação diacrônica da propriedade da transparência linguística, abordando, em particular, a transparência do sistema de expressão pronominal e desinencial do argumento-sujeito de primeira pessoa do plural no PB – o sistema EAS-1PP. Argumentamos, em termos teóricos, que a transparência desse sistema é influenciada pelo aumento diacrônico nos índices de uso do pronome *a gente* em detrimento do pronome *nós* e procuramos demonstrar, mediante cálculo quantitativo, o impacto específico dessa influência. Os resultados indicam que a mudança *nós/a gente*, em si, atuaria na direção de diminuição de transparência, ocasionando, da sincronia 19-1 à 21-1, uma queda de 32,6 pontos percentuais no grau de transparência do sistema – o que equivale a uma redução de 40,8% em relação ao grau de transparência do sistema no início do período observado (79,9%).

O resultado aqui descrito representa, obviamente, apenas uma parte do que ocorre diacronicamente com o sistema EAS-1PP, a parte relativa à influência que o sistema receberia da mudança *nós/a gente*. Como destacamos, o sistema também é afetado pelo aumento diacrônico no índice de expressão pronominal, que suplanta a não expressão pronominal na história do PB – ver, entre outros, Tarallo (1993), Duarte (1993, 1995), Guerra (2017, 2022) para dados quantitativos que atestam o domínio da expressão pronominal. Integrando a variação diacrônica *nós/a gente* com a variação expressão/não expressão pronominal, pode-se visualizar, grosso modo, que o sistema EAS-1PP chegaria a uma situação diferente daquela focalizada no presente trabalho: o predomínio, na sincronia 19-1, de construções de RS como $\emptyset$ *cantamos* e $\emptyset$ *cantávamos*, em vez de dar lugar, na sincronia 21-1, à prevalência de construções de RZ como $\emptyset$ *canta* $\emptyset$ e $\emptyset$ *cantava* $\emptyset$, daria lugar à predominância de outro tipo de RS, relativo a construções como *a gente canta* $\emptyset$ e *a gente cantava* $\emptyset$ – predominância decorrente do maior uso de *a gente* (em relação a *nós*), do maior uso de expressão pronominal (em detrimento da não expressão pronominal) e do fato de a maioria das formas verbais relativas a *a gente* não conter desinência número-pessoal explícita.

¹⁵ Há que se considerar que estudos que analisam a correlação entre a expressão do sujeito e a concordância de 1PP mostram que a ocorrência de $\emptyset$ *canta* $\emptyset$ é reduzida se comparada a *a gente canta* $\emptyset$ (por influência do aumento diacrônico das taxas de expressão do sujeito), o que indicaria que a mudança *nós/a gente* não teria correlação direta com uma passagem RS > RZ. No entanto, como destacamos em demais partes do artigo, o objetivo aqui é delimitar qual seria o papel específico da mudança *nós > a gente* sobre a transparência do PB. Trata-se de uma forma de contribuir para um entendimento particular do papel de cada fator envolvido na diacronia da transparência, o que não significa deixar de reconhecer que a mudança em questão compreende apenas parte do cenário completo que ocorre na língua.

Em Guerra (2017), discutimos justamente o impacto que a transparência do PB recebe da interação entre essas duas mudanças, em conjunto ainda com outras duas mudanças, o aumento diacrônico do uso do pronome *você* em detrimento de *tu*, e o aumento do uso de *vocês* em prejuízo de *vós*. Os resultados sugerem que a transparência do PB varia diacronicamente em conformidade com uma atuação contrabalanceada entre forças cognitivo-comunicativas que competem entre si na determinação da estrutura das línguas, sobretudo as conhecidas forças de *iconicidade* e *economia*, tradicionalmente tomadas, no campo do funcionalismo linguístico, como duas das principais motivações determinantes da estrutura gramatical das línguas.¹⁶ Com efeito, também o resultado aqui obtido parece igualmente condizente com a hipótese de que a variação de transparência da língua estaria ligada a um contrabalanceamento entre essas forças.

Note-se que, no cenário linguístico aqui delimitado, circunscrito à mudança *nós/a gente*, o sistema EAS-1PP deixaria de privilegiar construções de RS e passaria a fomentar construções de RZ. Construções de RZ são condizentes com a atuação da motivação de economia linguística – já que consistem na veiculação de um significado (o de 1PP) sem uso de nenhuma expressão linguística material (nem pronome, nem desinência verbal explícita) –, porém, não são condizentes com a força da iconicidade, que prevê pareamento biunívoco (transparente) entre forma e conteúdo. Por outro lado, num cenário mais completo, que contemple tanto a mudança *nós/a gente* quanto a mudança *expressão/não expressão pronominal*, a alteração no sistema seria compatível, ao mesmo tempo, com ambas as forças, iconicidade e economia.

A compatibilidade com a iconicidade verifica-se na medida em que o sistema, conforme apontado acima, deixaria de privilegiar um tipo de construção de RS (*Ø cantávamos*) para privilegiar outro tipo de construção também de RS (*a gente cantavaØ*). Assim, o sistema permaneceria afinado com a motivação da iconicidade, já que a RS pode ser vista justamente como uma construção icônica.

A compatibilidade com a economia também ocorre. No PB, a integração da mudança *nós/a gente* com a mudança *expressão/não expressão pronominal* impediu não só que o sistema EAS-1PP caminhasse para privilegiar a RZ (o que aconteceria caso ocorresse na língua apenas a mudança *nós/a gente*, como considerado no cálculo quantitativo focalizado no artigo), mas também impediu que o sistema viesse a privilegiar construções de RC: caso tivesse se processado na língua apenas o aumento diacrônico nos índices de expressão pronominal (não ocorrendo a mudança *nós/a gente*), o sistema, na sincronia 21-1, teria passado a privilegiar construções de RC como *nós cantamos*, *nós cantávamos*, *nós cantaríamos*, privilégio decorrente do predomínio da expressão pronominal (que suplantaria a não expressão), do predomínio de *nós* sobre *a gente* (resultante da não ocorrência da mudança entre esses dois pronomes) e do fato de que todas as formas verbais relativas ao pronome *nós* têm desinência número-pessoal explícita. Esse predomínio de RC, além de configurar-se como uma situação não icônica, seria também uma situação não econômica, já que a RC é uma estrutura redundante (é o uso de duas formas linguísticas para expressão de um mesmo significado). Portanto, o cená-

¹⁶ De acordo com autores como Haiman (1983) e Croft (2003), a motivação cognitivo-comunicativa da iconicidade prevê, dentre outras estruturas linguísticas, o pareamento biunívoco entre unidades de forma e unidades de significado da estruturação linguística sintagmática. Já a economia prevê pareamentos não biunívocos, como combinações entre: uma unidade de forma e duas de significado; uma unidade de significado e nenhuma unidade de forma (caso, por exemplo, do morfema zero). Nesse sentido, as construções transparentes aqui em estudo (RS) podem ser vistas como uma forma de iconicidade.

rio linguístico mais completo, que envolve tanto a mudança *nós/a gente* quanto a mudança expressão/não expressão pronominal, configura uma situação econômica (além de icônica) ao evitar o predomínio da RC.

Em resumo, o cenário aqui focalizado, circunscrito à mudança *nós/a gente*, seria compatível apenas com a força da iconicidade, não com a economia linguística, enquanto o cenário mais completo (e real), que abrange também a mudança *expressão/não expressão pronominal*, é condizente com ambas as forças. Nesse sentido, nossa visão é a de que, por essa razão (juntamente com outros fatores), o cenário circunscrito apenas à mudança *nós/a gente* não vem a ocorrer na língua efetivamente. A expressiva queda de transparência que o sistema EAS-1PP exibiria, privilegiando construções de RZ, pode não ter sido o caminho efetivamente instalado na língua pela falta de um maior alinhamento com o contrabalanceamento entre forças como iconicidade e economia.

A nosso ver, trata-se de uma hipótese concernente com os resultados aqui expostos e passível de ser explorada em outras pesquisas, que analisem *corpora* diversos, fenômenos distintos dos já estudados e novos recortes diacrônicos, em busca de uma compreensão cada vez mais completa e esclarecedora acerca do processo de variação diacrônica da propriedade da transparência linguística. Nossa expectativa é que o presente artigo possa contribuir para esse tipo de pesquisa, fornecendo indicações para a delimitação de temas de estudo, elementos para a elaboração de procedimentos metodológicos e dados empíricos para comparações futuras.

Referências

- ASSUMPÇÃO, L. Ilustríssimo Filho da mãe. In: ASSUMPÇÃO, L. *Onze peças de Leilah Assumpção*. Rio de Janeiro: Casa da Palavra, 2010, p. 575-603.
- BECHARA, E. *Moderna gramática portuguesa*. 37ª ed. Rio de Janeiro: Nova Fronteira, 2009.
- BORGES, P. R. S. *A gramaticalização de a gente no português brasileiro: análise histórico-social-linguística da fala das comunidades gaúchas de Jaguarão e Pelotas*. 2004. 206f. Tese (Doutorado em Linguística) – Universidade Federal do Rio Grande do Sul, 2004.
- BORGES, P. R. S. As dimensões sociais da mudança em peças de teatro de autores gaúchos: uma contribuição para o estudo da inserção e da propagação do pronome *a gente* no português brasileiro. *Cadernos do Instituto de Letras*, Porto Alegre, v. 1, n. 59, p. 71-88, 2019. DOI: <https://doi.org/10.22456/2236-6385.92310>.
- BARBOSA, M. C.; FALABELLA, M. Síndromes. In: BARBOSA, M. C.; FALABELLA, M. *Querido mundo e outras peças*. Rio de Janeiro: Lacerda Editores, 2003, p. 261- 292.
- BUARQUE, C. *Ópera do malandro*. São Paulo: Cultura, 1978.
- CAFEZEIRO, E.; GUERRA, R. Angélica e Firmino. In: CAFEZEIRO, E.; GUERRA, R. *Teatro completo de Araújo Porto-Alegre*. Rio de Janeiro: Funarte; Brasília: Ministério da Cultura, 1997. t. 2. p. 56-108.
- CÂMARA JR., J. M. *A Estrutura da língua portuguesa*. 19ª ed. Petrópolis: Vozes, 1989.
- CASTILHO, A. T. *Nova Gramática do Português Brasileiro*. São Paulo: Contexto, 2010.

- CASTILHO, A. T. (org.). *História do português brasileiro: corpus diacrônico do português brasileiro*. São Paulo: Contexto, 2019.
- CROFT, W. *Typology and universals*. Cambridge: Cambridge University Press, 2003.
- DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. A. (orgs.). *Português brasileiro: uma viagem diacrônica*. Campinas: Editora da Unicamp, 1993, p. 107-128.
- DUARTE, M. E. L. *A perda do princípio “evite pronome” no português brasileiro*. 1995. 151f. Tese (Doutorado em Ciências) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 1995.
- FRANÇA JÚNIOR, J. J. *Cahio o ministério*: comedia original de costumes em tres actos por França Junior. Rio de Janeiro: Livraria Popular de A. A. da Cruz Coutinho Livreiro Editor, 1883. Disponível em: <https://digital.bbm.usp.br/handle/bbm/7463>. Acesso em: 31 jan. 2025.
- GOMES, A. D. A invasão. In: GOMES, A. D. *A invasão; A revolução dos beatos*: teatro. Rio de Janeiro: Editora Civilização Brasileira, 1962, p. 1-125.
- GONÇALVES, S. C. L. *Projeto ALIP (Amostra Linguística do Interior Paulista)*: o português falado na região de São José do Rio Preto – constituição de um banco de dados anotado para o seu estudo. São José do Rio Preto: UNESP (Relatório científico, FAPESP), 2007.
- GONZAGA, A. *Ministro do Supremo*: comedia em tres actos. In: GONZAGA, A. *Comedias*. São Paulo: Livraria Teixeira, Vieira Pontes & Cia, 1940, p. 1-79.
- GRÁNDEZ ÁVILA, M. Language transparency in Functional Discourse Grammar: the case of Quechua. *Linguistics in Amsterdam*, Amsterdam, v. 4, n. 2, p. 22-56, 2011. Disponível em: <https://linguisticsinamsterdam.nl/home?issue=42>. Acesso em: 04 mar. 2024.
- GUERRA, A. R. *Diacronia do grau de transparência do sistema de referência por expressão pronominal e desinencial do argumento-sujeito de 1ª e 2ª pessoas no português brasileiro*. 2017. 170f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2017.
- GUERRA, A. R. Variação diacrônica do grau de transparência linguística do português brasileiro: o papel do aumento percentual da expressão pronominal do argumento-sujeito. *Cadernos de Estudos Linguísticos*, Campinas, v. 63, p. 1-16, e021011, 2021. DOI: <https://doi.org/10.20396/cel.v63i00.8657215>.
- GUERRA, A. R. Descrição diacrônica da expressão pronominal do argumento-sujeito no português brasileiro. *Revista do GEL*, v. 19, n. 3, p. 203-230, 2022. DOI: <http://dx.doi.org/10.21165/gel.v19i3.3449>.
- HAIMAN, J. Iconic and economic motivation. *Language*, v. 59, n. 4, p. 781-819, 1983. DOI: <https://doi.org/10.2307/413373>.
- HENGVELD, K. Transparency in Functional Discourse Grammar. *Linguistics in Amsterdam*, Amsterdam, v. 4, n. 2, p. 1-22, 2011a. Disponível em: <https://linguisticsinamsterdam.nl/home?issue=42>. Acesso em: 04 mar. 2024.
- HENGVELD, K. Epilogue: degrees of transparency. *Linguistics in Amsterdam*, Amsterdam, v. 4, n. 2, p. 110-114, 2011b. Disponível em: <https://linguisticsinamsterdam.nl/home?issue=42>. Acesso em: 04 mar. 2024.
- HENGVELD, K.; LEUFKENS, S. Transparent and non-transparent languages. *Folia Linguistica*, v. 52, n. 1, p. 139-175, 2018. DOI: <https://doi.org/10.1515/flin-2018-0003>.

- HENGEVELD, K.; MACKENZIE, J. L. *Functional Discourse Grammar*. New York: Oxford University Press, 2008.
- HOPPER, P. J. On some principles of grammaticalization. In: TRAUGOTT, E. C.; HEINE, B. (eds.) *Approaches to grammaticalization*. Philadelphia: John Benjamins Company, 1991. v. 1, p. 17-35.
- LEUFKENS, S. The transparency of creoles. *Journal of Pidgin and Creole Languages*, Amsterdam, v. 28, n. 2, p. 323-362, 2013. DOI: <https://doi.org/10.1075/jpcl.28.2.03leu>.
- LEUFKENS, S. *Transparency in language: a typological study*. Utrecht: LOT, 2015.
- LOPES, C. R. S. A inserção de “a gente” no quadro pronominal do português: percurso histórico. 1999. 181f. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, 1999.
- LOPES, C. R. S. O quadro dos pronomes pessoais. In: VIEIRA, S. R.; BRANDÃO, S. F. (orgs.). *Morfossintaxe e ensino de português: reflexões e propostas*. Rio de Janeiro: UFRJ, 2004. p. 151-178.
- MACEDO, J. M. *A torre em concurso: comédia burlesca em tres actos*. Rio de Janeiro: B. L. Garnier, Editor, 1863. Disponível em: <https://digital.bbm.usp.br/handle/bbm/3997>. Acesso em: 33 jan. 2025.
- NORDHOFF, S. Transparency in Sri Lanka Malay. *Linguistics in Amsterdam*, Amsterdam, v. 4, n. 2, p. 96-110, 2011. Disponível em: <https://linguisticsinamsterdam.nl/home?issue=42>. Acesso em: 11 mar. 2024.
- OMENA, N. P.; BRAGA, M. L. A gente está se gramaticalizando? In: MACEDO, A. T.; RONCARATI, C.; MOLLICA, M. C. (orgs.). *Variação e discurso*. Rio de Janeiro: Tempo Brasileiro, 1996. p. 75-84.
- PACHECO, C. S. A diacronia e a sincronia dos pronomes de primeira pessoa do plural *nós* e *a gente* no português brasileiro e no português uruguaio. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 26, n. 1, p. 221-253, 2018.
- PENA, L. C. M. *Os irmãos das almas: comédia em um acto*. Rio de Janeiro: [s.n.], 1847. Disponível em: <https://digital.bbm.usp.br/handle/bbm/4053>. Acesso em: 30 jan. 2025.
- RAPOSO, E. P. *Teoria da gramática: a faculdade da linguagem*. Lisboa: Editorial Caminho, 1992.
- RODRIGUES, N. Album de família. In: RODRIGUES, N. *Album de família e vestido de noiva*. Rio de Janeiro: Edições do Povo, 1946, p. 7-144.
- RUBIO, C. F.; GONÇALVES, S. C. L. Opções metodológicas no estudo de fenômenos variáveis relacionados à primeira pessoa do discurso no plural. *Gragoatá*, Niterói, v. 15, n. 29, p. 161-182, 2010. DOI: <https://doi.org/10.22409/gragoata.v15i29.33080>.
- SANTANA, A. M. B. *Nós e a gente: um retrato do português popular de Salvador*. 2014. 114f. Dissertação (Mestrado em Linguística) – Universidade do Estado da Bahia, 2014.
- SEARA, I. C. A variação do sujeito *nós* e *a gente* na fala florianopolitana. *Organon*, Porto Alegre, v. 14, n. 28-29, p. 179-194, 2000. Disponível em: <https://seer.ufrgs.br/index.php/organon/issue/view/1710>. Acesso em: 02 set. 2024.
- SEGUIN, L. Transparency and language contact: the case of Haitian Creole, French and Fongbe. *Linguistics in Amsterdam*, Amsterdam, v. 11, n. 2, p. 211-242, 2018. Disponível em: <https://linguisticsinamsterdam.nl/home?issue=112>. Acesso em 11 mar. 2024.
- SLOBIN, D. I. Language change in childhood and in history. In: MACNAMARA, J. (ed.). *Language learning and thought*. New York: Academic Press, 1977. p. 185-214.

TARALLO, F. Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX. In: ROBERTS, I.; KATO, M. A. (orgs.). *Português brasileiro: uma viagem diacrônica*. Campinas: Editora da Unicamp, 1993, p. 69-105.

VIANNA, J. B. S.; LOPES, C. R. S. A variação entre *nós* e *a gente*: uma comparação entre o português europeu e o português brasileiro. *Revista do GELNE*, Natal, v. 14, n. 1/2, p. 95-116, 2012. Disponível em: <https://periodicos.ufrn.br/gelne/issue/view/525>. Acesso em: 09 set. 2024.

Ensino de língua materna e edição de livros didáticos: políticas educacionais e linguísticas do Estado Novo brasileiro

Mother Tongue Teaching and Textbook Publishing: Educational and Linguistic Policies in Brazil

Cristian Imbruniz

Universidade Estadual Paulista (Unesp)

São José do Rio Preto | SP | BR

imbrunizc@gmail.com

<https://orcid.org/0000-0001-6506-9285>

Resumo: O objetivo deste artigo é analisar as relações entre políticas educacionais e linguísticas e um livro didático de português. Especificamente, analisar como concepções de linguagem ofereceriam pistas de políticas linguísticas implícitas. O trabalho se concentra nas políticas do Estado Novo para o ensino de português e no *Manual de Língua Portuguesa*, publicado, em 1945, pela Companhia Editora Nacional. Como arcabouço teórico, assume-se uma abordagem discursiva, de linha francesa, recorrendo, sempre que necessário, às contribuições da história da educação. Do ponto de vista metodológico, em uma pesquisa qualitativa e documental, analisa-se, a princípio, o *Manual* quanto à legislação educacional do Estado Novo, particularmente, o decreto-lei do livro didático, nº 1.006 de 30 de dezembro de 1938, em busca de políticas linguísticas explícitas. Em seguida, propõe-se uma análise linguístico-discursiva de suas concepções de linguagem, em busca de políticas linguísticas implícitas. Quanto às políticas explícitas, constatou-se que, submetido ao controle do Estado, o *Manual* cumpria exigências legais não somente nos conteúdos e métodos de ensino, mas também nos usos da língua, interditando “linguagens” consideradas “defeituosas” e adotando a ortografia oficial. Quanto às políticas implícitas, constatou-se que o livro se alinhava a um processo de organização, conservação e unificação nacional. Esse processo se marcava pela dupla expressão de uma concepção de linguagem como forma pura, ora pela afirmação da elegância, correção e pureza de



certa escrita literária, ora pela negação de usos que escapassem a essas expectativas.

Palavras-chave: livros didáticos de língua materna; Companhia Editora Nacional; políticas educacionais; políticas linguísticas; concepções de linguagem.

Abstract: The aim of this article is to analyze the relationships between educational and language policies and a Portuguese language textbook, examining how language conceptions may provide clues to implicit language policies. The study focuses on the policies of the *Estado Novo* for the teaching of Portuguese and on the *Manual de Língua Portuguesa*, published in 1945 by *Companhia Editora Nacional*. Grounded in a discursive approach and drawing, when necessary, on the history of education, the research is qualitative and documentary in nature. It first analyzes the *Manual* in light of the educational legislation of the *Estado Novo*, particularly Decree-Law no. 1,006 of December 30, 1938 (the textbook decree), in search of explicit language policies, and then conducts a linguistic-discursive analysis of its language conceptions to identify implicit policies. The findings indicate that, under state control, the textbook complied with legal requirements not only in its content and teaching methods but also in its uses of language, prohibiting “defective” forms and adopting the official orthography. Moreover, it aligned with a broader process of national organization, preservation, and unification, marked by a dual expression of language as pure form, both through the affirmation of elegance, correctness, and purity in literary writing and through the rejection of uses that diverged from these expectations.

Keywords: Portuguese language textbooks; Companhia Editora Nacional; educational policies; linguistics policies; conceptions of language.

1 Introdução

No Brasil, os livros didáticos participam das discussões sobre políticas educacionais¹. Não por acaso, eles têm sido, desde muito tempo, objeto de intervenções do Estado, sobretudo, por meio de sua avaliação, aquisição e distribuição (Cassiano, 2013, p. 49; Filgueiras, 2011, p. 19-25). No caso dos livros didáticos de português, essa participação se estende, também, às discussões sobre políticas linguísticas, pois eles desempenham um papel relevante no ensino de língua materna e, portanto, na sua manutenção política e cultural. Com efeito, políticas educacionais para o ensino de português e políticas linguísticas costumam aparecer conjuntamente em decretos, leis e portarias, com a ressalva de que, por vezes, as políticas linguísticas são evocadas de maneira implícita.

A coexistência desses dois tipos de políticas em documentos, que, oficial ou oficiosamente, regulam os livros didáticos enseja algumas observações. Por um lado, como já dito, os livros didáticos são uma parte relevante das políticas educacionais e linguísticas. Por outro, essas políticas seriam determinantes na produção, circulação e recepção desses materiais. A relação entre políticas educacionais e linguísticas e livros didáticos se inscreveria, portanto, na reversibilidade entre o estático e o dinâmico. Em outras palavras, políticas e livros asseguram, mutuamente, sua própria existência, instituindo processos de construção legítimos, práticas de ensino e aparelhos organizados (Maingueneau, 2006, p. 54). Como é de supor, essa relação não se limita ao cumprimento ou não cumprimento por autores e editores de decretos, leis e portarias referendadas por legisladores. Embora contemple a burocracia legal, essa relação a ultrapassa, registrando-se tanto no plano sociopolítico quanto no plano sócio-histórico do discurso². Na circulação cruzada de políticas educacionais e linguísticas e de livros de português,

¹ A relevância política, social e histórica dos livros didáticos pode ser medida de diferentes maneiras. Na vida prática, esses livros participam, desde muito tempo, da experiência escolar de alunos e professores. Na vida coletiva, são objetos recorrentes de debates políticos. Basta lembrar a recente tentativa do governo do estado de São Paulo de abandonar o uso de livros didáticos em favor de materiais digitais. Essa tentativa de abandono do livro didático causou reação da sociedade civil e ensejou uma calculada retratação do secretário de educação: “a não adesão [ao programa do livro didático] foi um dos grandes erros da minha vida” (Palhares, 2023). Na vida política, eles são objeto, no Brasil e em diferentes países, de longevas políticas públicas de aquisição e distribuição, como o Programa Nacional do Livro Didático. Esses e outros predicados dos livros didáticos justificam o interesse por sua investigação como objeto de pesquisa em diferentes disciplinas científicas. Nessas diferentes disciplinas, reconhece-se, em comum, o caráter multifacetado e complexo do livro didático, atrelado a funções curriculares, instrumentais, ideológicas e documentais (Choppin, 2020, p. 13). O campo de estudos dos livros didáticos, inscrito em diferentes disciplinas, procura, portanto, abordar essas variadas facetas e funções, que se traduzem em definições, finalidades e usos igualmente distintos (Batista, 1999, p. 567-589; Bruillard, 2005, p. 14; Bunzen, 2005, p. 27; Denizot, 2021, p. 167; Munakata, 2012, p. 58-62; Louichon, 2015, p. 19-20). Neste artigo, o livro didático é concebido como um objeto editorial inscrito na história de uma disciplina escolar, cuja produção, circulação e recepção podem dar pistas sobre debates e disputas sociopolíticas (Imbruniz, 2025, p. 18), como aqueles sobre as políticas educacionais e linguísticas.

² A diferença entre os planos sociopolítico e sócio-histórico do discurso é apenas metodológica. Em Guilhaumou e Maldidier (2014, p. 170), essa distinção está presente na sua proposta de história social dos textos. Para os autores, a análise do discurso se concentrou, desde suas etapas iniciais, nos *corpora* da política institucional e doutrinária. Essa escolha teria obliterado outras manifestações linguísticas que, com seus dispositivos textuais, instalavam o social no político. Trata-se, pois, de supor que, em diferentes arquivos e dispositivos, haveria a possibilidade de instalação do social no político. Para Guilhaumou e Maldidier, os variados arquivos e dispositivos permitiram a escrita de uma história social dos textos. Essa história não incluiria apenas os produtos da

se materializariam, de fato, expectativas políticas, sociais e históricas relativas à língua e ao seu ensino. De um lado, essas expectativas seriam institucionalizadas em gestos burocráticos e cristalizadas nos variados gêneros do aparato legal. De outro, elas seriam atravessadas por discursivos prévios que, como já ditos ou não ditos, poderiam levar à manutenção, à transformação ou à supressão desses enunciados institucionalizados (Maingueneau, 1997, p. 95; Pêcheux, 2010, p. 77). Essas constatações sugerem que, por meio da análise conjunta de políticas educacionais e linguísticas em livros de português, seria possível descrever e analisar, em períodos determinados, expectativas políticas, sociais e histórias quanto à língua e ao seu ensino.

No presente artigo, exploro essa hipótese, restringindo-me ao período do Estado Novo brasileiro (1937-1945) e a alguns de seus desdobramentos imediatamente posteriores. Sedimentado no tempo, o programa de Getúlio Vargas e de seus ministros é representativo dos três eixos abordados – políticas educacionais e linguísticas e livros didáticos –, instituindo-os, com clareza inédita, no campo das disputas políticas e burocráticas (Filgueiras, 2013, p. 165). Detenho-me, particularmente, em um livro de português, o *Manual de Língua Portuguesa*, de Artur de Almeida Torres, publicado pela Companhia Editora Nacional, entre 1945 e 1951. O *Manual* foi, possivelmente, o primeiro livro de português que, após a reforma Gustavo Capanema³, em 1942, a Nacional destinou ao ciclo final do secundário, chamado, então, de colegial (atual ensino médio). Esse fato lhe confere um lugar de destaque na história da editora que, durante boa parte do século XX, foi a maior produtora de livros didáticos do Brasil. Embora tenha sido publicado somente ao fim da ditadura de Getúlio Vargas, o livro foi ges-

política institucional e doutrinária, mas também diferentes materialidades linguísticas, como gritos populares, tratados, correspondências, obras literárias e debates em assembleias. Os questionamentos metodológicos de Guilhaumou e Maldidier (2014, p. 170) conduzem a uma primeira distinção entre sociopolítico e sócio-histórico. O plano sociopolítico do discurso contemplaria as manifestações da política institucional e doutrinária, realizando-se, preferencialmente, em enunciados burocráticos. Essas manifestações sociopolíticas seriam atravessadas por contradições políticas, sociais e históricas, como ocorre com todos os enunciados, mas se particularizariam por sua inscrição nas instituições das sociedades modernas. Essas instituições são conhecidas por suas tentativas de produzir dispositivos de neutralização do equívoco, como a própria noção de lei. São as contradições que atravessam o plano sociopolítico que, ao produzirem a falha nos dispositivos de controle do equívoco, constituem o plano sócio-histórico do discurso. Nos termos dessa distinção, a proposta de Guilhaumou e Maldidier consistiria em reconhecer que, para além dos enunciados da política institucional e doutrinária, outros tantos são atravessados pelo plano sócio-histórico do discurso e, portanto, são de interesse para o campo disciplinar. A distinção entre os planos sociopolítico e sócio-histórico do discurso ecoa, com efeito, a diferenciação feita por Pêcheux (2015, p. 22) entre discursos logicamente estabilizados e não logicamente estabilizados. Inscritos no plano sociopolítico e na idade teórica de suas instituições, os discursos logicamente estabilizados são aqueles que simulam uma homogeneidade lógica, ignorando o equívoco imposto pelos discursos não logicamente estabilizados. A distinção entre esses dois planos é, portanto, uma tentativa metodológica de capturar um funcionamento geral dos discursos por uma de suas manifestações particulares mais estáveis, a política institucional.

³ A reforma Capanema se refere às alterações promovidas, durante o Estado Novo, na educação secundária, constituindo-se em um dos marcos do projeto político do regime varguista (Montalvão, 2021, p. 6). Antes da reforma, entre 1931 e 1942, o ensino secundário se dividia em dois ciclos de frequência obrigatória, o fundamental, de cinco anos de duração, e o complementar, de dois anos de duração. Nesse período, acusava-se uma orientação cientificista da escola média, que escantearia as línguas vernáculas e clássicas. A reforma conduzida por Gustavo Capanema, em 1942, reordenou os ciclos do secundário em ginásio, de quatro anos de duração, e colégio, de três anos de duração, mantendo a frequência obrigatória ao longo de sete anos de curso. Como resposta às críticas de cientificismo e como gesto a sua base política, católica e conservadora, o ministro imprimiu uma orientação beletrista e humanista ao ensino secundário, com destaque para as línguas clássicas e vernáculas, como o latim e o português (Souza, 2009, p. 81; Imbruniz, 2025, p. 119-128).

tado sob a égide de um de seus projetos mais longevos, no qual a língua era um elemento incontornável: a reforma de ensino secundário (Bomeny, 1999, p. 151). Atenho-me, particularmente, à primeira edição do volume destinado ao segundo ano do colegial, representativa do projeto editorial investigado e de seu período de circulação.

Os objetivos deste trabalho são dois. De um lado, verificar as interações entre políticas educacionais e linguísticas em um livro didático de português, com destaque para a relação entre concepções de linguagem e políticas linguísticas implícitas. De outro, descrever e analisar, a partir dessas interações, expectativas do Estado Novo quanto à língua e ao seu ensino. Assumo, para tanto, uma perspectiva discursiva de linha francesa, em abordagem qualitativa e documental, recorrendo, sempre que necessário, às contribuições da história da educação e do campo disciplinar das políticas linguísticas. Esse arcabouço teórico-metodológico se ampara, sobretudo, na noção de concepção de linguagem, conforme desenvolvida em trabalhos anteriores (Imbruniz, 2022, p. 11). Compreendidas como o modo por meio do qual um sujeito concebe a linguagem e a língua para aqueles que considera seus iguais e seus outros, as concepções de linguagem⁴ são relevantes na medida que, em função de seu comprometimento intersubjetivo, remetem não só a fatores sociopolíticos, como políticas educacionais e linguísticas explícitas, mas a fatores sócio-históricos não ocultos, mas nem sempre visíveis (Foucault, 2008, p. 123), como as políticas linguísticas implícitas. A suposição é de que haveria um trânsito parcial entre concepções de linguagem e políticas linguísticas implícitas, hipótese a ser explorada na busca das expectativas do Estado Novo sobre a língua e o seu ensino.

Começo com uma breve discussão teórico-metodológica sobre políticas linguísticas explícitas e implícitas. Em seguida, apresento o *Manual de Língua Portuguesa* em termos de autoria, edição e políticas educacionais e linguísticas explícitas. Por fim, proponho uma análise de suas concepções de linguagem e de seu modo de interação com políticas linguísticas implícitas.

⁴ Em linguística aplicada, consolidou-se, desde os anos 1980, a categoria *concepção de linguagem*. À época, Geraldini (2011, p. 34-35) distinguiu, para a análise das práticas de ensino de língua portuguesa, três dessas concepções: (i) linguagem como expressão do pensamento; (ii) linguagem como instrumento de comunicação; e (iii) linguagem como processo de interação. Um marco do pensamento linguístico estaria associado, respectivamente, a cada uma dessas concepções: (i) normativo-gramatical; (ii) estruturalista-comunicacional; e (iii) linguístico-interacional. A partir da sistematização inicial de Geraldini (2011, p. 34), observou-se o uso crescente dessa categoria para análise de práticas, materiais e documentos de ensino de língua portuguesa (Imbruniz; Haynal, 2021, p. 175). Em trabalhos anteriores, sublinhei o limite dos usos da categoria de *concepção de linguagem* em gestos classificatórios (Imbruniz, 2022, p. 11). Em outras palavras, o uso da categoria que, ao encontrá-la ou não encontrá-la em uma prática, material ou documento, permite-se inscrever esses objetos em marcos do pensamento linguístico, produzindo uma história linear e cumulativa do ensino de português, equivalente àquela de um suposto desenvolvimento do pensamento linguístico (“as influências estruturalistas imprimem uma *concepção de linguagem* como instrumento de comunicação no livro didático”). O mesmo uso classificatório permite acusar contradições em práticas, materiais e documentos, que podem ser anunciados como tendo uma *concepção* e efetivando outra (“ele se diz interacionista, mas é normativista”). Em lugar desse uso classificatório da noção de *concepção de linguagem*, proponho sua repactuação teórico-metodológica. Essa repactuação consistiria em reconhecer que, em uma mesma prática, material ou documento, coexistiriam diferentes *concepções de linguagem*, materializadas na relação pretendida pelo sujeito com a língua que atribui a si e com a língua que atribui aos seus outros. Nessa proposta, as categorias tradicionalmente usadas em linguística aplicada seriam apenas formas de racionalização de um processo intersubjetivo instável e nunca concluído, cuja finalidade implícita seria conjurar os perigos do discurso e fixar uma história linear e cumulativa (Foucault, 1996, p. 8-9). Neste artigo, não recorro, portanto, a tais formas de racionalização para investigar as relações entre *concepções de linguagem* e políticas linguísticas implícitas. São, antes, as relações com a língua, travadas por um autor de livro didático e aqueles que representa como sendo seus iguais e seus outros que dão pistas para a investigação.

2 Políticas linguísticas explícitas e implícitas

Em *As políticas linguísticas*, Calvet (2018, p. 11-32) argumenta que, apesar dos esforços constantes de agentes de Estado para legislar sobre a língua, nem todas as suas tentativas podem ser consideradas políticas linguísticas. Segundo o autor, as políticas linguísticas, necessariamente acompanhadas de planejamento, englobam somente algumas iniciativas dos atores políticos. Essas iniciativas são, geralmente, relativas a decisões de larga escala sobre problemas linguísticos, como, por exemplo, a imposição de uma ou outra língua a certos grupos sociais. A partir das noções de política e de planejamento linguísticos, Calvet (2018, p. 11) delimita, nestes termos, o alcance pretendido no campo de pesquisa: as “determinações das grandes decisões referentes às relações entre as línguas e a sociedade [...] e sua implementação”.

Quanto aos instrumentos de planejamento linguístico, Calvet defende que, ao contrário do que normalmente se supõe, as línguas podem ser “equipadas” para exercer funções antes não previstas. O autor menciona o caso das línguas ágrafas, que, por não utilizarem a escrita, não poderiam ser objeto de uma campanha de alfabetização. Nesse caso, em um momento de necessidade (política ou social), uma dessas línguas poderia ser “equipada” com um sistema de escrita. Calvet (2018, p. 68) sublinha que uma suposição como essa sobre instrumentos de planejamento linguístico depende da crença na possibilidade de mudar as línguas. Essa crença permite ao autor distinguir dois tipos de gestão linguística, *in vivo* e *in vitro*. A gestão *in vivo* contemplaria as mudanças causadas por fatores internos e externos às línguas, mas não voluntários, velhos conhecidos daqueles que tiveram uma formação em linguística. A gestão *in vitro*, por sua vez, contemplaria as mudanças voluntárias, associadas a políticas linguísticas, nas quais estão envolvidos estudiosos da linguagem e agentes de Estado. Segundo o Calvet (2018, p. 71), os dois tipos de gestão de mudança da língua podem entrar em conflito, fazendo das políticas linguísticas um lugar de negociação entre “os objetivos do poder” e as “soluções intuitivas que são frequentemente postas em prática pelo povo”.

Nos termos de Calvet, a gestão *in vivo* seria, pois, produto da prática dos falantes, que, tacitamente, regulam o estatuto da mudança linguística. A gestão *in vitro* ocorreria por meio de intervenções do Estado com decretos, leis e portarias, cujos objetivos seriam variáveis, englobando a forma e o *status* das línguas. Para o autor, a característica distintiva da política linguística é, portanto, sua motivação e sua explicitude na solução de problemas que envolvem a língua. Submetida a esses critérios, a hipótese do presente trabalho exigiria um exame detido, pois, já mais de uma vez, me referi a políticas de língua “implícitas”, que não necessariamente obedecem a instrumentos intencionais de planejamento linguístico.

Schiffman (2002, p. 13) questiona se, de fato, as políticas linguísticas podem ser planejadas e aplicadas por agentes políticos por meio de intervenções explícitas. Embora reconheça essa possibilidade, o autor distingue as políticas planejadas e explícitas daquelas reguladas pelo que chama de *cultura linguística*. Estas seriam implícitas. Para o autor, as políticas reguladas pela *cultura linguística* estariam submetidas a formas culturais, como os comportamentos, as pressuposições, os estereótipos e as crenças populares. Spolsky (2004, p. 18) também distingue as políticas linguísticas explícitas das implícitas. Em relação às implícitas, ele argumenta que as práticas, as crenças e as ideologias linguísticas compõem a política de língua de uma comunidade de falantes. Segundo o autor, a proibição (como a de palavras de conotação sexual) ou o incentivo (como o de palavras sem viés de gênero) de certas expressões são um

exemplo de política linguística. Shohamy (2006, p. 45) argumenta que as políticas linguísticas de uma instituição não se reduzem àquelas explicitamente sistematizadas e disseminadas por seus atores. Para a autora, as políticas linguísticas “reais” se efetivariam nos “dispositivos que são utilizados para perpetuar práticas linguísticas, frequentemente de modo encoberto e implícito”⁵ (Shohamy, 2006, p. 46, tradução minha). Dentre esses dispositivos, estariam não só os já mencionados decretos, leis e portarias, mas também as políticas educacionais, as provas e os usos da língua no espaço público.

Essas considerações têm grande relevância para a ampliação do escopo da noção de política linguística e, portanto, para a sustentação dos pressupostos deste artigo. Em primeiro lugar, elas mostram que nem todas as políticas linguísticas se apresentam como leis sancionadas por entes nacionais ou supranacionais. Em segundo lugar, que nem toda política linguística declarada exerce, efetivamente, poder regulatório nas práticas dos falantes. Por fim, que nem toda ausência de política linguística declarada corresponde à ausência de regulação nas práticas linguísticas dos falantes. No entanto, as propostas desses autores não apenas ampliam o escopo da noção. Elas contemplam, de fato, as dimensões explícitas e implícitas das políticas. De acordo com Silva (2013, p. 306), elas sinalizam uma virada epistemológica no campo de pesquisa. Silva defende que, influenciadas por Schiffman e, posteriormente, desenvolvidas por Shohamy, as contribuições de Spolsky solidificaram uma transformação epistemológica que se desenrolava desde a década de 1980. Segundo o autor, essa transformação corresponde ao abandono de modelos a-históricos em favor de outros, “fortemente comprometido[s] com o social e o político” (Silva, 2013, p. 316). Silva observa que as discussões sobre a noção de “agência” tiveram um papel fundamental, já que seu questionamento permitiu a constatação de que não somente gestos intencionais e conscientes permitiriam a criação de políticas linguísticas. Ao contrário, outros gestos, quase sempre implícitos e alheios às intenções e à consciência dos agentes, concorreriam para o surgimento dessas políticas, o que ressaltaria sua dimensão social e histórica.

Ao pôr em questão a noção “agência”, a transformação epistemológica reconhecida por Silva (2013, p. 307) nos trabalhos Schiffman (2002), Spolsky (2004) e Shohamy (2006) oferece subsídios para a abordagem discursiva adotada neste trabalho, pois introduz, nesse campo disciplinar, a crítica ao predomínio dos gestos intencionais e conscientes nas políticas linguísticas. Há, porém, outro deslocamento necessário, relativo à subjetividade. Neste artigo, a subjetividade não é concebida como a manifestação de intenção. No entanto, ela tampouco corresponde à assimilação de práticas por meio de dispositivos culturais, direção para a qual parecem apontar os argumentos de Schiffman, Spolsky e Shohamy. Nos termos de Authier-Revuz (2008, p. 34) e Lemos (2000, p. 177), a subjetividade é concebida, antes, como a investigação da não coincidência do sujeito consigo mesmo. Authier-Revuz (2008, p. 40) explica que, estruturalmente dividido, o sujeito restaura, no imaginário, a ilusão de unidade por meio do retorno ao seu próprio dizer. Segundo Lemos (2000, p. 176-177), no entanto, o retorno ao dizer é, para o sujeito, a escuta de sua fala como a de um outro. Para as autoras, entre o dito e o retorno ao dito haveria, portanto, um intervalo que sublinha a relação do sujeito com a língua. Esse intervalo é teorizado por Authier-Revuz (2008, p. 40) como o estrato metaenunciativo do dizer e, por Lemos (2000, p. 176-177), como uma das estruturas que captura a criança na aquisição da linguagem. Essa compreensão da subjetividade par-

⁵ No original: “Devices that are used to perpetuate language practices, often in covert and implicit ways”.

ticipa da definição de concepção de linguagem, conceito mobilizado, neste trabalho, para a análise de políticas linguísticas implícitas (cf., aqui mesmo, nota 5). A busca por concepções de linguagem e de políticas linguísticas implícitas não corresponde, pois, à determinação do que um indivíduo “entende como língua” ao utilizar o “código” para falar sobre o “código” em um livro didático. Essas percepções teriam pouco valor teórico e metodológico e reinstaurariam a consciência e a intenção como critérios definidores das concepções de linguagem e das políticas linguísticas implícitas. Essa busca também não equivale à descrição da assimilação de atitudes e estereótipos em função da exposição a certos instrumentos e ferramentas, objetivo mais compatível com uma abordagem etnográfica.

As concepções de linguagem devem ser buscadas na relação, quase sempre conflituosa, dos sujeitos com as regras de formação da língua como *saber* (Foucault, 2008, p. 204; Imbruniz, 2022, p. 11). Trata-se de analisar como, ao falar da linguagem e da língua, um dado sujeito a concebe para si e para seus iguais e para os que considera seus outros. Ao atribuir posições ao mesmo e ao diferente, esse procedimento é capaz de, ilusoriamente, restaurar a unidade e a centralidade do sujeito na produção de sentido, sublinhando sua captura pelo funcionamento discursivo dos saberes sobre a língua (Imbruniz, 2022, p. 11). Assumida como um desses saberes, as políticas linguísticas implícitas se mostrariam, nos livros didáticos de português, como um epifenômeno do processo de subjetivação suposto pelas concepções de linguagem.

3 O *Manual de Língua Portuguesa*, de Artur de Almeida Torres (1945-1951)

Do início do século XIX ao início do século XX, é possível distinguir duas gerações de autores de livros didáticos, cujas relações com as editoras refletiam particularidades históricas, sociais e políticas do seu tempo (Bittencourt, 2004, p. 477).

A primeira geração corresponde àquela que, com a chegada da família real portuguesa ao Brasil, passou a se ocupar da organização dos cursos secundários e superiores, dedicando pouca atenção aos cursos elementares. Esse grupo reunia os sábios da corte, que compunham instituições políticas e culturais e se interessavam pela “formação moral” dos filhos das elites (Bittencourt, 2004, p. 480). No que diz respeito às editoras e a sua relação com esses autores, o fim do monopólio da imprensa régia, em 1822, foi um momento crucial. A partir de então, a estratégia adotada pelos editores foi a de se aproximar das figuras de poder, que garantiam a aprovação dos livros publicados junto às instâncias oficiais. Não causa espanto, pois, a presença constante de membros do Instituto Histórico Geográfico Brasileiro (IHGB) e de políticos nos catálogos das grandes editoras do século XIX.

Entre 1870 e 1880, com a expansão do ensino elementar nas escolas públicas⁶, o grupo dos autores de livros didáticos deixou, porém, de ser tão homogêneo (Bittencourt, 2004, p.

⁶ Em 1824, o governo imperial tornou a educação primária obrigatória. Nas décadas seguintes, não obstante os limites práticos e econômicos para uma efetiva expansão, muito se discutiu sobre a importância desse nível de ensino. Isso porque ele serviria para formar as populações livres e empobrecidas para o trabalho manual, retirando-a das ruas dos centros urbanos em crescimento (Schueler, 1999, p. 9). Na segunda metade do século XIX, os debates abolicionistas e seus desdobramentos legais, como a Lei do Ventre Livre, de 1871, encontraram, também, na educação elementar um meio de assimilar crianças recém-libertas ao trabalho manual e livre (Schueler,

483). O papel do sábio orientado à formação moral das elites foi posto em questão pela necessidade de escolarizar um público que, embora ainda restrito, era mais amplo do que aquele que compreendia apenas os filhos de proprietários rurais. Essa mudança não foi sem efeitos no ensino secundário e na relação entre as editoras e os seus autores. As empresas editoriais não tardaram em perceber as preferências de um novo mercado escolar em formação. No secundário, os livros que resultavam de trabalho em sala de aula caíram nas graças dos professores: “livros do nível secundário com maior sucesso de venda eram os provenientes ‘das aulas’ dos professores” (Bittencourt, 2004, p. 483). Foi nesse momento que os editores passaram a enfatizar, nos títulos e em outros elementos paratextuais, a experiência de seus autores como docentes, como em *Lições de História do Brasil*, de Joaquim Manuel de Macedo (Bittencourt, 2004, p. 478); ou ainda, em *Lições de Português*, de Otoniel Mota, e em *Lições de Arithmetica*, de Euclides Roxo. A partir de 1880, surge, portanto, uma segunda geração de autores de livros didáticos, dentre os quais a experiência pedagógica se tornou definitiva, preferencialmente em instituições reputadas, como o Colégio Pedro II e as escolas militares.

Na qualidade de professor suplementar de português do Colégio Pedro II, Artur de Almeida Torres fez parte de uma expressão tardia dessa segunda geração de autores de livros didáticos, tendo publicado, em 1936, o *Compêndio de Língua Portuguesa* e, em 1945, o *Manual de Língua Portuguesa*, ambos pela Companhia Editora Nacional. Nesse período, tanto o ensino secundário quanto a produção, a importação e a utilização dos livros didáticos foram institucionalizados, refletindo preocupações da burocracia educacional dos governos de Getúlio Vargas. Em um momento de mudança política, a atuação de Torres como professor em uma instituição prestigiosa e como filólogo nos círculos culturais e intelectuais do Rio de Janeiro lhe asseguraram um lugar no disputado catálogo da Nacional e permitiram que seus livros chegassem às mãos dos secundaristas. Essas credenciais já eram, em 1937, apresentadas com entusiasmo por certos impressos do Rio de Janeiro, ciosos da prata da casa:

Entre os que valentemente propugnam, na vanguarda dos defensores da língua portuguesa, no Brasil, está um moço fluminense que já conseguiu ver editados pela grande “Editora Nacional” alguns de seus livros: Arthur de Almeida Torres. [...]. Ele não se limita, como professor do Collegio Pedro II, a ensinar segundo o programma official. [...]. Manejando, com maestria e segurança, a “última flôr do Lacio”, elle conhece-lhe os segredos e é um guieiro seguro capaz de traçar o roteiro para os que necessitam aprender a expressar-se sem erros de linguagem (O Fluminense, 1937, p. 1).

Como muitos dos seus contemporâneos, Torres se formou em direito e, ainda jovem, ingressou no magistério. Paralelamente à docência, escreveu livros dedicados à literatura, à língua e ao seu ensino, alguns deles bem-recebidos pela crítica e pelo mercado. No entanto, algumas vezes, crítica e mercado se contradiziam. Em 1931, o “moço fluminense” publicou

1999, p. 8). Schueler (1999, p. 9-10) lembra, que, no Rio de Janeiro, nos anos 1870, o número de escolas públicas dobrou, passando, ao fim da década, de 45 para 95 estabelecimentos. O ensino primário era, portanto, um instrumento de assimilação social e profissional de livres e recém-libertos pertencentes às camadas empobrecidas da sociedade. O ensino secundário, que dispensava a frequência ao primário, permanecia sendo apanágio das classes dirigentes, ligadas ao trabalho intelectual e à direção do Estado. A expansão da educação primária, a partir dos anos 1870, traduz, pois, uma finalidade de assimilação social e profissional de grupos sociais marginalizados, daí a diversificação do perfil de alunos e, necessariamente, de autores de livros didáticos.

Regência Verbal, que lhe rendeu saudações na imprensa e uma menção honrosa no Prêmio Francisco Alves (Lamego, 1949, p. 1). Os escassos registros na imprensa mostram que, não obstante a boa acolhida, o livro teve longos intervalos entre suas reedições e foi impresso em pequenas tipografias⁷. No mesmo período, o autor publicou o *Compêndio de Língua Portuguesa*. Destinado a um público escolar em (tímida) expansão, o livro teve grande sucesso, traduzido em sucessivas reedições. Entre 1936 e 1958, o *Compêndio* somou 350 mil exemplares impressos⁸ com o selo da “grande Editora Nacional”. A recepção na imprensa foi, entretanto, irregular, mesmo desfavorável. Em sua coluna na *Gazeta de Notícias*, dedicada à língua portuguesa, Horácio Mendes não economizou reprimendas, antecipando um quadro avesso àquele exibido em *O Fluminense*. O “guieiro seguro” para a língua correta figurava como um cultuador de “frioleiras de linguagem” e o professor que ultrapassava os programas oficiais no Pedro II sequer conseguia, em seu *Compêndio*, atendê-los:

Como se sabe, esse programma [programa de português, de 1931] não se limita ao domínio da grammatica. Vae além, abrangendo a formação e a evolução da literatura luso-brasileira. O sr. Almeida Torres não pensou assim e deu-nos, em consequencia, um volume de mera gramatiquice, um resumo de frioleiras de linguagem (Mendes, 1936, p. 5).

Publicado em 1945, o *Manual de Língua Portuguesa* não causou reação na imprensa tampouco convenceu, com suas vendas, os editores. Comparado aos livros anteriores de Torres, foi um fracasso editorial, frustrando a crítica e o mercado. De produção lenta, com o terceiro e último volume publicado apenas em 1951, somou míseros 43 mil exemplares impressos, com longos intervalos entre as reimpressões. Para o mercado de didáticos, o resultado era desanimador, mesmo que se considerasse o pequeno contingente de alunos matriculados no segundo ciclo do secundário (atual ensino médio)⁹.

Apesar desses entraves, o *Manual de Língua Portuguesa* permanece sendo um documento relevante, pois registra, na consumação do fracasso, a memória de práticas editoriais, escolares e linguísticas. Particularmente, em relação ao tema deste trabalho – as políticas educacionais e as linguísticas –, ele conserva os modos de negociação, conflito e associação entre autor, editores e agentes do Estado.

4 Políticas linguísticas explícitas no *Manual de Língua Portuguesa*

No Diário Oficial da União, de 27 de setembro de 1947, o *Manual de Língua Portuguesa* consta, em duas entradas, na lista de livros aprovados pela Comissão Nacional do Livro Didático (CNLD), sob os números de processo 83.562-44 e 82.616-45. A Comissão foi criada em 1938, com sua regulamentação disposta no segundo capítulo da lei do livro didático, o decreto-lei nº 1.006, de 30

⁷ A primeira edição de *Regência verbal* foi publicada, em 1931, pela Tipografia São Benedito. A segunda edição, publicada em 1935, foi preparada pela Organização Simões.

⁸ Os dados dos livros didáticos da Nacional estão disponíveis no acervo da editora, sediado no Centro de Memória e Pesquisa Histórica, da Universidade Federal de São Paulo, *campus* de Guarulhos (SP).

⁹ Para mais detalhes sobre Torres e seus livros, bem como para análises de seus aspectos diversos, cf. Imbruniz, 2025, p. 128.

de dezembro de 1938. Apesar de alguns embaraços de funcionamento, o colegiado atuou regularmente e, sobretudo, participou de disputas pelo projeto de educação nacional (Filgueiras, 2013, p. 186). A presença do *Manual* na lista de aprovados da Comissão enseja três comentários, que podem elucidar sua relação com as políticas educacionais e linguísticas vigentes.

Em primeiro lugar, merece destaque o fato de que, no decreto-lei de 1938, estavam dispostos os critérios de avaliação de livros didáticos, que se dividiam em ideológicos e pedagógico-metodológicos. Na maioria dos casos, os livros eram reprovados por violações dos critérios pedagógico-metodológicos (Filgueiras, 2013, p. 167-168). Em segundo lugar, o fato de que, entre os critérios pedagógico-metodológicos, havia dois artigos que incidiam na relação entre livros didáticos, políticas educacionais e políticas linguísticas, a saber, os artigos 21 e 23:

Art. 21. Será ainda negada autorização de uso ao livro didático: a) que esteja escrito em linguagem defeituosa, quer pela incorreção gramatical, quer pelo inconveniente ou abusivo emprego de termos ou expressões regionais ou da gíria, quer pela obscuridade do estilo. [...] c) que esteja redigido de maneira inadequada, pela violação dos preceitos fundamentais da pedagogia ou pela inobservância das normas didáticas oficialmente adotadas, ou que esteja em desacordo com os preceitos essenciais da higiene da visão [...]. Art. 23. Não será autorizado o uso de livro didático que, escrito em língua nacional, não adote a ortografia estabelecida pela lei (Brasil, 1938, s/p, negritos meus).

Por fim, o fato de que a Nacional foi a editora que mais submeteu livros à Comissão e a que mais obteve aprovações. Esse alto nível de sucesso era possível, porque a editora e o colegiado compartilhavam um projeto de educação (Filgueiras, 2013, p. 175). Não por acaso, alguns de seus membros eram autores da Nacional. Livro didático publicado pela então maior editora do país, o *Manual* tinha toda chance de ser aprovado pela Comissão, como, de fato, foi. O cumprimento dessas exigências legais merece, contudo, um comentário, que permita a observação dos modos diretos e indiretos de interação entre autor, editor e Estado.

Como era de esperar, o livro de Torres respeita, estritamente, a norma-padrão (de orientação lusitana) e não utiliza quaisquer regionalismos ou gírias (art. 21), constatação alinhada com as críticas que, uma década antes, recebeu de Mendes (1936, p. 5) pelo seu apego às “frieiras de linguagem” e, ainda, com os encômios de *O fluminense* (1937, p. 1), que o considerava um dos “defensores da língua portuguesa”. Em relação aos programas de ensino (art. 21), logo após a folha de rosto, na primeira edição do volume destinado à segunda série, de 1945, autor e editores exibem o *Programa de português dos cursos clássico e científico*, da segunda série. A organização das lições do livro corresponde parcialmente a uma reprodução do programa, o que sugere um alinhamento quase completo. Esses indícios devem ser considerados cautelosamente, pois, como também observou Mendes (1936), Torres não se furtava de adaptar os programas oficiais, privilegiando seus temas de interesse. *O Fluminense* (1937) sinaliza a mesma atitude, mas, dessa vez, em chave positiva: o professor ultrapassava os programas. Quanto à nova ortografia (art. 23), a Nacional passou a adotá-la, desde 1938, conforme exigido por lei. É possível verificar essa adoção imediata por meio das fichas de edição disponíveis no acervo da editora, documento em que eram reunidos dados editoriais de cada um dos livros produzidos pela casa (para detalhes, cf. Imbruniz, 2023, p. 9094). No início da década de 1930, as fichas apresentavam o campo “ortografia”, a ser preenchido com “velha” ou “nova”, indicando aquela que foi utilizada na composição do livro. A partir de meados de 1943, o campo

“ortografia” deixou de aparecer nas fichas, sugerindo que, a partir de então, todos os livros da editora provavelmente passaram a adotar a ortografia oficial, o que dispensaria o registro dessa informação. Publicado em 1945, o *Manual* respeita essa exigência.

Esses três pontos, sumariamente abordados, ilustram como um livro didático, submetido à avaliação oficial, pode servir como móvel de políticas educacionais e linguísticas. A respeito da política educacional, é ponto pacífico que os livros didáticos são uma de suas principais formas de imposição. O alinhamento do *Manual* ao programa de português de Capanema (cf., aqui mesmo, nota 4) não causa espanto, mesmo que ele possa ser posto em suspeição pelos seus críticos e até mesmo pelos seus defensores. Seja como for, pelo cumprimento dos programas ou pelo convencimento dos avaliadores de suas adaptações, Torres e seu *Manual* receberam a chancela da oficialidade.

A respeito das políticas linguísticas, pode-se notar que elas regulam não só a relação entre autor, professor e aluno, mas a relação entre autor, editora e Estado, o que põe em destaque o alinhamento entre a Comissão Nacional do Livro Didático e a Companhia Editora Nacional (Filgueiras, 2013, p. 175). Imediatamente cumpridas pela Nacional, as exigências de uma norma linguística que recusa usos “defeituosos” e “regionais” e de uma ortografia oficial são parte de uma política linguística explícita do Estado Novo. Embora não seja uma política exclusivamente linguística, pois seu principal objetivo era regular a produção e a circulação de livros didáticos, ela é explícita por não se furtar à imposição do chamado “idioma nacional” em oposição a todas as outras manifestações linguísticas de então, como as línguas dos povos originários, as línguas de migração e mesmo as variações regionais. Não por acaso, o dito “idioma nacional” se confundia com a escrita e com a ortografia, alinhadas a uma representação da norma lusitana.

Ao recuperar, brevemente, a trajetória de Torres e do *Manual*, quis mostrar, portanto, que, entre os critérios de ascensão de um autor e os de publicação de um livro didático, as políticas educacionais e linguísticas são incontornáveis, principalmente em períodos de forte repressão estatal. No caso das políticas linguísticas, deve-se lembrar, porém, que todos os livros didáticos estavam submetidos a elas, fossem eles dedicados à língua materna ou não. Em outras palavras, como o *Manual*, um livro de ciências naturais ou de matemática também deveria adotar a ortografia oficial, usar “linguagem adequada” e evitar “expressões regionais”. É razoável se questionar, porém, se, em um livro de português, haveria lugares e meios menos aparentes de regulação dos usos e das expectativas em relação à língua, já que nele a linguagem é abordada como fenômeno social, representação coletiva, convenção gramatical e instrumento político e oficial.

Os legisladores estavam atentos ao potencial sociopolítico da língua materna e do seu ensino. Em sua exposição de motivos para a reforma educacional de 1942, o ministro Gustavo Capanema justificou, não sem interesse, a expansão dos estudos de português para todas as séries do secundário e o aumento da carga horária da disciplina. No texto, Capanema defende que “o conhecimento seguro da própria língua constitui para uma nação o primeiro elemento de organização e de conservação de sua cultura”, na medida que “o cultivo da língua nacional interessa à própria existência da nação, como unidade espiritual e como entidade independente e autônoma” (Capanema, 1952, p. 25). Para atingir os fins ligados à organização, à conservação e à unificação da nação pela língua, o livro didático de português pode ser considerado um objeto poderoso, pois, ao apresentar certas concepções de linguagem e excluir outras, esse artefato aprofunda, silenciosamente, um processo que, às claras, inicia-se com atos jurídicos ligados ao uso positivo de formas da língua ou à regulação do seu *status*.

5 Política linguísticas implícitas no *Manual de Língua Portuguesa*

Como se viu na seção anterior, o compromisso com as exigências da reforma Capanema, efetivo ou simulado, era uma das condições para a publicação do *Manual de Língua Portuguesa*. Isso porque a obediência aos programas oficiais estava não só no horizonte da Comissão Nacional do Livro Didático, mas, também, no dos editores, dos diretores de escola, dos professores e, necessariamente, no horizonte daqueles que comprariam os livros. Não surpreende, portanto, que a organização do *Manual* reproduza o programa de português do colegial, publicado na portaria ministerial nº 87, de 23 de janeiro de 1943. O livro tem uma seção dedicada à literatura, outra à gramática e, enfeixando as duas anteriores, uma reservada a textos antológicos. Pela oposição que exibem em relação à linguagem e à língua, traduzidas em concepções de linguagem a serem adotadas ou recusadas, que serão abordadas a seguir, res-trinjo-me às seções de literatura e de gramática¹⁰.

A seção dedicada à literatura conta com três unidades, cada uma delas com até quatro capítulos. As unidades localizam os temas literários no tempo histórico (era medieval, clássica ou moderna) e os capítulos apresentam as sucessivas escolas literárias (trovadorismo, humanismo, renascimento e assim sucessivamente). No interior dos capítulos, as características de estilo das escolas literárias são minuciosamente apresentadas, seguidas de um desfile de grandes autores, cujos dados biográficos são narrados aos alunos segundo sua importância na tradição literária. Camões recebe, portanto, mais espaço do que Júlio Moreira, que teve a biografia reduzida a menos de uma linha: “nasceu em 1854 e faleceu em 1911” (Torres, 1945, p. 38). Trata-se, pois, de uma abordagem tradicional do ensino de literatura, baseada na identificação de supostas características de estilo, na análise de filigranas textuais e na associação entre vida e obra, sempre de acordo com uma perspectiva cronológica e teleológica da história, conjunto de gestos que, no livro, dispensam até mesmo a leitura de excertos literários¹¹. Tradicionalmente, essa abordagem se associa à forma de racionalidade da linguagem como expressão do pensamento. Essa forma de racionalidade oculta, na condição de categoria clas-

¹⁰ Para uma análise da antologia, cf. Imbruniz; Haynal, 2021, p. 177.

¹¹ Para Cosson (2021, p. 76), a abordagem tradicional de ensino de literatura encontra eco nas práticas da *Ratio Studiorum*, modelo desenvolvido por padres jesuítas para seu propósito educacional e, sobretudo, religioso. Para ensinar latim e difundir a moral cristã, esses padres empregavam exercícios variados, como os de composição, explicação e reprodução de obras clássicas, expurgadas, como é de supor, de seu caráter pagão. A literatura clássica era, pois, subterfúgio e suporte para o ensino da língua. Ainda segundo o mesmo autor, a *Ratio Studiorum* orientou, desde o século XIX, a organização do ensino das literaturas vernáculas. Na condição de matéria, o ensino de literatura vernácula se subordinava, nesse período, ao de língua materna, sendo os textos literários abordados como fragmentos de obras canônicas e exemplos de usos modelares da língua. Essa abordagem tradicional se opõe, segundo Cosson (2021, p. 87), ao *letramento literário*, cujos pressupostos são o contato direto do aluno com o texto literário, a criação de comunidades de leitura e o desenvolvimento de competências literárias. Informada por um conceito de letramento como prática social e um de leitura literária como exercício de humanização e construção da subjetividade, a noção de *letramento literário* traduz esforços teóricos e metodológicos do presente para renovar esse objeto de ensino. A justaposição do ensino tradicional de literatura aos *letramentos literários* mostra, sobretudo, que os modos de ensinar e de conceber o ensino são devedores de condições sociais e históricas, cuja consideração é imprescindível para análise de materiais como livros didáticos. Em outras palavras, Torres e seu *Manual* são, na mesma medida, um retrato do ensino tradicional de literatura e da distância a ser percorrida por novas abordagens teóricas, metodológicas e pedagógicas desse objeto de ensino.

sificatória, a complexidade de concepções de linguagem concorrentes no mesmo livro, configuração que passo a explorar.

Esse modo de ensinar a literatura sugere um ponto de entrada pertinente para a observação de expectativas sobre a língua e o seu ensino. Não são raras as ocasiões em que, ao apresentar tais e tais literatos, Torres emprega adjetivos para classificar seus usos da língua. Esse procedimento levou, no curso do livro, à elaboração de um arcabouço de critérios de avaliação capaz de esclarecer, para os estudantes, quais seriam os bons e maus modelos. Samuel Usque tinha a linguagem “correta, pura e elegante” (Torres, 1945, p. 23). A linguagem do Frei Bernardo de Brito recebeu, no entanto, uma avaliação menos generosa: era “geralmente correta” e, no mais das vezes, “sóbria” e “fria” (Torres, 1945, p. 27). Os critérios empregados por Torres eram variados, discricionários e, mesmo, curiosos. Mais do que os pitorescos, são, no entanto, os critérios recorrentes aqueles de maior interesse: elegância, correção e pureza.

Não é casual que a elegância, a correção e a pureza sejam os critérios mais empregados para a avaliação dos usos linguísticos dos autores apresentados aos alunos. Pelo contrário, os três modificadores convergem para a mesma direção. Tanto a elegância quanto a correção podem ser reduzidas à pureza, como, de fato, ocorre no *Manual*. Torres concebe, portanto, a linguagem como uma forma que, embora pura, pode ser corrompida, por isso a fixação de critérios valorativos. Tradicionalmente, essa concepção de linguagem é encerrada na forma de racionalidade da linguagem como expressão do pensamento. Esse pensamento seria tão puro quanto fosse correto o uso da língua. Mais do que fixar uma forma de racionalidade, esses critérios estabelecem, contudo, uma oposição entre mesmo e diferente. De um lado, figurariam os modelos de linguagem, sempre de expressão escrita. Dentre esses modelos, estariam um certo conjunto de autores da tradição literária e o próprio Torres. De outro, figurariam as expressões desviantes da forma pura. Dentre esses “contramodelos”, estariam autores injustamente alçados à tradição literária e, como um não dito, o *povo*, coletividade abstrata e convocada por alusão.

Como se verá, uma concepção de linguagem como forma pura não é, entretanto, apenas expressão de uma postura conservadora individual. Ela é, antes disso, a materialização de expectativas coletivas. Ao supor a ascendência de um grupo social sobre o outro, o purismo linguístico serve a propósitos religiosos, étnicos e nacionalistas, apresentando-se como uma das mais poderosas expressões de política linguística implícita (Spolsky, 2004, p. 24). No presente caso, a adoção do purismo linguístico vai ao encontro do que pode ser considerada uma política linguística não oculta, mas não visível (Foucault, 2008, p. 123) do Estado Novo. Para o ministro Capanema, a ampliação do ensino de língua materna concorreria para a organização, a conservação e a unificação nacional. Esses predicados traduzem uma expectativa de purismo social e político, do qual a pureza da língua seria uma manifestação catalisadora. Esse alinhamento entre os motivos ministeriais e a concepção de linguagem de um livro didático de língua materna não se reduz, contudo, ao cumprimento de uma exigência legal, como no caso do expurgo dos regionalismos ou da adoção de certa ortografia. Estas últimas são, sem dúvida, expressões burocráticas de uma expectativa de purismo social realizadas conscientemente pelos legisladores. No caso em questão, por outro lado, o que se mostra são sujeitos que, em um tempo histórico particular, compartilhavam a mesma posição social. Essa posição definiria as condições para manifestações sociopolíticas e, sobretudo, para o retorno de já ditos e não ditos. O apego de Torres à pureza da língua e à defesa de Capanema da “unidade espiritual” da nação são formulações parafrásticas de um discursivo prévio (Pêcheux, 2010,

p. 77), que, quando observado apenas como uma manifestação sociopolítica, se dissimulam como gosto pessoal ou burocracia política. Como expressão sócio-histórica, no entanto, encontra-se o retorno de não ditos dirigidos à organização social e a sua relação com a língua. Trata-se de um processo de purificação do outro, que, mesmo não tendo sido iniciado no Estado Novo, ganhou, nesse período, uma feição particular, marcada em políticas educacionais e linguísticas e em livros didáticos. Não por outro motivo, as interdições, pela afirmação de usos modelares, se dirigem àqueles que, desde muito tempo, são total ou parcialmente excluídos da sociedade. Aludidos indiscriminadamente como *povo*, esses sujeitos congregariam classes menos abastadas, diferenciadas entre si, ainda, por outros marcadores sociais relevantes (raça, gênero etc.). O arcabouço de critérios de Torres não se limita, portanto, à classificação de “estilos literários”, à expressão de gosto estético ou ao respeito à legislação, tarefas cumpridas pelo autor em outras passagens do seu livro. Ao apresentar modelos de boa linguagem, esses critérios suscitam o retorno de um não dito que, sob a forma de concepção de linguagem, antecede, circunscreve e transcende o livro didático e certas formas de racionalidade capazes de expressar essa concepção de linguagem, mesmo à revelia de intenções. Uma língua elegante, correta e pura é aquela de um povo elegante, correto e puro.

A imposição de uma língua elegante, correta e pura a um povo que, supostamente, ostenta os mesmos predicados é, com efeito, um gesto de imposição de uma única concepção de linguagem, materializada nas relações dos sujeitos com seus iguais e seus diferentes e, também, em certas formas de racionalidade, como a linguagem como expressão do pensamento. O principal corolário dessa imposição é o apagamento de todas as outras manifestações linguísticas concorrentes àquela que se pretende dominante. Como faziam as políticas linguísticas explícitas, que interditavam regionalismos e variantes não padrão da língua, as concepções de linguagem que sustentam políticas linguísticas implícitas negam um lugar de existência para todas as línguas que não sejam o vernáculo elegante, correto e puro. Ignoram-se, portanto, línguas e povos originários, línguas e grupos de migração e variantes linguísticas e grupos diversos que compõem a sociedade. Os efeitos das políticas linguísticas explícitas e implícitas não são apenas simbólicos, mas também materiais. Isso porque, no caso ora analisado, servem para negar o direito de cidade a uma fatia expressiva da sociedade.

Se, na seção de literatura, as atenções se voltam para os modelos de boa linguagem, cujas contrapartes são interditadas pelo silêncio, os polos se invertem na seção de gramática. De menor extensão, a rubrica dedicada às normas convencionais da língua se reduz a duas unidades, que não ultrapassam dois capítulos. A primeira delas se dedica à concordância verbal e aos empregos do infinitivo pessoal e impessoal. Já a segunda, à regência verbal. Trata-se, uma vez mais, de uma reprodução fiel do programa do colegial.

Nessa seção, o recurso às práticas da tradição filológica é, como a classificação da linguagem literária, um ponto de entrada pertinente para a investigação sobre as expectativas em torno da língua e do seu ensino. Ao tratar da regência verbal, tema que lhe era caro, Torres seleciona e apresenta verbos de “intêresse sintático” (Torres, 1945, p. 63), isto é, aqueles que, recorrentemente, escapavam aos escreventes¹². Tanto a apresentação desses verbos quanto a discussão de suas particularidades ocorre segundo a lógica da abonação. Torres recupera os

¹² Refiro-me aos escreventes, porque, nos programas e nos livros do período, havia quase uma equivalência entre língua e escrita, questionada, desde muito tempo, nos estudos da linguagem.

verbos de interesse nas obras de gramáticos reconhecidos, que abonam ou desabonam usos da língua. Esse é o caso do verbo *pagar*:

Diz Mário Barreto: “Outro verbo que alguns empregam transitivamente, com acusativo de pessoa, é o verbo pagar: ‘Augusta não tinha com que pagar o mestre.’ [...] Construção de todo rejeitável. O regime de pessoa deve aqui aparecer como complemento indireto: ‘Dava-lhe dinheiro para pagar ao mestre de latim’” (Novos Estudos, p. 330) (Torres, 1945, p. 76).

Na citação de Barreto, destaca-se a caracterização do falante da variedade considerada por ele como não padrão: “alguns empregam”. Trata-se de um nome indefinido, que não lhe atribui característica alguma. O efeito imediato desse procedimento é a suspensão do escrevente fora do tempo e do espaço, substituindo-o por uma suposta avaliação pura do fato linguístico. Em seguida, o gramático afirma que a regência examinada seria de todo “rejeitável”, uso interessado, já que recupera, como não ditos, termos paradigmaticamente próximos, como reprovável, censurável, inaceitável, entre outros. Na formulação de Barreto, esses qualificativos incidem sobre a construção examinada (regência do verbo “pagar”). No entanto, quando citada em um livro didático de língua materna, tanto a seleção do trecho quanto os enunciados que ele evoca passam a remeter, também, ao escrevente imaginado, cuja conduta linguística, mas também social, deve ser corrigida. Em outras palavras, se a forma em questão é rejeitável, é também o escrevente que a coloca em circulação.

Mais adiante, Torres continua convocando gramáticos para discutir a regência do verbo “pagar”:

Também Cândido de Figueiredo assim se manifestara a respeito: “é verdade que o nosso velho Morais empregou as expressões: pagar as tropas, os criados, as dívidas. As dívidas, bem está; mas pagar os criados parece-me que foi precipitação ou lapso do respeitado dicionarista, que viveu no Brasil e se deixou, talvez insensivelmente, influenciar pela linguagem dos que o cercavam ou com êle viviam; tanto mais que, aduzindo textos clássicos, não citou um único texto em abono daquela sintaxe: pagar os criados” (O que se não deve dizer, III, p. 294) (Torres, 1945, p. 76).

Na citação de Cândido de Figueiredo, repete-se o procedimento de suspensão do escrevente. Ele não pertence a nenhum grupo social, como camponeses, escravizados ou artesãos. Essa figura compõe apenas a massa indistinta que habitou o tempo e o espaço do “respeitado dicionarista”, esse, sim, associado a um grupo social e profissional. Ao abordar a regência desviante nos usos de Morais, o autor a considera “precipitação” ou “lapso”. O desvio é, portanto, apresentado como uma atitude irrefletida ou um descuido da “língua pura”, diferentemente da natureza imanente do erro dos escreventes “comuns”. Mais interessante talvez seja a justificativa do “lapso” atribuído a Morais: a influência (ou contaminação?) da massa indistinta de brasileiros que o cercava. A corrupção da língua seria não só algo a se evitar, mas também algo de que o falante deveria, profilaticamente, proteger-se. O título e o subtítulo do livro de Cândido Figueiredo, do qual Torres extraiu o excerto citado, não esconde desejo de interdição do outro (“o que não se deve dizer”) e a remissão à doença (“enfermidades da linguagem”): *O que se não deve dizer: bosquejos e notas de filologia portuguesa ou consultório popular de enfermidades da linguagem*. Para liquidar a regência transitiva direta do verbo pagar

com objetos animados, o autor citado por Torres afirma que Moraes não “abonou” a construção com textos clássicos. São, portanto, os usos populares e a sua ausência no *corpus* literário que fazem dessa regência uma doença contagiosa e, portanto, uma forma impura.

Nas seções de literatura e de gramática, a concepção de linguagem como forma pura se manifesta como duas faces de uma mesma moeda. Na primeira seção, a elegância, a correção e a pureza dos modelos de linguagem concorrem para uma expectativa de purismo social e político. Na segunda, a identificação e a interdição das doenças e das impurezas de linguagem se prestam à mesma finalidade. A confluência dessas posições mostra, que, ao se ensinar língua materna, são postas em circulação, por objetos impressos, como livros didáticos, expectativas sociais e políticas de diferentes grupos, organizadas em relações antagônicas e hierárquicas. Esse modo de funcionamento discursivo não é próprio de Torres, de Capanema ou da Companhia Editora Nacional, sendo identificável, segundo certas particularidades, nos enunciados de outros sujeitos em outros períodos históricos. O retrato pintado por esse trio permanece sendo, contudo, uma ilustração da complexa relação entre políticas educacionais e linguísticas e livros didáticos.

6 Considerações finais

A partir de certas relações entre autores, editores e agentes de Estado, pretendi analisar as interações entre políticas educacionais e linguísticas e livros didáticos de língua materna. Esse objetivo foi conduzido por meio da investigação do *Manual de Língua Portuguesa*, de Artur de Almeida Torres, publicado, pela Companhia Editora Nacional, em respostas às políticas de educação e de língua que foram gestadas no Estado Novo brasileiro. A análise consistiu, de um lado, na confrontação do livro com a legislação vigente no período e, de outro, no exame de suas concepções de linguagem, noção que, de maneira crítica, tenho explorado em trabalhos recentes (Imbruniz, 2022, p. 11; 2025, p. 115). Como primeiro passo, descrevi e analisei os modos de cumprimento da legislação por autor e editores, particularmente, em relação aos programas de ensino, a usos de linguagem explicitamente interditados e à adoção da ortografia oficial. Num segundo passo, descrevi e analisei as concepções de linguagem presentes no *Manual*, sublinhando sua aproximação de pretensões sociopolíticas de setores do Estado Novo. A fabricação de uma nação organizada e unificada, a partir da conservação de costumes e valores dominantes, encontrava, pois, eco tanto no rigoroso cumprimento das exigências legais quanto numa concepção de linguagem como forma pura, capaz de exprimir uma expectativa de purismo social e político.

Esse percurso sugere que o livro didático de língua materna, quando analisado conjuntamente com políticas educacionais e linguísticas, é capaz de oferecer pistas sobre expectativas sociais e políticas atreladas à língua e ao seu ensino. Embora, tradicionalmente, reconheça-se a pertinência do livro didático na investigação das políticas educacionais e, por vezes, na de políticas linguísticas explícitas, as políticas implícitas raramente são postas em exame. Trata-se, pois, de uma primeira contribuição deste trabalho, que, segundo uma orientação teórico-metodológica discursiva, debruçou-se sobre aquilo que, embora não oculto, nem sempre é visível nas políticas educacionais e linguísticas (Foucault, 2008, p. 123). Uma segunda contribuição consiste na organização e na sistematização da produção de um autor e de uma editora que, em seu período de atuação, tiveram grande sucesso entre professores de língua materna e seus

alunos. Longe da biografia ou da memória institucional, esse esforço se orientou pela tentativa de inserir essas figuras na história social e política do Brasil, destacando, nessa história, o papel central conferido aos fenômenos de linguagem. Uma terceira e última contribuição se mostra no plano teórico-conceitual. A combinação da análise de documentos legislativos, editoriais e didáticos foi parcialmente orientada pela noção de concepção de linguagem, que, embora frequente no campo da linguística aplicada brasileira, apresenta delimitação pouco precisa, daí a necessidade de um trabalho crítico de definição e operacionalização em materiais particulares, como aqueles que foram investigados no presente artigo. Ancorado nos estudos da linguagem, este trabalho é, portanto, um esforço interdisciplinar de encontrar a língua, o ensino e a edição na história, no discurso, na sociedade e na política.

Finaciamento

Pesquisa financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Processos nº 2020/03933-1, 2022/13328-3 e 2025/08677-7.

Referências

- AUTHIER-REVUZ, J. O estrato meta-enunciativo, lugar de inscrição do sujeito em seu dizer: implicações teóricas e descritivas de uma abordagem literal. O exemplo das modalidades irrealizantes do dizer. *Matraga*, Rio de Janeiro, v. 15, n. 22, p. 33-63, 2008. Disponível em: <https://bit.ly/3wNIDrK>. Acesso em: 08 mai. 2025.
- BATISTA, A. Um objeto variável e instável: textos, impressos e livros didáticos. In: ABREU, Márcia (org.). *Leitura, história e história da leitura*. 1. ed. Campinas: Mercado das Letras/Fapesp/ALB, 1999. p. 529-576.
- BITTENCOURT, C. Autores e editores de compêndios e livros de leitura (1810-1910). *Educação e Pesquisa*, São Paulo, v. 30, n. 3, p. 475-491, 2004. DOI: <https://doi.org/10.1590/S1517-97022004000300008>. Acesso em: 19 jun. 2021.
- BOMENY, H. Três decretos e um ministério: a propósito da educação no Estado Novo. In: PANDOLFI, D. (org.). *Repensando o Estado Novo*. 1. ed. Rio de Janeiro: FGV, 1999. p. 137-166.
- BRASIL. Decreto nº 19.890, de 18 de abril de 1931. Dispõe sobre a organização do ensino secundário. *Diário Oficial da União*, Rio de Janeiro, RJ, 1 maio 1931. Disponível em: <https://bit.ly/38jVV4H>. Acesso em: 08 mai. 2025.
- BRASIL. Decreto-lei nº 1.006, de 30 de dezembro de 1938. Estabelece as condições de produção, importação e utilização do livro didático. *Diário Oficial da União*, Rio de Janeiro, RJ, 5 jan. 1939. Seção 1. Disponível em: <https://bit.ly/3n08FBn>. Acesso em: 08 mai. 2025.
- BRUILLARD, É. Les manuels scolaires questionnés par la recherche. In: BRUILLARD, Éric (org.). *Les manuels scolaires : regards croisés*. 1. ed. Caen: CRDP/Canopé, 2005. p. 3-22.
- BUNZEN, C. *Livro didático de língua portuguesa: um gênero do discurso*. 2005. 168f. Dissertação (Mestrado em Linguística Aplicada) — Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2009. DOI: <https://doi.org/10.47749/T/UNICAMP.2005.361369>.

- CAPANEMA, G. Exposição de motivos da lei orgânica do ensino secundário. In: BRASIL. *Ensino secundário no Brasil: organização, legislação vigente, programas*. 2. ed. Rio de Janeiro: Ministério da Educação e Saúde, 1952. p. 22-28.
- CALVET, L.-J. *As políticas linguísticas*. São Paulo: Parábola, 2018.
- CASSIANO, C. *O mercado do livro didático no Brasil do século XXI: a entrada do capital espanhol na educação nacional*. 1. ed. São Paulo: Editora Unesp, 2013.
- CHOPPIN, A. Contexto científico da pesquisa sobre a edição escolar local, nacional e mundial. Tradução e apresentação de Circe Bittencourt. *Educ. Pesqui.*, v. 46, e460100301, 2020. DOI: <https://doi.org/10.1590/S1517-97022020460100301>.
- CORREIO DA MANHÃ. *Correio da manhã*, Rio de Janeiro, 27 jun. 1948. p. 3. Disponível em: <https://bit.ly/3k458A4>. Acesso em: 08 mai. 2025.
- COSSON, R. Ensino de literatura, leitura literária e letramento literário: uma desambiguação. *Interdisciplinar*, v. 35, p. 73-92, 2021. DOI: <https://doi.org/10.47250/intrell.v35i1.15690>.
- DENIZOT, Nathalie. *La culture scolaire : perspectives didactiques*. 1. ed. Pessac Cedex: PUB, 2021.
- FILGUEIRAS, J. As avaliações dos livros didáticos na Comissão Nacional do Livro Didático: a conformação dos saberes escolares nos anos 1940. *Revista Brasileira de História da Educação*, Maringá, v. 31, n. 1, p. 159-192, 2013. Disponível em: <https://bit.ly/3q9ondg>. Acesso em: 08 mai. 2025.
- FILGUEIRAS, J. *Os processos de avaliação de livros didáticos no Brasil (1938-1984)*. 2011. 349 f. Tese (Doutorado em Educação) – Pontifícia Universidade Católica de São Paulo. Disponível em: <https://bit.ly/3vHSCvZ>. Acesso em: 08 mai. 2025.
- FOUCAULT, M. *A arqueologia do saber*. 7. ed. Rio de Janeiro: Forense Universitária, 2008.
- FOUCAULT, M. *A ordem do discurso*. 5. ed. Tradução: Laura Fraga de Almeida Sampaio. São Paulo: Loyola, 1996.
- GERALDI, João Wanderley. Concepções de linguagem e ensino de português. In: GERALDI, João Wanderley (org.). *O texto na sala de aula*. 1. ed. São Paulo: Ática, 2011. p. 33-38.
- IMBRUNIZ, C. Concepções de linguagem em livros didáticos de português (1950-1963): espaço de inter-incompreensão polêmica. *Investigações*, v. 35, n. 2, p. 1-26, 2022. DOI <https://doi.org/10.51359/2175-294x.2022.253201>. Disponível: <https://bit.ly/3ROZZhK>. Acesso em: 26 fev. 2026.
- IMBRUNIZ, C. O livro didático de português no Discurso e na História: as fichas de edição de O nosso idioma (1936-1950). *Fórum linguístico*, v. 20, n. 3, p. 9085-9101, 2023. DOI <https://doi.org/10.5007/1984-8412.2023.e77205>.
- IMBRUNIZ, C. *O mercado editorial e o livro didático de português no ensino secundário*. 2025. 387f. Tese (Doutorado em Filologia e Língua Portuguesa) — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2025. DOI: <https://doi.org/10.11606/T.8.2025.tde-03102025-152729>.
- IMBRUNIZ, C.; HAYNAL, T. A pureza e a riqueza da língua na antologia do Manual de Língua Portuguesa (1945-1951). In: CAMPOS, M. I.; SANTOS, T. J. F. (org.). *Discurso e ensino na linguística aplicada: propostas e interseções*. 1. ed. São Paulo: FFLCH/USP, 2021, v. 2, p. 172-205. DOI: <https://doi.org/10.11606/9786587621944>.

- GUILHAUMOU, J.; MALDIDIER, D. Efeitos do arquivo: a análise do discurso no lado da história. In: ORLANDI, E. (org.). *Gestos de leitura: da história no discurso*. 4. ed. Campinas: Editora Unicamp, 2014. p. 169-191.
- LAMEGO, L. Poetas de Resende. *O Fluminense*, Niterói, 11 out. 1949. p. 1. Disponível em: <https://bit.ly/3k4wMwO>. Acesso em: 08 mai. 2025.
- LEMONS, C. de. Questioning the notion of development: the case of language acquisition. *Culture & Psychology*, Londres, v. 6, n. 2, p. 169-182, 2000. DOI: <https://doi.org/10.1177/1354067X0062006>.
- LOUICHON, B. Pourquoi il faut continuer à s'intéresser aux manuels scolaires... *Le français aujourd'hui*, v. 3, n. 194, p. 107-118, 2016. DOI <https://doi.org/10.3917/lfa.194.0107>.
- MAINGUENEAU, D. *Discurso literário*. 2. ed. São Paulo: Contexto, 2006.
- MAINGUENEAU, D. *Novas tendências em análise do discurso*. 3. ed. Campinas: Pontes, 1997.
- MENDES, H. Sobre a língua portuguesa. *Gazeta de Notícias*, Rio de Janeiro, 10 mai. 1936, p. 5. Disponível em: <https://bit.ly/3ABpKhf>. Acesso em: 08 mai. 2025.
- MONTALVÃO, S. Gustavo Capanema e o ensino secundário no Brasil: a invenção de um legado. *História da educação*, v. 25, p. 1-31, 2021. DOI: <https://doi.org/10.1590/2236-3459/108349>.
- MUNAKATA, K. O livro didático como mercadoria. *Pro-posições*, v. 23, n. 3, p. 51-66, 2012. DOI <https://doi.org/10.1590/S0103-73072012000300004>.
- O FLUMINENSE. Estudos de português. *O Fluminense*, Niterói, 4 ago. 1937. p. 1. Disponível em: <https://bit.ly/3Axyapl>. Acesso em: 08 mai. 2025.
- PALHARES, I. Secretário de Tarcísio diz que abrir mão de livros didáticos foi um dos maiores erros de sua vida. *Folha de S. Paulo*, 27 set. 2023. Disponível em: <https://tinyurl.com/5e8wbf5z>. Acesso em: 26 fev. 2026.
- PÊCHEUX, M. Análise automática do discurso (AAD-69). In: GADET, F.; HAK, T. (orgs.). *Por uma análise automática do discurso*. 4. ed. Campinas: Editora Unicamp, 2010. p. 59-158.
- PÊCHEUX, M. *Discurso: estrutura ou acontecimento*. 7. ed. Campinas: Pontes, 2015.
- SCHIFFMAN, H. *Linguistic culture and language policy*. 1. ed. London: Routledge, 2002.
- SCHUELER, A. Crianças e escolas na passagem do Império para a República. *Revista Brasileira de História*, v. 19, n. 37, p. 1-14, 1999. DOI: <https://doi.org/10.1590/S0102-01881999000100004>.
- SHOHAMY, E. *Language policy: hidden agendas and new approaches*. 1. ed. Londres: Routledge, 2006.
- SILVA, E. A pesquisa em política linguística: histórico, desenvolvimento e pressupostos epistemológicos. *Trabalhos em Linguística Aplicada*, Campinas, v. 52, n. 2, p. 289-320, 2013. DOI: <https://doi.org/10.1590/S0103-18132013000200007>.
- SOUZA, R. A renovação do currículo do ensino secundário no Brasil: as últimas batalhas pelo humanismo (1920-1960). *Currículo sem fronteiras*, v. 9, n. 1, p. 72-90, 2009. Disponível em: <https://tinyurl.com/4ccy8x3z>. Acesso em: 22 fev. 2026.
- SPOLSKY, B. *Language policy*. 1. ed. Cambridge: Cambridge University Press, 2004.
- TORRES, A. *Manual de língua portuguesa: segunda série dos cursos clássico e científico*. 1. ed. São Paulo: Nacional, 1945.

Educação bilíngue para surdos em Portugal e Brasil: testemunhos de professores de Língua Gestual Portuguesa e Língua Brasileira de Sinais

Bilingual Education for Deaf People in Portugal and Brazil: Testimonies of Portuguese Sign Language and Brazilian Sign Language Teachers

Lodenir Becker Karnopp

Universidade Federal do Rio Grande do Sul (UFRGS) | Porto Alegre | RS | BR
lodenir.karnopp@ufrgs.br
<https://orcid.org/0000-0002-5370-5587>

Cristina Gil

Instituto Politécnico de Setúbal (IPS)
Setúbal | PT
Universidade Católica Portuguesa (UCP)
Lisboa | PT
<https://orcid.org/0000-0002-1356-998X>

Paulo Vaz de Carvalho

Universidade Católica Portuguesa (UCP)
Lisboa | PT
paulo.carvalho@ucp.pt
<https://orcid.org/0000-0001-5345-3896>

Resumo: A educação de Surdos apresenta desafios na implementação da modalidade de educação bilíngue, no âmbito das políticas linguísticas e educacionais vigentes. A partir da discussão sobre a configuração da educação bilíngue em espaços escolares em Portugal e no Brasil, este artigo analisa narrativas de teor testemunhal de docentes Surdos Professores de Língua Gestual Portuguesa (LGP) e Professores de Língua de Sinais Brasileira (LIBRAS), colocando-as em contraponto ao panorama das políticas educativas marcadas na legislação atual e desenvolvendo uma análise qualitativa e exploratória. Como resultado, identifica-se uma visão reducionista das políticas educacionais vinculadas à educação especial, que veem as Línguas Gestuais/de Sinais como instrumentos de apoio à comunicação, desconsiderando seu valor cultural e identitário; adverte-se que as escolas estão centradas no desenvolvimento da oralidade e terapias da fala, menorizando o papel da Língua Gestual/de Sinais e identificam-se esforços de defesa das Línguas Gestuais/Sinais e de presença da Cultura Surda enquanto *topos* de resistência. Conclui-se que a educação de Surdos requer revisões na política linguística e educacional, com um incremento do investimento na formação profissional docente (inicial e continuada) e com políticas públicas de valorização e divulgação das Línguas Gestuais/de Sinais. Para além da necessária revisão das políticas vigentes, faz-se fundamental uma monitorização da aplicação do modelo bilíngue para alunos Surdos no espaço escolar.



Palavras-chave: estudos surdos; educação bilíngue; inclusão; resistência; línguas gestuais/línguas de sinais.

Abstract: Deaf education presents challenges in the implementation of bilingual education, within the scope of current linguistic and educational policies. Drawing from the discussion on the configuration of bilingual education in school settings in Portugal and Brazil, this article analyzes testimonial narratives from Deaf teachers of Portuguese Sign Language (LGP) and Brazilian Sign Language (LIBRAS), contrasting them with the landscape of educational policies outlined in current legislation, while developing a qualitative and exploratory analysis. As a result, a reductionist view of educational policies linked to special education is identified, which regards signed languages as communication support tools, disregarding their cultural and identity value. It is noted that schools are focused on the development of oral language and speech therapies, downplaying the role of sign language, and efforts to defend signed languages and the presence of Deaf culture are identified as *topos* of resistance. The article concludes that Deaf education requires revisions in linguistic and educational policies, with increased investment in professional teacher training (both initial and ongoing) and public policies aimed at valuing and promoting signed languages. Beyond the necessary revision of current policies, it is crucial to monitor the implementation of the bilingual model for Deaf students in school settings.

Keywords: deaf studies; bilingual education; inclusion; resistance; sign languages.

1 Introdução

Este artigo propõe-se a analisar narrativas de teor testemunhal de docentes Surdos de Língua Gestual Portuguesa (LGP) e Língua Brasileira de Sinais (LIBRAS), sobre a educação bilíngue em Portugal e no Brasil¹, relacionando-as com a situação do enquadramento legal de ambos

¹ Advertimos desde já ao/à leitor/a que utilizaremos para cada país a terminologia nele utilizada. Por isso, quando falarmos sobre o Brasil, falaremos em LIBRAS: Língua Brasileira de Sinais, e, quando falarmos em Portugal, fala-

os países, com ênfase nos dados recolhidos através dos testemunhos de tais profissionais. Com esse ponto de partida, adotaremos as concepções teóricas abordadas em Ladd (2022) e em Svartholm (2010) para discutirmos as políticas para a educação bilíngue e os seus atritos com a educação especial, no atendimento aos alunos Surdos e como a presença das pessoas Surdas e os espaços que ocupam são, em si mesmos, pela presença inerente de Cultura Surda e Línguas Gestuais/de Sinais, *topos* de resistência² (Bourdieu, 1993; Gil, 2020).

Além disso, o artigo busca realizar uma crítica à visão reducionista de algumas políticas educacionais que encaram as Línguas Gestuais/Línguas de Sinais apenas como instrumentos de apoio à comunicação, desconsiderando elementos culturais e identitários fundamentais para a formação de crianças e jovens Surdos, diminuindo, assim, o seu estatuto de plenas línguas. Essas políticas educacionais resultam, frequentemente, em prejuízo à formação dos alunos, adotando-se uma abordagem pedagógica que subestima a riqueza e complexidade das Línguas Gestuais/de Sinais, e ignora frequentemente as potencialidades da Pedagogia Surda³ (Ladd, 2022), assim como o valor cultural que estas línguas desempenham no fortalecimento da identidade Surda e no acesso ao currículo (Ladd; Gonçalves, 2011).

Por fim, sistematizam-se os principais desafios enfrentados pelas docentes de Língua Gestual Portuguesa (doravante LGP) e Língua Brasileira de Sinais (doravante LIBRAS) nas escolas de Surdos em Portugal e no Brasil, com especial enfoque na Cultura e Identidades Surdas, na valorização da Língua Gestual/de Sinais na escola, no modo como os Estudos Surdos contribuem para a educação de Surdos e nas mudanças necessárias às práticas pedagógicas e linguísticas para melhorar a educação de Surdos. A reflexão articula não só o conceito de inclusão, vigente nas legislações atuais, mas também os testemunhos evidenciam as experiências daqueles que já foram alunos no sistema e que são agora professores e cujas práticas pedagógicas são pensadas visando promover uma educação que valorize a diversidade e a equidade, respeitando a autonomia, o futuro e os direitos dos alunos Surdos.

Em Portugal, podemos encontrar a gênese da educação bilíngue de alunos Surdos a partir da década de 70 do século XX, quando o Ministério da Educação celebrou em 1974 um protocolo de cooperação com a Suécia (Programa de Cooperação Luso-Sueco para a investigação e desenvolvimento). Até essa data, a educação de Surdos estava mergulhada numa visão

mos em Língua Gestual Portuguesa (LGP). O mesmo irá ocorrer sobre a terminologia dos níveis de ensino, no contexto brasileiro fala-se em Educação Infantil (creches para crianças de 0 a 3 anos e a pré-escola para crianças de 4 a 5 anos), Ensino Fundamental (anos iniciais do 1º ao 5º e anos finais do 6º ao 9º) e Ensino Médio (com duração de 3 anos: 1º, 2º e 3º ano para jovens entre os 15 e 17 anos). No contexto português, designa-se em Educação Pré-escolar (jardim de infância para crianças até os 6 anos), Ensino Básico (com 1.º ciclo: do 1º ao 4º ano para crianças de 6 a 10 anos; 2.º ciclo: o 5º e 6º anos e 3.º ciclo: do 7º ao 9º ano, para alunos dos 10 aos 15 anos) e Ensino Secundário (com 3 anos de duração, que corresponde ao 10º, 11º e 12º anos, para alunos entre 15 e 18 anos).

² Bourdieu (1993, p. 168) refere-se a “locais comuns”, para designar os locais ou obras, que, de forma abstrata, poderão ser encarados enquanto produções culturais, nas quais se condensam discursos onde o coletivo se encontra e se reconhece, pois já lhes são familiares. Nessa obra, Bourdieu também se debruça sobre o espaço de produção literária e artística, ou seja, a produção cultural, que qualifica como um “campo de forças” e um “campo de lutas” (p. 30) para os quais são estratégicos os posicionamentos patentes nas próprias produções culturais. A teoria de Bourdieu inspirou, posteriormente, Gil (2020) a designar como “*topos* de resistência” (p. 19) as produções culturais das Comunidades Surdas.

³ A Pedagogia Surda, de acordo com Paddy Ladd (2022), refere-se às abordagens educacionais desenvolvidas especificamente para atender às necessidades e características dos alunos Surdos, integrando de forma holística aspectos da Cultura Surda e das Línguas Gestuais/de Sinais. Essa pedagogia foi extraída a partir de metodologias que respondem às particularidades linguísticas, sociais e culturais presentes na educação de Surdos.

única, assentada no desenvolvimento da oralidade dos alunos Surdos, renegando a sua Língua Gestual e a sua Cultura, perspectiva que vinha sendo implementada em todo o mundo desde o Congresso de Milão de 1880, com a exceção dos EUA e da Suécia, que nunca ratificaram as resoluções desse congresso. O ensino de Surdos em Portugal dividia-se na época entre os dois grandes institutos de Surdos – O Instituto Araújo Porto, na cidade do Porto, e o Instituto Jacob Rodrigues Pereira (IJRP) da Casa Pia, de Lisboa, situado na cidade de Lisboa. Esses Institutos pertenciam ao Ministério do Trabalho e Previdência Social⁴. A celebração do protocolo luso-sueco, e a abertura do Ministério da Educação à área da educação de Surdos, trouxe a Portugal professores de Surdos suecos que apresentaram diversos modelos de educação alternativos ao modelo oralista, *i.e* Educação Bilíngue, Educação Bimodal, Comunicação Total. Deram-se assim os primeiros passos para uma mudança de paradigma na educação de Surdos em Portugal.

Sediado na Faculdade de Letras da Universidade de Lisboa, foi fundado um grupo de investigação, coordenado pela professora Raquel Delgado-Martins, que tinha como principais objetivos estudar a Língua Gestual Portuguesa e a Educação de Surdos. O trabalho desta equipe de investigação teve como resultados a publicação do primeiro dicionário de LGP (Prata, 1980) e a criação dos Núcleos de Apoio à Criança Deficiente Auditiva (NACDA) (designação da época) através da Lei 66/79, dando respostas à educação de Surdos em diversas regiões do território nacional.

Em articulação com esta equipe de investigação, em 1983, o professor Sérgio Niza, introdutor do Movimento da Escola Moderna em Portugal, e o presidente da Associação Portuguesa de Surdos, José Bettencourt, implementaram o primeiro projeto de educação bilíngue para alunos Surdos na escola A-da-Beja, baseado em 3 pilares – a Língua Gestual Portuguesa, o ensino da leitura e escrita do português e o respeito pelo momento de aprendizagem de cada aluno Surdo. Estas novas práticas, desenvolvidas pelo Ministério da Educação, vieram a ter influência direta no Instituto Jacob Rodrigues Pereira, que, através da nova diretora, a professora Maria Augusta Amaral, desenvolveu um estudo sobre as competências dos alunos Surdos sujeitos ao ensino oralista e cujos resultados demonstraram uma necessidade urgente de mudança de política na educação de Surdos (Amaral, 1996). Apoiada pela Universidade de Gallaudet, dos Estados Unidos da América, e pela Faculdade de Letras da Universidade de Lisboa, implementou-se a educação bilíngue de alunos Surdos no IJRP, monitorizada pelas universidades referidas.

Quanto à legislação relacionada com a educação bilíngue de alunos Surdos, em Portugal ocorreu com o Despacho 7520/98, com a implementação das Unidades de Apoio a Alunos Surdos (UAAS), que vieram substituir os NACDA, referidos anteriormente, um ano após o reconhecimento da LGP na Constituição da República Portuguesa, em 1997.

Neste enquadramento legal, prevê-se a contratação de docentes com formação especializada nas áreas de comunicação e linguagem e da deficiência auditiva, e, enquanto técnicos especializados, formadores⁵ de LGP, intérpretes de LGP e terapeutas da fala. Já se prevê, assim, a aprendizagem pela criança Surda da LGP enquanto primeira língua (L1) e a Língua Portuguesa enquanto segunda língua (L2), ao nível da escrita e da leitura e que contempla ainda uma prática

⁴ Designação da época.

⁵ A terminologia “formador” era utilizada na época, carregando uma conotação hierárquica que indicava uma posição considerada inferior, tanto académica quanto socialmente, em relação ao título de “professor”. Na época, estes profissionais eram contratados fora do enquadramento legal da contratação docente.

contínua de terapia da fala e treino auditivo. Nesse enquadramento existe ainda a previsão de colaboração com associações de pais, associações de Surdos e colaboração com a Comunidade Surda, visando a interação entre a Comunidade Surda e a comunidade ouvinte.

Dez anos volvidos, foram aprovados o Decreto-Lei 3/2008 e a Lei 21/2008, numa clara evidência de se posicionar e assumir enquanto ensino bilíngue, ao que logo à partida podemos evidenciar como se passaram a designar as escolas de Surdos: Escolas de Referência para a Educação Bilíngue de Alunos Surdos (EREBAS). No artigo 23º do Decreto-Lei 3/2008, os formadores Surdos já são designados enquanto docentes de LGP, evidenciando não só uma mudança no estatuto linguístico e social da LGP, assim como na valorização dos profissionais de ensino da LGP.

No que diz respeito à contratação de professores/as com formação em educação especial na área da surdez, observa-se também uma mudança na terminologia, com um claro afastamento do paradigma médico e da designação ‘deficiência auditiva’, transitando para o paradigma biocultural, no qual se faz referência à pessoa Surda ou à área da surdez. Neste contexto, enfatiza-se que, segundo o artigo 23º do Decreto-Lei 3/2008, não só o/a professor/a de educação especial, mas todos os/as docentes que trabalham com Surdos devem ser competentes em LGP. Este enquadramento legal mantém a colaboração da Comunidade Surda e das comunidades escolar e ouvinte. Outra inovação, que surgiu em 2008, na educação de Surdos, foi a introdução de uma língua estrangeira escrita no 3º ciclo do ensino básico ao ensino secundário.

Na atualidade, vigora o Decreto-Lei 54/2018, que revoga o Decreto-Lei 3/2008 e não reúne consensos nem o favoritismo entre professores de LGP e alguns professores de Surdos de outras disciplinas curriculares. Neste decreto, as escolas passaram a designar-se Escolas de Referência para a Educação Bilíngue (EREB) e no seu texto não consta qualquer menção à pessoa/criança/jovem Surdo/a, revelando um total apagamento político e social daqueles a quem o decreto diz respeito. Alvo de duras críticas (Carvalho; Mineiro, 2020; Carvalho, 2012; Sofiato; Carvalho; Coelho, 2021), apesar de ainda se designar como o âmbito de ação à educação bilíngue, e prevalecer a LGP enquanto L1, e o Português enquanto L2, existe uma total omissão da necessidade de competências em LGP da parte dos docentes com formação especializada em educação especial na área da surdez.

Apesar do surgimento, em 2007, do programa curricular de LGP para a educação pré-escolar e o ensino básico, da introdução, em 2008, do mesmo programa para o ensino secundário, da criação, em 2011, do programa curricular de português L2 para alunos Surdos, e da implementação, em 2018, do grupo de recrutamento para os docentes de LGP, permitindo-lhes concorrer à carreira docente tal como os/as demais professores/as (Decreto-Lei 16/2018), os/as professores de LGP não estão otimistas quanto ao retrato que fazem do ensino de crianças Surdas em Portugal na atualidade.

Em Portugal, a LGP passou a ser lecionada no Ensino Superior a partir de 1997, com a criação da Licenciatura em Tradução e Interpretação de Língua Gestual Portuguesa na Escola Superior de Educação do Instituto Politécnico de Setúbal (ESE/IPS), voltada para a formação de intérpretes de LGP. Posteriormente, essa formação também foi disponibilizada em 2001 na Escola Superior de Educação do Instituto Politécnico do Porto e, em 2005, na Escola Superior de Educação do Instituto Politécnico de Coimbra (ESEC). Além disso, a LGP passou a ser disponibilizada também no Ensino Superior no contexto das Licenciaturas para a formação de professores/as de LGP, nas extintas ESE/IPS, onde funcionam em regime noturno e numa edição

única no Instituto de Ciências da Saúde da Universidade Católica Portuguesa, e atualmente ainda é oferecida na ESEC. No entanto, há uma escassez de intérpretes no Ensino Superior em outros cursos, uma vez que não existe um enquadramento legal que contemple a presença de intérpretes de LGP nesse contexto.

No Brasil, a educação de Surdos é uma temática inquietante e com muitos desafios, conforme já indicava Cristina Lacerda, pesquisadora no campo dos Estudos Surdos, no final dos anos de 1990:

A educação de surdos é um assunto inquietante, principalmente pelas dificuldades que impõe e por suas limitações. As propostas educacionais direcionadas para o sujeito surdo têm como objetivo proporcionar o desenvolvimento pleno de suas capacidades; contudo, não é isso que se observa na prática. Diferentes práticas pedagógicas envolvendo os sujeitos surdos apresentam uma série de limitações, e esses sujeitos, ao final da escolarização básica, não são capazes de ler e escrever satisfatoriamente ou ter um domínio adequado dos conteúdos acadêmicos. Esses problemas têm sido abordados por uma série de autores que, preocupados com a realidade escolar do surdo no Brasil, procuram identificar tais problemas (Fernandes 1989, Trenché 1995 e Mélo 1995) e apontar caminhos possíveis para a prática pedagógica (Góes 1996 e Lacerda 1996) (Lacerda, 1998, p. 69).

Diante desse contexto, as primeiras investigações sobre educação bilíngue ocorreram no início da década de 1990, especialmente pelas influências internacionais que circulavam nas universidades brasileiras e em eventos na área de educação de Surdos. Pesquisadores das áreas de educação e linguagem, juntamente com a Federação Nacional de Educação e Integração de Surdos (FENEIS), assumiram a defesa da educação bilíngue a Surdos. No Decreto n.º 5.626/2005, o capítulo VI assegura a educação bilíngue a Surdos, sendo essa a primeira vez que tal proposta educacional surge em documentos legais, uma vez que prevê, na Educação Infantil e nos anos iniciais do Ensino Fundamental, uma educação sob a responsabilidade de professores/as bilíngues LIBRAS/Português. Já para os anos finais do Ensino Fundamental e Ensino Médio é prevista a presença de Tradutores e Intérpretes de LIBRAS - Língua Portuguesa (TILSP), a fim de atuarem nas escolas (Lodi, 2021).

Nos últimos 30 anos, ocorreram mudanças significativas na legislação, indicando os direitos linguísticos e educacionais dos Surdos. Destaca-se o reconhecimento da LIBRAS, em nível nacional, por meio da Lei 10.436 (Brasil, 2002) e das disposições presentes do Decreto 5.626 (Brasil, 2005), que regulamenta a referida Lei. Exemplos dessas mudanças foram a criação de cursos de graduação para a formação de professores/as de LIBRAS e tradutores-intérpretes (TILSP) e a inserção da disciplina de LIBRAS nos cursos superiores de licenciatura, Pedagogia e Fonoaudiologia.

A partir da legislação brasileira, no âmbito da educação, a educação de Surdos tem diferentes e contraditórios enfoques, em que se efetivam a educação bilíngue, a educação bilíngue inclusiva e a educação inclusiva. Entretanto, no âmbito da Educação Básica, recentemente ocorreu a aprovação da Lei nº 14.191 (Brasil, 2021), que dispõe sobre a modalidade de Educação Bilíngue de Surdos, alterando a Lei 9.394, de 20 de dezembro de 1996 – Lei de Diretrizes e Bases da Educação (LDB). Essa recente alteração foi considerada uma importante conquista para a implementação da educação bilíngue no território nacional. O capítulo V-A trata da Educação Bilíngue de Surdos e explicita que:

Art. 60-A. Entende-se por educação bilíngue de surdos, para os efeitos desta Lei, a modalidade de educação escolar oferecida em Língua Brasileira de Sinais (Libras), como primeira língua, e em português escrito, como segunda língua, em escolas bilíngues de surdos, classes bilíngues de surdos, escolas comuns ou em polos de educação bilíngue de surdos, para educandos surdos, surdo-cegos, com deficiência auditiva sinalizantes, surdos com altas habilidades ou superdotação ou com outras deficiências associadas, optantes pela modalidade de educação bilíngue de surdos.

§ 1º Haverá, quando necessário, serviços de apoio educacional especializado, como o atendimento educacional especializado bilíngue, para atender às especificidades linguísticas dos estudantes surdos.

§ 2º A oferta de educação bilíngue de surdos terá início ao zero ano, na educação infantil, e se estenderá ao longo da vida (Brasil, 2021, n. p.).

A partir dessas legislações, presenciou-se o impacto dessas mudanças na escolarização de Surdos, que vêm mobilizando ações na configuração de espaços que garantam a acessibilidade, a permanência, o aprendizado e a participação de Surdos nas escolas. Entretanto, acompanhar esses movimentos, suas tendências, desafios, (im)possibilidades e questionamentos tem sido a necessária tarefa de pesquisadores do campo dos Estudos Surdos em articulação a outras áreas e campos investigativos. Neste artigo, não será realizada uma análise detalhada dos aspectos históricos e legais, mas certamente serão pautados, quando necessário, tendo em vista as condições de possibilidade sobre as quais se desenvolvem as propostas educacionais.

Em investigações mais recentes, desenvolvidas por pesquisadoras do Grupo Interinstitucional de Pesquisa em Educação de Surdos, Karnopp, Klein e Lunardi-Lazzarin (2021) analisam a educação escolar bilíngue desenvolvida em escolas de Surdos no estado do Rio Grande do Sul, Brasil. Como resultados dessa investigação, indicam a potencialidade das escolas de Surdos a partir de três eixos: (1) as práticas de ensino bilíngues direcionadas aos estudantes Surdos, (2) o acesso e permanência na escola devido ao uso da Língua de Sinais e da Língua Portuguesa, (3) o encontro com diferentes Surdos e contato com artefatos da Cultura Surda. As autoras advertem que, nas escolas de Surdos, os significados culturais das Comunidades Surdas são cotidianamente negociados e (re)inventados, no conjunto das relações de poder/saber, em contato com outras experiências culturais. Evidencia-se um ambiente de permanente negociação entre línguas e significados culturais, que constituem a escola de Surdos como um espaço singular.

Mudanças na legislação, produção de pesquisas na área, formação de profissionais bilíngues, atuação de Surdos em diferentes espaços sociais e educacionais impactam potencialmente a educação de Surdos em diferentes níveis escolares. Esses aspectos, dentre outros, marcam os tempos em que nos situamos e servem como pistas às permanentes encruzilhadas investigativas.

2 Caminhos Investigativos: entrevistas cruzadas

2.1 Seleção e caracterização da amostra

Neste artigo, prioriza-se a seguinte questão de investigação: quais são os desafios atuais da educação de Surdos e como se configura a educação bilíngue em espaços escolares em

Portugal e no Brasil? A metodologia de pesquisa é qualitativa e exploratória, sendo que, para discutir a questão de investigação, foram realizadas entrevistas com seis docentes Surdos que atuam na educação de Surdos no Brasil e em Portugal. Primeiramente, foram convidadas três professoras de LGP e duas professoras e um professor de LIBRAS, que atuam na Educação Básica, nos seguintes espaços de escolarização: escolas bilíngues de Surdos, classes bilíngues de Surdos, escolas comuns ou em polos de educação bilíngue de Surdos. Os critérios para a seleção dos entrevistados foram: mais de dez anos de experiência em ensino de Língua Gestual/Língua de Sinais, mínimo de três anos de tempo de serviço na educação de Surdos; formação compatível com a docência (por exemplo, licenciatura em Letras-Libras/em Língua Gestual Portuguesa; profissionalização de Professores de LGP; pedagogia bilíngue) e atuação nos anos finais do Ensino Fundamental ou no Ensino Médio. E um critério fundamental: terem sido eles mesmos também, no passado, alunos Surdos, agora professores Surdos, dando ênfase à sua experiência enquanto narrativa de resistência, num ambiente educativo por vezes hostil à Cultura Surda. O Quadro 1, a seguir, apresenta informações gerais sobre os entrevistados.

Quadro 1 – Caracterização da amostra

Docente Surdo/ País	Formação	Ciclos/Séries de Ensino na atuação profissional	Tempo de serviço
Carla Benge/ PT	Professora de LGP	Professora de LGP em escola de Surdos. Primeiro ciclo. LGP como L1 e L2.	15 anos
Sandra Coelho/ PT	Professora de LGP	Professora de LGP desde a creche ao Ensino secundário, em diferentes escolas. Atualmente trabalha com o Ensino Secundário.	24 anos
Sónia Serras/ PT	Professora de LGP	Foi Professora de LGP em muitas escolas, de Norte a Sul, incluindo zonas bastante isoladas, desde o pré-escolar até ao Ensino Secundário, sendo a maior parte do tempo com os ciclos, entre o 5º e o 12º anos. No momento, está em licença saúde.	18 anos
Juliane Emmert/ Rio Grande do Sul, Brasil	Pedagogia. Letras-Libras Licenciatura. Pós-graduação: especialização em educação especial.	Professora regente em escola de Surdos, no 4º ano do Ensino Fundamental. No Ensino Médio, é professora da disciplina de Cultura Surda. Professora em curso de lideranças surdas e de LIBRAS para ouvintes.	12 anos
Cristiane Andrade/São Paulo, Brasil	Pedagogia. Letras-Libras Licenciatura. Pós-graduação: especialização em ensino de Surdos e em tradução.	Atuou como instrutora de Surdos durante 16 anos. Professora de LIBRAS para ouvintes. Após concluir Pedagogia, começou a trabalhar como professora em uma escola municipal bilíngue para Surdos, do 1º ao 9º ano do Ensino Fundamental.	23 anos
Fábio Escalace/ São Paulo, Brasil	Pedagogia bilíngue.	Trabalha com estudantes Surdos na educação infantil, no Ensino Fundamental e Educação de Jovens e Adultos. É concursado como instrutor de LIBRAS.	3 anos

Fonte: Elaboração dos autores.

As cinco professoras e o professor entrevistado são pessoas Surdas, apresentam formação compatível com a docência e atuam em escolas ou em polos de educação de Surdos com turmas específicas de Surdos. Trabalham como professor/as regentes, bem como no ensino de LGP ou LIBRAS.

Em Portugal, as três professoras possuem profissionalização em ensino de LGP e atuam em escolas de Surdos, há mais de 15 anos, com vasta experiência, que inclui o ensino desde a creche ao Ensino Secundário. No Brasil, as docentes possuem formação em curso de graduação (licenciatura em Letras-Libras ou Pedagogia Bilíngue) e em curso de pós-graduação *lato sensu* (especialização), com experiência de atuação na Educação Básica (desde a Educação Infantil ao Ensino Médio), nos seguintes espaços de escolarização: escolas bilíngues de Surdos, classes bilíngues de Surdos, escolas comuns ou em polos de educação bilíngue de Surdos.

O material produzido durante as entrevistas foi potente para o desenvolvimento de diversas análises em torno da discussão sobre educação de Surdos. Entretanto, para fins de análise, realizamos um recorte, neste artigo, a fim de indicar as problemáticas apontadas por profissionais que atuam na educação bilíngue e que apresentam histórias de vida relacionadas à educação de Surdos. Consideramos que essa experiência pessoal e profissional permite um olhar amplo a questões que se vinculam à escolarização, Comunidade Surda, Língua Gestual/de Sinais, inclusão e educação bilíngue, dentre outros temas.

2.2 Instrumento de recolha dos dados: entrevista semiestruturada

As entrevistas foram realizadas em formato público, *online* (plataforma *Zoom*), com a participação dos organizadores, das entrevistadas⁶ e de tradutores-intérpretes de LGP e LIBRAS. Os autores deste artigo, organizadores das entrevistas, enviaram o convite às professoras e explicitaram o propósito da entrevista e as questões de debate. A entrevista, em formato de conversa, está disponível no *YouTube* e os dados são públicos. Seguindo orientações da Resolução 510/2016, a pesquisa contou com o consentimento das convidadas e utilizou informações de acesso público. A duração das entrevistas foi de duas horas para cada grupo.

O roteiro para o desenvolvimento das entrevistas⁷ foi dividido em duas partes. Na parte inicial, houve uma entrevista para a caracterização da amostra e, por isso, contemplou a apresentação pessoal de cada participante, bem como um detalhamento com informações sobre a formação e atuação profissional, tempo de serviço na educação de Surdos, caracterização da escola e da turma onde trabalham. Na segunda parte, aplicou-se a entrevista semiestruturada, utilizada como instrumento de recolha de dados sobre educação e cultura, a partir das seguintes questões dirigidas: de que modo a Cultura Surda está presente na escola? Citar exemplos e práticas da (des)valorização da Cultura Surda e da Língua Gestual/de Sinais na escola. Quais são os desafios/obstáculos da educação bilíngue de Surdos? De que modo os

⁶ Como o grupo docente entrevistado constituiu-se por cinco professoras e um professor, optamos por adotar o gênero feminino quando nos dirigirmos a tal conjunto.

⁷ As entrevistas foram conduzidas pelos autores deste artigo, nos dias 6 e 7 de novembro de 2024, e promovidas pelo *Deaf Studies Lab* do Centro de Estudos de Comunicação e Cultura (CECC), da Universidade Católica Portuguesa, Lisboa, com o apoio do Grupo Interinstitucional de Pesquisa em Educação de Surdos (GIPES). Agradecemos a operacionalização técnica da *live* no *YouTube*, realizada pela estudante e monitora do GIPES, Luciana Feijó Caetani.

Estudos Surdos impactam/contribuem na educação de Surdos? Citar exemplos de mudanças necessárias às práticas pedagógicas e linguísticas para melhorar a educação de Surdos. Além dessas questões, surgiram outras, que foram propostas pelos organizadores e também pelas pessoas que estavam assistindo às *lives* pelo *YouTube*.

2.3 Procedimentos Metodológicos

A primeira sessão intitulou-se *O impacto dos Estudos Surdos na Educação de Surdos*, com foco no contexto português⁸, realizada no dia 06 de novembro, no horário das 18h às 20h (hora de Lisboa), pelas professoras Surdas de LGP Carla Benge, Sandra Coelho e Sónia Serras. A sessão está disponível *online* e acessível em LGP e Português, com a participação das tradutoras-intérpretes de LGP: Joana Sousa e Rafaela Cota Silva⁹.

A segunda sessão, *O impacto dos Estudos Surdos na Educação de Surdos*, com destaque para o contexto brasileiro¹⁰, foi o tema discutido no dia 07 de novembro, no horário das 18h30 às 20h30 (hora de Brasília), pelas docentes de LIBRAS, Juliane Emmert, Cristiane Esteves de Andrade e o docente Fábio Luis Escalace. A entrevista está disponível *online* e acessível em LIBRAS e Português, com a participação dos tradutores-intérpretes Camila Sogertz Vargas e Lucas Magnus¹¹.

Em relação à transcrição das entrevistas do *YouTube*, optou-se pela ferramenta *Transkriptor*, serviço *online* que viabiliza a conversão de conteúdo de áudio em texto. A transcrição foi feita a partir da tradução da LGP e da LIBRAS para o português, com posterior revisão pelos mesmos tradutores-intérpretes que atuaram nas *lives*. Na sequência, com o texto transcrito, foi necessária a revisão detalhada pelos pesquisadores, para ajustes finais da transcrição do áudio em texto. A transcrição das entrevistas do contexto brasileiro resultou em 30 páginas (14.555 palavras e 84.111 caracteres) e, do contexto português, 31 páginas (15.355 palavras e 90.786 caracteres).

3 Apresentação, Análise e Discussão dos Resultados

A partir das entrevistas com as docentes, sintetizamos algumas questões que contextualizam a escola de Surdos, o trabalho pedagógico das professoras de LGP/LIBRAS, a situação dos estudantes Surdos e de suas famílias, a fim de analisar as narrativas de teor testemunhal de docentes Surdos de LGP/LIBRAS sobre a educação bilíngue em Portugal e no Brasil. São evidentes os contrastes e atritos entre as políticas para a educação bilíngue e a educação especial no atendimento de estudantes Surdos, em ambos os países. Tais questões se vinculam

⁸ Disponível em: https://www.youtube.com/live/ziJkG77ke_8?si=MkhUHRcACzjWEBH. Acesso em: 11 mar. 2025.

⁹ O financiamento para o pagamento dos Tradutores-Intérpretes de Língua Gestual Portuguesa foi realizado pelo Centro de Estudos de Comunicação e Cultura (CECC).

¹⁰ Disponível em: https://www.youtube.com/live/yhngWCFn53g?si=9xy_AYli8GHJEyOk. Acesso em: 11 mar. 2025.

¹¹ O financiamento para o pagamento dos Tradutores-Intérpretes de LIBRAS foi realizado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/ PQ1D - Processo nº 317132/2021-0), taxa de bancada de Lodenir Karnopp.

a aspectos mais gerais, como a legislação no âmbito educacional e as representações sobre Surdos e Línguas Gestuais/de Sinais na atualidade, resultando em narrativas de resistência.

Quanto à escola, os relatos indicam que as docentes atuam em escolas/polos que adotam a educação bilíngue como principal projeto político-pedagógico da educação pública. Entretanto, a educação bilíngue se apresenta conceitualmente frágil e com uma prática escolar diversa, pois há escolas que se aproximam ao conceito de escola bilíngue e outras nem tanto. As experiências relatadas pelas docentes indicam a necessidade de mudança/revisão no conceito de educação bilíngue, nas políticas e práticas que envolvem o uso da Língua Gestual/de Sinais e do português. As entrevistadas questionaram, com ênfase na sua intervenção, a atuação de professores/as ouvintes sem competência na LGP na educação de Surdos, numa clara manifestação de resistência e descontentamento em relação ao *status quo*.

Muitos professores, que trabalham com o ensino bilíngue, não sabem a LGP, portanto, como é que podem ensinar? Como podem trabalhar com alunos Surdos? Como? Como? (Sandra Coelho, YouTube 2024, 00:58:03 a 00:58:25).

Há uma denúncia de que, nos espaços educacionais, há privilégio da Língua Portuguesa e a valorização da oralidade, bem como situações que envolvem linguicismo¹² (Skutnabb-Kangas; Phillipson, 1996) e neo-oralismo (Peluso, 2016). Em outras palavras, a prática da educação bilíngue está direcionada prioritariamente para o desenvolvimento da oralidade, conforme argumenta Sandra Coelho.

Então, a educação bilíngue é uma palavra maravilhosa, linda, linda. No entanto, na realidade, não é assim tão lindo. É normal, a maioria das pessoas dizem 'Ai meu Deus, educação bilíngue, que bom, oralidade, oralidade, oralidade, que bom'. Mas, se nós pensarmos na LGP enquanto primeira língua e, o português, como segunda língua escrita, é claro, para nós, não é? No entanto, isso não é a realidade do bilinguismo em Portugal: a realidade é o contrário, portanto, é oralidade em primeiro lugar (Sandra Coelho, YouTube 2024, 00:57:20 a 00:58:00).

Os desafios vão na direção de conjugar a língua gestual com oralidade; considerando a proposta político-pedagógica, bem como o uso balanceado das línguas. Nesse sentido, Carla Benge indica que:

Os desafios são a questão de conjugar a língua gestual com oralidade, porque, efetivamente, o nosso objetivo não é trabalhar a oralidade. Mas a verdade é que muitos pais querem que os filhos sejam muito bilíngues. É importante efetivamente que os filhos sejam bilíngues, mas coloca-se um enfoque muito na oralidade, esquecendo e menosprezando um pouco a aquisição da língua gestual e da aquisição dos conceitos em língua gestual, que é uma base fundamental para depois também desenvolver a escrita (Carla Benge, YouTube 2024, 00:19:44 a 00:20:15).

¹² O linguicismo refere-se à discriminação linguística ou ao tratamento desigual ou injusto de indivíduos ou grupos com base nas características da sua língua, sotaque, modalidade, vocabulário ou estrutura sintática, resultando em marginalização.

As críticas às formas como a educação bilíngue tem sido conduzida nas escolas estão centradas no privilégio que é dado à oralidade, em detrimento da Língua Gestual. Tal centralidade [na oralidade] traz prejuízos ao desenvolvimento linguístico da Língua Gestual, bem como da leitura e escrita da Língua Portuguesa. Além disso, linguicismo evidencia a assimetria entre as línguas, pois os depoimentos das professoras indicam que “os aplausos são direcionados aos que oralizam, mesmo que precariamente. Não há aplausos para quem é fluente na LGP”. Carla Benge pondera:

E o aluno, tendo a língua gestual e tendo a escrita, é também bilíngue e, portanto, não estamos a falar de duas línguas? É isso que é o bilinguismo, mas a preocupação das escolas, hoje em dia, é muito focada na oralidade, que os alunos desenvolvam oralidade (YouTube 2024, 00:20:40 a 00:21:08).

Isto transparece também porque muitos/as professores/as apenas avaliam os alunos Surdos com base no seu uso do Português, sem as competências necessárias para avaliar os alunos nas suas disciplinas por meio da LGP e, assim, dessa forma, é-lhes mais cômodo sonegar e minorizar a existência da língua gestual/de sinais no contexto escolar. Por isso, afirmar que a educação é bilíngue, por contar com a presença de um intérprete de Língua Gestual, é um argumento que cai por terra quando verificamos como é feita a avaliação dos alunos Surdos.

A desvalorização da LGP se evidencia tanto nas práticas pedagógicas quanto na atuação de professores/as que não são fluentes/sem LGP, que trabalham no ensino bilíngue de crianças Surdas ou que não apresentam formação pedagógica adequada para o ensino a Surdos. Novamente a pergunta, “sem LGP, como pode um Professor ensinar um aluno Surdo?”, que confirma a hegemonia da oralidade no ensino, em detrimento à LGP. A oralidade permanece como prioridade no currículo e indicador único de avaliação do estudante e do seu desempenho linguístico e cognitivo.

No Brasil, semelhante questão é abordada pelas docentes, quando manifestam o desejo de que a LIBRAS esteja em uso na escola de forma completa e fluente por todos. Cristiane Esteves de Andrade pondera:

Hoje eu enxergo a escola bilíngue (pelo menos onde eu estou) assim. Eu sonho com uma escola bilíngue, entre professores, ouvintes, surdos, que todo mundo sinalize de forma fluída com as crianças, todo mundo utilizando a língua de sinais, durante o almoço, todo mundo conversando, se divertindo. O pessoal, os funcionários da limpeza, o pessoal da direção, todo mundo em língua de sinais proficiente... mas na prática é um choque, é um choque cultural, porque, apesar de a cultura surda ser forte, a proficiência não está em sua totalidade dentro da escola (YouTube 2024, 00:27:57 a 00:28:40).

Cristiane defende um ambiente linguisticamente adequado para o desenvolvimento educacional dos estudantes Surdos, para que a LIBRAS seja a língua visualmente em uso pela comunidade escolar. Afirma que a escola precisa garantir a aquisição da Língua de Sinais e combater o isolamento linguístico que os alunos vivenciam.

Além das questões destacadas até o momento, há outros desafios/obstáculos da educação bilíngue de Surdos, tais como: discussão da política de educação bilíngue articulada à inclusão; contratação de intérpretes para trabalho em duplas e para atendimento da comunidade escolar; necessidade de produção de materiais para a educação bilíngue; atualizações

do programa curricular de LGP com a atualidade dos Estudos Surdos e a promoção de um ambiente bilíngue no espaço escolar e não apenas na sala de aula.

A preocupação das docentes está centrada em políticas e ações que promovam a preservação da LGP e sua sustentabilidade linguística, por meio da atualização dos programas, da participação e presença de estudantes em eventos da Comunidade Surda. Destacam também a necessidade de ensinar Português como L2 para alunos Surdos com interpretação de LGP, bem como Língua Gestual como língua adicional para funcionários/as da escola e familiares de Surdos.

Existe a noção da necessidade de resistir e preservar a sustentabilidade da LGP, com a recuperação da LGP da clandestinidade e do perigo de retrocesso, especialmente em contexto de legislação direcionado às políticas de inclusão. As professoras de LGP trazem exemplos e práticas de valorização da Cultura Surda e da Língua Gestual/de Sinais na escola, embora sejam atividades raramente oferecidas. Com a percepção de que o enfoque da escola é na oralidade, as professoras Surdas tentam criar estratégias para que os alunos desenvolvam e valorizem a identidade e Cultura Surdas. Um dos exemplos é o projeto de preservação da LGP: clube de desenvolvimento da LGP, proposto por Sônia Serras:

Era o clube de desenvolvimento da Língua Gestual Portuguesa, da LGP. Eu queria que isso fosse uma salvaguarda e uma defesa para a LGP dentro da escola, para evitar que a língua gestual se perdesse nas escolas e a cultura também (YouTube, 2024, 00:34:13 a 00:34:50).

O Clube de LGP é um claro esforço de resistência pela sustentabilidade da Cultura Surda e da LGP em contexto educativo e teve o objetivo de preservar a língua, trazer reflexões amplas sobre a língua, sobre o passado, presente e futuro da comunidade, dentre outras atividades que envolvem produção de vídeos e testemunhos. Entretanto, as atividades foram encerradas, devido à inexistência na escola de outros/as professores/as no desenvolvimento do Clube de LGP.

No Brasil, as docentes apresentaram relatos semelhantes, no que se refere à importância da preservação da Língua de Sinais e de sua transmissão intergeracional. Nesse sentido, garantir a aquisição da linguagem (LIBRAS), por meio do convívio entre Surdos de diferentes idades, é uma das estratégias adotadas por alguns docentes. Juliane Emmert salienta que o convívio entre Surdos de diferentes idades proporciona uma integração entre eles e a percepção de diferentes formas de usar a LIBRAS.

Então é muito interessante que eles percebam, através dessa integração, as diferentes formas de se comunicar em língua de sinais e de se expressar também. E aí a gente vê, às vezes, crianças e jovens falando em LIBRAS muito rápido, sinalizando muito rápido, adolescentes se desenvolvendo também e desenvolvendo gírias, sua forma também de se comunicar através de classificadores. [...] Isso é muito interessante e esse espaço é muito importante pra valorização da língua de sinais, pra cultura e pra valorização da cultura surda também como um todo (Juliane Emmert, YouTube 2024, 00:14:00 a 00:15:00).

Juliane Emmert destaca a potencialidade da escola bilíngue e a compara a uma terra fértil, onde a Língua de Sinais brota e floresce. É nesse espaço escolar que os estudantes suprem suas necessidades de informação sobre os acontecimentos diários. Assim, a escola assume muitas funções, além da escolarização, como informar e responder questões do cotidiano (especialmente aquelas veiculadas nos meios de comunicação), pois a maioria dos pais não se comunica com os seus filhos em Língua de Sinais.

Esse espaço é como se fosse uma sementeira, a gente vai plantando essas sementes nessa terra fértil, que é a escola. E a língua de sinais vai brotando e vai florescendo. E a gente vai vendo esse espaço lindo e rico que é a escola, que tem a língua de sinais como primeira língua, que é a escola bilíngue. O ambiente bilíngue na escola de Surdos traz inúmeros desafios. Muitas vezes a escola precisa suprir necessidades de comunicação da criança que extrapolam o ambiente de sala de aula. Em casa, a maioria não tem acesso completo às informações, porque os pais não usam a Língua de Sinais. Eles chegam com perguntas sobre o que eles viram na televisão, mas eu não sei do que se trata, então eles explicam claramente com classificadores e eu consigo entender (Juliane Emmert, YouTube 2024, 00:15:57 a 00:17:45).

As informações sobre acontecimentos diários, notícias veiculadas na televisão – que em sua maioria estão sem legendas e sem intérpretes de Língua Gestual/de Sinais –, questões pessoais, emocionais ou familiares, muitas vezes demandam atenção e discussão em sala de aula. Entretanto, situações como essa frequentemente entram em confronto com as orientações políticas de escolas que adotam uma perspectiva conteudista. As docentes defendem que a escola deve atender aos questionamentos dos estudantes e que muitos assuntos precisam ser tratados, em um diálogo informativo e adequado à idade e escolaridade dos estudantes. Cristiane, por exemplo, defende que a escola não é uma linha de produção, não deve ter um enfoque apenas conteudista, mas precisa de um olhar sensível e contextualizado. Acrescenta ainda que o tipo de trabalho realizado em sala de aula depende da aquisição e desenvolvimento da Língua Gestual/de Sinais de cada aluno.

O coordenador pedagógico precisa ter esse olhar sensível e precisa pensar nas políticas educacionais da Secretaria de Educação, porque às vezes é aquela preocupação em nota e conteúdo, aquela perspectiva muito conteudista, de que precisa seguir corretamente o mesmo conteúdo do dia. E, nesse cronograma, [o professor] tem que colocar o check list ali direitinho, né? Não é uma linha de produção, a escola não é uma linha de produção. Isso não pode acontecer. A prática tem que ser diferente (Cristiane Esteves, YouTube 2024, 1:43:30 a 1:44:15).

As docentes destacam que a presença de professor bilíngue em sala de aula traz contribuições para uma escola flexível e acolhedora. Além disso, estudantes Surdos se identificam com professores/as Surdos/as e, por isso, sentem-se mais à vontade para compartilhar dúvidas pessoais, questionamentos e impasses. Salientam também a importância de maior variedade de convívio com outros modelos Surdos, por meio, por exemplo, de atividades organizadas em colaboração com associações de Surdos. Nesse sentido, relatam que promovem atividades de passeio e contato com Surdos de associações, em atividades culturais e de lazer, sempre que possível.

E a presença do professor bilíngue é muito importante e é maravilhoso para que eles possam se desenvolver. Então, na Educação Infantil, ter esse acompanhamento do professor bilíngue é muito importante (Fábio Escalace, YouTube 2024, 00:55:55 a 00:56:20). E eu acho que o professor também é isso. É muito importante que nós professores possamos ser modelos... para que estes alunos possam se desenvolver futuramente (Carla Bengue, YouTube 2024, 00:16:31 a 00:17:00).

Professores/as Surdos/as transmitem a herança cultural da Comunidade Surda, por meio da Língua Gestual/de Sinais e da Literatura Surda, que passa de geração a geração,

por meio do compartilhamento de experiências e convívio com pessoas Surdas de diferentes idades. Temas sensíveis e difíceis de serem trabalhados nas escolas, tais como enfrentamento ao preconceito, discriminação, desprezo, dentre outros, são abordados na direção de promover o fortalecimento da identidade e Cultura Surda, para que não percam o interesse pela própria escola e pela comunidade. As entrevistadas defendem que, quando estudantes Surdos/as convivem com Surdos/as adultos, aprendem formas de resistência ao desprezo, ao menosprezo e à indiferença, pois os contextos sociais discriminam pessoas Surdas e a Língua Gestual/de Sinais. Desse modo, destacam elas que a Cultura Surda sempre existiu na escola de Surdos, pois este é o seu berço, proporcionando a transmissão intergeracional das Línguas Gestuais/de Sinais, bem como dos elementos fundamentais que sustentam a Cultura Surda, como um *Mythomoteur Surdo* (Gil, 2019), que alimenta a sua coesão, num esforço coletivo de resistência enquanto minoria linguística e cultural, e sendo assim mantida pela atuação de professores/as Surdos/as e de membros da Comunidade Surda (colegas, profissionais etc.).

Além disso, as docentes indicam a potência da articulação entre instituições e a construção de um trabalho coletivo e colaborativo entre escola, família e comunidade. Fábio destaca as orientações e os cursos para formação de professores/as, desenvolvido pela Universidade Federal de São Carlos (UFSCar), o que contribui para o desenvolvimento do trabalho pedagógico na escola em que atua.

Então eu tenho contato com a orientação da universidade e é muito importante para poder atender essas crianças, saber como adaptar os materiais, como melhorar sempre as atividades, o que funciona e o que não funciona. E também pensar em estratégias criativas para que essas crianças consigam se desenvolver (Fábio Escalace, YouTube 2024, 00:58:15 a 00:58:50).

A colaboração entre investigações desenvolvidas em Universidades, no âmbito dos Estudos Surdos, deve ser desenvolvida de acordo com aquilo que é a realidade das escolas, ponderam as docentes. A investigação em Estudos Surdos, executada por ouvintes sem a participação da Comunidade Surda e/ou do investigador Surdo, provoca uma ruptura entre produção de conhecimento e engajamento político, o que traz prejuízos para a educação de Surdos. Neste sentido, é preciso investigar com, produzir conhecimento com, mais do que investigar para/sobre temáticas vinculadas aos Surdos e às Línguas Gestuais/de Sinais.

A representação social, que investigadores têm da pessoa Surda, prejudica a participação em pesquisas, pois há desconfiança quanto à capacidade das pessoas Surdas em ler e escrever em português, além do fato de que geralmente a investigação conduzida por uma pessoa Surda é desvalorizada em relação a uma pessoa ouvinte. Em muitos casos, Surdos não dispõem de tradutores-intérpretes para o desenvolvimento de investigações, para acompanhamento em pesquisas de campo e para tradução de materiais da LGP/Libras para o português, nas instituições em que desenvolvem as suas pesquisas.

Efetivamente os estudos são uma prova de que os métodos podem contribuir para o desenvolvimento da educação. Se, por exemplo, estes desafios que vão surgindo, se nós não tivermos unidos – todas as pessoas – se as pessoas não trabalharem em conjunto e a própria área dos estudos Surdos não estiverem com as escolas, com a realidade da educação, o próprio ensino certamente não se irá desenvolver, não haverá desenvolvimento. Portanto, eu acho que falta aqui uma colaboração entre os estudos Surdos e aquilo que é a realidade das escolas (Carla Benge, YouTube 2024, 00:22:30 a 00:23:00).

Outro problema é que a produção de conhecimento pelos Estudos Surdos não está contemplada na formação da maioria dos docentes. Em geral, docentes têm contato com as Línguas Gestuais/de Sinais, mas não há formação que aborde temáticas vinculadas à educação bilíngue, metodologias de ensino a Surdos, ensino de Português como L2, currículo, avaliação em Línguas Gestuais/de Sinais, adaptações curriculares, dentre outros temas.

As lutas da Comunidade Surda persistem, sendo necessário um trabalho pedagógico colaborativo para atender as necessidades e a diversidade de estudantes Surdos, especialmente quando estão isolados e carecem de contato com outros Surdos e com as associações de Surdos; quando são estrangeiros e necessitam de programa de língua estrangeira, em articulação com sua língua materna; quando ingressam tardiamente nas escolas e requerem programas e adequações curriculares, dentre outras possibilidades.

4 Conclusões

As questões que envolvem os desafios da educação de Surdos estão vinculadas à construção de políticas educacionais e linguísticas que também carecem, por sua vez, de implementação nos espaços escolares, especialmente com investimento na formação inicial e continuada de professores/as, na promoção e valorização das Línguas Gestuais/de Sinais, tanto na escola quanto fora dela, na produção de materiais bilíngues, no tipo de relação entre educação especial e educação de Surdos nos espaços educacionais, na promoção e divulgação da História e Cultura Surdas como um movimento de resistência contra a discriminação e segregação.

A docência de professores/as que atuam na educação de Surdos foi destacada e está resumidamente indicada neste artigo, a constituir ainda tópicos que necessitam de uma abordagem mais detalhada. As narrativas e discursos recolhidos permitem destacar que o/a professor/a de LGP/LIBRAS desenvolve atividades em torno do desenvolvimento da LGP/LIBRAS e da Cultura Surda – o que é em si mesmo um ato de resistência. Ao coexistirem nos espaços educacionais, as pessoas Surdas, que gestuam/sinalizam, tornam visível o *topos* de resistência, pois subsiste ainda a falta de rigor nas atribuições das funções dos/as professores/as de Educação Especial, dos/as intérpretes e professores/as de LGP/LIBRAS.

As participantes da pesquisa salientaram que não existem alunos/as bilíngues sem professores/as bilíngues; que faltam professores/as bilíngues para o ensino de diversas áreas científicas; que há carência de abordagem sobre a educação de Surdos, de pedagogia Surda, em cursos de formação inicial e contínua de professores/as e que capacite os/as profissionais. Por fim, elas indicaram que os concursos públicos e a colocação de professores/as nas escolas não apresentam uma política de gestão escolar coerente com as atribuições necessárias para a atuação na educação bilíngue de Surdos.

Elas consideram que a educação bilíngue deve privilegiar o ensino da Língua Gestual/Língua de Sinais como língua primeira e, a Língua Portuguesa, como segunda na modalidade escrita, e com a sua oralidade com cariz facultativo, contrastando epistemologicamente com as abordagens mais centradas em adaptações, reduções curriculares e estratégias voltadas para o desenvolvimento da oralidade, da reabilitação e terapia, típicas da abordagem clínica da educação especial. As docentes indicam a necessidade de autonomização da educação bilíngue de Surdos da Educação Especial, visto que, de forma generalizada, a supremacia da

língua majoritária ouvinte continua a ser a que dita e impõe leis e práticas, contradizendo o que ela própria plasmou na lei: a inclusão.

Embora experiências com educação bilíngue tenham antecedido o reconhecimento, enquanto modalidade de ensino a Surdos em legislação vinculada às políticas educacionais, nos dois países em estudo percebe-se que um longo caminho de lutas empenhadas pela comunidade Surda e pela comunidade escolar ainda se fez necessário, tanto no âmbito do reconhecimento e implementação de políticas mais eficazes, quanto na formação de professores/as e gestão da educação bilíngue. Neste sentido, o reconhecimento das Línguas Gestuais/de Sinais, como línguas das pessoas Surdas (L1), e a língua portuguesa como língua segunda (L2), foi um marco importante no reconhecimento dos direitos linguísticos das pessoas Surdas enquanto minoria linguística.

Tal reconhecimento buscou implementar ou aprimorar a educação de Surdos nos espaços de escolarização. Entretanto, é necessário ainda construir colaborativamente propostas para que a educação bilíngue atenda aos estudantes Surdos em sua aprendizagem ao longo da vida. Para isso, é necessário que exista uma monitorização de qualidade da aplicação dos modelos de educação bilíngue à realidade(s) do terreno e que as leis não sejam apenas um conjunto de intenções não realizáveis. Essa monitorização só será eficaz se quem a efetue tenha conhecimento da área que está a monitorizar. Essas equipes de monitorização devem ser constituídas por Surdos e ouvintes com formação na área da educação bilíngue para alunos Surdos e competentes em ambas as línguas – LGP/LIBRAS e português –, permitindo a diferenciação pedagógica tão necessária para a heterogeneidade da população escolar surda.

As entrevistas com as docentes Surdas, que atuam na educação bilíngue, indicam que a implementação da educação bilíngue não é tarefa simples e exige trabalho colaborativo entre instituições, agentes educacionais e com a participação ativa da Comunidade Surda. A questão central é o reconhecimento da importância da aquisição e desenvolvimento das Línguas Gestuais/de Sinais na educação pelas crianças Surdas na mais tenra idade. A escola é o berço da Cultura Surda e, em muitos casos, o primeiro espaço em que os Surdos têm a possibilidade de adquirir pleno desenvolvimento linguístico e pedagógico. Isso requer a atuação de docentes Surdos, tradutores intérpretes de LGP/LIBRAS e docentes bilíngues na educação de Surdos.

Uma crítica recorrente é quando professores/as ouvintes atuam na educação de Surdos sem competência suficiente em Língua Gestual/de Sinais, e desconhecimento da Cultura Surda e de metodologias da pedagogia Surda (como anteriormente referida na introdução), comprometendo significativamente o processo de ensino/aprendizagem do estudante Surdo. Desse modo, a educação de Surdos requer política linguística e educacional, com investimento na formação profissional (inicial e continuada) de docentes e com políticas públicas de valorização e divulgação das Línguas Gestuais/de Sinais. Isso pressupõe uma revisão das políticas educativas no que concerne à educação bilíngue e à sua complementaridade com as propostas de educação inclusiva nas políticas públicas direcionadas aos estudantes Surdos, que problematize o atual privilégio da oralidade, em detrimento das Línguas Gestuais/de Sinais, e que questione a prevalência do neo-oralismo (Peluso, 2016) e linguicismo (Skutnabb-Kangas; Phillipson, 1996) (referidos anteriormente na análise e discussão de resultados) nos esforços de resistência pelo direito à educação de Surdos.

Declaração de autoria

Durante a realização da investigação, Lodenir Karnopp indicou e convidou as pessoas Surdas do Brasil, bem como os/as tradutores/as-intérpretes de Libras, para participarem da entrevista. Cristina Gil e Paulo Vaz de Carvalho indicaram e convidaram as professoras Surdas e os tradutores-intérpretes de LGP de Portugal para a realização das entrevistas. Os três autores atuaram durante as *lives*, realizaram a transcrição e revisão dos dados. Os documentos e informações referentes a cada país foi providenciado de acordo com a nacionalidade dos autores. A escrita do artigo, a elaboração dos referenciais teórico-metodológicos e o desenvolvimento das análises foram realizados pelos três autores de modo colaborativo.

Referências

- AMARAL, M. A. O Instituto Jacob Rodrigues Pereira Contribuições para a sua História Recente. *Para Além do Silêncio*, Revista da Associação Portuguesa de Professores e Técnicos de Reabilitação de Crianças e Jovens Surdos, p. 25-33, 1996.
- BOURDIEU, P. *The Field of Cultural Production*. Oxford: Columbia University Press, 1993.
- BRASIL. Lei no 10.436, de 24 de abril de 2002: Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências. *Diário Oficial da União*, Brasília, n. 79, seção 1, p. 23, 25 abr. 2002.
- BRASIL. Decreto no 5.626, de 22 de dezembro de 2005. Regulamenta a Lei no 10.436, de 24 de abril de 2002, que dispõe sobre a Língua Brasileira de Sinais – Libras e o art. 18 da lei no 10.098, de 19 de dezembro de 2000. *Diário Oficial da União*, Brasília, n. 246, seção 1, p. 28, 23 dez. 2005.
- BRASIL. Ministério da Educação. *Política Nacional de Educação Especial na Perspectiva da Educação Inclusiva*. Brasília, 2008. Disponível em: <http://portal.mec.gov.br/arquivos/pdf/politicaeduc ESPECIAL.pdf>. Acesso em: 15 mar. 2023.
- BRASIL. Lei nº 13.146, de 6 de julho de 2015. Lei brasileira de inclusão da pessoa com deficiência (estatuto da pessoa com deficiência). *Diário Oficial da União*, Brasília, 2015. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 15 mar. 2025.
- BRASIL. Conselho nacional de saúde. *Resolução nº 510/2016*. Brasília, 2016. Disponível em: <http://conselho.saude.gov.br/resolucoes/2016/reso510.pdf>. Acesso em: 10 mar. 2025.
- BRASIL. Ministério da Educação. *Resolução nº 19, de 18 de março de 2019*. Aprova o Regulamento do Atendimento Educacional Especializado (AEE) no âmbito do Instituto Federal de Educação, Ciência e Tecnologia Baiano. Salvador, 2019. Disponível em: <https://www.ifbaiano.edu.br/unidades/lapa/files/2019/12/resolucao-19-2019.pdf>. Acesso em: 15 mar. 2025.
- BRASIL. Lei nº 14.191, de 3 de agosto de 2021. Dispõe sobre a educação bilíngue de surdos e altera dispositivos da Lei nº 9.394, de 20 de dezembro de 1996, e dá outras providências. *Diário Oficial da União*: seção 1, Brasília, DF, ano 158, n. 146, p. 1-5, 4 ago. 2021. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14191.htm. Acesso em: 7 out. 2025

- CARVALHO, P. V. A educação bilíngue para surdos: uma experiência portuguesa. In: CONGRESSO INTERNACIONAL, 11; SEMINÁRIO NACIONAL DO INES, 17, 2012. *Anais [...]*. 2012. Disponível em: <https://www.porsinal.pt/index.php?ps=artigos&idt=artc&cat=7&idart=308>. Acesso em: 10 mar. 2025.
- CARVALHO, P. V.; MINEIRO, A. Políticas linguísticas na educação de surdos em Portugal. *Educação Unisinos*, São Leopoldo, v. 24, p. 1-11, 2020. DOI: <https://doi.org/10.4013/edu.2020.241.19456>.
- GIL, C. “O Mythomoteur Surdo” em Línguas de Sinais: Cultura Educação Identidade. In: CORREIA, I. S. C.; CUSTÓDIO, P. B.; CAMPOS, R. M. R. *Línguas de Sinais: Cultura Educação Identidade*. Lisboa: Ex-Libris, 2019. p. 75-92.
- GIL, C. *Deaftopia: Utopian Representations and Community Dreams by the Deaf*. 2020. Tese (Doutoramento) – Universidade Católica Portuguesa, Lisboa, 2020.
- KARNOPP, L. B.; KLEIN, M.; LUNARDI-LAZZARIN, M. L. Cultura Surda na Contemporaneidade: (re) significações. *Revista Espaço*, Rio de Janeiro, INES, n. 56, p. 61-77, 2021.
- LACERDA, C. B. F. Um pouco da história das diferentes abordagens na educação dos surdos. *Cadernos Cedes*, Campinas, v. 19, n. 46, p. 68-80, 1998. DOI: <https://doi.org/10.1590/s0101-32621998000300007>.
- LADD, P. *Seeing Through New Eyes: Deaf Culture and Deaf Pedagogies - the unrecognized curriculum*. San Diego: Dawn Sign Press, 2022.
- LADD, P.; GONÇALVES, J. C. Cultura Surda e o Desenvolvimento de Pedagogias Surdas. In: KARNOPP, L. B.; KLEIN, M.; LUNARDI-LAZZARIN, M. L. (Org.). *Cultura Surda na Contemporaneidade: negociações, intercorrências e provocações*. Canoas: Ulbra, 2011. p. 296-329.
- LODI, A. C. “Educação em língua brasileira de sinais: um direito dos surdos a ser assegurado”. *Cadernos de Linguagem e Sociedade*, v. 22, n. 2, p. 316-330, 2021. DOI: <https://doi.org/10.26512/les.v22i2.40916>.
- PELUSO, L. La lengua de señas uruguaya y las políticas lingüísticas. *Revel*, v. 14, n. 26, p. 120-146, 2016. Disponível em: <http://www.revel.inf.br/files/b8ab9d9f392bbe6d21545e30d2521438.pdf>. Acesso em: 10 mar. 2025.
- PORTUGAL. Despacho nº 7520/98. *Diário da República*, 2.ª série, número 104, 6 de maio, 1998. Disponível em: <https://diariodarepublica.pt/dr/detalhe/despacho/7520-1998-3149475>. Acesso em: 11 mar. 2025.
- PORTUGAL. Decreto-lei n.º 3/2008. *Diário da República*, 1.ª série, número 4, 7 de janeiro, 2008a. Disponível em: <https://diariodarepublica.pt/dr/detalhe/decreto-lei/3-2008-386871>. Acesso em: 11 mar. 2025.
- PORTUGAL. Lei n.º 21/2008. *Diário da República*, 1.ª série, número 91, 12 de maio, 2008b. Disponível em: <https://diariodarepublica.pt/dr/detalhe/lei/21-2008-249230>. Acesso em: 15 mar. 2025.
- PORTUGAL. Decreto-lei n.º 16/2018. *Diário da República*, 1.ª série, número 47, 7 de março, 2018a. Disponível em: <https://diariodarepublica.pt/dr/detalhe/decreto-lei/16-2018-114825662>. Acesso em: 17 mar. 2025.
- PORTUGAL. Decreto-lei n.º 54/2018. *Diário da República*, 1.ª série, número 129, 6 de julho, 2018b. Disponível em: <https://diariodarepublica.pt/dr/detalhe/decreto-lei/54-2018-115652961>. Acesso em: 11 mar. 2025.
- PRATA, I. *Mãos que Falam*. Laboratório de Fonética da Universidade de Lisboa & Divisão do Ensino Especial da Direção Geral do Ensino Básico do Ministério da Educação. Lisboa, 1980.

SKUTNABB-KANGAS, T.; PHILLIPSON, R. Linguicide and Linguicism. Kontaktlinguistik. Contact Linguistics. Linguistique de Contact. Ein Internationales Handbuch Zeitgenössiger Forschung. V. 1. In: GOEBL, H.; NELDE, P. H.; STARÝ, Z. K.; WÖLCK, W. (ed.). *An International Handbook of Contemporary Research*. Manuel International Des Recherches Contemporaines. Berlin; New York: Walter de Gruyter, 1996. p. 667-675.

SOFIATO, C. G.; CARVALHO, P. V.; COELHO, O. Experiências de educação bilíngue para surdos: entrelaçamentos entre Brasil e Portugal. In: SILVA, R. A. F.; HOLLOSI, M (org.). *Educação de surdos, linguagens e experiências*. Uberlândia: Navegando Publicações, 2021. DOI: <https://doi.org/10.29388/978-65-86678-60-4-0>.

SVARTHOLM, K. Bilingual Education for Deaf Children in Sweeden. *International Journal of Bilingual Education and Bilingualism*, London, 2010.

A tipologia de tópicos no português brasileiro na interface entre sintaxe e estrutura informacional

The Topic Typology in Brazilian Portuguese at the Interface Between Syntax and Information Structure

Renato Lacerda

Universidade de São Paulo (USP) | São Paulo | SP | BR

renatolacerda@usp.br

<https://orcid.org/0000-0001-9118-4949>

Resumo: Este artigo investiga a distribuição sintática e as condições de licenciamento de tópicos *aboutness* (*-shifting*), tópicos de informação dada e tópicos contrastivos no português brasileiro, a partir da tipologia de Frascarelli e Hinterhölzl (2007) e Bianchi e Frascarelli (2010). A investigação busca determinar quais tipos de tópicos podem aparecer nos seguintes contextos sintáticos: em posição canônica (i.e., sem alteração da ordem básica dos constituintes da sentença), deslocado para a periferia esquerda da sentença, deslocado para a posição pós-verbal na periferia de vP e deslocado à direita em posição final. Com base nas observações empíricas, o texto discute as condições de licenciamento de cada tipo de tópico e argumenta que apenas os tópicos do tipo *aboutness* (*-shifting*) estão sujeitos a uma bipartição da sentença em tópico-comentário, ao passo que tópicos de informação dada e tópicos contrastivos são licenciados com base em sua posição relativa ao foco da sentença. Ao final, o artigo propõe um mapeamento das posições marcadas de tópicos no português brasileiro e defende uma visão contextual para o fenômeno.

Palavras-chave: topicalização; tipologia de tópicos; periferia esquerda; periferia de vP.

Abstract: This article investigates the syntactic distribution and the licensing conditions of aboutness (*-shifting*) topics, given topics, and contrastive topics in Brazilian Portuguese, based on the topic typology of Frascarelli and Hinterhölzl (2007) and Bianchi and Frascarelli (2010). The investigation seeks to determine which types of topics can appear in the following syntac-



tic contexts: in canonical position (i.e., without a change in the basic order of the constituents of the sentence), dislocated to the left periphery of the sentence, dislocated to the postverbal position in the periphery of vP, and right-dislocated in a final position. Based on the empirical observations, the text discusses the licensing conditions of each type of topic and argues that only aboutness(-shifting) topics are subject to a bipartition of the clause as topic-comment, whereas given topics and contrastive topics are licensed based on their position relative to the focus of the sentence. In the end, the article proposes a mapping of the marked topic positions in Brazilian Portuguese and defends a contextual view of the phenomenon.

Keywords: topicalization; topic typology; left periphery; periphery of vP.

1 Introdução

Nos estudos do português brasileiro, tanto gerativistas como funcionalistas, há uma vasta literatura a respeito de construções de topicalização envolvendo a periferia esquerda da sentença (veja, e.g., Guesser; Mioto, 2016; Kato, 1989, 1998; Kato; Raposo, 2007; Macedo-Costa; Andrade, 2015; Pontes, 1987; Mioto, 2001; Silva, 2017; entre muitos outros). Na linha dos estudos da Escola de Praga (Firbas, 1964; Mathesius, 1983[1929]), a maioria destes trabalhos assume (explícita ou implicitamente) que a topicalização articula (i.e., particiona) a sentença em um formato de *tópico-comentário* (ou *tema-rema*), como ilustrado em (1). Enquanto o elemento deslocado *este livro* é interpretado como o tópico, aquilo sobre o que se fala, a sentença *eu ainda não li (ele)* é interpretada como o comentário, aquilo que se fala sobre o referente denotado pelo tópico.

- 1) a. Este livro, eu ainda não li (ele).
b. [Este livro]_{TÓPICO} [eu ainda não li (ele)]_{COMENTÁRIO}

Para além de construções como aquela vista em (1), diversos trabalhos de linha gerativista dos últimos vinte anos têm mostrado (i) que a categoria de tópico não é uniforme — isto é, há diferentes tipos de tópicos (e.g., Bianchi; Frascarelli, 2010; Büring, 2003, 2016; Frascarelli; Hinterhölzl, 2007; Lacerda, 2020, 2021) — e (ii) que tópicos não estão restritos a posições deslocadas na periferia esquerda da sentença — isto é, também podem aparecer em posições internas à sentença (e.g., Belletti 2001, 2004; Frey, 2004; Jayaseelan, 2001; Lacerda, 2020, 2021). Apesar destes avanços, falta ainda na literatura sobre o português brasileiro um trabalho que leve a fundo em consideração estes dois fatores conjuntamente na investigação da topicalização. Este artigo, portanto, se propõe a preencher essa lacuna.

Para atingir esse objetivo, adotaremos, em relação ao ponto (i) acima, a tipologia tripartite de tópicos de Bianchi e Frascarelli (2010) (baseada em Frascarelli; Hinterhölzl, 2007), que apontam, com base em dados do italiano, do inglês e do alemão, que cada tipo de tópico está associado a uma distribuição sintática e um evento tonal distintos. Buscaremos, neste trabalho, determinar a distribuição sintática em português brasileiro dos tipos de tópicos apontados em (2) abaixo, investigando-os individualmente e comparativamente.

- 2) a. Tópico *aboutness(-shifting)* (veja, e.g., Heim, 1982; Reinhart, 1981)
- b. Tópico de informação dada (veja, e.g., Givón, 1983; Schwarzschild, 1999)
- c. Tópico contrastivo (veja, e.g., Büring, 2003; Wagner, 2012)

Em relação ao ponto (ii) acima, investigaremos se cada tipo de tópico pode ser licenciado em cada um dos quatro contextos sintáticos descritos em (3) e representados em (4), respectivamente.

- 3) a. Em posição canônica (i.e., sem alteração da ordem básica dos constituintes);
 - b. Deslocado para a periferia esquerda da sentença (CP);
 - c. Deslocado para a posição pós-verbal na periferia de vP;
 - d. Deslocado à direita em posição final.
-
- 4) a. *sujeito verbo objeto*
 - b. tópico, *sujeito verbo objeto*
 - c. *sujeito verbo*, tópico, *objeto*
 - d. *sujeito verbo objeto*, tópico

Ao determinar a distribuição de cada tipo de tópico, discutiremos suas condições de licenciamento na interface entre sintaxe e estrutura informacional. Argumentaremos que a tradicional articulação da sentença em tópico-comentário, como em (1) acima, é um requerimento apenas de tópicos do tipo *aboutness(-shifting)*, enquanto tópicos de informação dada e tópicos contrastivos são licenciados posicionalmente em relação ao foco, sem exigir uma bipartição especial da sentença. Discutiremos ainda como os diferentes tipos de tópicos estão associados a diferentes conjuntos de condições de licenciamento, avaliadas em relação ao contexto sintático.

O artigo está organizado da seguinte forma. Na Seção 2, apresentaremos as propriedades definidoras de cada tipo de tópico da tipologia adotada; na Seção 3, discutiremos a distribuição sintática de cada tipo de tópico no português brasileiro; na Seção 4, discutiremos as relações hierárquicas entre os três tipos de tópicos ora analisados e proporemos um mapa de suas posições marcadas na sentença; na Seção 5, concluímos o artigo com a posição de que é imprescindível considerar, no estudo da topicalização, uma tipologia de tópicos adequada.

2 Tipologia

Para apresentar nossa tipologia de tópicos, adotaremos como pano de fundo duas noções fundamentais acerca do papel da estrutura informacional (Chafe, 1976) na conversação. Primeiro, assumimos com Krifka (2008) que o terreno comum (*common ground*; Stalnaker, 1974) partilhado entre falantes se constrói em duas partes complementares: *conteúdo* e *gerenciamento*. O conteúdo do terreno comum inclui, sobretudo, as proposições mutuamente aceitas como verdadeiras pelos interlocutores. Já o gerenciamento do terreno comum diz respeito ao uso pragmático das expressões linguísticas e se manifesta nos objetivos comunicativos do falante com relação à organização e ao desenvolvimento do conteúdo do terreno comum. Segundo, assumimos com Reinhart (1981) e Heim (1982) que o discurso se organiza a partir de elementos referenciais, isto é, tópicos potenciais. Assim, as proposições que fazem parte do terreno comum se organizam a partir dos referentes a que dizem respeito.

Adotadas essas duas noções, entendemos que a topicalização diz respeito ao gerenciamento do terreno comum, na medida em que serve para apresentar (i.e., introduzir ou reintroduzir) referentes para a organização da conversa. Assim, cada tipo diferente de tópico deve ser entendido como exercendo uma função diferente no gerenciamento do terreno comum. É sob esse prisma que definiremos nesta seção a tipologia de tópicos adotada neste artigo a partir dos trabalhos de Frascarelli e Hinterhölzl (2007) e de Bianchi e Frascarelli (2010).

2.1 Tópicos aboutness(-shifting)

Na tradicional topicalização *aboutness*, a bipartição da sentença em *tópico-comentário* estabelece a relação de que o comentário *é sobre* o referente denotado pelo tópico sentencial, definido por Reinhart (1981) como a entidade sobre a qual a sentença fala. A partir de Reinhart (1981), estabeleceu-se na literatura o teste “me fale sobre X” para identificar o tópico *aboutness*. Como ilustrado no exemplo do português brasileiro em (5), o falante A cria um contexto para que o falante B empregue um tópico *aboutness*, fazendo um comentário sobre o livro.¹ A proposição *o final dele é surpreendente* é adicionada ao terreno comum como sendo sobre *este livro*.

- 5) A: Me fale alguma coisa sobre o livro que você acabou de ler.
B: *Este livro*_{TA}, o final dele é surpreendente.

Frascarelli e Hinterhölzl (2007) e Bianchi e Frascarelli (2010) refinam a noção de tópico *aboutness* de Reinhart (1981) para capturar o fato de que este tipo de tópico pode não somente introduzir um referente sobre o qual se comenta, mas também mudar a direção da conversa para um novo referente. As autoras introduzem, assim, a noção de *aboutness-shifting*. Em português brasileiro, uma estratégia de *aboutness-shifting* pode ser identificada pelo uso da partícula *já* (quando associada a um tópico), como ilustrado no diálogo em (6), em que o falante B muda o tópico da conversa da França para a Espanha.²

¹ Tópicos *aboutness* serão indicados pelo subscrito TA.

² Para uma discussão de conectores discursivos, entre eles o *já*, em português brasileiro, veja Miranda e Silva (2015).

- 6) A: A França é linda. O Pedro gosta muito de lá.
B: Já a Espanha_{TA}, ele ainda não conheceu. Ele precisa muito ir lá.

Bianchi e Frascarelli (2010) afirmam que os tópicos *aboutness* “identificam a entrada única sob a qual a proposição assertada deve ser armazenada no terreno comum proposicional” (p. 63).³ Transpondo essa definição para a famosa metáfora da ficha de arquivo (*filecard*) de Heim (1982), que compara referentes discursivos a entradas em um arquivo, podemos entender tópicos *aboutness* como exercendo a função de dar ao interlocutor a instrução de selecionar uma única ficha de arquivo, na qual a proposição denotada pelo comentário deve ser armazenada. Observe o diálogo em (7).

- 7) A: Você tem notícias do pessoal da faculdade?
B: O João_{TA}, ele se mudou pra Europa. Já a Maria_{TA}, ela foi pro Canadá.

Em (7), o falante B primeiro introduz o referente de *o João*, selecionando sua ficha de arquivo, e nela adiciona a proposição *ele se mudou pra Europa*; em seguida, em uma estratégia de *aboutness-shifting*, o falante B introduz o referente de *a Maria*, acessando uma nova ficha de arquivo, na qual então adiciona a proposição *ela foi pro Canadá*.

2.2 Tópicos de informação dada

Tópicos de informação dada, também chamados na literatura de tópicos familiares, têm a função de retomar referentes que já foram introduzidos no terreno comum (i.e., informação dada; Schwarzschild, 1999), indicando uma estratégia de continuidade de tópico (Givón, 1983). Isso é ilustrado pelo diálogo em (8), em que os referentes de *a Maria* e de *a festa de Natal* são tomados como informação dada pelo falante B, após serem ativados no terreno comum pelo falante A.⁴ Ao topicalizar *a Maria*, o falante B indica que este referente deve permanecer ativo na conversa.

- 8) A: O que você deu pra Maria na festa de Natal?
B: A Maria_{TD}, eu dei *dois livros*_F (pra ela) (na festa de Natal).

Ao retomar um referente saliente no discurso, o tópico de informação dada pode indicar que a conversa continuará a ser organizada em torno daquele referente, de modo que as novas proposições a serem adicionadas ao terreno comum continuem a ser relacionadas ao mesmo referente.⁵ Assim, a topicalização de *a Maria* em (8B) facilita uma sequência do diálogo como aquela vista em (9).

³ “[Aboutness] topics identify the unique entry under which the asserted proposition must be stored in the propositional [common ground].”

⁴ Tópicos de informação dada serão indicados pelo subscrito TD, enquanto focos serão indicados pelo subscrito F.

⁵ Como aponta um parecerista anônimo, a função de continuidade de tópico também é atribuída na literatura ao chamado *continuing aboutness topic*. Neste trabalho, não adotamos esta nomenclatura por entender que tópicos *aboutness* e tópicos de informação dada são significativamente distintos, como se mostrará ao longo do texto; assim, adotamos uma tipologia que os separa, ao invés de aproximá-los.

- 9) Ela adorou os livros. Ela gosta muito de ler.

Além disso, a topicalização de *a Maria* em (8B), ao reforçar a seleção de sua “ficha de arquivo” correspondente (proposta pelo falante A) naquele momento da conversa, pode servir também para preparar o terreno para uma futura mudança de tópico, como em (10).

- 10) A: O que você deu pra Maria na festa de Natal?
B: A *Maria*_{TD}, eu dei *dois livros*_F (pra ela) (na festa de Natal).
(Ela adorou os livros. Ela gosta muito de ler.)
...*Já pro João*_{TA}, eu dei *uma camiseta*_F.

Como afirmam Bianchi e Frascarelli (2010), tópicos de informação dada “simplesmente se referem ao conteúdo do terreno comum existente com uma função resumptiva” (p. 59).⁶

2.3 Tópicos contrastivos

Bianchi e Frascarelli (2010) adotam a definição de tópico contrastivo proposta por Büring (2003), para quem tópicos contrastivos indicam uma “estratégia” de resposta parcial a uma pergunta. Ao invés de responder à *pergunta em discussão* (*question under discussion*; Roberts, 2012 [1996]) em uma dada conversa, o uso de topicalização contrastiva cria uma família de sub-perguntas. Observe o diálogo em (11) abaixo, traduzido e adaptado de Büring (2003, p. 525).⁷ Ao invés de responder completamente à pergunta do falante A, o falante B a subdivide em duas sub-perguntas, vistas em (12), e responde apenas à sub-pergunta SubQ1, assim respondendo parcialmente à pergunta em (11A).

- 11) A: O que os pop stars vestiram? *pergunta em discussão*
B: As *pop stars mulheres*_{TC} vestiram *cafetãs*_F. *resposta parcial*

- 12) Q: O que os pop stars vestiram?
SubQ1. O que *as pop stars mulheres* vestiram?
SubQ2. O que *os pop stars homens* vestiram?

A não-resolução da pergunta em discussão é uma propriedade importante da topicalização contrastiva (*unresolvedness*, segundo Büring, 2016) e gostaríamos de apontá-la como sendo crucial para distingui-la da topicalização *aboutness*. Comparemos os diálogos em (13) e (14). Observe que a topicalização contrastiva em (13B) é pragmaticamente feliz para criar uma família de sub-perguntas como, por exemplo, aquela vista em (15) e assim responder parcialmente à pergunta (implícita) em (15Q), deixando a pergunta em discussão SubQ1 não-resolvida (importante: neste momento da conversa). Por outro lado, o uso de uma estratégia de *aboutness-shifting* (evidenciado pela partícula *já*) em (14B1) é infeliz. Ao contrário da topicalização contrastiva, a estratégia de *aboutness-shifting* requer o endereçamento (ou resolução)

⁶ “They simply refer to the existing [common ground] content with a retrieval function.”

⁷ Tópicos contrastivos serão indicados pelo subscrito TC.

da sub-pergunta correlata anterior, como se vê pelo contraste em (14). Uma vez endereçada a sub-pergunta sobre o Pinker em (14B2), o uso de *já* (no enunciado seguinte) se torna aceitável.⁸

- 13) A: Quantos livros do Pinker você leu pro curso de sintaxe?
 B: Bem, *do Chomsky*_{TC}, eu li *três livros*_F.
- 14) A: Quantos livros do Pinker você leu pro curso de sintaxe?
 B1: #*Já do Chomsky*_{TA}, eu li *três livros*_F.
 B2: Só um. *Já do Chomsky*_{TA}, eu li *três livros*_F.
- 15) Q: Quantos livros (de diferentes autores) você leu pro curso de sintaxe?
 SubQ1. Quantos livros *do Pinker* você leu pro curso de sintaxe? ✕
 SubQ2. Quantos livros *do Chomsky* você leu pro curso de sintaxe? ✓
 SubQ3. Quantos livros *do Kayne* você leu pro curso de sintaxe? ✕

Vimos nesta seção que os diferentes tipos de tópicos de nossa tipologia desempenham funções distintas no gerenciamento do terreno comum; em outras palavras, o tópico não configura uma categoria uniforme. Veremos na próxima seção que também é distinta a distribuição sintática dos diferentes tipos de tópicos em português brasileiro.

3 Distribuição

Investigaremos detalhadamente nesta seção a distribuição sintática de cada um dos três tipos de tópicos apresentados na seção anterior. Em particular, buscaremos determinar se cada tipo de tópico pode ser realizado (i) em posição canônica (i.e., sem alteração da ordem básica dos constituintes da sentença), (ii) deslocado para a periferia esquerda da sentença (i.e., no domínio do CP), (iii) deslocado para a posição pós-verbal na periferia de vP e (iv) deslocado à direita em posição final. Estes contextos sintáticos são representados em (16a–d), respectivamente.

- 16) a. *sujeito verbo objeto*
 b. tópico, *sujeito verbo objeto*
 c. *sujeito verbo*, tópico, *objeto*
 d. *sujeito verbo objeto*, tópico

⁸ É importante apontar que a resolução de uma pergunta não se confunde com sua resposta. Como visto em (i) abaixo, apontar que uma pergunta não tem uma resposta (conhecida) é suficiente para resolvê-la e prosseguir com a mudança de tópico.

i) A: Quantos livros do Pinker você leu pro curso de sintaxe?
 B: Não sei. *Já do Chomsky*_{TA}, eu li *três livros*.

Nesta seção, trataremos da distribuição de cada tipo de tópico separadamente, para depois, na Seção 4, tratar das construções com tópicos múltiplos.

3.1 Tópicos aboutness(-shifting)

O tópico *aboutness(-shifting)* corresponde ao tipo canônico de tópico, identificado pela tradição da Escola de Praga. Assim sendo, este tópico ocorre em posição que se conforma à tradicional articulação da sentença em *tópico-comentário*, de modo que pode ser realizado na periferia esquerda da sentença, mas não na posição pós-verbal na periferia de vP, como mostra o contraste entre (17B1) e (17B2). Enquanto (17B1) soa natural, (17B2) não é compatível com topicalização *aboutness*, uma vez que introduz informação nova (i.e., *eu comi*) à esquerda do tópico.⁹

17) A: Me conta alguma coisa sobre a feira renascentista que você foi ontem!

B1: *Na feira renascentista*_{TA}, eu comi várias comidas típicas.

B2: #Eu comi, *na feira renascentista*_{TA}, várias comidas típicas.

Ainda que o teste “me fale sobre X” tenha se tornado praxe na literatura para identificar tópicos *aboutness*, é importante apontar que a menção prévia ao referente relevante (e.g., *feira renascentista* em (17A) acima) torna o contexto compatível também com a topicalização de informação dada, em que o tópico simplesmente retoma o referente cuja ficha de arquivo já foi selecionada, ao invés de indicar tal seleção por si só. Diante disso, a noção de *aboutness-shifting* de Frascarelli e Hinterhölzl (2007) e Bianchi e Frascarelli (2010) é de extrema importância, uma vez que esta se apresenta como uma estratégia não-ambígua para elicitar um tópico *aboutness*, ao introduzir um referente discursivamente novo e não-contrastivo. Assim, se adaptarmos o diálogo em (17) para um contexto de *aboutness-shifting* com o uso da partícula *já*, como em (18), o contraste entre as respostas se torna ainda maior, na medida em que (18B1) é aceitável, enquanto (18B2) é agramatical — e não simplesmente pragmaticamente infeliz.

18) A: Você aproveitou bem pouco a festa junina, não é?

B1: Verdade. *Já na feira renascentista*_{TA}, eu comi várias comidas típicas.

B2: Verdade. *Eu comi, *já na feira renascentista*_{TA}, várias comidas típicas.

O contraste em (18) confirma, portanto, a conclusão que tiramos a partir do contraste em (17) acima: tópicos *aboutness(-shifting)* podem ser realizados na periferia esquerda da sentença, mas não na posição pós-verbal. Aqui devemos nos perguntar se o licenciamento de tópicos *aboutness* requer uma estrutura marcada de deslocamento à esquerda ou se basta que o elemento relevante esteja em posição inicial, na medida em que ambas as hipóteses são compatíveis com a restrição vista acima.

⁹ Um falante cooperativo pode tentar acomodar (à força bruta) a informação à esquerda do tópico em (17B2) como uma informação já presente no terreno comum, o que geraria uma outra estrutura em que na feira renascentista se comporta como um tópico de informação dada (ver Seção 3.2). Em todo o texto, os julgamentos são dados para a interpretação indicada, sem acomodações pragmáticas.

Para responder a essa pergunta, façamos uma breve excursão ao italiano, que é uma língua *pro-drop*. Nesta língua, os sujeitos nulos (*pro*) buscam seus antecedentes no tópico *aboutness* atualmente ativo na conversa (veja Calabrese, 1986; Frascarelli, 2007). Sujeitos pré-verbais, como mostra o exemplo de Rizzi (2010) em (19) (tradução nossa), podem funcionar como antecedentes para sujeitos nulos (*pro*). A depender de se (19c) segue (19a) ou (19b), o referente do sujeito nulo será o velho ou o rapaz, respectivamente. Isso demonstra que sujeitos iniciais na estrutura SVO podem ser interpretados como tópicos *aboutness* (veja também Rizzi 2018). Nas palavras de Reinhart (1981, p. 62) sujeitos são os tópicos “não-marcados”.

- 19) a. Un vecchio ha insultato un ragazzo.
 ‘Um velho insultou um rapaz.’
 b. Un ragazzo è stato insultato da un vecchio.
 ‘Um rapaz foi insultado por um velho.’
 c. Poi, *pro* se n’è andato.
 ‘E então se foi.’ (Rizzi 2010: 26)

Com os sujeitos pré-verbais estando em posição canônica no início da sentença, a articulação sujeito-predicado pode ser interpretada na interface como uma articulação tópico-comentário, como representado em (20). Assim, em português brasileiro, a ordem canônica SVO (estrutura não-marcada) permite a interpretação do sujeito como um tópico *aboutness*, inclusive na estratégia de *aboutness-shifting*, como em (21B1), com a estrutura marcada de deslocamento e resumpção vista em (21B2) também sendo uma possibilidade.

20) [Sujeito]_{TÓPICO} [Predicado]_{COMENTÁRIO}

- 21) A: O Pedro leu dez livros do Chomsky pra esse curso.
 B1: Já o João_{TA} não leu nenhum.
 B2: Já o João_{TA}, ele não leu nenhum.

Se qualquer outro elemento da sentença for deslocado para a periferia esquerda, como a expressão-*qu* nos exemplos em (22), a posição canônica de sujeito deixa de permitir a interpretação de tópico *aboutness(-shifting)*, de modo que a estrutura marcada de deslocamento (com resumpção) se torna a única possível, por devidamente colocar o elemento a ser interpretado como tópico *aboutness* na posição inicial da sentença.¹⁰

¹⁰ A obrigatoriedade do resumptivo em (22B2–B3) se dá por razões independentes (e ortogonais à presente discussão), uma vez que o português brasileiro não é uma língua *pro-drop* e o interrogativo deslocado bloqueia uma estrutura de sujeito nulo por *topic-drop* (o grau de marginalidade de (22B2) pode estar sujeito a variação dialetal, como aponta um parecerista anônimo). Note que não há a princípio nenhum problema em um tópico cruzar um interrogativo, como mostra (i) abaixo.

i) A: O Pedro deu dois livros pra Joana.
 B: Já pra Maria, quantos livros ele deu?

22) A: O Pedro leu dez livros do Chomsky pra esse curso.

B1: *O que *já o João*_{TA} leu?

B2: ?**Já o João*_{TA}, o que _ leu?

B3: *Já o João*_{TA}, o que ele leu?

Diante da restrição acima, qualquer elemento cuja posição canônica não seja inicial deverá recorrer a uma estrutura marcada de deslocamento para ser interpretado como tópico *aboutness*, o que é ilustrado pelo objeto direto em (23).

23) A: O Pedro teve que ler o Programa Minimalista pra esse curso.

B1: *Já o Barriers*_{TA}, o Pedro não teve que ler (ele).

B2: *O Pedro não teve que ler *já o Barriers*_{TA}.

Tendo visto que tópicos *aboutness* em português brasileiro podem aparecer na periferia esquerda e em posição canônica (neste caso, se for um sujeito em posição inicial) e não podem aparecer na posição pós-verbal na periferia de vP, finalmente vemos em (24) que a estrutura de deslocamento à direita também não é compatível com a topicalização *aboutness* (*-shifting*), independentemente da resumpção.

24) A: O João leu três livros pro curso de sintaxe.

B1: *Já o Pedro*_{TA} leu só dois livros.

B2: *Já o Pedro*_{TA}, ele leu só dois livros.

B3: *Leu só dois livros, *já o Pedro*_{TA}.

B4: *Ele leu só dois livros, *já o Pedro*_{TA}.

Presumivelmente, a estrutura de deslocamento à direita não particiona a sentença adequadamente no formato tópico-comentário, que é rígido no sentido de que o tópico deve preceder o comentário e por isso é restrito à posição inicial da sentença.

3.2 Tópicos de informação dada

Antes de falarmos sobre a topicalização de informação dada, precisamos discutir o estatuto da informação dada no gerenciamento do terreno comum. Primeiro, note que elementos da sentença previamente introduzidos no discurso podem aparecer em suas posições canônicas, como em (25).¹¹ Como se sabe, o foco em português brasileiro pode ser realizado *in situ* e identificado apenas pelo seu acento prosódico (e.g., Lacerda, 2024, 2025; Quarezemin, 2009), com a respectiva desacentuação dos elementos dados (cf. Schwartzchild, 1999), sem exigir uma alteração da ordem básica dos constituintes como nas línguas de foco final.¹²

¹¹ Quando necessário, informação dada não-topicalizada é indicada pelo subscrito D.

¹² Ao contrário do que ocorre em português brasileiro, em línguas de foco final os constituintes que fazem parte da informação dada devem preceder o foco, que aparece na última posição, como ilustrado em (i) com dados do português europeu.

i) A: Quem é que partiu a janela?

B1: #O *Paulo*_F partiu_D a janela_D.

B2: Partiu_D a janela_D o *Paulo*_F. (adaptado de Costa, 2000, p. 197)

- 25) A: O que você deu pra Maria na festa de Natal?
B: Eu_D dei_D dois livros_F pra Maria_D na festa de Natal_D.

A fim de evitar repetições desnecessárias, a informação dada que não tem proeminência no discurso é frequentemente apagada, de modo que as respostas curtas em (26B1) e (26B2) abaixo são mais naturais do que a resposta completa em (25B) acima.

- 26) A: O que você deu pra Maria na festa de Natal?
B1: Eu_D dei_D dois livros_F.
B2: Dois livros_F.

Diante disso, o não-apagamento de informação dada confere aos elementos relevantes alguma proeminência no discurso (isto é, têm alguma função a desempenhar no gerenciamento do terreno comum). Para ilustrar esse efeito, comparemos o cenário em (27) abaixo com aquele em (25)/(26) acima. No momento em que o falante B retorna ao falante A com a resposta de sua pergunta, a informação de que o João deu algo para a Maria continua sendo dada (i.e., presente no terreno comum), mas pode não estar mais ativa e/ou acessível para os participantes da conversa. Desse modo, a resposta curta em (27B2) pode não ser suficiente para reativar a pergunta contextualizadora em (27A), exigindo-se a resposta completa em (27B3), na qual os elementos dados são mais naturais se pronunciados com entonação normal, sem desacentuação.

- 27) A: Eu estava no exterior durante as festas de fim de ano.
Você consegue descobrir o que o João deu pra Maria na festa de Natal?
B1: Vou descobrir.
Horas depois, o falante B retorna ao falante A e diz:
B2: #Dois livros_F.
B3: O João_D deu_D dois livros_F pra Maria_D na festa de Natal_D.

É em face da aquisição de proeminência discursiva que ocorre a topicalização de informação dada. Quando um elemento é topicalizado em uma posição mais alta na sentença, o seu referente se torna mais acessível no discurso, indicando que a conversa deve continuar a respeito deste e não de outro referente. Como os exemplos em (28) mostram, tanto a periferia esquerda da sentença como a periferia de vP podem ser empregadas para tal fim em português brasileiro.¹³ Uma vez que o referente de *Maria* tem uma alta ativação no discurso, ele indica a direção da conversa, facilitando uma continuação como (29).

- 28) A: O que você deu pra Maria na festa de Natal?
B1: Pra Maria_{TD}, eu dei dois livros_F na festa de Natal.
B2: Eu dei pra Maria_{TD} dois livros_F na festa de Natal.

- 29) ...Eu gosto muito dela, eu queria dar algo legal de presente pra ela.

¹³ Um parecerista anônimo corretamente aponta que o PP *pra Maria* pode ter também interpretação de tópico contrastivo. Neste momento, estamos analisando somente a interpretação em que *pra Maria* tem a função de continuar o tópico da conversa (o que fica mais claro na topicalização de DP com resumpção, como em (8) na Seção 2.2). Optamos pelo uso do PP no exemplo (28) para que a comparação com a posição pós-verbal seja possível. Sobre a topicalização contrastiva, veja as seções 2.3 e 3.3.

Uma vez que o tópico de informação dada adquire alta proeminência discursiva, ele deve respeitar uma condição de boa-formação em relação a outro elemento de alta proeminência: o foco da sentença. Como as sentenças em (30) mostram, uma vez deslocado, o tópico de informação dada deve necessariamente preceder o foco, o que podemos entender como sendo o efeito de uma restrição conhecida na literatura como *given before new* ('dado antes de novo') (veja, e.g., Bresnan *et al.*, 2007; Gundel, 1988; Lacerda, 2017, 2020). Enquanto em (28) acima o foco é o objeto direto, permitindo a topicalização de *pra Maria* tanto na periferia esquerda como na periferia de vP, em (30) o foco é o sujeito, de modo que apenas a topicalização na periferia esquerda é possível.

- 30) A: Quem deu um presente pra Maria na festa de Natal?
 B1: *Pra Maria*_{TD}, *o João*_F deu um presente na festa de Natal.
 B2: **O João*_F deu *pra Maria*_{TD} um presente na festa de Natal.

O tópico de informação dada também pode ser realizado em construções de deslocamento à direita, como mostram os dados em (31), e neste caso, curiosamente, não está sujeito à restrição *given before new*. Comparados aos tópicos deslocados à esquerda, podemos conjecturar razoavelmente que o tópico deslocado à direita tem menor proeminência discursiva, uma vez que é pronunciado desacentuado (note que estruturas de deslocamento à direita têm entonação final descendente), de modo que não “entra em conflito” com o foco da sentença e pode segui-lo ao invés de precedê-lo, tal como informação dada não-topicalizada (cf. e.g. (25) e (27) acima).

- 31) A: O João leu três livros pro curso de sintaxe.
 B1: Costa muito de sintaxe, *o João*_{TD}.
 B2: Ele gosta muito de sintaxe, *o João*_{TD}.

De acordo com Silva (2016), a função do tópico deslocado à direita é simplesmente retomar o tópico do discurso, e “basta que o elemento deslocado à direita seja conhecido dos participantes” para que seja licenciado (p. 318). Assim, o tópico de informação dada, quando deslocado à direita, não tem proeminência discursiva suficiente para ativar o efeito *given before new*.

Entendemos, enfim, que a informação dada no português brasileiro está sujeita a uma escala de proeminência, representada em (32), que regula o impacto e/ou a função de informação dada no gerenciamento do terreno comum, a depender das intenções comunicativas do falante. Sob o quadro teórico adotado de Krifka (2008), podemos entender as construções marcadas como envolvendo maior impacto no gerenciamento do terreno comum do que as construções não-marcadas (canônicas).

- 32) *Escala de proeminência de informação dada no português brasileiro:*
- | | |
|-------------------------------------|---------------------------|
| apagamento | <i>menor proeminência</i> |
| desacentuação in situ | |
| não-desacentuação in situ | |
| deslocamento à direita | |
| topicalização no meio de campo | |
| topicalização na periferia esquerda | <i>maior proeminência</i> |
- ↓

Vimos, então, que informação dada no português brasileiro pode ser realizada em todos os quatro contextos sintáticos investigados neste trabalho, com diferentes graus de proeminência discursiva.

3.3 Tópicos contrastivos

Por fim, trataremos dos tópicos contrastivos. Além de apontar as posições da sentença em que estes podem ocorrer, discutiremos também sua posição em relação ao foco, apontando os contextos em que devem preceder o foco e os contextos em que podem sucedê-lo.

Diferentemente dos tópicos *aboutness(-shifting)* e assim como os tópicos de informação dada, tópicos contrastivos em português brasileiro podem ser deslocados tanto para a periferia esquerda da sentença, como em (33B1), como para a posição pós-verbal na periferia de vP, como em (33B2), desde que precedam o foco da sentença. Enquanto nos exemplos em (33B1) e (33B2) o foco é o objeto direto, nos exemplos em (34B1) e (34B2) o foco recai sobre o sujeito; neste caso, apenas o tópico deslocado para a periferia esquerda da sentença é lícito, uma vez que o tópico deslocado para a periferia de vP não precede o foco.

- 33) A: Quantos livros você leu pro curso de sintaxe?
B1: Bem, *do Chomsky*_{TC}, eu li *três livros*_F.
B2: Bem, eu li, *do Chomsky*_{TC}, *três livros*_F.
- 34) A: Quem resenhou os livros pro curso de linguística?
B1: Bem, *do Chomsky*_{TC}, *eu*_F resenhei os livros.
B2: *Bem, *eu*_F resenhei, *do Chomsky*_{TC}, os livros.

Tópicos contrastivos também podem aparecer em posições canônicas, e neste caso sua posição em relação ao foco da sentença é crucial para determinar a aceitabilidade da sentença e as possíveis interpretações. Se, na ordem canônica dos constituintes da sentença, o tópico contrastivo preceder o foco, o tópico contrastivo pode introduzir um conjunto de alternativas (contrastivas) no discurso e gerar as implicaturas relevantes. Vejamos como isso se dá no diálogo em (35). Supondo que o falante A pergunta apenas sobre o livro *Barriers*, o falante B pode marcar fonologicamente *o Barriers* como tópico contrastivo e gerar a implicatura de que há outros livros que o João possa ter dado para outras pessoas. Em outras palavras, a marcação de *o Barriers* como tópico contrastivo é pragmaticamente feliz neste cenário.

- 35) A: Pra quem o João deu o *Barriers*?
B: Bem, ele deu *o Barriers*_{TC} *pra Maria*_F.

Por outro lado, se na ordem canônica o tópico contrastivo não precede o foco, como em (36B), a sentença falha em introduzir um conjunto de alternativas contrastivas relevantes. Neste caso, em que o falante A pergunta apenas sobre a Maria, a marcação de *pra Maria* como tópico contrastivo é pragmaticamente infeliz e a sentença não pode gerar a implicatura de que há outras pessoas para quem o João possa ter dado livros.

- 36) A: Que livro o João deu pra Maria?
B: #Bem, ele deu *o Barriers*_F *pra Maria*_{TC}.

Sentenças com foco polar independentemente corroboram o contraste apontado acima, como nos exemplos a seguir, adaptados de Büring (2016). Em (37), o falante A pergunta apenas sobre o gato; neste caso, a marcação de *pro gato* como tópico contrastivo pelo falante B é pragmaticamente infeliz, uma vez que *pro gato* está à direita do verbo focalizado, e a sentença não pode gerar a implicatura de que há outros animais para quem o falante B possa ter dado comida. Ao contrário de (37), no cenário em (38) um conjunto de animais é dado no terreno comum; assim, o elemento marcado como tópico contrastivo (*pro gato*) não precisa preceder o foco (*dei*) e a sentença (38B) na ordem canônica é pragmaticamente feliz.

- 37) A: Você deu comida pro gato?
B: #Bem, eu *dei_F* comida *pro gato_{TC}*.

- 38) A: Você deu comida pros animais?
B: Bem, eu *dei_F* comida *pro gato_{TC}*.

Quando não há um conjunto de alternativas contrastivas relevantes no terreno comum, um tópico contrastivo gerado (na ordem canônica) abaixo do foco deve ser deslocado para uma posição acima do foco, para que possa gerar as implicaturas relevantes. Nos casos em que o foco recai sobre o verbo (mais precisamente, a polaridade do verbo), o deslocamento para a periferia esquerda da sentença é possível, como em (39B1), mas o deslocamento para a periferia de vP não é, como em (39B2), uma vez que o tópico contrastivo deslocado não precede o foco — compare com os exemplos (33B1) e (33B2) acima, em que o foco recai sobre o objeto direto.

- 39) A: Você deu comida pro gato?
B1: Bem, *pro gato_{TC}*, eu *dei_F* comida.
B2: *Bem, eu *dei_F* *pro gato_{TC}*, comida.

Enquanto deslocamentos para a esquerda (para a periferia de vP ou de CP) podem funcionar para posicionar um tópico contrastivo à esquerda do foco, o deslocamento à direita de um tópico contrastivo será sempre infeliz, como em (40B), na medida em que tópicos contrastivos deslocados devem sempre preceder o foco.

- 40) A: Quantos livros a Maria leu pro curso de sintaxe?
B: #Bem, leu *três livros_F*, *o João_{TC}*.

Assim, vemos que a distribuição de tópicos contrastivos em português brasileiro é dependente de sua posição em relação ao foco da sentença, em alguns casos podendo suceder o foco e em outros devendo precedê-lo, sendo possível o seu licenciamento tanto em posições canônicas como em posições deslocadas à esquerda (para a periferia de vP ou de CP), em contextos bem definidos. Em geral, vimos que o lugar preferencial dos tópicos contrastivos é à esquerda do foco, mas a ordem foco-tópico também é possível na ordem canônica se o conjunto de alternativas relevantes já estiver dado no contexto.

4 Mapeamento

Frascarelli e Hinterhölzl (2007) e Bianchi e Frascarelli (2010) argumentam (contra Rizzi, 1997), a partir de dados do italiano, do inglês e do alemão, que as posições de tópico na periferia esquerda da sentença não são livremente reiterativas, mas sim obedecem a uma hierarquia estrita, e propõem o seguinte mapeamento das posições de tópico na periferia esquerda da sentença:

41) *Hierarquia de tópicos na periferia esquerda da sentença:*

Tópico *aboutness(-shifting)* > Tópico contrastivo > Tópico de informação dada

42) $[_{A-TopP} \text{tópico-A } [_{C-TopP} \text{tópico-C } [_{D-TopP} \text{tópico-D } [_{TP} \dots]]]]]$

Nestas línguas, os autores observam uma hierarquia tripartite entre os tópicos, em que o tópico *aboutness(-shifting)* é o mais alto, precedendo o tópico contrastivo, que por sua vez precede o tópico de informação dada. A título de exemplo, observe o contraste entre (43a) e (43b) em inglês; o DP deslocado *Rosa* (que pode ser marcado por *as for* e é retomado por um pronome resumptivo) é interpretado como tópico *aboutness* e portanto deve preceder o tópico contrastivo *my next book* (“meu próximo livro”) (que não pode ser marcado por *as for* e não é retomado por um pronome resumptivo).

43) a. (As for) Rosa, my next book I will dedicate to her.

[(Quanto a) Rosa, meu próximo livro eu dedicarei a ela.]

b. *My next book, Rosa, I will dedicate to her.

[(Meu próximo livro, Rosa, eu dedicarei a ela.)]

(Bianchi; Frascarelli, 2010, p. 62)

Nesta seção, com base na análise de sentenças com tópicos não só na periferia esquerda da sentença (CP), mas também na posição pós-verbal na periferia de vP, argumentaremos que, em português brasileiro, o que se observa não é uma hierarquia em três níveis, mas sim uma hierarquia em dois níveis, com o tópico *aboutness(-shifting)* de um lado e o os tópicos de informação dada e contrastivo de outro — estes não estando sujeitos a uma hierarquia entre si.

Primeiro, analisemos as possibilidades de coocorrência entre os três tipos de tópicos. Como já apontado na Seção 3, o tópico *aboutness* deve estar em posição inicial; assim, quando este coocorrer com um tópico de informação dada ou um tópico contrastivo, deve precedê-lo, como em (44) e (45), respectivamente.¹⁴

44) A: A Maria doou cinco livros pro departamento.

B1: Já o João_{TA}, pro departamento_{TD} ele doou dez livros_F.

B2: Já o João_{TA} doou pro departamento_{TD} dez livros_F.

B3: Já o João_{TA}, ele doou pro departamento_{TD} dez livros_F.

¹⁴ Nos exemplos desta seção, usamos novamente a estratégia mudança de tópico, indicada pela partícula *já*, para que o elemento relevante seja interpretado como tópico *aboutness(-shifting)*, sem sofrer acomodação para outro tipo de tópico.

- B4: **Pro departamento*_{TD}, *já o João*_{TA} doou *dez livros*_F
 B5: **Pro departamento*_{TD}, *já o João*_{TA}, ele doou *dez livros*_F

- 45) A: A Maria leu cinco livros do Pinker.
 B1: *Já o João*_{TA}, *do Chomsky*_{TC}, ele leu *dez livros*_F
 B2: *Já o João*_{TA} leu, *do Chomsky*_{TC}, *dez livros*_F
 B3: *Já o João*_{TA}, ele leu, *do Chomsky*_{TC}, *dez livros*_F
 B4: **Do Chomsky*_{TC}, *já o João*_{TA} leu *dez livros*_F
 B5: **Do Chomsky*_{TC}, *já o João*_{TA}, ele leu *dez livros*_F

Os exemplos em (46) mostram que o tópico de informação dada e o tópico contrastivo não exibem efeitos de hierarquia entre si, podendo aparecer em qualquer ordem um em relação ao outro. Além disso, podem livremente ativar tanto a periferia esquerda da sentença (CP) como a posição pós-verbal na periferia de vP (desde que precedam o foco, como discutido na Seção 3).

- 46) A: Quantos livros do Pinker a Maria doou pro departamento?
 B1: Bem, *do Chomsky*_{TC}, *pro departamento*_{TD} ela doou *dez livros*_F
 B2: Bem, *pro departamento*_{TD}, *do Chomsky*_{TC}, ela doou *dez livros*_F
 B3: Bem, ela doou, *do Chomsky*_{TC}, *pro departamento*_{TD}, *dez livros*_F
 B4: Bem, ela doou *pro departamento*_{TD}, *do Chomsky*_{TC}, *dez livros*_F
 B5: Bem, *do Chomsky*_{TC}, ela doou *pro departamento*_{TD} *dez livros*_F
 B6: Bem, *pro departamento*_{TD} ela doou, *do Chomsky*_{TC}, *dez livros*_F

Sentenças em que os três tipos de tópicos aparecem também são possíveis (ainda que possam soar um tanto marcadas), desde que o tópico *aboutness(-shifting)* preceda o tópico de informação dada e o tópico contrastivo (estes novamente podendo aparecer em qualquer ordem entre si), como em (47).

- 47) A: A Maria doou cinco livros do Pinker pro departamento.
 B1: *Já o João*_{TA} doou *pro departamento*_{TD}, *do Chomsky*_{TC}, *dez livros*_F
 B2: *Já o João*_{TA} doou, *do Chomsky*_{TC}, *pro departamento*_{TD}, *dez livros*_F
 B3: ?*Já o João*_{TA}, *pro departamento*_{TD}, *do Chomsky*_{TC}, ele doou *dez livros*_F
 B4: ?*Já o João*_{TA}, *do Chomsky*_{TC}, *pro departamento*_{TD} ele doou *dez livros*_F
 B5: **Pro departamento*_{TD}, *já o João*_{TA} doou, *do Chomsky*_{TC}, *dez livros*_F
 B6: **Do Chomsky*_{TC}, *já o João*_{TA} doou *pro departamento*_{TD} *dez livros*_F

Agora, analisemos a possibilidade de reiteração de cada tipo de tópico. Frascarelli e Hinterhölzl (2007) e Bianchi e Frascarelli (2010) apontam que no italiano, no inglês e no alemão só pode haver um tópico *aboutness(-shifting)* por enunciado, na medida em que este tipo de tópico seleciona o único referente sobre o qual se comenta (conforme discutido na Seção 2.1). Em português brasileiro, a situação não é diferente, como se vê em (48) (cf. (44) acima). Os nossos exemplos de mudança de tópico em (48) corroboram a observação de Bastos-Gee (2011) de que a múltipla introdução de tópicos *aboutness* é impossível nesta língua, como em (49).

- 48) A: A Maria doou cinco livros pro departamento.
 B1: **Já o João_{TA}, já pra biblioteca_{TA}, ele doou dez livros_F*
 B2: **Já o João_{TA}, já pra biblioteca_{TA}, ele doou dez livros_F*
- 49) a. *O livro, a Maria, (foi) o João (que) comprou ele pra ela.
 b. *A Maria, o livro, (foi) o João (que) comprou ele pra ela.
 (Bastos-Gee, 2011, p. 19)

Ao contrário do tópico *aboutness(-shifting)*, o tópico de informação dada e o tópico contrastivo podem reiterar em português brasileiro. Observemos primeiro a reiteração de tópicos de informação dada, ilustrada em (50). Os elementos *na festa de Natal* e *pra Maria*, que retomam referentes previamente dados no contexto em (50A), podem ambos ser deslocados (em qualquer ordem) para a periferia esquerda da sentença e/ou para a região pós-verbal na periferia de vP.

- 50) A: O que você deu pra Maria na festa de Natal?
 B1: *Na festa de Natal_{TD}, pra Maria_{TD}, eu dei dois livros_F*
 B2: *Pra Maria_{TD}, na festa de Natal_{TD}, eu dei dois livros_F*
 B3: *Eu dei, na festa de Natal_{TD}, pra Maria_{TD}, dois livros_F*
 B4: *Eu dei, pra Maria_{TD}, na festa de Natal_{TD}, dois livros_F*
 B5: *Na festa de Natal_{TD}, eu dei pra Maria_{TD} dois livros_F*
 B6: *Pra Maria_{TD}, eu dei na festa de Natal_{TD} dois livros_F*

Da mesma maneira, tópicos contrastivos também são reiterativos em português brasileiro, podendo também ativar a periferia de CP e/ou a periferia de vP, como em (51).

- 51) Contexto:
 Enquanto o João lê alguns livros de linguística em suas edições originais em inglês, ele mais frequentemente lê livros traduzidos para o português.
 A: O João leu quais livros do Pinker em edições traduzidas?
 B1: *Bem, do Chomsky_{TC}, na edição original_{TC}, ele leu só o Barriers_F*
 B2: *Bem, na edição original_{TC}, do Chomsky_{TC}, ele leu só o Barriers_F*
 B3: *Bem, ele leu, do Chomsky_{TC}, na edição original_{TC}, só o Barriers_F*
 B4: *Bem, ele leu, na edição original_{TC}, do Chomsky_{TC}, só o Barriers_F*
 B5: *Bem, do Chomsky_{TC}, ele leu, na edição original_{TC}, só o Barriers_F*
 B6: *Bem, na edição original_{TC}, ele leu, do Chomsky_{TC}, só o Barriers_F*

Uma vez observadas as possibilidades de coocorrência e reiteração dos tópicos *aboutness(-shifting)*, de informação dada e contrastivos, chegamos ao mapeamento das posições marcadas de tópicos no português brasileiro esquematicamente representado em (52), em que o elemento X-TopP* indica a possibilidade de projeção reiterativa e em qualquer ordem tanto de um tópico de informação dada (D-TopP) como de um tópico contrastivo (C-TopP).

- 52) *Mapeamento de tópicos marcados em português brasileiro:*

$$[_{A-TopP} \text{tópico-A} [_{X-TopP*} \text{tópico-D/C} [_{TP} [_{X-TopP*} \text{tópico-D/C} [_{vP} \dots]]]]]]$$

Vimos nesta seção que, na periferia esquerda da sentença, o tópico *aboutness(-shifting)* pode ser seguido por tópicos de informação dada e/ou tópicos contrastivos (possivelmente reiterativos, o que é indicado pelo asterisco), sendo assim possíveis estruturas como aquelas em (53). Na região pós-verbal na periferia de vP, tópicos de informação dada e tópicos contrastivos podem ocorrer em qualquer ordem e possivelmente reiterar, sendo assim possíveis estruturas como aquelas em (54).

- 53) a. $[_{A-TopP} [_{D-TopP^*} [_{C-TopP^*} [_{TP} \dots]]]]$ tópicos na periferia esquerda
 b. $[_{A-TopP} [_{C-TopP^*} [_{D-TopP^*} [_{TP} \dots]]]]$

- 54) a. $[_{CP} [_{TP} [_{D-TopP^*} [_{C-TopP^*} [_{vP} \dots]]]]]]$ tópicos na periferia de vP
 b. $[_{CP} [_{TP} [_{C-TopP^*} [_{D-TopP^*} [_{vP} \dots]]]]]]$

Considerando ainda uma posição de tópico deslocada à direita, restrita a tópicos de informação dada (como discutido na Seção 3), podemos adicionar ao nosso mapeamento a estrutura em (55).¹⁵

- 55) $[_{CP} [_{D-TopP} [_{TP} \dots] \text{tópico-D}]]]$

Em conclusão, as possibilidades de coocorrência e reiteração de tópicos em português brasileiro apontam para uma divisão de comportamento entre o tópico *aboutness(-shifting)*, de um lado, e os tópicos de informação dada e contrastivo, de outro — o que, a nosso ver, reforça a nossa posição de que apenas o tópico *aboutness(-shifting)* é licenciado através de uma bipartição da sentença em tópico-comentário, enquanto o tópico de informação dada e o tópico contrastivo são licenciados em relação ao foco da sentença.

5 Considerações finais

Na conclusão de seu artigo sobre topicalização contrastiva, Büring (2003) afirma:

It is of course possible that additional concepts like Topic-Comment or Rheme-Theme have a role to play in the explanation of other phenomena, but for those discussed here, I submit, they are irrelevant [...].¹⁶ (Büring, 2003, p. 538-539)

Tanto pela sua função exercida no gerenciamento do terreno comum (cf. Krifka, 2008) como pela sua distribuição sintática, podemos concluir que o tópico contrastivo é licenciado de maneira distinta do tradicional tópico *aboutness(-shifting)* na interface entre sintaxe e estrutura informacional, em particular ao não requerer uma bipartição da sentença em tópi-

¹⁵ Outra possibilidade seria derivar o tópico deslocado à direita com topicalização à esquerda seguida do movimento do TP para uma posição acima do tópico no CP, como em (i) abaixo. Neste trabalho não nos comprometemos com a exata estrutura das construções de deslocamento à direita.

(i) $[_{CP} [_{XP} [_{TP} \text{sujeito verbo objeto}]_i [_{X'} X^o [_{D-TopP} \text{tópico-D} [_{D-TopP'} D-Top^o t_i]]]]]]$

¹⁶ “É claramente possível que conceitos adicionais como Tópico-Comentário ou Rema-Tema tenham um papel a desempenhar na explicação de outros fenômenos, mas para aqueles discutidos aqui, eu mantenho, eles são irrelevantes.”

co-comentário. Neste trabalho, vimos que os dados do português brasileiro vão ao encontro da afirmação de Büring (2003) citada acima, na medida em que o tópico contrastivo tem suas condições de licenciamento descritas a partir de seu posicionamento não em relação a um comentário, mas sim em relação ao foco da sentença. Ao não requerer a posição inicial, o tópico contrastivo têm nesta língua distribuição mais ampla do que o tópico *aboutness*(–*shifting*). *Mutadis mutandis*, podemos afirmar o mesmo a respeito do tópico de informação dada.

Comparando diferentes tipos de tópicos em português brasileiro, vimos neste trabalho que suas condições de licenciamento devem ser entendidas de maneira contextual, considerando que:

- i) tópicos podem aparecer tanto em posições canônicas da sentença (isto é, sem requerer o reordenamento da ordem básica de constituintes) como em posições marcadas (deslocadas), em contextos bem definidos;
- ii) o licenciamento de um tópico vai além de seu aparecimento em uma “posição de tópico” e depende de seu posicionamento relativo a outros elementos com os quais se relaciona na estrutura informacional (comentário ou foco);
- iii) para os tópicos de informação dada e contrastivo, há mais de uma posição marcada na estrutura sentencial, que podem ser ativadas a depender do posicionamento do foco.

Ainda que todos os tipos de tópicos tenham em comum o fato de atuarem no gerenciamento do terreno comum a partir de elementos referenciais, conforme a tradição de Reinhart (1981) e Heim (1982), podemos concluir, de maneira geral, que é imprescindível considerar, no estudo da topicalização, uma tipologia de tópicos adequada, uma vez que se trata de uma categoria não-uniforme. Uma vez que cada tipo de tópico exerce uma função diferente no gerenciamento do terreno comum, podemos esperar que cada tipo de tópico tenha distribuição e condições de licenciamento distintas na interface entre sintaxe e estrutura informacional, o que é de fato observado nos dados do português brasileiro. Longe de tentar oferecer uma palavra definitiva, este trabalho almejou demonstrar que é frutífera a investigação da topicalização sob esta perspectiva.

Agradecimentos

Agradeço a Jairo Nunes e aos pareceristas anônimos por suas valiosas contribuições. O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil (processo nº 2020/15877-9). As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do autor e não necessariamente refletem a visão da FAPESP.

Referências

BASTOS-GEE, A. C. Information structure within the traditional nominal phrase: The case of Brazilian Portuguese. 284 f. 2011. Tese (Doutorado em Linguística) – Departamento de Linguística, University

of Connecticut, Estados Unidos, 2011. Disponível em: <https://digitalcommons.lib.uconn.edu/dissertations/AA13510503>. Acesso em 14 nov 2025.

BELLETTI, A.. Inversion as focalization. In: HULK, A. C. J.; POLLOCK, J.-Y. (eds.). *Subject inversion in Romance and the theory of Universal Grammar*. Nova lorque: Oxford University Press, 2001. p. 60-90.

BELLETTI, A. Aspects of the low IP area. In: RIZZI, L. (ed.). *The structure of CP and IP: the cartography of syntactic structures*, vol. 2. Nova lorque: Oxford University Press, 2004. p. 16-51.

BIANCHI, V.; FRASCARELLI, M. Is topic a root phenomenon? *Iberia*, Sevilla, v. 2, n. 1, p. 43-88, 2010. Disponível em: <https://revistascientificas.us.es/index.php/iberia/article/view/86>. Acesso em: 14 nov 2025.

BRESNAN, J.; CUENI, A.; NIKITINA, T.; BAAYEN, R. H. Predicting the dative alternation. In: BOUMA, G.; KRÄMER, I.; ZWARTS, J. (eds). *Cognitive foundations of interpretation*. Amsterdã: Royal Netherlands Academy of Science, 2007. p. 69-94.

BÜRING, D. On D-trees, beans, and B-Accents. *Linguistics & Philosophy*, Dordrecht, n. 26, p. 511-545, 2003. DOI: <https://doi.org/10.1023/A:1025887707652>

BÜRING, D. (Contrastive) topic. In: FÉRY, C.; ISHIHARA, S. (eds.). *The Oxford handbook of information structure*. Oxford: Oxford University Press, 2016. p. 64-85.

CALABRESE, A. Some properties of the Italian pronominal system: an analysis based on the notion of *thema* as subject of predication. In: STAMMERJOHANN, H. (ed.). *Tema-rema in Italiano*. Tübingen: Gunter Narr Verlag, 1986. p. 25-36.

CHAFE, W. L. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In: LI, C. N. (ed.). *Subject and topic*. Nova lorque: Academic Press, 1976. p. 27-55.

COSTA, J. Focus *in situ*: evidence from Portuguese. *Probus*, Berlim, v. 12, n. 2, p. 187-228, 2000. DOI: <https://doi.org/10.1515/prbs.2000.12.2.187>

FIRBAS, J. On defining the theme in functional sentence perspective. *Travaux Linguistiques de Prague*, Praga, v. 1, p. 267-280, 1964.

FRASCARELLI, M. Subjects, topics and the interpretation of referential pro: An interface approach to the linking of (null) pronouns. *Natural Language & Linguistic Theory*, Berlim, v. 25, n. 4, p. 691-734, 2007. DOI: <https://doi.org/10.1007/s11049-007-9025-x>

FRASCARELLI, M.; HINTERHÖLZL, R. Types of topics in German and Italian. In: WINKLER, S.; SCHWABE, K (eds.). *On information structure, meaning and form*. Amsterdam: John Benjamins, 2007. p. 87-116.

FREY, W. A medial topic position for German. *Linguistische Berichte*, Hamburgo, v. 198, p. 153-190, 2004. DOI: https://doi.org/10.46771/9783967696967_2

GIVÓN, T. Topic continuity in discourse: an introduction. In: GIVÓN, T. (ed.). *Topic continuity in discourse: a quantitative cross-language study*. Amsterdã: John Benjamins, 1983. p. 5-41.

GUESSER, S.; MIOTO, C. Notas sobre a topicalização à esquerda sem retomada pronominal (explícita) no PB. In: GUESSER, S. (org.). *Linguística: pesquisa e ensino*. Boa Vista: EDUFRR, 2016. p. 79-102.

GUNDEL, J. K. *Universals of topic-comment structure*. Amsterdã: John Benjamins, 1988.

HEIM, I. R. The semantics of definite and indefinite noun phrases. 263 f. 1982. Tese (Doutorado em Linguística) - Departamento de Linguística, University of Massachusetts, Estados Unidos,

1982. Disponível em: <https://semanticsarchive.net/Archive/jA2YTJmN/Heim%20Dissertation%20with%20Hyperlinks.pdf>. Acesso em 14 nov 2025.

JAYASEELAN, K. A. IP-internal topic and focus phrases. *Studia Linguistica*, Oxford, v. 55, n. 1, p. 39-75, 2001. DOI: <https://doi.org/10.1111/1467-9582.00074>

KATO, M. A. Tópico e sujeito: duas categorias em sintaxe?. *Cadernos de Estudos Lingüísticos*, Campinas, v. 17, p. 109-131, 1989. DOI: <https://doi.org/10.20396/cel.v17io.8636803>

KATO, Mary A. Tópicos como alçamento de predicados secundários. *Caderno de Estudos Linguísticos*, Campinas, v. 34, p. 67-76, 1998. DOI: <https://doi.org/10.20396/cel.v34io.8637052>

KATO, M. A.; RAPOSO, E. Topicalization in European and Brazilian Portuguese. In: CAMACHO, J., FLORES-FERRÁN, N.; SÁNCHEZ, L.; DÉPREZ, V.; CABRERA, M. J. (eds.). *Romance linguistics 2003: selected papers from the 36th Linguistic Symposium on Romance Languages*. Amsterdã: John Benjamins, 2007. p. 213-226.

KRIFKA, M. Basic notions of information structure. *Acta Linguistica Hungarica*, Budapeste, v. 55, n. 3, p. 243-276, 2008. DOI: <https://doi.org/10.1556/aling.55.2008.3-4.2>

LACERDA, R. Information Structure in child English: Contrastive topicalization and the dative alternation. In: LAMENDOLA, M.; SCOTT, J. (eds.). *Proceedings of the 41st annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 2017. p. 387-400. Disponível em: <https://www.lingref.com/buclid/41/BUCLD41-31.pdf>. Acesso em 14 nov 2025.

LACERDA, R. Middle-field Syntax and Information Structure in Brazilian Portuguese. 284 f. 2020. Tese (Doutorado em Linguística) – Departamento de Linguística, University of Connecticut, Estados Unidos, 2020. Disponível em: <https://ling.auf.net/lingbuzz/005669>. Acesso em 14 nov 2025.

LACERDA, R. The middle field of Brazilian Portuguese and the size of the verbal domain. In: LASZAKOVITS, S.; SHEN, Z. (eds.). *The size of things I: structure building*. Berlin: Language Science Press, 2021. p. 285-304. Disponível em: <https://langsci-press.org/catalog/book/323>. Acesso em 14 nov 2025.

LACERDA, R. On the absence of low focus movement in Brazilian Portuguese. *Natural Language & Linguistic Theory*, Berlim, v. 42, n. 1, p. 209-245, 2024. DOI: <http://doi.org/10.1007/s11049-023-09582-5>

LACERDA, R. Focalization in situ and the absence of covert focus movement in Brazilian Portuguese. *Probus: International Journal of Romance Linguistics*, Berlim, v. 37, n. 2, 2025. p. 233-261. DOI: <https://doi.org/10.1515/probus-2025-0004>

MACEDO-COSTA, T.; ANDRADE, A. de. Caracterização semântico-pragmática de tópicos pendentes no português brasileiro. *Letrônica*, Porto Alegre, v. 8, n. 2, p. 391-408, 2015. DOI: <https://doi.org/10.15448/1984-4301.2015.2.20333>

MATHESIOUS, V. Functional linguistics. In: VACHEK, J.; DUŠKOVÁ, L. (eds.). *Praguiana: some basic and less known aspects of the Prague Linguistic School*. Amsterdã: John Benjamins, 1983[1929]. p. 121-142.

MIOTO, C. Sobre o sistema CP no português brasileiro. *Revista Letras*, Curitiba, v. 56, p. 97-139, 2001. DOI: <https://doi.org/10.5380/rel.v56io.18409>

MIRANDA, W.; SILVA, F. R. da. Formal similarities and distinctions between the contrastive markers *mas* (but), *já* (already) and *agora* (now) in Brazilian Portuguese. *ReVEL*, Porto Alegre, v. 13, n. 9, p. 120-

138, 2015. Disponível em: <https://www.revel.inf.br/files/729d5bf91a8d7794a57b881a3555ada1.pdf>. Acesso em 14 nov 2025.

PONTES, E. *O tópico no português do Brasil*. Campinas: Pontes Editores, 1987.

QUAREZEMIN, S. Estratégias de focalização no português brasileiro: uma abordagem cartográfica. 198 f. 2009. Tese (Doutorado em Linguística) - Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Brasil, 2009. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/92341>. Acesso em 14 nov 2025.

REINHART, T. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, Ghent, v. 27, n. 1, p. 53-94, 1981. DOI: <https://doi.org/10.21825/philosophica.82606>

RIZZI, L. The fine structure of the left periphery. In: HAEGEMAN, L. (ed.). *Elements of grammar*. Dordrecht: Kluwer, 1997. p. 260-318.

RIZZI, L. On some properties of criterial freezing. In: PANAGIOTIDIS, E. P. (ed.). *The complementizer phase: subjects and operators*. Oxford: Oxford University Press, 2010. p. 17-32.

RIZZI, L. Subjects, topics and the interpretation of pro. In: PETROSINO, R.; CERRONE, P; van der HULST, H. (eds.). *From sounds to structures: beyond the veil of Maya*. Berlin: De Gruyter Mouton, 2018. p. 510-529.

ROBERTS, C. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, Lexington, v. 5, art. 6, p. 1-69, 2012[1996]. DOI: <http://dx.doi.org/10.3765/sp.5.6>

SCHWARZSCHILD, R. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, Berlim, v. 7, p. 141-177, 1999. DOI: <https://doi.org/10.1023/A:1008370902407>

SILVA, F. R. da. Tópico em posição pós-verbal no português brasileiro. *Estudos Linguísticos*, São Paulo, v. 44, n. 1, p. 306-320, 2016. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/891>. Acesso em: 14 nov 2025.

SILVA, F. R. da. Deslocamento de tópico e foco no português brasileiro. 148 f. 2017. Tese (Doutorado em Linguística) – Departamento de Linguística, Universidade de São Paulo, Brasil, 2017. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8139/tde-18072017-181108/pt-br.php>. Acesso em 14 nov 2025.

STALNAKER, R. Pragmatic presuppositions. In: MUNITZ, M. K.; UNGER, P. K. (eds.). *Semantics and philosophy*. Nova Iorque: New York University Press, 1974. p. 197-214.

WAGNER, M. Contrastive topics decomposed. *Semantics & Pragmatics*, Lexington, v. 5, art. 8, p. 1-54, 2012. DOI: <https://doi.org/10.3765/sp.5.8>

Um estudo das representações linguísticas de descendentes germânicos em Niterói

A Study of the Linguistic Representations of German Descendants in Niterói

**Anderson Lucas da Silva
Macedo**

Universidade Federal Fluminense (UFF)
Niterói | RJ | BR
andersonlucas@id.uff.br
<https://orcid.org/0000-0002-6552-2200>

**Mônica Maria Guimarães
Savedra**

Universidade Federal Fluminense (UFF)
Niterói | RJ | BR
msavedra@id.uff.br
<https://orcid.org/0000-0003-1056-9391>

Resumo: Este estudo investiga as representações linguísticas dos descendentes de imigrantes alemães no município de Niterói observando como esses sujeitos constroem, interpretam e valorizam a língua alemã no contexto contemporâneo. A pesquisa insere-se no campo da sociolinguística de contato e parte de um panorama histórico da imigração germânica no Brasil (Bezerra, 2015; Lang, 2002). A fundamentação teórica mobiliza o conceito de representação linguística, discutido a partir de autores como Petitjean (2009), Gueunier (2003) e Calvet (1998), enfatizando o caráter simbólico e social das percepções linguísticas. Metodologicamente, o estudo adota uma abordagem qualitativa baseada em entrevistas sociolinguísticas semiestruturadas, realizadas com doze descendentes de alemães residentes em Niterói. Os resultados indicam uma relação predominantemente distante com a língua e a cultura alemãs, marcada pela perda intergeracional do idioma, pelo pouco contato com parentes europeus e por uma desconexão cultural reforçada por fatores como barreiras linguísticas e limitações econômicas. Ainda assim, observa-se entre os participantes um forte desejo de preservar a memória familiar, assim como uma valorização simbólica das raízes germânicas. A análise evidencia como as representações linguísticas articulam identidade, memória e pertencimento, revelando as tensões entre herança cultural e assimilação linguística no contexto brasileiro.

Palavras-chave: representação linguística, imigração alemã em Niterói, identidade.



Abstract: This study investigates the linguistic representations of descendants of German immigrants in the municipality of Niterói, examining how these individuals construct, interpret, and ascribe value to the German language in contemporary contexts. The research is situated within the field of contact sociolinguistics and draws on a historical overview of German immigration to Brazil (Bezerra, 2015; Lang, 2002). The theoretical framework mobilizes the concept of linguistic representation, discussed by authors such as Petitjean (2009), Gueunier (2003), and Calvet (1998), emphasizing the symbolic and social dimensions of linguistic perceptions. Methodologically, the study adopts a qualitative approach based on semi-structured sociolinguistic interviews conducted with twelve descendants of Germans residing in Niterói. The results indicate a predominantly distant relationship with the German language and culture, marked by intergenerational language loss, limited contact with relatives in Europe, and a cultural disconnection reinforced by factors such as linguistic barriers and economic constraints. Nevertheless, participants expressed a strong desire to preserve family memory, as well as a symbolic appreciation of their German heritage. The analysis highlights how linguistic representations articulate identity, memory, and belonging, revealing tensions between cultural heritage and linguistic assimilation in the Brazilian context.

Keywords: Language representation, German immigration in Niterói, identity.

1 Introdução

A imigração alemã no Brasil, iniciada na segunda década do século XIX, teve o estado Rio de Janeiro como um dos primeiros lugares de acolhimento. O movimento migratório da época resultou na formação de diversas comunidades que, ao longo do tempo, desenvolveram dinâmicas culturais e linguísticas próprias (Savedra e Rosenberg, 2019). Em uma publicação sobre o tema, Silva & Mello (2010) afirmam que entre 1886 e 1930 cerca de 250 mil alemães chegaram em território. No âmbito de nossa investigação, observamos que os descendentes de imigrantes germânicos no município de Niterói a partir de 1824. É importante mencionar que tal chegada não se concentrou num único período apenas; houve diversas chegadas ao longo das décadas.

Os descendentes dos imigrantes falantes de alemão mantêm uma relação singular com a língua alemã, que oscila entre a preservação como símbolo de herança cultural e o afastamento diante do predomínio do português. Nesse contexto, a língua alemã assume um papel complexo na construção da identidade desses indivíduos, ao mesmo tempo em que reflete dinâmicas de representatividade linguística e cultural.

Pesquisar a presença alemã no município citado não é uma tarefa muito simples, pois os escritos a esse respeito não abundam. Destacamos, ainda assim, a excelente pesquisa de Bezerra (2015) intitulada “Britânicos e Alemães em Niterói: um estudo de migração urbana” a qual conta atentamente o percurso desses europeus na localidade referida mediado por um viés histórico. Nosso intuito é complementar os estudos já realizados sobre imigração na história niteroiense pela perspectiva da sociolinguística de contato (Savedra *et al.*, 2021). Para tal, consideramos o conceito de identidade e representação linguísticas para compreender como a língua alemã é percebida, valorizada ou marginalizada pelos descendentes de imigrantes germânicos em Niterói.

O método escolhido para compreender o fenômeno linguístico citado foi dar ouvidos às narrativas contadas pelos próprios descendentes de Niterói. Trata-se de uma prática consolidada há décadas nos estudos sociolinguísticos, destacada por Tarallo (1997, p. 23), para quem “a narrativa de experiência pessoal é a mina de ouro que o pesquisador-sociolinguista procura”. A partir das histórias compartilhadas pelos entrevistados, é possível explorar como as relações com a língua alemã são construídas, negociadas e reinterpretadas ao longo do tempo e das gerações.

O artigo está organizado em quatro momentos distintos. O primeiro objetiva introduzir o tema da imigração alemã em Niterói a fim de contextualizar o cenário histórico e sociocultural que serve como pano de fundo para a pesquisa. O segundo momento trará uma discussão sobre o conceito de representação linguística, que é o alicerce teórico selecionado para a discussão dos dados mostrados aqui. Logo após, apresentaremos a metodologia utilizada na pesquisa, com destaque para o caráter qualitativo e a aplicação de entrevistas semiestruturadas como instrumento de coleta de dados. Por fim, o quarto momento foca na análise e discussão dos dados, trazendo trechos das entrevistas que ilustram como os descendentes de alemães em Niterói percebem a língua alemã e seu papel na formação de suas identidades.

2 Contexto sócio-histórico

Geralmente, quando se pensa em imigração alemã no Brasil, faz-se referência às grandes comunidades desses europeus que se instalaram nos estados do Sul do país, como Blumenau, Joinville, Nova Hamburgo etc. Sabemos, no entanto, que diversas áreas brasileiras receberam imigrantes alemães. A esse respeito nos mostram Silva & Mello (2010, p. 418).

Embora concentrados em maior número nas regiões sulinas, os imigrantes e seus remanescentes encontram-se espalhados por todo território brasileiro. Junto trouxeram suas línguas de origem, seus hábitos, seus costumes, sua cultura, imprimindo uma nova fisionomia à sociedade e à cultura nacional.

O estado do Rio de Janeiro, incluindo o município de Niterói, recebeu um número expressivo desses indivíduos¹, cuja presença marcou diversos aspectos da vida econômica e cultural fluminense.

Ao tratar desta de fluxo migratório, Michahelles (2002) esclarece que, nos centros urbanos do Rio de Janeiro, as comunidades alemãs eram muito reduzidas numericamente se comparadas com aquelas que foram formadas no interior. A autora informa que, em 1860, a população alemã era de cerca de 2600 pessoas, já em 1920, o número chegou a 5000. Nesta década, as chegadas de imigrantes alemães aumentaram significativamente. Por outro lado, diversos desses europeus, por vários motivos, decidiram regressar à sua terra natal. Havia aqueles que planejaram fazer fortuna aqui e voltar à Alemanha e os que planejaram viver definitivamente no Brasil.

A imigração alemã para o Brasil no século XIX foi impulsionada pelas políticas governamentais que visavam o povoamento de áreas pouco habitadas e o desenvolvimento da agricultura, promovendo a ocupação organizada e o fortalecimento econômico dessas regiões. Michahelles (2002, p. 1) acrescenta que os imigrantes alemães foram “atraídos pelas possibilidades de desenvolver atividades comerciais em decorrência da abertura dos portos”. Bezerra (2015) nos lembra que a situação econômica² era muito difícil na Alemanha³ em decorrência do fim das guerras napoleônicas. As próprias autoridades europeias encorajaram a emigração com a finalidade de diminuir as responsabilidades sociais e econômicas que pesavam sobre elas (p. 19). Os que vieram no começo do século XX ainda foram motivados pelas negativas consequências da 1ª Guerra Mundial. No Rio de Janeiro, os alemães desembarcaram em diferentes momentos do século referido. O município de Petrópolis, por exemplo, idealizado por D. Pedro II e fundado em 1843, pôde contar com a presença e influência dos imigrantes alemães em sua construção e urbanização. Savedra e Mazzelli (2020, p. 107) mencionam que “o município de Petrópolis situado a 59 km da capital Rio de Janeiro é conhecido como a cidade Imperial e por muitos também chamada de a *Versailles Brasileira*”. Destaca-se também Nova Friburgo, que começou como colônia suíça em 1819, mas que foi abrigo de alemães nos anos subsequentes.

Em se tratando especificamente de Niterói, alvo da nossa investigação, é correto afirmar que, devido à proximidade com a cidade do Rio de Janeiro (que foi capital do país até 1960), foi um lugar interessante e estratégico não só para os imigrantes alemães, mas também para os italianos e portugueses. A vinda desses sujeitos a Niterói geralmente está associada à busca por oportunidades econômicas, com destaque às áreas urbanas e periurbanas. Além disso, diversos visitantes à cidade do Rio de Janeiro cruzavam a baía de Guanabara simplesmente para conhecer Niterói. Isto inclui a nobre viagem de D. João VI, que se hospedou em São Domingos para lidar com o luto da morte de sua mãe (Bezerra, 2015).

¹ Os primeiros que desembarcaram na então província do Rio de Janeiro foram, na verdade, os suíços, que chegaram entre 1819 e 1820. Eles se acomodaram no município de Nova Friburgo como consequências de atos políticos assinados por D. João VI em 1818 (Bezerra, 2015, p. 20)

² Importante mencionar que as adversidades financeiras vividas na Europa no século XIX foram lembradas em algumas de nossas entrevistas quando os informantes nos explicavam os motivos da imigração de seus antecessores.

³ A unificação alemã foi um processo histórico ocorrido na segunda metade do século XIX, com seu marco final em 1871. Apesar de a Alemanha como estado-nação ainda não existir antes disso, utilizamos o termo “Alemanha” de forma anacrônica para fins de comunicação e referência a territórios e povos que posteriormente compuseram o país unificado.

O perfil profissional desses imigrantes que se instalaram no estado do Rio de Janeiro era frequentemente de artesãos, comerciantes, operários qualificados. Sobre este último afirma Michahelles (2002, p. 1): “a imigração se caracterizou por uma política específica, dirigida para a captação de famílias formadas principalmente por agricultores”. Na comunidade alemã também havia negociantes ligados ao ramo da exportação e importação. O Brasil consumia muitos produtos europeus como vinho, azeite, queijo, além de manufaturados como a seda e a lã, ferramentas, roupas de luxo, móveis, entre outros. Savedra e Mazzelli (2020) apontam a participação de alemães na construção civil “construíram estradas e limpavam o terreno para construir casas e plantar hortas; trabalharam também nas obras públicas” (p. 108). A atuação desses alemães contribui diretamente para o desenvolvimento econômico do município. Sobre esse tema, nos esclarece Bezerra (2015, p. 22):

Vale a pena esclarecer, todavia, que os primeiros imigrantes alemães a chegarem ao Brasil não foram colonos, mas indivíduos que se instalaram na cidade do Rio de Janeiro como comerciantes e artífices, atraídos pelas perspectivas econômicas advindas da abertura dos portos, em 1808. Assim, desde o início do século XIX, já era possível encontrar mais de vinte empresas alemãs na cidade do Rio de Janeiro. E, embora estes imigrantes não fossem expressivos em termos numéricos, fundaram, em 1821, a primeira associação germânica do país: a Gesellschaft Germania.

No campo religioso, os alemães protestantes se reuniam para suas práticas de culto, nas quais coexistiam harmonicamente as orientações luterana e calvinista. Bezerra (2015) destaca os desafios enfrentados por esses imigrantes na prática de sua fé, em função das “limitações impostas pelo texto constitucional de 1824” (p. 26). Essa legislação permitia a liberdade de culto, mas restringia as manifestações religiosas não católicas a espaços privados, proibindo a construção de templos visíveis ou qualquer forma exterior de expressão religiosa.

Durante o Estado Novo, o presidente Getúlio Vargas implementou políticas nacionalistas que restringiram o uso de línguas de imigração, incluindo o alemão, e limitaram práticas culturais estrangeiras. A campanha de abasileiramento visava integrar os imigrantes que haviam se estabelecido permanentemente no Brasil. Entre suas ações, destacaram-se o fechamento de escolas com currículo estrangeiro e a supressão de publicações na língua materna, como jornais e revistas, enfraquecendo significativamente as expressões culturais das comunidades imigrantes (Lang, 2002).

Devido o advento do regime nazista na Alemanha, os imigrantes deste país experimentaram ondas de discriminação e até perseguições violentas. “Em Niterói, assim como em boa parte do país, tal campanha teve graves consequências, pois o governo obteve apoio popular na repressão aos imigrantes” (Lang, 2002, p. 261). Tal repressão repercutiu diretamente na manutenção da identidade germânica em Niterói e contribui para o aprofundamento do fenômeno psicológico conhecido como assimilação. Lembremos que havia, no município, uma escola alemã (localizada no centro da cidade). A instituição precisou fechar as portas em 1942 devido às políticas mencionadas, além do pressão popular relacionada ao clima de guerra existente que se direcionava em forma de hostilidade contra tudo que fosse associado à Alemanha.

Para encerrar esta seção, apresentamos uma fotografia que foi exibida no evento “Imigração alemã no Rio de Janeiro” que ocorreu em 9 de novembro de 2024 no Museu da Imigração da Ilha das Flores, celebrando os 200 anos da imigração alemã no Brasil. Tal evento foi vinculado à UERJ e teve como objetivo principal a divulgação da comunidade alemã no

Estado citado. Na imagem, podemos ver um ponto de entrada no Rio de Janeiro para os imigrantes. O cartaz a seguir escrito em português, alemão e línguas eslavas, transmite uma breve mensagem de acolhimento aos recém-chegados ao Brasil. A multiculturalidade e o multilinguismo chamam a atenção, assim como as poucas bagagens que, às vezes, as famílias traziam em busca de uma vida melhor.

Imagem 1 – Chegada e acolhimento aos imigrantes alemães



Fonte: <https://www.pmsg.rj.gov.br/sao-goncalo-recebe-exposicao-imigracao-alema-no-rio-de-janeiro/>

3 Representação linguística

Nesta seção, objetivamos levantar um breve debate acerca do conceito de representação linguística, utilizado em pesquisas sociolinguísticas desde a década de 1960, mesmo que com nomes diferentes. O que diremos a esse respeito não limita ou esgota o que o tema pode gerar. Tampouco ofereceremos uma definição única ou definitiva. Não pretendemos estabelecer “a” maneira correta de abordar a representação linguística, mas, sim, contribuir com reflexões que iluminem suas múltiplas facetas e promovam novos diálogos sobre essa questão complexa e interdisciplinar.

“Representação” é um termo que abre múltiplos caminhos interpretativos. Em seu sentido mais amplo, refere-se ao ato de tornar algo presente por meio de um símbolo, imagem, descrição ou substituto. O dicionário online de língua portuguesa Priberam⁴ define essa palavra, entre outras acepções, como “exposição, exibição” e “exposição verbal ou escrita do

⁴ <https://dicionario.priberam.org>

que temos na mente”. Já o Dicionário online de português⁵, traz as definições “ato ou efeito de representar, de mostrar com clareza” e também “conceito, ideia ou imagem que criamos do mundo ou de alguma coisa”.

No campo dos estudos sociolinguísticos, a concepção de representação linguística passou a ser empregada para descrever a maneira como línguas, variedades e práticas linguísticas são percebidas, descritas e simbolizadas em contextos sociais, culturais e políticos. Digno de nota é o fato de que o uso do termo “representação” foi emprestado das ciências sociais, em especial da antropologia e psicologia social (Gueunier, 2003; Calvet, 2006). Partindo do entendimento de que os falantes possuem algum tipo de relação com as línguas/variedades com que já tiveram contato, este conceito desempenha um papel crucial na análise e compreensão das identidades individuais e coletivas, evidenciando as interseções entre linguagem, sociedade e cultura.

Petitjean (2008) trata desta temática de modo cuidadoso ao lidar com variedades linguísticas da língua francesa. A pesquisadora investigou como os falantes percebem e avaliam variações linguísticas regionais, destacando os aspectos simbólicos e sociais atribuídos a essas características de fala. Já na sua tese de doutorado, Petitjean (2009) discute amplamente o conceito de representação linguística no campo da sociolinguística, explorando tanto seus fundamentos teóricos quanto metodológicos. Ao longo do trabalho, a autora apresenta reflexões relevantes sobre o tema, oferecendo uma base sólida para embasar nosso próprio debate.

De acordo com a autora, o surgimento, desenvolvimento e uso do presente conceito passou por três períodos que se solidificaram gradualmente. A **primeira fase**, dentre as décadas 1960 a 1980, é caracterizada pelo advento das ideias incipientes à representação. Em outros termos, a representação linguística ainda não era um conceito claramente definido, mas já aparecia de forma implícita nos estudos sociolinguísticos. Petitjean (2008) traz como exemplo o trabalho de Trugdill (1974) no qual trouxe a noção de *opinião dos falantes*⁶, sentimentos linguísticos (Bernstein, 1971), e o conceito de insegurança linguística (Labov, 1966). Neste período, a maneira como os falantes pensam suas práticas linguísticas começa a ganhar atenção e forma na literatura, embora os trabalhos não se preocupassem com o refinamento de uma sólida definição acadêmica para o fenômeno citado. É por isso que no período referido encontram-se diversos termos como atitude, visão, imagem, ideologia, valor afetivo, entre outros, os quais foram revisitados e conceituados nas décadas seguintes. Pereira e Costa (2012) abordam esse período e explicam que:

Os estudos que levam em conta indivíduos, línguas e seu uso tiveram início nos anos de 1960, sob a ótica da atitude linguística. Esta é definida como uma disposição a reagir de maneira favorável ou não a um dado objeto, no caso, a uma língua. Especialistas em aprendizagem têm a representação como um conceito fundamental e, por isso, aprofundam esta noção para uma perspectiva didática. O fato é que a noção de representação está cada vez mais ligada aos estudos linguísticos, tratando da representação que os locutores fazem acerca das línguas, de suas normas, de suas características, ou de seus status face às outras línguas, influenciando as estratégias que envolvem o uso e o aprendizado (Pereira; Costa, 2012, p. 173).

⁵ <https://www.dicio.com.br/representacao/>

⁶ No texto em francês, a autora usa o termo “*l’opinion des locuteurs*”.

A segunda fase, a partir da década de 1980, é fortemente marcada pelo levantamento de questões/reflexões metodológicas. Pode-se citar os trabalhos de Tabouret-Keller (1981) e Houdebine (1982) que refletem um grande avanço no refinamento nos métodos de análise. A partir de 1990, temos a última fase na qual ocorreu um amadurecimento tanto da noção quanto o uso da representação linguística, além da apropriação do emprego do termo em si. Para ilustrar esse terceiro momento, podemos citar a contribuição de Calvet (1998). Para este sociolinguista, ao lado das representações está a maneira como os falantes refletem sobre os usos linguísticos, posicionam-se em relação a outros falantes e práticas linguísticas, e como percebem sua própria língua em comparação com outras línguas presentes em seu contexto social.

Existe uma próxima relação entre o conceito de representação⁷ linguística e a psicologia social. Este ramo da ciência psicológica, segundo Petitjean (2008) busca compreender como o indivíduo está inserido e atua no social, e, reciprocamente, como o social estrutura o indivíduo. Ora, essa definição dialoga bastante com a própria definição de sociolinguística – um ramo da linguística que estuda a relação entre língua e contexto social incluindo as formas como aquela varia e muda. Dentro da psicologia social, o conceito de representação social desempenhou um papel central e trata de como os indivíduos e grupos constroem e compartilham significados sobre o mundo social. A representação social é elaborada através da comunicação, da construção da realidade e integração e controle dos atores sobre seu ambiente. Nas palavras de Petitjean (2008) “a representação social estabelece, portanto, uma relação entre um sujeito e um objeto social” (p. 31) e “uma representação é qualificada como social na medida em que se desenvolve a partir de processos de trocas e interações, permitindo a constituição de conhecimentos comum a um grupo” (p. 31).

O diálogo bem estruturado entra a psicologia social e a sociolinguística nos leva a recordar a interdisciplinaridade de nossa ciência, estabelecida desde seus mais antigos estudos labovianos. Essa relação transdisciplinar faz com que observemos a representação linguística de maneira dupla: por um lado, a representação social da língua e, por outro, a representação na língua. As representações social e linguística, embora provenham de campos de estudos diferentes, se conectam intensamente ao investigarem como indivíduos/grupos compreendem e simbolizam o mundo à sua volta.

Para compreender como as representações linguísticas se formam é fundamental considerar que elas não são baseadas em conhecimentos científicos. Elas podem ser mais bem descritas como construções sociais compartilhadas.

Uma representação linguística seria um conjunto de conhecimentos não científicos, socialmente elaborados e partilhados, de natureza fundamentalmente interativa e discursiva, com um grau mais ou menos elevado de julgamento e fixidez, e permitindo ao(s) falante(s) desenvolver(em) uma construção comum de realidade linguística, isto é, a(s) língua(s) da comunidade ou a(s) língua(s) das comunidades exógenas, e gerir as suas atividades linguísticas dentro desta interpretação comum da realidade linguística⁸. (Petitjean, 2008, p. 33)

⁷ Segundo Gueunier (2003) o termo “representação” foi emprestado das ciências sociais.

⁸ Une représentation linguistique serait un ensemble de connaissances non scientifiques, socialement élaborées et partagées, fondamentalement interactives et de nature discursive, disposant d'un degré plus ou moins élevé de jugement et de figement, et permettant au(x) locuteur(s) d'élaborer une construction commune de la réalité linguistique, c'est-à-dire de la ou des langues de la communauté ou de la ou des langues des communautés exogènes, et de gérer leurs activités langagières au sein de cette interprétation commune de la réalité linguistique.

Gueunier (2003) não discorda da opinião apresentada acima e afirma que uma representação é aquilo que difere do conhecimento conceitualizado. A autora ressalta que esse termo se relaciona a outros conceitos sociais como crença, preconceito, ideologia e atitude. Em uma primeira leitura, essas noções podem parecer bastante parecidas e até mesmo trazer alguma confusão. Para ilustrar essa diferenciação, a autora apresenta critérios claros para distinguir *atitude linguística* de *representação*. Dois parâmetros básicos são apontados: a **discursividade** e a **figuratividade**. O primeiro parâmetro indica que a atitude motiva diretamente um comportamento observável, enquanto a representação “só é acessível por meio do discurso que a formula” (p. 42). Em outras palavras, a representação é identificada pela observação discursiva dos falantes. A pesquisadora apresenta um exemplo claro de como tais ideias podem ser discernidas:

Por exemplo, atualmente, os pais franceses geralmente preferem que seus filhos aprendam inglês como segundo idioma em vez de outros idiomas vivos, europeus ou não. Essa atitude pode ser medida de acordo com seu comportamento na escolha que fazem entre os vários idiomas propostos pela escola. Mas as representações subjacentes a essas atitudes só podem ser descobertas a partir de seu discurso, que pode ser mais ou menos restrito (respostas a questionários) ou livre (conversas comuns), no qual elas são expressas: “O inglês é mais falado do que o alemão/italiano”, “é o idioma mais útil”, “é fácil de aprender”, “não tem gramática” etc⁹. (Gueunier, 2003, p. 42)

Ao se concentrar na ideia de figuratividade, Gueunier (2003) nos mostra que as atitudes podem ser avaliadas quantitativamente, isto é, com informações numéricas. Isto pode ser realizado comparando o comportamento linguístico de um falante em situações comunicativas distintas ou com métodos como testes e questionários. Por outro lado, as representações linguísticas, que são mais subjetivas, podem ser medidas qualitativamente através da busca por conteúdos simbólicos e discursivos nas percepções linguísticas. A pesquisadora ainda traz exemplos de certas imagens/metáforas usados na sociedade para representar ideias. No texto, Gueunier menciona a “limpeza nas roupas” que, em alguns contextos, pode ser simbolicamente associada à ideia de brancura. Semelhantemente, a ortografia padrão (no caso, a francesa) é tradicionalmente vinculada a imagens de clareza e pureza, como se de fato houvesse uma maneira superior de escrita. Esse exemplo nos desafia a pensar como uma representação social/linguística não é apenas uma ideia isolada, mas algo que compartilha fronteiras com outras representações.

Pereira e Costa (2012, p. 175) propõem explicar a diferença entre atitude e representação. Para as autoras, a atitude linguística reflete a atitude social dos indivíduos, sendo caracterizada por se concentrar e se referir tanto à língua em si quanto à forma como ela é utilizada no contexto social. Por outro lado, ao explicarem o que é representação, elas afirmam que “dizem respeito ao modo com que os falantes pensam as suas práticas, como eles se situam em relação aos demais”.

⁹ For instance, French parents generally, at present, prefer their children to learn English as a second language rather than other living languages, European or otherwise. This attitude can be measured according to their behavior in the choice they make between the various languages proposed by the school. But the representations underlying these attitudes can only be discovered from their discourse, which may be more or less constrained (replies to questionnaires) or free (ordinary conversations) in which they are expressed: “English is more widely spoken than German/Italian,” “it is the most useful language,” “it is easy to learn,” “it has no grammar,” etc.

Na França, as representações linguísticas parecem estar profundamente conectadas à escrita, em especial à ortografia, em detrimento dos aspectos orais da língua. Essa percepção cultural dá uma importância maior à correção na escrita e na leitura, enquanto aspectos como sotaque ou dificuldades na expressão oral são considerados menos graves. Gueunier (2003) apresenta este exemplo de como a representação linguística ocorre empiricamente:

Para a pergunta “o que são palavras?” A.L., de 11 anos, que acabou de entrar na sexta série, responde: “grupos de letras que significam algo”. Para mim, isso confirma como a consideração das representações linguísticas na França, mais do que em outros países europeus, está ligada à ortografia em detrimento dos elementos fônicos. “Ter um sotaque” e “não saber como se expressar em público” são certamente percebidos como desvantagens, mas muito mais leves do que aquelas relacionadas à leitura e à escrita. Elas podem ser reconhecidas sem vergonha, com a sensação de que são principalmente um constrangimento na vida pública e profissional, enquanto o analfabetismo (não incomum em culturas escritas) e até mesmo a mera dificuldade de ortografia serão sistematicamente mantidos ocultos do reconhecimento público. (Gueunier, 2003, p. 43)

Gueunier (2003) também destaca como as representações sociais associadas ao latim em contextos europeus vão além de justificativas puramente linguísticas, enraizando-se em construções identitárias e sociais. “French comes from Latin” (p. 45) é uma sentença bastante usada para apresentar uma boa justificativa para a centralidade do latim na educação (pelo menos em décadas atrás), no entanto a autora adverte para o fato de que “tais representações, associadas ao latim, têm uma relação tanto na identidade europeia quanto na românica, e na identidade social” (p. 45). O conhecimento desta língua (que simbolizou poder e prestígio) parece, ainda, ser um elemento importante na formação escolar das elites, de acordo com Gueunier.

4 Procedimentos metodológicos

Neste trabalho¹⁰, adotamos uma perspectiva qualitativa para observar e compreender os fenômenos sociolinguísticos relacionados ao conceito de representação. Tal perspectiva se justifica porque o conceito referido envolve subjetividades, valores e percepções coletivas e individuais. Também, a pesquisa qualitativa se nos mostrou adequada porque lidamos com aspectos históricos, sociais e culturais moldam as representações. Assim, este tipo de abordagem nos permite entender essa complexidade, explorando as perspectivas dos participantes em profundidade.

No que tange à coleta de dados, utilizamos a técnica de entrevista sociolinguística. As entrevistas foram realizadas de maneira semiestruturada, o que nos permitiu explorar os tópicos abordados no projeto e, ao mesmo tempo, explorar mais profundamente temas que surgissem durante a conversa (o que, de fato, ocorreu). Essa flexibilidade foi importante para deixar os participantes relatarem suas experiências pessoais. As entrevistas nos possibilitaram ter foco no entrevistado e em suas relações com a língua e cultura alemã. Consideramos essa técnica pertinente a esse tipo de pesquisa, pois pudemos levantar informações espontâneas que não teriam sido identificados através de formulários. Como pesquisadores, atuamos de

¹⁰ Nossa pesquisa foi devidamente aprovada pelo Comitê de Ética – CAAE 72865923.7.0000.8160

modo responsável e crítico, tentando sempre deixar os informantes o mais confortáveis possível. Tal técnica tem sido usada, de maneira bastante ampla, em pesquisas sociolinguísticas.

Todas as entrevistas foram gravadas com a devida permissão dos participantes. Importante ressaltar que algumas entrevistas ocorreram na modalidade online e outras, presencialmente.

No total, doze descendentes foram entrevistados. Nossa primeira ideia era de abarcar um grupo maior de entrevistados, contudo não foi muito fácil encontrar indivíduos que preenchessem os critérios necessários (ser descendente de alemão e ser morador do município de Niterói). Vale destacar que as experiências dos entrevistados em relação à cultura e à língua alemã são muito semelhantes às de seus irmãos e irmãs (segundo os relatos coletados), de modo que entrevistar um único membro da família proporcionou uma compreensão provavelmente representativa de todo o núcleo familiar.

A internet foi o principal meio pelo qual encontramos os participantes. Utilizamos as seguintes estratégias: i) fizemos anúncios em comunidades referentes a Niterói no Facebook e postamos a necessidade de encontrar os descendentes em contas do Instagram; ii) levantamos sobrenomes alemães e os buscamos em redes sociais e, a partir daí, fizemos contatos com tais indivíduos; iii) enviamos e-mails para clubes e associações relacionadas a alemães no Brasil e iv) fizemos anúncios em grupos de Whatsapp de moradores em Niterói. A estratégia (i) se mostrou a mais eficaz; os contatos obtidos nos levaram a outros informantes.

As entrevistas ocorreram no segundo semestre de 2024 e, média, duraram 25 minutos. As perguntas foram cuidadosamente elaboradas; incluímos questões referentes à ancestralidade, relação com os parentes alemães, uso do alemão e proficiência da língua, práticas e atitudes linguísticas, bem como relatos de histórias passadas de geração em geração. Ao longo das entrevistas, os participantes também foram incentivados a refletir sobre questões identitárias e intergeracionais ligadas à língua. A fim de proporcionar ao leitor uma compreensão mais abrangente da natureza das perguntas feitas em nossas entrevistas, selecionamos algumas para ilustrar:

Quadro 1 – Exemplos de perguntas empregadas nas entrevistas

“Eu me sinto em casa quando ouço a língua alemã”. Numa escala de 0 a 10, quão verdadeira/significativa é essa frase para você?

Como o alemão influencia na maneira como você se define e se identifica?

Você consome produções culturais da cultura alemã como livros, filmes, séries, festivais?

Você tem alguma história ou experiência interessante para me contar relacionada à cultura ou à língua alemã?

Você mantém algum vínculo com familiares europeus?

Fonte: elaboração dos autores

Após a coleta de dados, iniciamos o processo de transcrição. Utilizamos o “transkriptor”¹¹ (uma ferramenta online) para fazê-lo, o que acelerou significativamente essa etapa. Com as transcrições feitas, o processo de leitura e interpretação dos dados foi iniciado. Discussões adicionais sobre nosso percurso metodológico encontram-se em Savedra *et al.* (2025).

6 O perfil dos entrevistados

De forma geral, os descendentes de alemães entrevistados mantêm uma relação distante com a Alemanha, sua língua e sua cultura. Dentre todos os entrevistados, por exemplo, apenas um fala alemão pelo fato de ter antepassados europeus (o outro caso é de um homem que foi estudar na Suíça, por isso aprendeu a língua). Além da baixa manutenção da língua, há uma notável desconexão entre esses brasileiros e seus familiares na Alemanha. A maioria dos entrevistados não tem contato com familiares na Alemanha ou em outros países europeus de origem, o que contribui para uma menor integração cultural e linguística com a Europa. Alguns informantes revelaram a tentativa de aproximação de alguns familiares brasileiros com os europeus; algumas tentativas foram exitosas e outras não.

Mais da metade dos nossos participantes nunca teve a chance de visitar a Alemanha. Isto reforça a distância linguístico-cultural que percebemos nas entrevistas. No entanto, os informantes apresentaram grande interesse e desejo de conhecer suas raízes europeias. Os motivos apontados por nunca terem ido à Europa foram: não saberem a língua alemã e não terem recurso financeiro para tal empreendimento. O primeiro motivo se mostra interessante para nós pelo fato de representar uma barreira linguística tão forte que desencoraja a ida desses sujeitos à Alemanha. Por não fazerem esta visita, a desconexão cultural é reforçada, certamente. O trecho de uma das entrevistas a seguir ilustra o que afirmamos:

Entrevistador: E você tem algum vínculo ou contato com seus parentes alemães?
Entrevistado 5M: Eu não tenho, porque assim, eu nunca fiz contato com ninguém assim, porque é muito distante, né? Bisavô, né? Eu nunca tive, não.

Entretanto, dos entrevistados que puderam ir ao país referido, destacamos que o motivo da viagem foi relacionado ao fato de descenderem desses europeus, o que revela a importância dada à preservação da memória familiar. Ressaltamos também que, dentre as inúmeras opções de turismo que poderiam ter sido escolhidas, a seleção pela Alemanha mostra uma ligação emocional (o fato de pisar no território em questão já é um ato cheio de simbolismos), como podemos constatar no fragmento a seguir:

Entrevistador: E você já foi à Alemanha?
Entrevistada 2F: Já.
Entrevistador: E por que razão você foi?

¹¹ O Transkriptor é uma ferramenta online que facilita a transcrição automática de áudios e vídeos para texto. Utilizando tecnologias de reconhecimento de fala, o site permite que os usuários façam upload de seus arquivos de áudio ou vídeo, os quais são processados para gerar transcrições precisas em diversas línguas. É uma plataforma bastante conveniente para pesquisadores.

Entrevistada 2F: Para conhecer, porque a Alemanha é um país que chama a minha atenção e por conta dessa história da família também. Foi a primeira viagem internacional com os meus pais e a gente fez questão de passar lá. Acabou que a gente não foi à região da minha família, mas a gente fez questão de ir à Alemanha para a parte da minha mãe e minha tia conhecerem também.

Entrevistador: Mas vocês não conseguiram estabelecer nenhum vínculo com parentes de vocês lá, né?

Entrevistada 2F: Não, mas tem outras pessoas da família que já foram lá, numa cidadezinha bem pequena, e conheceram pessoas da família, foram numa igreja e as pessoas vão vendo os contatos, né?

No que tange à importância dada à preservação das histórias familiares, os entrevistados acreditam firmemente que é crucial transmitir a história e a língua aos membros das futuras gerações. Há uma preocupação significativa em manter viva a memória cultural e garantir que os descendentes futuros conheçam e valorizem suas origens. Alguns participantes lamentaram não terem aprendido a língua alemã com seus avós e mencionaram que seria muito positivo se pudessem estudá-la em algum momento. Outros se dedicaram ao estudo da língua inglesa ao invés de estudar alemão por razões profissionais e oportunidade de ascensão socioeconômica.

7 Explorando a representação linguística: perguntas como ferramentas de análise

A análise das entrevistas seguiu uma perspectiva qualitativa, que se concentra em explorar e entender as perspectivas subjetivas dos entrevistados.

Com o objetivo de investigar como os participantes percebem a relação entre língua e identidade, selecionamos três perguntas-chave (utilizadas nas entrevistas) para explorar o papel da ascendência alemã na formação de sua identidade. As perguntas, escolhidas minuciosamente, servirão como ponto de partida para analisar atitudes, autoidentificações e sentimentos dos falantes em relação às línguas que usam. As perguntas se conectam diretamente às discussões sobre a representação linguística desenvolvidas neste texto. A tabela a seguir mostra de maneira ordenada as perguntas escolhidas ao lado da relação e desdobramento do conceito de representação.

Tabela 1 –

Perguntas selecionadas	Relação com representação linguística
Como o alemão influencia na maneira como você se define e se identifica?	<i>Influência das línguas na definição e identidade pessoal</i>
Você se considera um brasileiro/alemão?	<i>Autoidentificação como brasileiro/alemão</i>
“Eu me sinto em casa quando ouço a língua alemã”. Numa escala de 0 a 10, quão verdadeira/significativa é essa frase para você?	<i>Sentimento de pertencimento e conexão com a língua e cultura alemã</i>

Fonte: elaboração dos autores.

7.1 Influência das línguas na definição e identidade pessoal

Ao introduzirmos a pergunta “Como o alemão influencia na maneira como você se define e se identifica?” tivemos o objetivo entender de que maneira os informantes percebem a influência da língua alemã na construção de suas identidades pessoais e sociais. As respostas a esse tipo de questão abrem espaço para observar se o alemão é visto como um elemento central na definição de quem eles são ou pelo menos representa uma parte de sua identidade cultural.

De modo geral, os descendentes em Niterói demonstram uma relação ambivalente com a língua e a cultura alemã. Embora haja interesse em se conectar a esse legado, segundo as entrevistas, a influência do alemão na construção identitária muitas vezes aparece de forma indireta, mediada por valores culturais e práticas transmitidos pela família. Na maioria dos casos, o alemão já não é mais falado nos ambientes familiares, mas é percebido como uma herança cultural específica. Esta afirmação está apoiada pelo “sentimento de orgulho” que percebemos durante as narrativas ouvidas. Sendo assim, esta língua representa um componente importante (porém não exclusivo) na construção da identidade desses descendentes. Utilizando a imagem de um mosaico, o alemão seria uma peça.

O relato do entrevistado 1M, cujo fragmento está apresentado a seguir, é revelador no sentido que destaca a influência cultural alemã naquela família em particular. As práticas de leitura (em mais de uma língua), o desenvolvimento de habilidades práticas, a busca e apreço pelo conhecimento científico são percebidas pelo informante como transmissão direta de seus antecessores alemães. Na visão do informante, o gosto pelos estudos foi passado por esses familiares. Neste caso língua/cultura alemã influencia a visão de mundo através de um conjunto de valores transmitidos pela família.

Uma outra reflexão a ser feita é que a influência linguística e cultural dos antecessores alemães vai além da “necessidade de falar alemão”, antes é percebido em relatos como esse apresentado. A prática de leitura em múltiplas línguas, por exemplo, não apenas demonstra um legado cultural, mas também simboliza um capital cultural que reforça a percepção de pertencimento a um grupo diferenciado. Em outros termos, a língua alemã pode exercer um grande poder simbólico, mesmo não sendo falada no cotidiano dos descendentes.

Entrevistador: E como o alemão, essa cultura, essa ascendência, influencia na maneira como você se define e se identifica hoje?

Informante 1M: Eu não sei até que ponto isso pode influenciar na minha vida e tudo mais, mas uma coisa que eu sempre entendi, que eu via como por exemplos dentro da família do meu pai, que todos tinham muito conteúdo, muito estudo científico. Meu avô, por exemplo, ele lia muita coisa em alemão, em inglês, português e passou isso para os filhos. Então, nós temos um desenvolvimento metodológico, científico muito grande, que nos trouxe a capacidade de desenvolvimento de estudo, de fazer várias coisas que não são comuns de se fazer. Por exemplo, desde consertar carro, a eletrodoméstico, a fazer obras de alvenaria, essas coisas que hoje em dia não são comuns.

O informante 3M traz um ponto de vista diferente e mais abrangente comparado ao informante 1M no quesito construção de identidade. 3M adota uma maneira de incluir outras etnias/culturas na descrição de sua identidade e destaca a generosa receptividade do Brasil em relação aos povos estrangeiros que aqui vêm buscar moradia. Tal abordagem é revela-

dora porque dialoga intensamente com as falas sobre construção de identidade trazidas por Hall (1992), Jenkins (2006), Edwards (2009) e reforçada por Macedo (2024). De acordo com os autores, a identidade é um processo dinâmico, em constante construção. Os autores se afastam da ideia de que a identidade é entidade fixa, moldada pela história. O informante destaca essa característica ao reinterpretar sua identidade em um contexto multicultural, o que faz bastante sentido em um país como o Brasil devido sua combinação de muitas culturas ao longo da história.

Para o participante 3M, a herança cultural alemã tem o seu papel e seu valor em sua individualidade, que está acompanhada de outras influências. Notemos, por exemplo, a menção aos escoceses e libaneses, que indicam a preferência do informante por uma narrativa que privilegia a pluralidade e não o protagonismo alemão. Uma reflexão mais cuidadosa do fragmento a seguir pode nos levar a conceber o papel da língua alemã como simbólico e não necessariamente funcional ou central na prática cotidiana dos indivíduos.

Entrevistador: E o fato de você ser um descendente alemão influencia na maneira como você se define e como você se identifica hoje?

Informante 3M: Então, se a gente parar para pensar, a curto prazo, eu vim ali dos alemães. Se a gente buscar mais informações ou abrir a árvore, pai e mãe, a gente tem pai vindo da Alemanha e mãe diretamente vindo da Escócia, escocês. Então, você começa a misturar. Isso é que eu sinto no Brasil. Mistura de raças, e tanto é que é um país que é muito receptivo para qualquer cultura, língua, seja judeu, seja árabe, seja libanês, como a gente está recebendo agora, é um país maravilhoso nesse sentido não faz acepção de pessoas.

Outros informantes não se posicionaram de maneira tão plural e optaram pelo destaque da cultura e língua alemã em suas famílias, o que nos indica, mais uma vez que a construção da identidade é um fenômeno relativo, diverso e variado.

7.2 Autoidentificação como brasileiro/alemão

Quando perguntamos aos participantes se eles se consideravam brasileiros/alemães buscávamos investigar possíveis casos de identidades múltiplas e/ou híbridas. Consideramos parte da exploração à representação linguística (e social, neste caso) saber se os indivíduos se veem integrantes de uma identidade “transnacional” e como a veiculam através de práticas linguísticas.

Entrevistador: E você se considera um brasileiro/alemão?

Entrevistado 1M: Não. Eu sou um brasileiro nato. Ou seja, eu sou miscigenado. Eu tenho alemão, eu tenho indígena, eu tenho português. Devo ter algum traço de negro também. (...) Eu acho que todos os brasileiros, a maior parte deles, todos não, mas a maior parte dos brasileiros são miscigenados.

Entrevistador: O senhor se considera um brasileiro/alemão?

Entrevistado 2M: Não. Não porque não tenho muito vínculo. Hoje eu me considero brasileiro/português, porque eu consegui a cidadania portuguesa e tenho realmente muito mais vínculo com Portugal do que com a Alemanha. Mas eu sempre tive muita admiração pela Alemanha, pelo povo alemão, pela história.

Notamos que a fala do informante 1M evidencia uma visão essencialista de identidade nacional, na qual ser brasileiro é, para ele, sinônimo de miscigenação e integração cultural, em oposição à ideia de uma identidade dual ou transnacional. Ele rejeita a possibilidade de se identificar como brasileiro e alemão simultaneamente, destacando que sua identidade é moldada por múltiplas origens (alemã, indígena, portuguesa e possivelmente africana), mas sem reivindicar uma conexão específica e ativa com nenhuma delas. Mais uma vez, a identidade é vista como uma entidade fluida e divisível.

Como visto no trecho acima, o informante M2 prefere não se identificar como um brasileiro/alemão. O motivo é a ausência de vínculos tanto com seus familiares alemães quanto ao não uso (e conhecimento) da língua referida. No caso deste participante especificamente, os laços com o passado português são mais evidentes, reforçado pela obtenção da cidadania portuguesa. A resposta reflete como laços culturais (e também “administrativos” - como a cidadania - podem influenciar fortemente a percepção de identidade social. Para M2, a conexão com a Alemanha é mais admirativa e histórica, enquanto a com Portugal é mais concreta e atual, baseada em vínculos reais e tangíveis. É fundamental lembrar aqui que Niterói também recebeu um alto número de imigrantes portugueses, cuja presença influenciou significativamente a composição cultural e social da cidade, convivendo com as comunidades alemãs.

À luz de Petitjean (2009), podemos compreender que as respostas dos entrevistados não expressam apenas uma autoavaliação identitária, mas constituem representações sociais que organizam como cada indivíduo se posiciona em relação às origens familiares, às línguas associadas a essa herança e aos grupos culturais envolvidos. Para a autora, representações linguísticas e identitárias se constroem socialmente e se atualizam no discurso, funcionando como formas de classificação simbólica do mundo social.

7.3 Sentimento de pertencimento e conexão com a língua e cultura alemã

Através da terceira pergunta selecionada (“‘Eu me sinto em casa quando ouço a língua alemã’. Numa escala de 0 a 10, quão verdadeira/significativa é essa frase para você?”), almejamos conhecer a relação emocional dos entrevistados com a língua alemã. A escala utilizada foi uma maneira de quantificar a intensidade da conexão emocional que os entrevistados podem ter com a língua alemã, permitindo uma análise mais ou menos precisa das percepções individuais sobre seu significado cultural e identitário. Ao usar a escala de 0 a 10, buscamos obter uma medida que vai desde uma total desconexão (0) até uma forte identificação emocional com a língua (10). Desta maneira, pudemos estudar não apenas a ausência ou presença de uma carga emocional, mas explorar as nuances dessa relação. Além disso, muitos entrevistados, além da nota introduziram um “porque”.

Entrevistador: Eu me sinto em casa quando ouço a língua alemã/inglesa. Numa escala de 0 a 10, quão verdadeira/significativa é essa frase para você?

Entrevistado 2M: Não, falso.

Entrevistador: Numa escala de 0 a 10, quão verdadeira/significativa é essa frase para você?

Entrevistada 1F: Eu acho que 0 ou no máximo 1.

Entrevistador: (nome do entrevistado), eu vou te falar uma frase e aí você me diz, numa escala de 0 a 10, quão verdadeira ou quão significativa é essa frase pra você, tá bom?

Entrevistado 3M: É dois. Primeiro, porque você não está entendendo o que está sendo dito. Então, você é um estranho ali. Se fosse algo familiar, eu diria 10, porque você está compreendendo aquela língua e você pode replicar, você pode ter uma comunicação efetiva ali, né?

Observamos um padrão de distanciamento emocional dos entrevistados em relação à língua alemã através das notas baixas dadas nas respostas. O informante 2M ao responder simplesmente com a palavra “não” enfatiza sua ausência de identificação com a língua em questão. De maneira semelhante, a entrevistada 1F atribui nota 0 ou 1, o que reforça essa desconexão quase absoluta. O participante 3M atribui uma nota baixa que pode ser explicada pela barreira linguística; ele utilizou a palavra “estranho” que é bastante reveladora neste contexto. A representação do alemão é mediada pelo critério da compreensão, ou seja, familiaridade cognitiva. Para o informante, “sentir-se em casa” está relacionado ao fato de poder participar do mundo social da língua; isso se ancora nos pressupostos de Petitjean (2009), que afirma que a representação linguística é apoiada em relações de proximidade e prática.

Os dados revelam uma correlação direta entre a dificuldade em compreender o alemão e a falta de identificação com a língua. Para os participantes, a capacidade de usar a língua no dia a dia é fundamental para se sentir parte de uma cultura. A ausência dessa conexão prática reforça a percepção de que o alemão é “uma peça no mosaico identitário” e não um componente central ou de destaque.

Segundo Petitjean (2009), as representações linguísticas não descrevem a língua em si, mas refletem a forma como os sujeitos significam socialmente essa língua, situando-se na interseção entre memória, afetos, valores e identidades. Quando examinamos as respostas dos entrevistados, fica claro que o alemão não aparece como um simples código linguístico, mas como um objeto simbólico, cuja representação está associada a pertencimento e familiaridade. Em outros termos, embora os informantes não saibam a língua alemã, esta faz parte do conjunto complexo de suas representatividades sociais.

A não familiaridade com a língua alemã não equivale, com base no que já apresentamos até aqui, uma completa desconexão ou ruptura com a ancestralidade e preservação da memória familiar dos nossos informantes. A língua, inevitavelmente, atua como um fator poderoso de fazer com essas pessoas se sintam parte (ou não) de um outra cultura, entretanto, mesmo sem saber tal língua, os descendentes de Niterói demonstram grande valorização e respeito por sua história.

8 Considerações finais

Neste artigo, abordamos a questão da representatividade linguística em um contexto específico: os descendentes de alemães no município de Niterói, Rio de Janeiro. Discutimos a relevância da imigração alemã para o Brasil e argumentamos que, embora numericamente menos expressiva em comparação a outras regiões do país, a presença desses imigrantes na cidade deixou marcas, contribuindo para a formação do patrimônio cultural e linguístico local.

Apresentamos e discutimos o conceito de representatividade linguística com base nos trabalhos de Petitjean (2008; 2009) e Gueunier (2003), analisando sua construção teórica ao longo do tempo e destacando sua importância para a compreensão das relações entre língua, representação e práticas sociais. No que tange à coleta de dados, realizamos uma série de entrevistas sociolinguísticas com descendentes de alemães, analisando como eles constroem suas identidades sociais com base em sua ancestralidade linguística. Essas interações permitem explorar as percepções, atitudes e práticas relacionadas à língua, destacando como a herança cultural e linguística influencia suas representações identitárias. Incluímos um breve seção comentando o perfil sociolinguístico e cultural dos entrevistados. A parte final, uma análise das respostas a três perguntas utilizadas nas entrevistas, revelou como as narrativas construídas em torno da língua alemã são permeadas por um entrelaçamento de fatores históricos, culturais e sociais. Embora a presença da comunidade alemã em Niterói não seja tão amplamente documentada nem tenha produzido impactos estruturais duradouros, como ocorreu em estados do Sul do Brasil, as referências à origem e à língua alemãs foram mobilizadas discursivamente pelos entrevistados no plano da memória e da identificação simbólica.

Ainda que tenhamos um foco linguístico, nosso estudo é capaz de dialogar com outros trabalhos realizados por pesquisadores não apenas da sociolinguística, mas de outras ciências, como a história, sociologia e antropologia, que têm interesse em explorar a relação entre língua, identidade e memória cultural em contextos de imigração.

Nossa pesquisa se destaca pelo fato de (re)contar parte da história niteroiense através de uma perspectiva teórica não muito difundida e que enriquece os estudos socioculturais ao lançar luz sobre dimensões linguísticas e identitárias frequentemente negligenciadas.

Digno de menção é o fato de que este trabalho integra um projeto mais amplo, conforme apresentado na introdução, e que outras entrevistas, assim como os resultados decorrentes, ainda estão em processo de análise.

Esperamos que as contribuições trazidas com esse estudo possam fomentar a discussão sobre a relação representação linguística e formação de identidade cultural, incentivando novos olhares interdisciplinares e aprofundando a compreensão sobre os impactos sociolinguísticos da imigração em diferentes contextos. Esperamos também que este presente artigo sirva como ponto de partida para futuras pesquisas.

Agradecimentos

Este trabalho foi realizado com o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), por meio das bolsas Pós-Doutorado Nota 10 (processo nº E-26/200.191/2024) e Cientista do Nosso Estado (CNE). Agradecemos também o fomento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através da bolsa de produtividade.

Declaração de autoria

Anderson Lucas da Silva Macedo foi responsável pela coleta dos dados junto aos descendentes de imigrantes alemães em Niterói, pela análise sociolinguística e pela redação da versão

inicial do manuscrito. Mônica Maria Guimarães Savedra contribuiu para a fundamentação teórica, especialmente no que se refere ao conceito de representatividade linguística, bem como para a interpretação dos resultados. Ambos os autores foram responsáveis pela metodologia aplicada, realizaram revisão crítica e o aprimoramento da versão final do texto.

Referências

BEZERRA, Maria Cristina Caminha. *Britânicos e Alemães em Niterói: um estudo de migração urbana*. 2015. 350 f. Tese (Doutorado) – Universidade Federal Fluminense, Niterói, 2015.

CALVET, Louis-Jean. Insécurité linguistique et représentations: approche historique. Une ou des normes? In: CALVET, Louis-Jean; MOREAU, Marie-Louise (orgs.). *Insécurité linguistique et normes endogènes en Afrique francophone*. Paris: Didier Erudition, 1998. p. 9-17.

CALVET, Louis-Jean. *Towards an Ecology of World Languages*. Cambridge: Polity Press, 2006.

EDWARDS, Jonh. *Language and Identity*. Cambridge: Cambridge University Press, 2009.

GUEUNIER, Nicole. Attitudes and representations in sociolinguistics: theories and practice. *International Journal of the Sociology of Language*, n. 160, p. 41-62, 2003.

HALL, Stuart. Who needs 'identity'? In: HALL, Stuart; DU GAY, Paul (ed.). *Questions of Cultural Identity*. London: Sage, 1996. p. 1-17.

LANG, Agnes Cristina Wiedemann. Comunidade de luteranos alemães de Niterói: memória e identidade. *Cadernos do CEOM*, ano 16, n. 15, UNOCHAPECÓ, jun. 2002.

JENKINS, Richard. *Social Identity*. Londres/Nova Iorque: Routledge press, 2004.

MACEDO, Anderson Lucas. *Ethno-Linguistic Identity of the Post-Apartheid Generation*. Berlin: Peter Lang, 2024. DOI: 10.3726/b21579

MICHAHELLES, Marina. A "Colônia Alemã" do Rio de Janeiro pelas lentes do clube Germânia. *TRAVESSIA - revista do migrante*, [S. l.], v. 15, n. 44, p. 35-40, 2002. DOI: 10.48213/travessia.i44.858. Disponível em: <https://travessia.emnuvens.com.br/travessia/article/view/858>. Acesso em: 20 nov. 2024.

PETITJEAN, Cécile. *Representations linguistiques et accents régionaux du français*. Journal of Language Contact, 2008.

PETITJEAN, Cécile. *Représentations linguistiques et plurilinguisme*. 2009. 488 f. Tese (Doutorado em Linguística) – Institut des Sciences du Langage et de la Communication, Université de Neuchâtel, Neuchâtel, 2009.

SAVEDRA, Mônica M.; ROSENBERG, Peter. Deutsche Einwanderung in Brasilien: Sprachrevitalisierung und Transkulturalisierung. In: JOHNEN, Thomas; SAVEDRA, Monica; SCHRÖDER, Ulrike (org.). *Sprachgebrauch im Kontext: Die deutsche Sprache im Kontakt, Vergleich und in Interaktion mit Lateinamerika/Brasilien (Kultur – Kommunikation – Kooperation)*. Frankfurt am Main: Peter Lang, 2019. p. 19-54.

SAVEDRA, Mônica Maria Guimarães; MAZZELLI, Leticia. Variedades linguísticas da imigração germânica no Brasil: vitalidade, glotopolítica e território. *Revista Feira de Santana*, v. 21, n. 1, p. 105-131, jan./abr. 2020.

SAVEDRA, M. M. G.; CHRISTINO, B.; PUPP SPINASSÉ, K.; ARAUJO, S. S. de F. Studies in contact socio-linguistics in Brazil: ethnolinguistic diversity in focus. *Cadernos de Linguística*, [S. l.], v. 2, n. 1, p. e315, 2021. DOI: 10.25189/2675-4916.2021.v2.n1.id315. Disponível em: <https://cadernos.abralin.org/index.php/cadernos/article/view/315>. Acesso em: 15 feb. 2025

SAVEDRA, Mônica Maria Guimarães; MACEDO, Anderson Lucas; SILVA, Maria Eduarda; TÁVORA, Robert Gomes. Entre práticas linguísticas e imigração: o caso dos germânicos de Niterói. In: SAVEDRA, Mônica Maria Guimarães; PEREIRA, Telma; MACEDO, Anderson Lucas (Orgs.). *LABPEC UFF: contato, glotopolítica e etnicidade em movimento*. São Carlos: Pedro & João Editores, 2025. p. 171-192. DOI: 10.51795/9786526521410.

SILVA, Sidney de Souza; MELLO, Heloísa Augusta Brito de. A Colônia do Rio Uvã: várias histórias em uma só. *Signum: Estudos de Linguística*, Londrina, n. 13/2, p. 417-452, dez. 2010.

TARALLO, Fernando. *A pesquisa sociolinguística*. São Paulo: Ática, 1997.

TRUDGILL, Peter. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press, 1974.

A alteridade na aquisição da língua materna: um olhar retrospectivo

Alterity in the Acquisition of the Mother Tongue: A Retrospective Look

Giovane Fernandes Oliveira

Universidade Federal de Pelotas (UFPel)

Pelotas | RS | BR

giofoliver@gmail.com

<https://orcid.org/0000-0001-8251-8353>

Resumo: Este artigo tem por objetivo delinear um breve histórico do problema da relação criança-outro no campo da aquisição de linguagem. Para tanto, inicialmente, introduz o debate sobre duas visões do problema, uma visão dual e uma visão triádica. Em seguida, examina duas propostas, cada qual representativa de uma visão: a proposta de Jerome Bruner, representativa da visão dual (criança-outro), e a proposta de Cláudia de Lemos, representativa da visão triádica (criança-outro-língua). Os resultados conduzem à conclusão de que as duas visões, embora converjam na tomada do diálogo como unidade analítica, divergem tanto em termos metodológicos (elegendo para análise distintos períodos da trajetória linguística da criança) quanto em termos teóricos (assumindo diferentes concepções de linguagem/língua, cultura, aquisição, sujeito, outro e interpretação).

Palavras-chave: diálogo; interação; alteridade; aquisição.

Abstract: This article aims to outline a brief history of the problem of the child-other relationship in the field of language acquisition. To this end, it initially introduces the debate on two views of the problem, a dual view and a triadic view. It then examines two proposals, each representing a different view: Jerome Bruner's proposal, representing the dual view (child-other), and Cláudia de Lemos' proposal, representing the triadic view (child-other-language). The results lead to the conclusion that the two views, although converging in taking dialogue as the analytical unit, diverge both in



methodological terms (electing different periods of the child's linguistic trajectory for analysis) and in theoretical terms (assuming different conceptions of language, culture, acquisition, subject, other and interpretation).

Keywords: dialogue; interaction; alterity; acquisition.

1 Considerações iniciais

Estímulo, input, mediação, imitação, interlocução, comunicação, conversação, cooperação, convivência, partilha, díade, ação e atenção conjuntas, dialogismo, processos dialógicos, diálogo, interpretação, alteridade, intersubjetividade, alocação. Não são poucos os termos que dizem da relação criança-outro no campo da aquisição de linguagem, em sua diversidade de perspectivas teóricas, abordagens metodológicas e interesses temáticos.

Menos numerosas, contudo, são reflexões que visem a uma desnaturalização desses termos, vale dizer, a uma interrogação das acepções que assumem no interior de uma pesquisa e das consequências teórico-metodológico-analíticas dessas acepções em tal pesquisa. Com efeito, da circulação dos termos em pauta nos estudos aquisicionais, poderíamos dizer o mesmo que diz Benveniste (2023 [1966/1967], p. 223) sobre a circulação do termo *signo* nos estudos linguísticos: “Não se pode deixar de ficar admirado por ver tantos autores manipularem inocentemente este termo ‘signo’ sem discernir o que ele contém de restrições para quem o adota e em que ele o compromete a partir daí”.

Neste artigo, exponho os resultados da primeira parte de um estudo a respeito da alteridade na aquisição da língua materna. Mais especificamente, tenho por objetivo delinear um breve histórico do problema da relação criança-outro no campo da aquisição de linguagem. Tal histórico será apresentado em três seções: inicialmente, introduzo o debate sobre duas visões do problema nesse campo, uma visão dual e uma visão triádica; em seguida, discuto a visão dual; na sequência, abordo a visão triádica. Na conclusão, faço uma síntese do trajeto percorrido e anuncio os próximos passos da investigação aqui iniciada.

2 A relação criança-outro nos estudos em aquisição de linguagem¹

No campo aquisicional, a relação criança-outro passou a ser mais problematizada em um terceiro momento de estudos do referido campo (M. T. G. de Lemos,² 2002). Inicialmente,

¹ Esta seção retoma e aprofunda parte da apresentação do livro que organizei juntamente às Profas. Dras. Marlete Sandra Diedrich (UPF) e Alessandra Del Ré (UNESP-Araraquara) (Diedrich; Oliveira; Del Ré, 2023).

² Neste artigo, serão citadas duas autoras de sobrenome “Lemos”: Cláudia Thereza Guimarães de Lemos e Maria Tereza Guimarães de Lemos. A fim de evitar confusão entre as duas pesquisadoras, a primeira (da qual serão mencionados diversos estudos) será referida como “Lemos” e a segunda (da qual será mencionado um estudo de 2002), como “M. T. G. de Lemos”.

na década de 1950, em que o empirismo dominava a cena psicológica (Skinner) e a cena linguística (Bloomfield), a psicolinguística emergiu analisando o dito comportamento verbal do adulto e da criança à luz da linguística estrutural norte-americana. Em seguida, na década de 1960, em que o racionalismo (Chomsky) passou a dominar a cena dos estudos linguísticos e psicolinguísticos, os estudiosos lançaram-se na busca por uma gramática infantil, a ser alcançada mediante a identificação de regularidades formais da fala da criança. Posteriormente, a partir da década de 1970, antagonizando tanto com o empirismo quanto com o racionalismo, uma terceira posição epistemológica começa a se destacar no campo: o interacionismo.

Interacionistas: assim se tornaram conhecidos estudos muito diversos entre si, seja em termos de diversidade teórica (perspectivas cognitivas, pragmáticas, linguístico-psicanalíticas, enunciativas, discursivas, multimodais e, mesmo, interfaces entre tais perspectivas), seja em termos de diversidade analítica (processamento linguístico, formatos comunicativos, posições estruturais, operações enunciativas, movimentos dialógicos, matriz multimodal). Apesar dessa diversidade teórico-analítica, esses trabalhos podem ser agrupados em três grandes classes de acordo com a unidade que elegem para análise (Lemos, 1986a).

Na primeira classe, situam-se os trabalhos acerca do “manhês”, cuja unidade analítica é o enunciado da mãe em um momento inicial e o enunciado da criança em um momento posterior. Trata-se de enunciados isolados e não de uma relação entre eles, o que compromete a qualificação desses estudos como *interacionistas* e os mostra preocupados menos com a interação adulto-criança do que com a relação do aprendiz com o *input* linguístico. Segundo Lemos (1986a), tal preocupação dá a ver uma *visão facilitativa* da interação na aquisição, o que aproxima essa primeira classe mais do inatismo (racionalismo) do que do interacionismo.

Na segunda e na terceira classes, situam-se os trabalhos que se afastam da visão facilitativa e buscam uma *relação explicativa* entre a interação social e a aquisição de linguagem, explicação definida em termos de *continuidade funcional* ou *estrutural* do dito período pré-linguístico (não verbal) para o período linguístico (verbal).

Na segunda classe, situam-se os trabalhos acerca da “comunicação”, cuja unidade analítica consiste no comportamento comunicativo infantil no período dito pré-linguístico e/ou no período dito linguístico inicial (holofrástico ou de um vocábulo). A principal semelhança desses estudos é a tomada de tal comportamento – gestual, vocal, prosódico – como realização não linguística de uma intenção comunicativa da criança. Já a principal diferença reside nas soluções divergentes que encontram ao tentarem explicar a continuidade entre o pré-linguístico e o linguístico, alguns remetendo a explicação ao desenvolvimento cognitivo e outros, à maturação do organismo. Para Lemos (1986a), a despeito de sua contribuição descritiva, as investigações dessa segunda classe seguiram analisando a produção da criança dissociada da produção do outro, não conseguindo demonstrar como a interação social opera sobre o curso da aquisição de linguagem.

Na terceira classe, situam-se os trabalhos acerca da “interação” propriamente dita, cuja unidade de análise é o diálogo criança-outro, interpretado como esquema interacional por alguns e como alteridade estruturante por outros. Dentre os pesquisadores do primeiro grupo, destaca-se Jerome Bruner, cuja hipótese de continuidade funcional entre o pré-linguístico e o linguístico assenta-se na defesa da linguagem como reflexo da cognição e da interação como pré-requisito da aquisição. Dentre os estudiosos do segundo grupo, destacam-se pesquisadores que, como Cláudia de Lemos, voltaram-se para o papel da “sintaxe vertical” (produzida no diálogo adulto-criança) na estruturação da “sintaxe horizontal” (produzida pela criança).

Em suas origens, nos anos 1980, a proposta da autora ficou conhecida pelos termos *sociointeracionismo* e *socioconstrutivismo* (M. T. G. de Lemos, 2002, p. 10), devido à sua proximidade de autores da psicologia do desenvolvimento (Vygotsky e Piaget). Todavia, nos anos 1990, essa proposta passou a ser chamada de *interacionismo*, *interacionismo brasileiro* ou, ainda, *interacionismo estrutural*, devido à sua filiação ao estruturalismo francês (Saussure e Jakobson) e à psicanálise (Lacan). Filiação que especifica a proposta de Lemos no conjunto de outras propostas não menos interacionistas, mas vinculadas a outros teóricos, sejam teóricos da aquisição (Bruner, Tomasello), sejam teóricos da linguagem aos quais se filiam aquisicionistas (McNeill, Kendon, Bakhtin e o Círculo, Benveniste, Authier-Revuz, Pêcheux).

A esse respeito, embora reconheçam a inclusão do outro como elemento comum às diferentes propostas interacionistas, Lier-DeVitto e Carvalho (2024 [2008]) optam por distinguir os termos *sociointeracionismo* e *interacionismo*, associando o segundo rótulo à perspectiva de Cláudia de Lemos (voltada aos aspectos estruturais da interação) e o primeiro rótulo, às demais perspectivas (voltadas aos aspectos sociais e cognitivos da interação).

No entanto, as autoras não mencionam o fato de os estudos iniciais de Cláudia de Lemos reivindicarem o rótulo *sociointeracionismo* e, mesmo, o rótulo *socioconstrutivismo*. Além disso, a emergência, a partir dos anos 2000, de estudos aquisicionais filiados a teorias enunciativas e discursivas – como a teoria benvenistiana (cf., por exemplo, Oliveira; Silva, 2023) –, obriga-nos a relativizar a distinção *sociointeracionismo/interacionismo* feita por Lier-DeVitto e Carvalho (2024 [2008]) por duas razões. Em primeiro lugar, a maioria desses estudos não parece aderir aos pressupostos teóricos do sociointeracionismo como pelas autoras caracterizado (dentre os quais a concepção instrumental de linguagem e a concepção psicológica de sujeito). Em segundo lugar, apesar de nem todos esses trabalhos se dizerem *interacionistas*, se tivessem de ser inseridos em uma das três grandes posições do campo (empirismo, racionalismo e interacionismo), não há dúvidas de que seria a terceira aquela que os abrigaria; afinal, é de interação – na acepção mais geral de *relação criança-outro* – que, ao fim e ao cabo, todos tratam ao tratarem de aquisição, enunciação e discurso.

Voltando à divisão interna da posição interacionista, no interior de sua terceira classe, pode ser operada, ainda, uma clivagem mais compatível com a configuração atual dos estudos que tomam a relação criança-outro como unidade analítica. Trata-se de uma clivagem entre duas visões de diálogo,³ cada qual reúne perspectivas teóricas que, embora distintas, partilham traços em comum: de uma parte, uma *visão dual*; de outra parte, uma *visão triádica*.

A fim de proceder à caracterização dessas visões, devido aos limites de espaço, impõe-se a necessidade de dois recortes. Primeiro recorte: em vez de serem referidos panoramicamente diversos autores e estudos, foram eleitas, para um exame mais detido, duas propostas, cada qual representativa de uma visão: a proposta de Jerome Bruner, representativa da visão dual, e a proposta de Cláudia de Lemos, representativa da visão triádica. Segundo recorte: tal exame não abrangerá toda a obra de cada um desses dois estudiosos, mas somente os trabalhos que nos permitam discernir os contornos das visões de diálogo antagônicas por eles representadas. Esse recorte se justifica em razão de que tanto Bruner quanto de Lemos publicaram estudos sobre outros temas além da aquisição de linguagem.

³ Uma terceira grande visão parece não se enquadrar nessa clivagem: trata-se da visão de diálogo tributária do Círculo de Bakhtin e da noção de *heteroglossia dialogizada* como “o contínuo processo de encontros e desencontros, de aceitação e recusa, de absorção e transmutação das vozes sociais” (Faraco, 2005, p. 219). Para reflexões (elas próprias diferentes entre si) acerca dessa visão em estudos de aquisição de linguagem, ver Lemos (1994) e Del Ré *et al.* (2022), no âmbito do vir a ser falante, e Capristano (2023), no âmbito do vir a ser escrevente.

A título de contextualização, cabem breves comentários relativos aos dois teóricos.

Jerome Seymour Bruner (1915-2016) foi um psicólogo estadunidense, professor e pesquisador das Universidades de Harvard e de Oxford. Um dos líderes da chamada Revolução Cognitiva, Bruner adotou um enfoque interdisciplinar em suas investigações acerca do desenvolvimento humano, inspirando-se em aportes da filosofia da linguagem, da biologia evolutiva e da antropologia. Dentre os eixos temáticos por ele investigados, destacam-se a ação, o pensamento, a percepção, a educação e a linguagem. No âmbito da aquisição de linguagem, aliando-se às perspectivas pragmáticas de Austin, de Searle e de Grice, produziu uma teorização sobre a ontogênese dos atos de fala no seio da interação adulto-criança, teorização ancorada na observação longitudinal e naturalística de interações de mães com bebês nos primeiros dezoito meses de vida destes (Bruner, 2007 [1983]; Linaza, 2007 [1984]).

Cláudia Thereza Guimarães de Lemos (1934-2026) foi uma linguista e psicanalista brasileira, professora e pesquisadora aposentada da Universidade Estadual de Campinas. Uma das pioneiras do campo da aquisição de linguagem no Brasil, Lemos implementou na Unicamp, em 1976, o Projeto de Aquisição da Linguagem. No quadro desse projeto, foi constituído um banco de dados com gravações de crianças de um a cinco anos, acompanhadas em coletas naturalísticas, em interações com as mães, outros membros da família e investigadoras. Como já dito, a trajetória da autora é marcada por uma mudança teórica, a qual a leva a falar não só sobre aquisição – o que faz de maneira renovada –, mas também sobre a obra saussuriana, a relação linguística-outras ciências, a relação língua-discurso, a poética, a psicanálise (Lemos, 2002; Pereira De Castro, 2005).

É interessante notar que Bruner e Lemos mantiveram certo contato durante algum tempo, o que é reconhecido por ambos.

No prefácio de seu livro *Como as crianças aprendem a falar* (2007 [1983]), Bruner agradece à estudiosa brasileira em dois momentos: “*Colegas convidados de inestimável contribuição: David Olson de Toronto, Robert Grieve de St. Andrews, Katherine Nelson de Yale, Cláudia de Lemos de Campinas e Virginia Volterra de Roma*”; “*Durante um ano de estudo no Netherlands Institute for Advanced Study na Holanda, beneficiei sobretudo de discussões com David Olson, Melissa Bowerman, Cláudia de Lemos*” (Bruner, 2007 [1983], p. 11-12).

Em artigo de 2002, intitulado “Das vicissitudes da fala da criança e de sua investigação”, Lemos comenta ser significativo que “o título do Projeto [de Aquisição da Linguagem], implementado em 1976, era ‘Relações entre desenvolvimento lingüístico, desenvolvimento cognitivo e desenvolvimento comunicativo’, colado, assim, a tudo o que [se] fazia na época”, ao que acrescenta em nota de rodapé: “Ou a exemplo de projetos como o do Instituto de Psicologia do CNR, de Roma, de que participavam Elizabeth Bates, Luigia Camaioni, Virginia Volterra e Francesco Antinucci, e do Departamento de Psicologia da Universidade de Oxford, coordenado por Jerome Bruner, projetos com os quais mantivemos relações estreitas até meados dos anos 80” (Lemos, 2002, p. 45). As razões que suspenderam tais relações deverão ficar claras nas seções seguintes.⁴

⁴ Antes, porém, de passar a essas seções, gostaria de fazer dois esclarecimentos em resposta a um comentário do Avaliador 1 deste artigo, comentário de acordo com o qual haveria uma disparidade em termos numéricos no que concerne aos estudos de Jerome Bruner e de Cláudia de Lemos aqui discutidos (“três publicações de Bruner e dezessete de Lemos”). O primeiro esclarecimento diz respeito ao fato de que, das três publicações de Bruner mencionadas pelo Avaliador 1, uma é um artigo e duas são livros que reúnem estudos publicados pelo autor em diferentes periódicos entre a década de 1970 e o início da década de 1980. Trata-se de seis artigos reunidos no

3 A visão dual do diálogo na aquisição de linguagem

Como Bruner relata no prefácio de *Como as crianças aprendem a falar* (2007 [1983]), foi a mudança para a Universidade de Oxford, em 1972, que o levou a voltar-se para a linguagem. Ali, encontrou uma atmosfera dominada pela teoria dos atos de fala, o que o lançou em uma investigação sobre a ontogênese destes. Seguiu-se uma década de intenso trabalho, com coletas de dados, leituras, discussões e a produção dos primeiros textos sobre aquisição. Publicados separadamente, tais textos foram revisados e reunidos no livro publicado em 1983.

Esse livro conta com um prefácio, dois capítulos teóricos iniciais, três capítulos analíticos e um capítulo teórico final.

No prefácio, Bruner amarra os dois fios condutores da obra, ressaltando a centralidade, em sua proposta, da interação adulto-criança. O primeiro fio, de caráter externo e cultural, tem a ver com o modo como “*a comunidade linguística providencia encontros de fala, de modo a que o jovem aspirante a locutor possa aprender como tornar as suas intenções comunicativas claras, e como penetrar as intenções dos outros*”, sendo o principal veículo dessa ajuda “*o formato, as situações padronizadas que permitem ao adulto e à criança cooperarem na ‘transmissão’ de uma linguagem*” (Bruner, 2007 [1983], p. 10, itálicos e não itálico do original).⁵ O segundo fio, de caráter interno e processual, tem a ver com o modo como “*a intenção comunicativa é sucessivamente transformada, através da negociação, em processos linguísticos cada vez mais poderosos*” (Bruner, 2007 [1983], p. 10, itálicos do original).

Amarrados, esses dois “fios” formam o “nó” que constitui o núcleo teórico do livro: a criança e o adulto são sujeitos intencionais, cooperativos e negociadores; a linguagem é um instrumento de comunicação transparente; e a sua aquisição é um processo de aprendizagem de expressão das intenções próprias e de acesso às intenções alheias.

livro *Como as crianças aprendem a falar* (Bruner, 2007 [1983]) e de dez textos compilados na obra *Ação, pensamento e linguagem* (2007 [1984]). Portanto, também dezessete são os trabalhos presentes nas três referidas publicações de Bruner. O segundo esclarecimento diz respeito ao fato de que, na seleção dos escritos a serem analisados no presente artigo, outros textos de Bruner e Lemos foram lidos, mas foram selecionados apenas aqueles que tratavam de aquisição de linguagem. E, no interior desse conjunto, foram eleitos os estudos que permitem discernir os contornos das visões de diálogo antagônicas por eles representadas – conforme o segundo recorte explicitado anteriormente. Ficaram de fora, então, tanto estudos sobre outros temas (no caso de Bruner, temas como a ação, o pensamento, a percepção e a educação; no caso de Lemos, temas como a obra saussuriana, a relação linguística-outras ciências, a poética e a psicanálise) quanto estudos que, mesmo tratando de aquisição de linguagem, não acrescentam ideias novas à visão de diálogo de cada teórico. É por isso que as ideias de Jerome Bruner são aqui abordadas a partir do livro *Como as crianças aprendem a falar* (2007 [1983]), no qual constam as suas principais proposições sobre a aquisição de linguagem, em geral, e sobre a relação criança-outro nessa aquisição, em particular – proposições que, a meu ver, não sofreram substanciais mudanças teóricas nas reflexões seguintes do autor. Dentre tais reflexões, constam aquelas presentes em Bruner (1997 [1986], 1997 [1990], 1991 e 2000 [1997]), nenhuma das quais aborda a aquisição de língua materna mais enfaticamente nem avança nesse terreno. A aquisição não é um tema transversal à obra de Jerome Bruner, diferentemente de Cláudia de Lemos, que fez da aquisição o principal foco investigativo de sua produção intelectual – o que justifica, também, o fato de o número de estudos da linguista aqui examinados ser maior do que aquele do psicólogo.

⁵ Com vistas à fluidez da leitura, não serão sinalizadas informações como “destaques meus/do original” e “tradução minha”, nem serão registradas, em rodapé, as versões originais das citações traduzidas. Quanto aos destaques, aqueles em itálicos são meus (salvo exceções sinalizadas, como nas citações do prefácio acima), ao passo que os demais (negritos, não itálicos, aspas, sublinhados, maiúsculas) são dos textos citados. Quanto às traduções, são todas de minha responsabilidade.

No Capítulo I, intitulado “Introdução”, Bruner fala em três facetas da linguagem necessariamente *inseparáveis* e *interdependentes* na aquisição desta: a sintaxe (forma), a semântica (sentido) e a pragmática (referência). Porém, ainda nesse capítulo, o autor afirma que, *antes* de a criança produzir um discurso léxico-gramatical, a mãe e ela “criam um formato prévio de interação”, por meio do qual “a criança adquire, então, o domínio da gramática, como referir e significar e como expressar as suas intenções comunicativamente” (Bruner, 2007 [1983], p. 14-15). Afinal, a semântica e a pragmática são inseparáveis/interdependentes da sintaxe ou anteriores a esta? Tal indefinição concernente ao estatuto da relação entre as três facetas persiste nos capítulos seguintes, mas a resposta parece se inclinar, já nesse capítulo, à anterioridade do semântico e do pragmático relativamente ao sintático. Isso porque a condição da aquisição de linguagem é, para Bruner, a posse, pela criança, de

um conjunto único e propenso de capacidades de aprendizagem da linguagem – algo similar ao que Noam Chomsky designou como o Dispositivo de aquisição de linguagem, DAL (LAD – *Language Acquisition Device*). Mas o dispositivo de aquisição de linguagem da criança não poderia funcionar sem a ajuda de um adulto, que entra com ela num formato transaccional. Esse formato, inicialmente sob o controlo do adulto, fornece um sistema de suporte à aquisição de linguagem, SSAL (*Langage Acquisition Support System* – LASS). Ele enquadra ou estrutura a entrada da linguagem e da interação no dispositivo de aquisição de linguagem, de maneira a “fazer o sistema funcionar”. Resumindo, é a interação entre o DAL e o SSAL que torna possível à criança entrar na comunidade linguística – e, ao mesmo tempo, na cultura a que a linguagem dá acesso. (Bruner, 2007 [1983], p. 15).

No Capítulo II, intitulado “Da comunicação à fala”, Bruner aborda a passagem da comunicação pré-linguística à comunicação linguística, defendendo que “A linguagem é o meio para a interpretação e a regulação da cultura” (Bruner, 2007 [1983], p. 18), o que requer “um papel bastante mais activo do adulto na ajuda à aquisição da linguagem pela criança, do que o de simples ‘modelo’”, auxiliando-a “a tornar as intenções claras e a sua expressão adequada às condições e aos requisitos da ‘comunidade de fala’, isto é, a cultura” (Bruner, 2007 [1983], p. 32). Nessa direção, “a entrada na linguagem é uma entrada no discurso, que exige que ambos os membros de um *par dialogante* interpretem uma comunicação e a sua intenção”, sendo essa mútua interpretação definida como uma “*perfeita sintonização*”: “Os pais falam a um nível em que as crianças os podem compreender” (Bruner, 2007 [1983], p. 32-33).

Se, no capítulo anterior, a cultura é situada como algo diferente da comunidade linguística, nesse capítulo, a cultura é situada como a própria comunidade linguística/de fala. Em ambos os capítulos, a linguagem figura como a porta de entrada na cultura/comunidade, porta cuja chave é fornecida pelo adulto à criança em uma relação perfeitamente sintonizada.

No Capítulo III, intitulado “Brincadeiras, jogos e linguagem”, Bruner relata estudos empíricos sobre a ontogênese das brincadeiras e dos jogos da infância, como o esconde-esconde. Dentre as características dos jogos, três ressaltam a linguagem e a interação. Em primeiro lugar, “eles fornecem a primeira ocasião para o *uso sistemático da linguagem da criança com um adulto*”, possibilitando a ela “explorar como conseguir *fazer coisas com palavras*” (Bruner, 2007 [1983], p. 38). Em segundo lugar, “Tais jogos incluem, igualmente, uma outra característica de delineação da linguagem – a atribuição de *papéis alternados que são intercambiáveis*. Há um descobridor e um escondido, um actor e um paciente da experiência” (Bruner, 2007 [1983], p. 38). Em terceiro lugar, “os jogos fornecem uma oportunidade para *distribuir a atenção*

por uma sequência ordenada de acontecimentos. O jogo é o *tópico* sobre o qual cada uma das jogadas pode ser considerada um *comentário*” (Bruner, 2007 [1983], p. 39).

Tais características dão a ver o primado, na teorização de Bruner, da instrumentalidade da linguagem (primeira característica), da simetria criança-outro (segunda característica) e da anterioridade da cognição à linguagem (terceira característica), sendo as estruturas linguísticas vistas como reflexos das estruturas cognitivas, a exemplo da estrutura tópico-comentário, a qual refletiria a estrutura da atenção enquanto síntese do todo e análise das partes.

No Capítulo IV, intitulado “O crescimento da referência”, o autor relata estudos sobre as raízes psicológicas da referência, sustentando quatro pressupostos ligados à interação adulto-criança. O primeiro pressuposto é o de que “os indivíduos podem *assinalar um ao outro* que têm uma intenção referencial ou indicativa” (Bruner, 2007 [1983], p. 57). O segundo pressuposto é o de que a criança e o adulto “podem referir-se ao ‘mesmo’ tópico com graus amplamente diferentes de precisão”, mas esse tópico é “um *referente conjunto* que não apenas existe em tais *conversações assimétricas*, mas que pode ser desenvolvido pelo seu valor de verdade e pela sua definição”, podendo a “*negociação dialogal* [...] até *aumentar* a definição da criança” (Bruner, 2007 [1983], p. 58). O terceiro pressuposto é o de que “a referência é uma forma de interação social, que tem a ver com a gestão da *atenção conjunta*” (Bruner, 2007 [1983], p. 58, *itálico do original*). O quarto pressuposto é o de que “há uma estrutura de objetivo no referir”, embora alcançar o objetivo referencial tenha “pouco a ver com a concordância quanto a um único referente definitivo. É suficiente que as partes numa *troca referencial* saibam que partilham *suficiente coincidência* na sua atenção focal para tornar válida a sua manutenção” (Bruner, 2007 [1983], p. 58).

Embora o psicólogo mencione *conversações assimétricas*, trata-se de uma “assimetria na referência” (Bruner, 2007 [1983], p. 78), entendida em termos de conhecimento de mundo desigual entre adulto (que conhece muito) e criança (que conhece pouco). Trata-se de uma assimetria superável na negociação dialogal que, na troca referencial, vai conjugando a atenção de cada membro do par para o referente em questão. Em outras palavras, trata-se de uma assimetria cuja conversão em simetria é garantida.⁶

No Capítulo V, intitulado “O desenvolvimento do pedido”, Bruner relata estudos sobre a aquisição dos atos de pedir, na qual destaca o papel de intérprete das demandas da criança

⁶ A esse propósito, o Avaliador 1 fez um interessante comentário, que aqui reproduzo de maneira a respondê-lo: “As ideias de ‘simetria criança-outro’ e ‘lugares simétricos’ parecem constituir-se mais como uma leitura de Bruner feita pelo/a próprio/a autor/a do artigo e por M. T. G. de Lemos, a quem ele/a cita, a partir de uma perspectiva teórica própria, do que como uma formulação de Bruner. No artigo, afirma-se que Bruner postula uma *simetria* entre criança e adulto. No entanto, conceitos centrais em sua teoria que não foram discutidos, como os de *tutela* e de *scaffolding*, reforçam a ideia de assimetria constitutiva desta relação. Cabe ressaltar que a leitura de uma proposta teórica a partir de outra traz inevitáveis deslocamentos conceituais, já que os termos utilizados podem coincidir, mas não necessariamente se referem aos mesmos conceitos”. Gostaria de fazer duas ressalvas quanto a esse comentário: 1) Entendo que críticas de uma teoria a outra são legítimas desde que respeitadas as suas diferenças epistemológicas. Por exemplo, como linguista, critico a visão dual do diálogo de Bruner, na qual não há lugar para a língua enquanto terceiro na relação criança-outro (conforme o desenvolvimento e a conclusão deste artigo). De igual forma, creio que psicólogos têm total legitimidade para criticarem linguistas que, como eu, não tratam da cognição. O que, a meu ver, não seria epistemologicamente válido seria exigir que psicólogos abordassem a língua como linguistas ou, então, que linguistas abordassem a cognição como psicólogos – ou, ainda, que, no interior de cada área, linguistas abordassem a língua todos da mesma maneira e que psicólogos abordassem a cognição todos da mesma maneira. Se assim não fosse, não veríamos os grandes teóricos tecendo críticas entre si, como, dentre outros, Bruner relativamente a Chomsky, Chomsky relativamente a Skinner, Lemos relativamente a Bruner. 2) Discordo do Avaliador 1 não somente em termos epistemológicos de

atribuído à mãe e as mudanças de tal papel, solidárias das mudanças de um tipo de pedido a outro. Três são os tipos descritos: (a) pedido de objeto; (b) pedido de convite; (c) pedido de ação de suporte. Quanto ao primeiro tipo, trata-se de um pedido “elaborado desde uma fase precoce, em que se dirige exclusivamente a um objeto presente, próximo e visível, *frequentemente na posse de um adulto* [...] até formas posteriores, dirigidas a objectos remotos mas visíveis fora do alcance; e finalmente a uma forma destinada a pedir objectos que não estão à vista” (Bruner, 2007 [1983], p. 83). Quanto ao segundo pedido, trata-se de um “pedido para um adulto partilhar uma *relação de papel* numa brincadeira ou num jogo” (Bruner, 2007 [1983], p. 83, *itálico do original*). Quanto ao terceiro tipo, trata-se de um pedido em que “a criança tenta *recrutar a aptidão ou a força de um adulto* para a ajudar a conseguir o objectivo desejado” (Bruner, 2007 [1983], p. 84). Segundo Bruner, o papel da mãe muda conforme o tipo de pedido: no primeiro, ela interpreta qual objeto quer a criança; no segundo, ela interpreta para que é o convite da criança; no terceiro, ela interpreta a ajuda de que necessita a criança. Em todos os tipos, a mãe “é o ‘*agente da cultura*’, por assim dizer, e é ela que impõe as condições de adequabilidade no pedir” (Bruner, 2007 [1983], p. 84).

Desse papel de intérprete da mãe como agente da cultura, podemos tirar duas conclusões. A primeira é a de que, em Bruner, a cultura envolve ação, interação e convenção. A segunda é a de que, em Bruner, não só à cognição, mas também à cultura subordinam-se a linguagem, sua aquisição e sua utilização. Logo, na proposta do autor, tanto o cognitivo quanto o cultural antecedem e condicionam o linguístico.

Contra-argumentos à última conclusão poderiam ser duas alegações. A primeira alegação seria a de que, para Bruner, se há relação de subordinação, é da cultura e da cognição relativamente à linguagem – alegação subjacente a formulações antes mencionadas sobre a linguagem como meio de acesso, interpretação e regulação da cultura, bem como a formulações presentes no Capítulo VI, intitulado “Aprendendo a falar”, a exemplo da seguinte: “Quanto ao *papel constitutivo* da linguagem, o seu papel na ‘criação’ do mundo no qual a criança entra, os formatos de jogo que examinámos são certamente constitutivos, no sentido mais profundo”, pois “Os jogos são, literalmente, produtos do que se diz, como se diz e em que contexto” (Bruner, 2007 [1983], p. 115-116). A segunda alegação seria a de que, para Bruner, a relação da linguagem com a cultura e com a cognição é de contiguidade e não de subordinação – alegação subjacente a outras formulações desse mesmo capítulo final, como a seguinte: “Procurei apresentar uma perspectiva da aquisição da linguagem, que a torna *contígua e dependente* da aquisição pela criança da sua cultura. A cultura é constituída por processos simbólicos, conceitos e distinções, que só podem ser efectuados na linguagem” (Bruner, 2007 [1983], p. 123).

Como se vê, há, no livro de Bruner, margem para as três leituras: (a) a da subordinação da linguagem à cultura e à cognição; (b) a da subordinação da cultura e da cognição à linguagem; (c) a da contiguidade da relação cultura-cognição-linguagem. Todavia, a despeito de formulações pontuais, seu projeto global e seus encaminhamentos teórico-analíticos conver-

diálogo entre teorias, mas igualmente em termos teóricos de interpretação da teoria de Bruner. A despeito de este salientar o caráter assimétrico da relação criança-outro, predomina, em sua proposta, a conversão de tal assimetria em simetria, do que dão testemunho diversos termos e ideias ao longo de seus textos, conforme passagens anteriores e, sobretudo, passagens posteriores deste artigo – termos como *troca referencial*, *suficiente coincidência*, *par dialogante*, *perfeita sintonização*, *papéis intercambiáveis*, *negociação dialogal*, *atenção conjunta*; ideias como as de *cooperação*, *transação*, *mentes semelhantes*, *mútuo reconhecimento de intenções*.

gem muito mais para a primeira leitura do que para as duas outras, do que dão testemunho diversos trechos do capítulo final, a exemplo destes:

A aquisição de linguagem parece ser um subproduto (e um veículo) da transmissão da cultura. As crianças aprendem a usar a linguagem, inicialmente (ou os seus precursores pré-linguísticos) para conseguir o que querem, para jogar jogos, para se manterem em contacto com aqueles de que dependem. Ao fazê-lo, elas encontram limitações que prevalecem na cultura à sua volta, incorporadas nas restrições e convenções dos seus pais. O motor que conduz o processo não é a aquisição da linguagem per se, mas a necessidade de dar seguimento às exigências da cultura. [...] Os pais ajudam-nas num espírito semelhante: eles querem ajudá-las a tornarem-se seres humanos “civilizados”, não apenas locutores de linguagem (Bruner, 2007 [1983], p. 93-94).

Defendi claramente que a “a intenção de referir” não é aprendida, assim como não o é o reconhecimento dessa intenção nos outros. [...] *Deve existir alguma base para a intersubjectividade referencial, antes do aparecimento da linguagem propriamente dita. [...] É uma premissa, que as “outras mentes” são tratadas como semelhantes à nossa mente. Uma outra premissa é que existe um mundo “em volta”, que é partilhado com outros (Bruner, 2007 [1983], p. 111).*

Relativamente à sensibilização para formas gramaticais que correspondem a categorias do conhecimento do mundo real, [...] Esta é basicamente uma questão da *relação entre semântica e sintaxe* e que aponta para o *possível papel “formador” da primeira em relação à última.* (Bruner, 2007 [1983], p. 115).

Um formato é uma interacção rotinizada e repetida, em que um adulto e uma criança *fazem coisas a e com o outro.* Uma vez que *tais formatos emergem antes do discurso léxico-gramatical, eles são veículos cruciais na passagem da comunicação à linguagem.* [...] os formatos inserem as intenções comunicativas da criança numa matriz cultural; eles são instrumentos para transmitir a cultura, bem como a sua linguagem (Bruner, 2007 [1983], p. 121-122).

Tais trechos exibem um último aspecto interessante a ser destacado nesse capítulo final: a vinculação entre linguagem, cultura e convencionalização. “Como é que se inicia o processo de convencionalização?”, indaga-se Bruner, ao que responde: “Proponho que, antes da linguagem propriamente dita entrar em cena, a criança possui maneiras ‘naturais’ de inserir os seus gestos e vocalizações em contextos de acção e interacção”, de forma que “muito do desenvolvimento comunicativo inicial consiste em converter modos naturais de contextualização para padrões convencionais” (Bruner, 2007 [1983], p. 119). Essa conversão é, para o psicólogo, um processo mais civilizatório do que linguístico.

Como já informado, o livro de Bruner aqui examinado atualiza um conjunto de textos publicados na década de 1970, dentre os quais o famoso “The ontogenesis of speech acts” (1975). Tal artigo é alvo de uma detida crítica por parte de M. T. G. de Lemos (2002), a qual mira justamente na visão dual do diálogo que preside as reflexões do psicólogo. Apesar de a autora não se debruçar sobre o livro de 1983, seus argumentos podem ser considerados relativamente a essa obra.

O primeiro argumento ataca o termo *intersubjectividade*, empregado quatro vezes por Bruner no artigo de 1975.⁷ Um desses empregos é o seguinte: “A criança apreende a ‘ideia’ de

⁷ Salvo engano de minha parte, no livro de 1983, consta apenas uma ocorrência do termo em questão, na seguinte citação: “Deve existir alguma base para a *intersubjectividade referencial*, antes do aparecimento da lin-

intersubjetividade – que os outros têm *intenções*” (Bruner, 1975, p. 8). De acordo com M. T. G. de Lemos (2002), “A intersubjetividade de Bruner é, assim, uma função de ‘acordo’ entre os sujeitos falantes. [...] O outro de Bruner é um outro sujeito, isto é, um outro sujeito como eu”, o que dimensiona “o outro e o sujeito como *semelhantes*, daí ser possível a identificação do outro por suas intenções. Isso supõe que a dupla mãe-criança seja constituída por dois lugares *intercambiáveis* e *simétricos*” (M. T. G. de Lemos, 2002, p. 164). Embora o termo *intersubjetividade* não seja frequente no livro de 1983, a visão dual do diálogo por ele subsumida no artigo de 1975 retorna em ocorrências terminológicas como *par dialogante*, *perfeita sintonização*, *papéis intercambiáveis*, *negociação dialogal*, *atenção conjunta*.

O segundo argumento ataca a noção de *formatos* como esquemas interacionais rotineiros entre a mãe e a criança. Conforme M. T. G. de Lemos (2002), o formato é, para Bruner, uma estrutura pré-linguística, preenchida não por elementos significantes ou sígnicos, mas por papéis enquanto instâncias psicológicas ou sociais anteriores à linguagem e representadas por esta, como representáveis seriam as intenções da criança: “A atribuição de intenção à criança é problemática porque a própria existência dos *formats* [...] dá a ver uma assimetria e indica que, antes de mais nada, é necessário que a mãe faça ‘entrar’ a criança num lugar de sujeito do sentido” (M. T. G. de Lemos, 2002, p. 168). Também a qualificação dos formatos como pré-linguísticos é criticável, pois, sendo estruturas em que a mãe interpreta as demandas da criança, “como supor demanda sem linguagem? Ora, a ubiquidade da interpretação da mãe mostra que a linguagem já está em pleno exercício” (M. T. G. de Lemos, 2002, p. 168).⁸

O terceiro argumento ataca o não reconhecimento, por Bruner, do real da língua, entendido como a ordem própria desta e sua anterioridade lógica ao sujeito. Para M. T. G. de Lemos, ao supor segmentos de ação como posições ocupadas em uma sequência por atos variados e substituíveis, a noção de *formato* aproxima-se da de *estrutura*, mais precisamente daquela do estruturalismo americano. Porém, o formato é caracterizado como espelho que reflete relações cognitivas: “Mas como poderia uma estrutura desse tipo refletir relações como tópico-comentário, agente-objeto-recipientes, se é feita de lugares vazios, se é puro movimento de substituição? O que está em ação aqui, a mãe e a criança ou o *format*?” (M. T. G. de Lemos, 2002, p. 167).

guagem propriamente dita” (Bruner, 2007 [1983], p. 111).

⁸ Acerca dessa crítica de M. T. G. de Lemos (2002), o Avaliador 1 ressaltou que, tanto em Bruner quanto na psicologia do desenvolvimento de modo mais geral, o termo *pré-linguístico* “faz referência ao período anterior à produção, pela criança, de palavras ou estruturas reconhecíveis, do ponto de vista da língua (estrutura ou sistema). Não diz respeito, portanto, à ausência de linguagem (ou ao fato de que a linguagem não estaria em ‘pleno exercício’)”. De minha parte, gostaria de chamar a atenção para um ponto mais preciso do argumento de M. T. G. de Lemos (2002): sua crítica incide não sobre a carência de formas léxico-gramaticais por parte da criança, mas sim sobre a qualificação dos formatos como estruturas pré-linguísticas. Ora, uma vez que a mãe – falante constituída – é parte não só integrante, mas também (e sobretudo) constitutiva do *format*, como poderia uma tal estrutura ser pré-linguística se nela reina a língua instanciada nas incessantes interpretações maternas? Ademais, embora reconheça a importância do papel do outro, o próprio Bruner, em muitos trechos, fala como se houvesse um momento em que a linguagem, antes ausente, passaria a se fazer presente, como neste trecho recém-citado: “Proponho que, antes da linguagem propriamente dita entrar em cena, a criança possui maneiras ‘naturais’ de inserir os seus gestos e vocalizações em contextos de ação e interação” (Bruner, 2007 [1983], p. 119). Se se sustenta a relação criança-outro como fundante, não há como supor um momento em que a linguagem estaria “fora de cena”.

O quarto argumento ataca a relação de analogia por Bruner estabelecida entre os planos da intersubjetividade e da linguagem. Segundo M. T. G. de Lemos (2002), ao sustentar que as regras gramaticais são aprendidas por analogia com as regras de ação e de atenção, o estudioso toma os planos da intersubjetividade e da linguagem como “ordens paralelas e, assim, independentes”, um “ponto cego” de sua argumentação que “não só afeta como acaba por anular a importância que ele pretendia atribuir à intersubjetividade, na aquisição da linguagem pela criança” (M. T. G. de Lemos, 2002, p. 169). Aqui, porém, tenho um entendimento distinto do da autora: a meu ver, tal relação analógica não envolve paralelismo e independência, mas, ao contrário, anterioridade e dependência. Anterioridade da intersubjetividade, em termos de precedência das regras de ação e de atenção conjuntas – a cultura e a cognição sendo construídas conjuntamente. Dependência da linguagem, em termos de subordinação das regras gramaticais – o linguístico sendo antecedido e condicionado pelo cultural e pelo cognitivo.

A despeito das críticas, M. T. G. de Lemos (2002) reconhece duas contribuições de Bruner aos estudos em aquisição de linguagem. A primeira contribuição envolve a introdução do outro nesses estudos, o que o distingue de muitos pesquisadores “que se valiam da obra de Piaget, na qual a atividade tinha um lugar de destaque, mas era compreendida apenas como interação com o ‘meio’, e não com o outro” (M. T. G. de Lemos, 2002, p. 163). A segunda contribuição envolve o fato de que Bruner “fez dessubjetivar a língua, deslocando-a de dentro do sujeito para sua exterioridade. Exterioridade pouco definível, ainda que suspostamente ‘exercida’ pelo outro materno”, ao que acrescenta a autora: “O que interessa, entretanto, é que foi Bruner quem encontrou os termos necessários para realmente subverter a demanda, ainda em vigor, de escrever uma gramática da criança: nada mais podia alimentá-la depois de reconhecido que a língua, afinal, é do Outro” (M. T. G. de Lemos, 2002, p. 170). Contudo, tais contribuições de Bruner são ameaçadas pelo fato de que sua teoria

tem todos os elementos necessários para apagar sua própria descoberta e, assim, não realizar esse reconhecimento [da anterioridade e da exterioridade da língua]. E isso não apenas por causa da analogia, pois se esta reduz a linguagem a um espelho, a língua ressurgiu, triunfante, em outro lugar, no *format*. O problema maior é que o *format* de Bruner não poderia revelar a alteridade da ordem da língua, pois ele é dominado por essa convergência, essa afinação dos sujeitos na intersubjetividade. Na sua identificação com o Outro materno, Bruner se deixou fascinar pela complementariedade do olhar [entre mãe e bebê] e não pôde nomear o que, no entanto, percebeu brilhantemente: que mãe e criança não fazem dois, mas Um. (M. T. G. de Lemos, 2002, p. 170-171).

Da visão dual do diálogo, representada por Jerome Bruner, com seus conceitos de *formato* e de *intersubjetividade*, os estudos de Cláudia de Lemos se aproximam e se afastam – inicialmente, com seu conceito de *processo dialógico*; posteriormente, com sua visão triádica do diálogo.

Em texto de 1982, intitulado “Sobre a aquisição de linguagem e seu dilema (pecado) original”, Lemos introduz dois dos processos que, a seu ver, dariam conta tanto das relações entre o enunciado da criança e o enunciado do adulto quanto das relações entre elementos combinados em um único enunciado ou turno do diálogo. O primeiro é “o processo de especularidade ou de incorporação pela criança de parte ou de todo o enunciado adulto no nível segmental” (Lemos, 1982, p. 113). O segundo é o processo de complementaridade, o qual se instancia de duas maneiras: como “complementaridade interturnos, em que a resposta da

criança preenche um lugar ‘semântico’, ‘sintático’ e ‘pragmático’ instaurado pelo enunciado imediatamente precedente do adulto”; e como “complementaridade intraturnos, em que o enunciado da criança resulta da incorporação de parte do enunciado adulto imediatamente precedente [...] e de sua combinação com um vocábulo complementar” (Lemos, 1982, p. 113). Se, inicialmente, a instanciação desses processos dá a ver uma dependência da criança relativamente à fala do adulto e à situação imediata, posteriormente, os processos passam a ser instanciados com relativa independência de tais fatores, quando construções linguísticas passam a ser “decontextualizadas” e utilizadas em outras situações: “[...] a capacidade que a criança vai adquirindo de combinar vocábulos e fragmentos do discurso adulto independentemente do enunciado do interlocutor e da situação, *depende*, em grande parte, do desenvolvimento de sua *capacidade de representar as intenções, a atenção e o conhecimento do seu interlocutor*” (Lemos, 1982, p. 113). Tal reflexão conduz Lemos (1982) a concluir que “é através da linguagem enquanto AÇÃO SOBRE O OUTRO (ou procedimento comunicativo) e enquanto AÇÃO SOBRE O MUNDO (ou procedimento cognitivo) que a criança constrói a linguagem enquanto OBJETO sobre o qual vai poder operar” (Lemos, 1982, p. 119-120).

Nesse artigo, chamam a atenção dois pontos. O primeiro ponto envolve as definições de *especularidade* e de *complementaridade*, as quais enfatizam a ação da criança de espelhar e de complementar a fala da mãe em um movimento que pareceria unidirecional. O segundo ponto envolve a dependência da linguagem e de sua “construção” como “objeto” (capacidade de combinação de vocábulos e de fragmentos alheios) relativamente à comunicação e à cognição (capacidade de representação das intenções, da atenção e do conhecimento do outro).

Em texto de 1985, intitulado “On specularity as a constitutive process in dialogue and language acquisition”, esses dois pontos também se destacam, mas com diferenças. Quanto ao primeiro ponto, os processos de especularidade e de complementaridade são redefinidos por Cláudia de Lemos no sentido de mostrar que a mãe igualmente espelha e complementa a fala da criança: “[...] ambos os parceiros preenchem seus turnos no diálogo incorporando pelo menos parte do enunciado anterior do outro”, de modo que “a coesão e a progressão do diálogo são os resultados dessa *mútua incorporação*” (Lemos, 1985, p. 25). Quanto ao segundo ponto, a autora introduz um terceiro processo dialógico: trata-se da reciprocidade, concebida como “*reversibilidade de papéis* no que diz respeito ao diálogo adulto-criança, visto que ele [tal processo] se refere à gradual assunção, por parte da criança, do papel previamente assumido pelo adulto”, o papel de iniciar “a interação, constituindo o Outro como interlocutor ou como aquele que deve ocupar o próximo turno, *atribuindo intenções, conhecimento e crenças*” (Lemos, 1985, p. 26). Tanto a redefinição da especularidade e da complementaridade quanto a introdução da reciprocidade levam Lemos a aproximar-se da noção de *imitação recíproca* de James Baldwin como relida por Jean Piaget: “[...] A *imitação recíproca*, isto é, precisamente o reflexo (no sentido próprio) de si em outrem e de outrem em si” (Piaget, 1928, p. 168 *apud* Lemos, 1985, p. 28).

Mútua incorporação, reversibilidade de papéis, atribuição de intenções, conhecimento e crenças, imitação recíproca: termos que dizem do alinhamento de Cláudia de Lemos, nesse período, a estudos como os de Jerome Bruner, centrados na simetria criança-outro e na intercambiabilidade de seus lugares. Deve-se pontuar, entretanto, uma diferença significativa já entre os estudos iniciais da linguista e aqueles do psicólogo: a linguagem, embora vista como dependente da comunicação e da cognição nesses estudos iniciais, tem neles mais espaço tanto na reflexão teórica quanto – e sobretudo – na análise de dados.

Em texto de 1986, intitulado “A sintaxe no espelho”, a autora reflete sobre a relação entre os processos dialógicos e a emergência de estruturas sintáticas. Nessa reflexão, confronta o termo *aquisição* (que designa os atos/gestos da criança em direção ao objeto a ser adquirido) e o termo *input* (que designa o fornecimento de dados linguísticos primários para processamento pela criança), observando que ambos os termos supõem uma capacidade prévia de análise por parte da criança e reduzem o papel do adulto ao de “apresentador do objeto, ou de demonstrador de sua utilidade” (Lemos, 1986b, p. 9). Como alternativa a tais suposição e redução, Lemos (1986b, p. 10) afirma que não basta “recorrer ao diálogo enquanto recorte empírico nem à *comunicação* enquanto processo, que nele se dá. Em vez disso, seria necessário recuperá-lo enquanto *interação* e na sua *oralidade*”. Recuperar o diálogo enquanto interação é recuperá-lo “enquanto *inter-ação* ou *atividade conjugada*” ou, ainda, “enquanto *matriz de significação*”: “No diálogo adulto-criança é apenas mais transparente essa *atividade inter-subjectiva* (ou *trans-subjectiva*?) pela qual é regulada a produção de significação ou as formas sob as quais ela se dá enquanto tensão entre linguagem e mundo” (Lemos, 1986b, p. 10). Recuperar o diálogo na sua oralidade é “atender ao próprio ato de dar forma ao som em sua materialidade e à tensão entre o ‘continuum’ e o discreto que existe nessa materialização”, o que põe o investigador “diante de *formas cujos contornos não são tão definidos* como da palavra-objeto da escrita” (Lemos, 1986b, p. 10). Analisando dados à luz dessa noção de *diálogo*, a autora mostra um percurso que parte dos processos dialógicos, passa por processos de análise e de síntese de objetos linguísticos e chega à formação de “esquemas abstratos”, percurso que “fala por si próprio a favor da substituição da metáfora da aquisição pela metáfora da construção” (Lemos, 1986b, p. 12). Para a pesquisadora, porém, a construção da linguagem não tem, como requisito, uma capacidade analítica da criança, pois mesmo o processo de especularidade é inicialmente instanciado pela mãe, a qual atribui sentido ao fluxo comportamental da criança, possibilitando que, mais tarde, esta realize tais espelhamento e interpretação: “No espelho do Outro, convivem dialeticamente a fusão e a diferenciação, como Wallon (1942) fez muita questão de ressaltar”, sendo nesse espelho possível “subjetivar-se, objetivando-se no Outro” (Lemos, 1986b, p. 207).

Desse estudo, ressalto quatro aspectos. O primeiro aspecto concerne à linguagem, a qual assume ainda mais protagonismo teórico e analítico: é uma linguista que vemos aqui falando, ainda que permaneça a interlocução com a psicologia do desenvolvimento. O segundo aspecto concerne à relação especular criança-adulto: se, no texto de 1982, a criança espelhava a fala do interlocutor e se, no texto de 1985, ambos espelhavam a fala um do outro, nesse texto de 1986, o espelhamento da criança pelo outro é condição do posterior espelhamento do outro pela criança. O terceiro aspecto concerne à natureza da constituição da criança enquanto sujeito e, por conseguinte, da relação da linguagem com a comunicação e com a cognição: se, no texto de 1982, tal constituição era vista como construção da linguagem enquanto objeto subordinado à comunicação e à cognição, nesse texto de 1986, tal constituição é vista como construção de “esquemas abstratos” cuja gênese se encontra nos processos dialógicos e cujo desenvolvimento se encontra nos processos de análise e de síntese dos objetos linguísticos; ou seja, se a comunicação permanece no ponto de partida, a cognição é deslocada para o ponto de chegada, como produto da linguagem. O quarto aspecto concerne ao estatuto dessa comunicação, redefinida como *inter-ação*, *atividade conjugada*, *matriz de significação*, termos que remetem à *intersubjetividade* de Bruner, expressão que igualmente

aparece, mas para ser questionada: *atividade inter-subjetiva* ou *trans-subjetiva*?, pergunta-se Lemos, sugerindo a presença de algo a mais na relação criança-adulto.

Para M. T. G. de Lemos (2002), se o conceito de *processo dialógico* teve o mérito de focalizar as relações não entre sujeitos – interações adulto-criança –, mas entre estruturas linguísticas – interações tanto entre fala materna e fala infantil quanto entre elementos linguísticos no interior da fala infantil –, tal conceito foi enfraquecido por duas fragilidades. A primeira fragilidade, de caráter metodológico, concerne à insuficiência do conceito de *processo dialógico* na análise de dados mais tardios, nos quais a fala do outro não determina do mesmo modo a fala da criança. Conforme M. T. G. de Lemos (2002, p. 201), “A relação entre esses processos [...] com a ordem própria da língua – isto é, sua Alteridade – é que fica por realizar e, nessa ausência, a porta ficou aberta para os reducionismos que anulam, retroativamente, a própria posição que se instaurara” (M. T. G. de Lemos, 2002, p. 201). A autora refere-se, aqui, à diferença criada pelo conceito em questão entre os estudos de Cláudia de Lemos e outros do campo da aquisição de linguagem – diferença que, para ser sustentada, obrigaria a que se respondesse mais positivamente pelo linguístico. A segunda fragilidade, de caráter teórico, concerne à ideia de *reciprocidade*, que aproximaria Cláudia de Lemos de Jerome Bruner, pois “não é outra coisa que a intersubjetividade”, a qual “é a porta por onde o sujeito psicológico entra”, sendo “incompatível com o caráter constitutivo da especularidade no diálogo (isto é, na criança e na mãe)” (M. T. G. de Lemos, 2002, p. 201-202). Essa incompatibilidade deve-se ao fato de a reciprocidade associar-se mais à noção piagetiana de *descentramento* (que supõe um sujeito no controle de sua atividade linguístico-cognitiva) e a especularidade, mais à noção lacaniana de *alienação* (que supõe um sujeito que, mais do que falar a língua, é falado pela língua). Tal exame crítico conduz M. T. G. de Lemos (2002, p. 211) à seguinte conclusão:

Essa discussão nos permite colocar uma questão relativa ao modo como o conceito de processo dialógico deve ser entendido se, de fato, a hipótese sociointeracionista assume aquilo que está na base de sua descoberta: a ligação entre outro e ordem da língua. Se os processos dialógicos são constitutivos – como indica Lemos – então *não se trata mais de falar de criança e adulto tão simplesmente – mas de uma topologia que inclui a ordem da língua*. Isso significa assumir, a nosso ver, uma interpretação que chamaríamos estrutural do processo dialógico, em oposição a uma interpretação dinâmica ou mesmo funcional, que acreditamos ter sido a dominante na literatura sociointeracionista. Apenas desse modo se poderia sair de um impasse, e não repetir mais uma vez o sintoma da área dos estudos em aquisição de linguagem, pois essa interpretação estrutural possibilita ir mais além dos efeitos imaginários da relação com o outro, *ao incluir a ordem da língua como elemento assimetrizável na dupla*. Isso significa a queda da intersubjetividade na teoria.⁹

Publicado como livro em 2002, o trabalho de Maria Tereza Guimarães de Lemos foi originalmente realizado como tese de doutoramento, defendida em 1994. Em seu traba-

⁹ Com efeito, essa queda é atestada pela quase completa ausência do termo *intersubjetividade* em estudos mais recentes vinculados à proposta de Cláudia de Lemos. Digo “quase completa ausência”, pois o termo figura, por exemplo, no estudo de Figueira (2022), no qual igualmente comparece o termo *relações intersubjetivas*. É preciso, no entanto, registrar que tais ocorrências terminológicas devem-se ao recurso da autora, nesse trabalho, às contribuições de Émile Benveniste e de Oswald Ducrot, linguistas em cujos textos os termos *intersubjetividade* e *relações intersubjetivas* são empregados com acepções que em nada se aproximam daquelas que têm em estudos cognitivistas como os de Jerome Bruner. A esse propósito, cabe pontuar que o psicólogo estadunidense

lho, a autora investiga os primeiros estudos de Cláudia de Lemos, publicados nos anos 1980. Entretanto, como já antecipamos, nos anos 1990 e 2000, essa estudiosa rompe com uma visão dual do diálogo e assume uma visão triádica deste, conforme discussão a seguir.¹⁰

4 A visão triádica do diálogo na aquisição de linguagem

Diferentemente de Jerome Bruner, Cláudia de Lemos não publicou um livro em que apresentasse suas principais ideias sobre a aquisição de linguagem. Assim, qualquer um que se proponha a examinar, no curso de sua teorização, a trajetória de uma noção ou de um conceito, precisa mapear tal noção/conceito em diferentes escritos seus, publicados em distintos lugares, ao longo de diversos anos. É o nosso caso, aqui, relativamente à noção de *diálogo/ interação/ alteridade*. Na seção anterior, examinamos o tratamento dessa noção em trabalhos da autora da década de 1980. Nesta seção, examinaremos o modo como tal noção é tratada em estudos seus das décadas de 1990 e de 2000.

Em texto de 1992, intitulado “Os processos metafóricos e metonímicos como mecanismos de mudança”, Lemos revisita a reflexão de Saussure sobre as relações associativas e sintagmáticas enquanto eixos da língua como sistema, assim como a reflexão de Jakobson sobre os processos metafóricos e metonímicos enquanto leis de composição interna da linguagem e polos de seu funcionamento na fala individual. Propondo o processo metafórico (a substituição, na cadeia significante, de um termo por outro) e o processo metonímico (a combinação, na cadeia significante, entre os termos e suas posições) como mecanismos de mudança na aquisição de linguagem, a autora define o outro “enquanto discurso ou instância de funcionamento da língua constituída”, cuja “*interpretação* consiste em submeter os significantes da criança a processos metafóricos e metonímicos”, os quais produzem, como efeito, uma “*ressignificação* através de relações com outros significantes” (Lemos, 1998a [1992], p. 161). Essa resignificação como efeito interpretativo acentua o papel estruturante do outro e do texto na aquisição de linguagem, pois os processos metafóricos e metonímicos “Emergem no interior de textos pelos quais o adulto interpreta o comportamento das crianças dentro de um particular domínio discursivo” (Lemos, 1998a [1992], p. 195). Ou ainda:

Na aquisição inicial da linguagem, o enunciado da criança é ouvido e resignificado pelo enunciado do adulto, já que seus significantes [os da criança] são formas isoladas, independentes, cuja significação não vem dada por sua posição numa estrutura oracional ou textual. É só na medida em que os processos metafóricos e metonímicos se cristalizam em redes de relações que a criança passa a ouvir/ressignificar seus próprios enunciados e, além da posição de interpretado [sic], pode assumir a

cita Benveniste e Ducrot ao falar em intersubjetividade nos textos “Los formatos de la adquisicion des lenguaje” (1982) e “Pragmatica des lenguaje y lenguaje de la pragmática” (1984), ambos presentes no livro *Ação, pensamento e linguagem* (2007 [1984]). Trata-se, contudo, de uma leitura psicologizante – e, por isso mesmo, comprometida em sua gênese – de formulações de Benveniste e de Ducrot, linguistas sempre críticos ao psicologismo (cfe., dentre outros, Benveniste, 2005 [1966], 2006 [1974]; Ducrot, 1987 [1984], 1990).

¹⁰ Em Lemos (2002), a autora apresenta uma detida revisão das mudanças operadas em sua proposta teórica. Devido aos limites do presente artigo, ressaltarei apenas as implicações dessas mudanças em suas concepções de diálogo/interação/alteridade.

de intérprete de si mesmo [sic] e de outro. As chamadas autocorrekções são sintomas dessa mudança de posição.

A partir desse ponto de vista, a “experiência com a própria linguagem” implica uma interação, mas ao mesmo tempo é *incompatível* com uma noção de comunicação que se assente no acesso direto às intenções e significados do outro e de si mesmo. (Lemos, 1998a [1992], p. 167).

Nesse texto, sublinho cinco deslocamentos, um epistemológico e quatro teóricos. O deslocamento epistemológico consiste no distanciamento de autores da psicologia do desenvolvimento e na filiação ao programa estruturalista. O primeiro deslocamento teórico consiste na substituição do conceito de *processo dialógico* pelo conceito de *processos metafóricos e metonímicos* na descrição e na explicação da trajetória de *infans* a falante. O segundo deslocamento teórico consiste na qualificação dessa trajetória não mais como construção de “esquemas abstratos” a partir de processos dialógicos e analítico-sintéticos, mas agora como passagem de interpretada pelo outro a intérprete de si mesma e do outro. O terceiro deslocamento teórico consiste na redefinição da noção de *interpretação*, vista não mais como atribuição ao outro de intenções, conhecimentos e crenças, mas agora como escuta e fala que ressignificam o enunciado ouvido, ressignificação em termos seja de estruturação (quando o outro insere as formas isoladas da criança em uma cadeia), seja de autocorreção (quando a criança passa a se corrigir). O quarto deslocamento teórico consiste na definição do interlocutor da criança (antes chamado principalmente de *adulto*) como *discurso* ou *instância de funcionamento da língua constituída*, definição que afasta uma perspectiva empirista da interação, em que o outro é um outro indivíduo empírico como a criança.

Em texto de 1995, intitulado “Língua e discurso na teorização sobre aquisição de linguagem”, além das linguísticas saussuriana e jakobsoniana, Cláudia de Lemos aproxima-se da linguística benvenistiana, da análise do discurso peucheutiana e da psicanálise lacaniana, abordando as restrições linguístico-discursivas que se impõem à criança e dando novos contornos ao *alter* na aquisição. Acerca das restrições, trata-se, por um lado, de restrições mais estritamente linguísticas que cada segmento do texto impõe aos segmentos que o sucedem e que o precedem e, por outro lado, de restrições que “o próprio texto enquanto *figuração de uma totalidade*” impõe às suas partes, restrições que “não deixam de se fazer sentir também no diálogo que, nesse e em outros sentidos, *funciona como um texto*”: “Até mesmo a criança pequena, em seus diálogos iniciais com o adulto, parece sensível a essas restrições: ela não diz nem faz qualquer coisa em resposta a seu interlocutor ou a determinada situação” (Lemos, 1995, p. 19). Acerca dos novos contornos do *alter*, trata-se tanto de um aprofundamento da definição do outro (com inicial minúscula) quanto da introdução do Outro (com inicial maiúscula). Além de reiterar a definição de 1992 do outro “enquanto instância de funcionamento linguístico-discursivo”, a linguista acrescenta que, “Assim definido, o outro não é mais tomado enquanto individualidade, [...] passando a ser visto a partir da *posição subjetiva*, efeito de funcionamento linguístico-discursivo, que lhe permite *interpretar* a criança. A saber, colocar a criança – sua fala, gesto, olhar, movimento – num texto” (Lemos, 1995, p. 24). Segundo a autora, tal interpretação fornece uma ancoragem textual-discursiva ao enunciado da criança, restringindo sua indeterminação de forma e de sentido, o que lhe confere um estatuto fonológico, morfológico e sintático e, conseqüentemente, contém a sua deriva. Já o Outro é introduzido ao final do artigo, em que ecos da afetação de Cláudia de Lemos pela teoria psicanalítica reverberam em dois momentos. Primeiramente, quando a pesquisadora observa que estudos aquisicionais que excluem questões textuais e discursivas e que adotam a noção de *desenvolvimento* situam

“a criança como um sujeito diante da língua como objeto, impedindo a apreensão do *processo de aquisição como um processo de subjetivação*, ainda que votado à heterogeneidade e à divisão” (Lemos, 1995, p. 27). Em seguida, quando a estudiosa salienta que, nas análises apresentadas nesse texto, buscou mostrar “como a criança pode sair da posição de interpretada pela fala do outro, atuante em sua própria fala, para a posição em que é *a língua, enquanto Outro*, que a desloca e ressignifica” (Lemos, 1995, p. 27).

Quatro são as observações suscitadas pelo artigo em questão. Em primeiro lugar, se o trabalho de 1992 mobilizava uma noção empírica de *texto*, tal estudo de 1995 assume uma noção mais teórica, definindo texto como *figuração de uma totalidade* e associando seu funcionamento àquele do diálogo como espaço em que vigem as restrições linguístico-discursivas. Em segundo lugar, o deslocamento de uma noção psicológica a uma noção linguística de *interpretação* é ainda mais acentuado, com o ato interpretativo do outro sendo concebido como âncora linguístico-discursiva que restringe a indeterminação formal e semântica da fala da criança. Em terceiro lugar, a aquisição de linguagem, qualificada no texto anterior como passagem da criança de interpretada pelo outro a intérprete de si mesma e do outro, é nesse texto qualificada, pela primeira vez (embora sem maiores detalhes), como *processo de subjetivação*. Em quarto lugar, também a língua é aí qualificada, pela primeira vez (apesar de igualmente sem detalhamento), como *Outro*, com inicial maiúscula.¹¹

Em texto de 1997, intitulado “Native speaker’s intuitions and metalinguistic abilities: what do they have in common from the point of view of language acquisition?”, Cláudia de Lemos responde à pergunta do título sustentando que, implícito à noção gerativista de *intuição do falante nativo* e à noção psicolinguística de *habilidades metalinguísticas*, está o pressuposto da mudança de posição da criança em relação à língua. Se, no caso da noção gerativista, a mudança de posição diz respeito à transição de um estado zero de conhecimento a um estado estável,¹² no caso da noção psicolinguística, a mudança de posição diz respeito à transição de um uso inconsciente da língua a um uso consciente. A ideia de *mudança de posição* é cara à estudiosa, a qual sustenta, contudo, uma visão alternativa. De fato, é nesse texto de 1997 que Lemos esboça a sua proposta de três posições: (a) a *primeira posição*, em que a fala da criança incorpora fragmentos da fala do outro, daí porque as formas aparentemente corretas que apresenta; (b) a *segunda posição*, em que a fala da criança se distancia da fala do outro, submetendo-se ao funcionamento da língua e irrompendo em erros tanto previsíveis quanto imprevisíveis; (c) a *terceira posição*, em que a fala da criança se torna mais estável e homogênea, voltando a convergir com a fala do outro. Tais estabilidade e homogeneidade devem-se à possibilidade de a criança interpretar sua própria fala e corrigir-se, mas não podem ser referidas à ideia de *consciência metalinguística*, pois esta não implica “a suposição de que a aquisição

¹¹ Nos textos da autora da década de 1980, o termo “Outro”, com inicial maiúscula, também aparece, mas não relacionado à língua nem revestido de um estatuto teórico mais aparente.

¹² Para reflexões mais detalhadas sobre a leitura de ideias chomskianas por Cláudia de Lemos, ver, além de Lemos (1997), Lemos (1998c, 1999b, 2008, 2014). Mais especificamente acerca da ideia de passagem de um estado zero de conhecimento a um estado estável, pontua a autora em texto de 2014: “Tanto investigadores da aquisição de uma mesma língua como investigadores que se reuniram para comparar dados da aquisição de línguas diferentes se viram diante de falas que exibiam, ao contrário do esperado, uma forte resistência à sistematização. Tentativas de superá-la resultaram em uma verdadeira deriva teorizante e, por fim, no balde de água fria representado pela posição dissonante de Chomsky ao afirmar que os estados (não estágios) intermediários entre o estado 0 e o estado estável eram irrelevantes do ponto de vista linguístico (1986, p. 24-25)” (Lemos, 2014, p. 959).

da linguagem é um processo de subjetivação *cujo produto é o falante sendo dividido* nas posições de ser interpretado e de ser o intérprete”, ao que se acrescenta a linguista: “Se se pode dizer que *a criança é primeiramente interpretada pelo outro/adulto e depois pela língua*, sua divisão como falante mostra que ela permanece na posição de ser interpretada pela língua, da qual o intérprete (o outro e ela mesma) é um reflexo” (Lemos, 1997, p. 12). Conforme a autora, assumir essa proposta teórica impossibilita tomar “a interação adulto-criança como uma troca comunicativa envolvendo (pelo menos dois) indivíduos homogêneos cujas intenções e significados são acessíveis e transparentes um ao outro”, devendo-se atribuir a tal interação o papel de “*lôcus do funcionamento linguístico-discursivo* onde as relações estruturantes ocorrem e o *infans* inicia sua jornada como falante” (Lemos, 1997, p. 12-13).

Tal artigo dá a ver uma reiteração e quatro avanços. A reiteração diz respeito à crítica à perspectiva empirista da interação adulto-criança, definida pela transparência de suas intenções e de seus significados. O primeiro avanço diz respeito ao delineamento de uma explicação para o vir a ser falante que, representada pelas três posições, é teoricamente mais robusta do que aquela representada, inicialmente, pelos processos dialógicos e do que aquela representada, posteriormente, pelos processos metafóricos e metonímicos. O segundo avanço diz respeito à concepção de aquisição de linguagem como processo de subjetivação, ao qual é agora atribuído um produto: a divisão do falante entre as posições de interpretado e de intérprete. O terceiro avanço diz respeito justamente ao papel de intérprete, antes remetido apenas ao adulto e à criança, mas agora estendido à língua. O quarto avanço diz respeito à noção de *diálogo/interação*: anteriormente definido como *unidade de análise, inter-ação, atividade conjugada, matriz de significação, atividade inter- ou trans-subjetiva*, ele passa a ser definido como *lôcus do funcionamento linguístico-discursivo*. Esse último avanço é interessante por duas razões: (a) o termo *lôcus* (bem como o termo *posição*) parece ir ao encontro da reivindicação de M. T. G. de Lemos (2002) de uma interpretação estrutural do diálogo como topologia que inclua a ordem da língua; (b) assim definida, a noção de *diálogo/interação* remete à definição do outro como *instância de funcionamento linguístico-discursivo* – seria, então, possível pensar que, para Lemos, o diálogo como *lôcus* comporta instâncias como *posições*?

Em texto de 1998, intitulado “Sobre a aquisição da escrita: algumas questões”, um dos raros momentos nos quais se volta para o vir a ser escrevente, Cláudia de Lemos define “a interação como *lugar de transformação*” e a transformação como “*mudança* que se opera através do funcionamento simbólico” (Lemos, 1998b, p. 21). Propondo-se a “discernir o outro da/na aquisição da escrita” (Lemos, 1998b, p. 28), a autora argumenta que, assim como fragmentos da fala do adulto retornam na fala da criança, aspectos gráficos de textos escritos retornam em sua escrita inicial, sendo tais fragmentos e aspectos significantes de situações discursivas anteriores que, entrando em relação com fragmentos e aspectos que significam outras situações, podem ressignificar-se para a criança. Tal processo de ressignificação incidiria não sobre unidades isoladas como palavras e sílabas, letras e fonemas, mas sobre cadeias de textos-discursos em que tais unidades se tornam significantes. Nesse processo, “o papel do outro seria, como na aquisição da linguagem oral, o de *intérprete*”, ao que acrescenta: “Lendo para a criança, interrogando a criança sobre o sentido do que ‘escreveu’, escrevendo para a criança ler, o alfabetizado, como *outro que se oferece ao mesmo tempo como semelhante e como diferente*, insere-a no movimento lingüístico-discursivo da escrita” (Lemos, 1998b, p. 28). Adverte a autora, no entanto, que “essa interpretação *não tem origem no adulto* mas no discurso em que ele próprio, submetido ao funcionamento lingüístico-discursivo, é significado” (Lemos, 1998b, p. 27).

Duas observações podem ser feitas a partir desse texto. A primeira observação refere-se às definições de *interação* e de *transformação*: a relação criança-outro passa a ser sede não só do funcionamento simbólico (linguístico-discursivo), mas também da mudança por ele operada. A segunda observação refere-se a nuances de que se reveste o papel de intérprete atribuído ao outro, cuja interpretação é marcada por semelhanças e por diferenças relativamente à criança, bem como é vista enquanto efeito do funcionamento linguístico-discursivo que submete o próprio adulto, significando-o e o levando-o a significar a criança. Se, na publicação de 1997, à língua era atribuído o papel de intérprete da criança e do adulto, nessa publicação de 1998b, tal papel é igualmente atribuído ao *discurso* – termo não definido em nenhuma das duas publicações, mas em ambas relacionado ao termo *texto* e à análise do discurso peuceutiana.

Em texto de 1999, intitulado “Sobre o interacionismo”, Cláudia de Lemos relata que, inicialmente, buscou na psicologia um lugar para o outro na aquisição de linguagem: “Bruner era, naquele momento, quem conduzia essa linha de pesquisa, já cunhada então de interacionista”, por privilegiar “os ‘formats’, em que se podia emoldurar o que se apresentava à primeira vista como regularidades enunciativas da relação mãe-criança” (Lemos, 1999a, p. 13). Todavia, pontua a autora que ela sempre colocou em primeiro plano o linguístico e não o comunicativo, o que exigiu recortar o diálogo como unidade de análise (e não de comunicação) e interrogar-se “sobre o efeito da linguagem, através da fala do outro, na constituição da criança como sujeito” ou, ainda, sobre a “passagem da fala da criança enquanto fala não-analisada do outro para uma fala articulada, no sentido em que uma fala articulada supõe certos termos e posições que os determinam” (Lemos, 1999a, p. 13). Se o rótulo *interacionista* nunca foi abandonado, para diferenciar a sua proposta das demais assim rotuladas, “foi por causa do ‘outro’”, o qual, “ainda que redefinido, *despojado do saber que lhe era suposto como intérprete da criança*, ali permanecia como lugar primordial de inserção da criança no funcionamento da língua” (Lemos, 1999a, p. 12). Ao fim do artigo, Lemos questiona-se: “O que revela [...] essa fala que, determinada pela língua, é indeterminada enquanto ação e enquanto ato, essa fala que, tendo na língua sua condição de possibilidade, com ela não coincide?”, ao que responde: “Talvez ela não revele mas apenas aponte para uma *instância subjetiva* que está aquém ou além do que se pode saber sobre a língua. É essa *instância em constituição* que se faz ver na fala da criança” (Lemos, 1999a, p. 16).

Desse texto, vale ressaltar a insistência de Cláudia de Lemos em distanciar-se dos estudos filiados às teorias psicológicas, do que dá testemunho a sua não adesão à noção de *comunicação* enquanto relação entre semelhantes e às implicações dessa noção, como a subordinação do linguístico ao comunicativo e a suposição do outro como intérprete das intenções da criança¹³. Vale ressaltar, ainda, a definição da criança como *instância subjetiva em constituição*, o que me permite responder positivamente à pergunta que antes fiz sobre a possibilidade de se entender, no pensamento da autora, o diálogo como *lócus* que comporta instâncias

¹³ Essa noção de *comunicação* da qual se afasta Cláudia de Lemos já havia sido contestada em dois artigos anteriores. O primeiro texto é o de 1998a (1992) sobre os processos metafóricos e metonímicos, no qual a autora afirma que “a ‘experiência com a própria linguagem’ implica uma interação, mas ao mesmo tempo é *incompatível* com uma noção de comunicação que se assente no acesso direto às intenções e significados do outro e de si mesmo” (Lemos, 1998a [1992], p. 167). O segundo texto é o de 1997a, no qual a linguista defende a impossibilidade de se tomar “a interação adulto-criança como uma troca comunicativa envolvendo (pelo menos dois) indivíduos homogêneos cujas intenções e significados são acessíveis e transparentes um ao outro” (Lemos, 1997a, p. 12-13).

como *posições*: na aquisição de linguagem, se o outro se posiciona, no diálogo, como instância de funcionamento da língua constituída, a criança é neste posicionada como instância de funcionamento da língua em constituição.

Em texto de 2000, intitulado “Desenvolvimento da linguagem e processo de subjetivação”, a estudiosa novamente reforça seu distanciamento da noção de *comunicação* e da noção de *desenvolvimento*. Sobre a primeira noção, ao tratar do retorno da fala materna na fala infantil, Cláudia de Lemos afirma que, nesse retorno, à aparente coincidência entre os significantes da mãe e os da criança, opõe-se “a não-coincidência no nível do significado”, o que “mostra a impossibilidade de tratar a relação da criança com a fala do outro como comunicação, como *relação entre semelhantes*” (Lemos, 2000, p. 61). Sobre a segunda noção, ao questionar a concepção de aquisição de linguagem como aquisição de um conhecimento sobre a língua e o pressuposto de que tal conhecimento implica desenvolvimento, a linguista defende que “Falar em processo de subjetivação significa colocar a *anterioridade lógica da linguagem* relativamente a um *corpo pulsional* que é por ela *capturado e significado*” (Lemos, 2000, p. 55). Ainda nesse artigo, Lemos chama a atenção para a alternância entre duas alteridades quando significantes da mãe evocam, na criança, textos ausentes, de outras situações, como quando o significante *telefone* na fala da mãe evoca o significante *alô* na fala da criança: “[...] à *presença efetiva do outro*, de seu corpo e de sua voz, do qual depende cada enunciado da criança e a progressão do diálogo, se alterna o *outro enquanto texto*, fragmento de um texto ausente”, em “uma relação entre significantes que resulta na obliteração dessa voz, desse corpo, do momento e do lugar de sua instanciamento” (Lemos, 2000, p. 62). Essa alternância dá a ver uma primeira forma de separação, que “aponta para uma *dialética de alienação e separação* relativamente à fala do outro ou ao outro enquanto instância de funcionamento lingüístico-discursivo” (Lemos, 2000, p. 59). Uma segunda forma de separação ocorre em relação não só ao outro, mas também à língua, com a divisão do sujeito “entre a instância que fala e a que escuta, *instâncias essas que não coincidem* nem na criança nem no adulto, conforme nos revelou Freud através do chiste (1905/1988) e do lapso (1901/1997)” (Lemos, 2000, p. 69).

Tal artigo merece cinco comentários. Primeiro comentário: a relação criança-outro se define em termos não de relação entre semelhantes, mas de relação entre semelhantes e diferentes, em uma dialética de *alienação* e de *separação* – termos lacanianos que dizem do diálogo cada vez mais estreito de Cláudia de Lemos com a psicanálise. Segundo comentário: as duas formas de separação têm a ver com duas alteridades – o outro como outro corpo/voz e o outro como outro texto/fragmento, o que remete, respectivamente, à alteridade com o interlocutor e à alteridade com a língua. Terceiro comentário: a primeira forma de separação concerne à mudança da primeira para a segunda posição, em que a fala da criança se afasta da dominância da fala do outro e dos fragmentos que dela incorpora; já a segunda forma de separação concerne à mudança da segunda para a terceira posição, em que a fala da criança se afasta da dominância do funcionamento da língua e dos erros que este torna visíveis, dando lugar à dominância da relação fala-escuta e à possibilidade de autocorreção facultada por essa relação. Quarto comentário: tal relação entre o falar e o escutar é uma relação entre duas instâncias subjetivas não coincidentes, o que suscita uma dúvida – seria lícito vislumbrar, nessa divisão do sujeito, uma outra alteridade? Quinto comentário: as características de anterioridade lógica, de captura e de significação conferidas à linguagem relativamente à criança como *corpo pulsional* (outro termo que, embora não definido, remete à teoria psicanalítica) enfatizam a alteridade do simbólico.

Em texto de 2001, intitulado “Sobre o estatuto linguístico e discursivo da narrativa na fala da criança”, Cláudia de Lemos avança em sua teorização inspirada, na linguística, por Saussure e por Jakobson (sobretudo, mas também por Benveniste) e, na psicanálise, por Lacan. Se, em estudos anteriores, o termo *discurso* não era definido, nesse trabalho, encontramos uma definição do “discurso enquanto fala (*parole*) que faz texto” (Lemos, 2001, p. 29). Situando um quadro teórico então mais consolidado, a pesquisadora concebe “a criança, enquanto corpo pulsional, como capturada pelo funcionamento da língua na qual é significada, por um outro, como falante, antes mesmo de o ser”, captura que a coloca “em uma *estrutura* em que comparece o *outro como instância de interpretação* e o *Outro como depósito e rede de significantes*. Essa estrutura é a mesma em que se move o adulto (que é também o outro da criança), enquanto sujeito falante também submetido ao Outro” (Lemos, 2001, p. 29). Mais adiante, ao mencionar a terceira posição, a linguista a caracteriza um pouco mais detidamente: “É na terceira posição que a criança enquanto sujeito falante se divide entre aquele que fala e aquele que escuta sua própria fala, sendo capaz de retomá-la, reformulá-la e *reconhecer a diferença* entre sua fala e a fala do outro, *entre a instância que fala e a instância que escuta de um lugar outro*” (Lemos, 2001, p. 29).

Do desenho dessa estrutura com três polos – a criança, o adulto como outro e a língua como Outro –, relevam três aspectos. Em primeiro lugar, o fato de a reflexão de Cláudia de Lemos apartar-se, definitivamente, de uma visão dual do diálogo, agora subvertida por uma visão triádica deste. Em segundo lugar, o fato de o terceiro elemento dessa tríade estrutural (a língua), referido em 1995 como “Outro” e em 2000 como “outro enquanto texto”, passar a ser designado enquanto “Outro como depósito e rede de significantes”. Em terceiro lugar, o fato de a terceira posição ser caracterizada não somente pela divisão subjetiva, mas igualmente pelo reconhecimento, pela criança, da diferença tanto entre a fala própria e a fala alheia quanto entre a instância que fala e a instância que escuta de um “lugar outro” – o que talvez autorize uma resposta afirmativa à pergunta anterior sobre a licitude de se perscrutar, nessa não coincidência entre as duas instâncias subjetivas, uma outra alteridade. Se assim for, não parece ir longe demais distinguir essa alteridade das duas anteriores (com o outro e com o Outro) em termos de *condição* e de *efeito* da aquisição de linguagem como processo de subjetivação: a alteridade com o outro materno e a alteridade com a língua são as condições de tal processo, ao passo que a divisão/alteridade entre uma instância que fala e uma instância que escuta é dele efeito.

Em texto de 2002, intitulado “Das vicissitudes da fala da criança e de sua investigação”, a autora radicaliza a visão triádica do diálogo e a inclusão da língua como elemento assimetrizável na estrutura (porque a ela submetidos a criança e o outro). Partindo da ideia de ordem própria da língua e da exclusão do sujeito psicológico como pontos comuns a Saussure (que afirma a impossibilidade de intervenção do falante na língua) e a Chomsky (que afirma a impossibilidade de aprendizagem da língua por generalização indutiva devido à não observabilidade das propriedades linguísticas), Lemos (2002) propõe:

O que a mim pareceu, então, coerente com essa autonomia e alteridade radical da língua foi dar a ela, à língua, a função de *captura*, entendida como estenograma ou abreviatura (sobre esse conceito metodológico, ver Milner 1989) de processos de subjetivação. Considerada sua anterioridade lógica relativamente ao sujeito, [a língua] o precede e, considerada em seu funcionamento simbólico, poder-se-ia inverter a relação sujeito-objeto, conceber a criança como capturada por um funcionamento linguístico-discursivo que não só a significa como lhe permite significar outra coisa, para além do que a significou. (Lemos, 2002, p. 55).

Ainda nesse artigo de 2002, encontramos duas novas formulações relacionadas à alteridade na teorização de Cláudia de Lemos, ambas tributárias da formação psicanalítica a que, à época, vinha se submetendo a linguista (Pereira de Castro, 2005). A primeira formulação é uma definição mais explícita de *corpo pulsional*: “No lugar da criança referida como organismo – corpo biológico sob a égide da necessidade – passo a colocar, portanto, o corpo pulsional”, qualificando-o “como o ‘que *demand*a interpretação’, isto é, corpo que, ‘*articulado na e pela linguagem*’, se acha no regime da demanda e do desejo” (Lemos, 2002, p. 64). A segunda formulação é uma releitura da terceira posição a partir do reconhecimento de que a fala da criança revela mais sobre aspectos subjetivos do que sobre aspectos linguísticos: “A escuta de sua própria fala, assim como a escuta da fala do outro em sua diferença, adviriam, então, não de um movimento da língua sobre si mesma”, mas “do *grande Outro*, definido por Lacan (1998 [1966]) como tesouro de significantes, do qual o emissor recebe sua própria mensagem invertida”, isto é, “onde a mensagem repercute e de onde retorna uma resposta divergente, uma resposta outra” (Lemos, 2002, p. 64). Subjacente a ambas as formulações, parece estar um novo deslocamento da noção de *interpretação*: se o deslocamento anterior havia sido de uma noção psicológica nos trabalhos da década de 1980 (interpretação como mútua atribuição de intenções, conhecimentos e crenças entre criança e outro) a uma noção linguística nos estudos da década de 1990 (interpretação como ancoragem linguístico-discursiva da fala infantil e contenção de sua deriva semântico-formal), o novo deslocamento é de uma noção linguística a uma noção psicanalítica nesse artigo da década de 2000 (interpretação como articulação na e pela linguagem de um corpo que demanda e deseja ser significado).

Em texto de 2006 (2000), intitulado “Uma crítica (radical) à noção de desenvolvimento na Aquisição de Linguagem”, Cláudia de Lemos se detém em um parágrafo no qual Saussure se refere à aquisição e declara ser falsa a ideia de que o estudo da linguagem em suas *origens* é mais simples do que o estudo dela em suas *condições permanentes*, expressão para a qual a autora oferece a seguinte leitura:

Parece plausível dizer que as “condições permanentes” que estão presentes mesmo no instante – mítico ou não – da origem são:

- o *ato do falante* como lugar em que emerge uma diferença;
- os processos através dos quais o *falante é identificado e se identifica com o outro*, cuja alteridade é *obliterada em favor de um assemelhamento*, obliteração essa que incide sobre diferenças linguísticas;
- e, finalmente, *la langue como sistema de relações internas*, que oblitera tanto a cena ou o ato individual quanto a semelhança como efeito de processos de identificação. (Lemos, 2006 [2000], p. 27, *itálico em “la langue” do original*).

Reencontramos, nessa citação, os três polos da estrutura que, para Lemos, é constitutiva da aquisição de linguagem: a criança, o outro e a língua. Segundo a linguista, as ditas condições permanentes barram concepções de linguagem/língua como objeto de conhecimento e de criança como sujeito epistêmico que adquiriria tal objeto por ter propriedades perceptuais e cognitivas precedentes e determinantes dessa aquisição. Ao contrário, “É a linguagem, ou melhor, *le langage* – e nela está incluído o *outro enquanto semelhante e, na sua diferença, enquanto ‘outro’* – que precede e determina a transição da criança do estado de *infans* para o de falante”, ao que acrescenta: “[...] a criança é capturada por *le langage*, atravessada e significada como é pela *parole* do outro, matriz de sua identificação como *semelhante* – e

membro da comunidade lingüística e *cultural* – e como *dessemelhante*, referido a uma subjetividade figurada como individual” (Lemos, 2006 [2000], p. 27, itálicos em “le langage” e “parole” do original). Também nesse texto, Lemos reanalisa episódios representativos das três posições, descortinando aspectos subjetivos próprios a cada uma. No caso da primeira posição, trata-se da identificação da criança com a fala da mãe (alienação) e da não coincidência entre os significados de seus enunciados (separação), o que mostra que “‘*semelhança*’ e ‘*diferença*’ são faces alternativas de uma subjetividade emergente” (Lemos, 2006, p. 29). No caso da segunda posição, trata-se da impermeabilidade (não escuta ou alienação ao movimento da língua/Outro) da criança à correção do adulto/outro. No caso da terceira posição, por um lado, trata-se “do reconhecimento, pela criança, de que seu enunciado afeta o interlocutor e, conseqüentemente, do reconhecimento da *alteridade do interlocutor*” e, por outro lado, trata-se da não coincidência entre a instância “daquele que fala e a daquele que escuta seu próprio enunciado como *fala de um outro*. Em outras palavras, a ‘subjetividade’ implicada em meu trabalho é aquela introduzida pela Psicanálise” (Lemos, 2006, p. 31).

Desse texto de 2006 (2000), atento a três pontos relativos à alteridade. O primeiro ponto refere-se ao desdobramento (ou à divisão?) do outro com inicial minúscula, ideia esboçada no texto de 1998b e aqui mais lapidada: enquanto semelhante, trata-se da matriz dos processos identificatórios que aproximam a criança tanto do adulto quanto da comunidade lingüística e cultural que este representa para ela; enquanto dessemelhante, trata-se igualmente de uma representação, porém não mais da comunidade e sim de uma subjetividade que figura, para a criança, como uma individualidade diferente da sua própria, com significados não coincidentes com os seus. O segundo ponto refere-se ao emprego da grafia “outro” (com aspas) para distinguir o adulto enquanto diferente do adulto enquanto semelhante (outro, sem aspas), uma marca tipográfica que se soma a uma já empregada anteriormente, nos textos de 1995, de 2001 e de 2002: a inicial maiúscula (Outro) para distinguir a alteridade da língua da alteridade do interlocutor. O terceiro ponto refere-se à introdução da *cultura* na reflexão de Cláudia de Lemos, noção que, embora não definida diretamente, parece pela autora vista em seu aspecto simbólico, mas no sentido não de *convencional* e sim de *estruturante*, estruturação resultante dos processos identificatórios.

Em conferência de 2018, intitulada “Do encontro com a fala da criança”, Lemos volta a falar de cultura, vinculando-a à linguagem/língua e ao devir humano:

Do ponto de vista epistemológico, pude, então, virar de avesso a expressão “aquisição de linguagem”, identificando processos implicados nessas mudanças estruturais como a captura da criança pelo funcionamento lingüístico-discursivo, o termo “captura” servindo como um estenograma de processos de subjetivação. Em palavras mais simples, *em vez de adquirir, a criança é adquirida pela linguagem e, através dela, pela cultura e pelo humano*. (Lemos, 2018, 21min57s-22min33s).

[...] vou terminar com duas coisas. Primeiro, com o final da vida e da obra do Lacan, em que ele fala de *lalíngua* e da mãe como *lalíngua*, da mãe como língua, *na transmissão da linguagem... E da hominização! Do se tornar humano...* (Lemos, 2018, 40min15s-40min43s).

Em três textos mais recentes, fortemente ancorados na psicanálise, Cláudia de Lemos introduz novas hipóteses atinentes à alteridade. De 2008, o primeiro texto intitula-se “Entre o falante ideal e o sujeito falante: por onde se move a pesquisa lingüística e/ou por onde circula

o lingüista”, sendo um ensaio no qual a agora linguista-psicanalista (ou linguista/psicanalista?) argumenta que, ao analisar uma fala, o linguista não o faz fora de sua própria condição de falante e, portanto, de um outro dessa fala que a submete a seu ato de interpretação, o que obriga o reconhecimento de que “a condição do falante, indissociável da posição do lingüista, é, antes de tudo, a de *objeto do desejo do Outro e, portanto, efeito de haver linguagem*”, ao que acrescenta: “[...] em meu trabalho sobre fala de crianças tenho argumentado contra a visão de que a criança *adquire* linguagem e a favor da *captura* da criança pela língua em ato na cadeia signifiicante em que poderá vir a emergir como sujeito” (Lemos, 2008, p. 7-8, itálicos em “adquire” e “captura” do original). De 2014, o segundo texto intitula-se “A criança e o linguista: modos de habitar a língua?”, sendo um texto no qual a autora leva adiante a discussão anterior, sustentando que “a criança deve recalcar a condição de *objeto do Outro* e as relações com a *fala do outro* que dessa condição decorrem, para se alçar como ator da fala de que é produto” (Lemos, 2014, p. 962). Além do “efeito performativo” que consiste nesse “poder alçar-se como quem fala ainda que falada pelo Outro” (Lemos, 2014, p. 960), tal recalque seria responsável pelo desaparecimento, na fala da criança, daquilo que reaparece nas formações do inconsciente (lapso, chiste, ato falho, sonho), na poesia, no dizer psicótico, a saber, “procedimentos linguageiros [...] em que se destacam os recursos à homonímia e aos efeitos de sentido responsáveis pela figurabilidade no sonho e na ficção” (Lemos, 2014, p. 961). De 2015, o terceiro texto intitula-se “*Lalangue*: acontecimento e transmissão”, sendo uma reflexão na qual a estudiosa interroga-se sobre a transmissão de lalíngua (*lalangue*) à criança em sua passagem de *infans* a falasser (*parlêtre*) e toma “*lalangue/lalíngua* como acontecimento, como inconsciente em ato”, transmitida nos “primeiros jogos entre mãe e bebê, jogos de ressonância – em que a mãe faz eco aos sons que vêm do corpo do filho, com sílabas que podem vir estruturadas em palavras”, o que põe a nu “o jogo a que estamos submetidos como seres de linguagem, jogo de fazer e desfazer o signo, de fazê-lo soar e ressoar, suspendendo o sentido, jogo este que faz de *lalíngua* o próprio movimento do pensar, dizer e escutar” (Lemos, 2015, p. 48).

5 Considerações finais

Como antecipado na introdução, este artigo consiste na primeira parte de um estudo a respeito da alteridade na aquisição da língua materna. Busquei, aqui, delinear um breve histórico sobre o problema da relação criança-outro no campo da aquisição de linguagem. Após um sobrevoo panorâmico pelo campo, debruicei-me sobre as ideias de duas figuras representativas de visões antagônicas do diálogo no vir a ser falante. A título de conclusão, sublinharei os traços gerais das duas visões confrontando suas principais diferenças metodológicas e teóricas.

Em termos de diferenças metodológicas, embora tanto Jerome Bruner quanto Cláudia de Lemos tenham privilegiado o diálogo como unidade de análise e o acompanhamento de crianças *in vivo* e não *in vitro*¹⁴, o psicólogo e seu grupo acompanharam bebês nos primeiros dezoito meses de vida, enquanto a linguista e seu grupo acompanharam crianças de um a cinco anos. Pode-se, assim, dizer que o não reconhecimento, por Bruner, da alteridade da língua deve-se não só ao fato de ele não ser linguista, mas também ao tipo de dado que esco-

¹⁴ As expressões *in vivo* e *in vitro* são utilizadas por Jerome Bruner para referir-se, metaforicamente, a coletas naturalísticas (*in vivo*) e a coletas experimentais (*in vitro*).

lheu analisar, pois “seus sujeitos estavam ainda no chamado período pré-linguístico, em que a ‘estranheza’, a diferença da fala da criança ainda estava por aparecer” (M. T. G. de Lemos, 2002, p. 171). Nessa direção, o psicólogo foi favorecido por sua escolha metodológica e se deixou encantar pela aparente cumplicidade mãe-bebê em termos de ação e de atenção conjuntas, a ponto de postular uma “‘sintonização perfeita’ na aquisição da linguagem” (Bruner, 2007 [1983], p. 112). Em contrapartida, os dados mais tardios de Lemos mostravam tanto a dependência da fala da criança relativamente à fala da mãe em distintos aspectos linguístico-discursivos (prosódicos, sintáticos, narrativos, argumentativos) quanto o distanciamento da fala materna indiciado por “‘erros’ de natureza estritamente lingüística, resultantes da substituição de verbos causativos por incoativos e vice-versa – objeto específico da investigação de Rosa Attité Figueira (1986) –, que mostraram uma possível porta de saída da relação dual em que os processos dialógicos pareciam detidos” (Lemos, 2002, p. 49). Essa porta de saída não era senão o reconhecimento da língua em sua radical alteridade.

Em termos de diferenças teóricas, estas concernem a divergentes concepções de linguagem/língua, cultura, aquisição, sujeito, outro e interpretação.

De um lado, a visão dual do diálogo representada por Jerome Bruner ergue-se sobre uma concepção de linguagem enquanto instrumento de comunicação, cuja aquisição consiste em um processo de aprendizagem de caráter tanto cultural quanto cognitivo. Cultural, porque, ao aprender a falar, a criança aprende atos de fala pragmaticamente adequados às convenções de sua cultura enquanto comunidade de fala constituída por processos simbólicos (no sentido de *convencionais*), conceitos e distinções. Cognitivo, porque, ao aprender a falar, a criança aprende a expressar suas intenções comunicativas prévias e acessar as intenções alheias. A interpretação reduz-se, aí, à atribuição de intenções mutuamente reconhecíveis pela criança e pelo adulto, tomados, ao mesmo tempo, como indivíduos empíricos e como sujeitos psicológicos que, assimétricos em termos de conhecimento, encontram a simetria na negociação dialógica (ação e atenção conjuntas, intercambialidade de papéis, alternância de turnos), negociação promovida pelos formatos enquanto rotinas comunicativas anteriores ao discurso léxico-gramatical. Essa visão dual do diálogo consiste na interpretação dinâmica/funcional deste que M. T. G. de Lemos (2002) defende ser predominante na literatura sociointeracionista em aquisição de linguagem. Conforme a autora, os psicolinguistas que, nos anos 1970 e 1980, integravam essa literatura aliavam-se, de uma parte, à psicologia e, de outra parte, à pragmática, opondo-se não somente à gramática gerativa, mas igualmente (apesar de maneira menos explícita) à linguística mesma, visto haver “um duplo movimento nesse [à época] novo projeto: ao mesmo tempo em que se introduz o outro, a questão da língua submerge” (M. T. G. de Lemos, 2002, p. 161). Trocando em miúdos, a língua como ordem própria aí não encontra espaço.

De outro lado, a visão triádica do diálogo representada por Cláudia de Lemos após sua virada teórica assenta-se em uma concepção de linguagem enquanto funcionamento simbólico (no sentido de *linguístico-discursivo*), que inclui a língua como sistema de relações internas, a fala como língua em ato na cadeia significante e o discurso como fala que faz texto enquanto figuração de totalidade. O diálogo/interação deixa de ser lido, então, como relação comunicativa adulto-criança e é relido como relação estrutural da fala da criança com a fala do interlocutor (outro/alteridade) e com a língua (Outro/alteridade radical) – “estrutural”, porque se trata de uma estrutura com três polos assimétricos, estando o polo da criança e o polo do outro (eles próprios já assimétricos entre si) submetidos ambos ao polo da lín-

gua. Essa releitura da noção de *diálogo/interação* subverte o empirismo de uma “relação com o outro materno como relação dual, provedora de fala, porém, não atravessada pela língua enquanto terceiro” (Lemos, 2002, p. 49). Essa subversão se estende à concepção de aquisição de linguagem, que passa a ser concebida como processo de subjetivação cuja condição é a alteridade tanto com o outro (primeira posição) quanto com o Outro (segunda posição) e cujo efeito é a divisão do sujeito entre a instância que fala e a instância que escuta a sua fala como a fala de um outro (terceira posição). Essa subversão se estende, igualmente, à concepção de interpretação, tomada tanto em uma dimensão linguístico-discursiva (como ancoragem da fala da criança e contenção de sua deriva semântica e formal) quanto em uma dimensão psíquica (como articulação na e pela linguagem de um corpo pulsional que demanda e deseja ser significado). Essa subversão se estende, ainda, à concepção de cultura, a qual, se é por Lemos vista em termos de comunidade linguística, como por Bruner, é por ela vista também – e sobretudo – em termos de processos identificatórios (mais do que civilizatórios), processos dependentes da linguagem e solidários do devir humano; logo, se a relação linguagem-cultura pode ser abreviada, no pensamento do psicólogo estadunidense, pelo termo *convencionalização*, no pensamento da linguista e psicanalista brasileira, o termo que abrevia tal relação é *hominização*. Assim configurada, essa visão triádica do diálogo consiste na interpretação estrutural deste que M. T. G. de Lemos (2002) anuncia como alternativa à sua interpretação dinâmica/funcional. Trocando em miúdos, a língua como ordem própria é aí reconhecida teórica e analiticamente.

Entretanto, em vez das expressões *interpretação dinâmica/funcional* e *interpretação estrutural*, opto pelas expressões *visão dual* e *visão triádica*. Isso porque, além da proposta de Cláudia de Lemos – inspirada por Ferdinand de Saussure, Roman Jakobson e Jacques Lacan –, há duas outras propostas que – inspiradas por Émile Benveniste – sustentam uma visão triádica do diálogo na aquisição. Trata-se da proposta de Carmem Luci da Costa Silva (2009), no âmbito do vir a ser falante (Silva, 2009), e de minha própria proposta, no âmbito do vir a ser escrevente (Oliveira, 2022). Na segunda parte deste estudo, em que o olhar retrospectivo dará lugar a um olhar prospectivo, deter-me-ei nessas duas propostas benvenistianas, a fim de retomar e de avançar a partir de suas ideias sobre a alteridade na aquisição da língua materna.

Agradecimentos

Este artigo foi produzido durante meu Estágio de Pós-Doutoramento no Instituto de Estudos da Linguagem (IEL) da Universidade Estadual de Campinas (UNICAMP), sob a supervisão da Profa. Dra. Maria Fausta Pereira de Castro e sob o financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Agradeço à professora pela interlocução teórica e ao CNPq, pela Bolsa de Pós-Doutorado Júnior (PDJ). Agradeço, ainda, aos Avaliadores 2 e 3 pelos elogios ao presente texto. Agradeço, especialmente, ao Avaliador 1, pela detalhada leitura e pelos preciosos apontamentos, que qualificaram este estudo tanto em termos de esclarecimentos de determinados pontos quanto em termos de diálogo crítico travado no corpo do texto e em notas de rodapé. Sinto-me lisonjeado por uma leitura tão minuciosa.

Referências

- BENVENISTE, É. *Problemas de linguística geral II*. 5. ed. Campinas: Pontes Editores, 2023 [1966/1967]. 387 p.
- BENVENISTE, É. *Problemas de linguística geral II*. 2. ed. Campinas: Pontes Editores, 2006. 294 p.
- BRUNER, J. S. The ontogenesis of speech acts. *Journal of Child Language*, Cambridge/Inglaterra, v. 2, n. 1, p. 1-19, abr. 1975. DOI: <https://doi.org/10.1017/S0305000900000866>
- BRUNER, J. S. *Como as crianças aprendem a falar*. Instituto Piaget: Lisboa, 2007 [1983]. 135 p.
- BRUNER, J. S. *Acción, pensamiento y lenguaje*. Compilación de José Luis Linaza. Alianza Editorial: Madrid, 2007 [1984]. 232 p.
- BRUNER, J. S. *Realidade mental, mundos possíveis*. Porto Alegre: Artes Médicas, 1997 [1986]. 211 p.
- BRUNER, J. S. *Atos de significação*. Porto Alegre: Artes Médicas, 1997 [1990]. 130 p.
- BRUNER, J. S. The narrative construction of reality. *Critical Inquiry*, Chicago, v. 1, n. 18, p. 1-21, 1991. DOI: <https://doi.org/10.1086/448619>
- BRUNER, J. S. *A cultura da educação*. Lisboa: Edições 70, 2000 [1997]. 240 p.
- CAPRISTANO, C. C. A escrita infantil nos primeiros anos de escolarização: alteridade e rasura. In: PRADO, G. do V. T.; SERODIO, L. A.; RODRIGUES, N. C. (orgs.) *Da indiferença à não indiferença: como viver a amorosidade?* São Carlos: Pedro & João Editores, 2023. p. 998-1010.
- DEL RÉ, A.; VIEIRA, A. J.; HILÁRIO, R. N.; GOUVÊA, C. P. A incongruência como um dos ingredientes do humor: um olhar dialógico-discursivo para os dados de crianças brasileiras ouvintes e surdas. *Cuadernos de la ALFAL*, v. 2, n. 14, p. 127-147, nov. 2022. DOI: 5935/2218-0761.2022007
- DUCROT, O. *O dizer e o dito*. Campinas: Pontes, 1987.
- DUCROT, O. *Polifonía y Argumentación*. Conferencias del Seminario Teoría de la Argumentación y Análisis del Discurso. Cali: Universidad del Valle, 1990.
- DIEDRICH, M. S.; OLIVEIRA, G. F.; DEL RÉ, A. *Língua, discurso e suas relações na aquisição da linguagem*. São Carlos: Pedro & João Editores, 2023. 322 p. Disponível em: <https://pedroejoaoeditores.com.br/produto/lingua-discurso-e-suas-relacoes-na-aquisicao-da-linguagem/>. Acesso em: 03 nov. 2025.
- FARACO, C. A. Interação e linguagem: balanço e perspectivas. *Calidoscópio*, São Leopoldo, v. 3, n. 3, p. 214-221, 2005. Disponível em: <https://revistas.unisinos.br/index.php/calidoscopio/article/view/6244>. Acesso em: 02 abr. 2026.
- FIGUEIRA, R. A. Os matizes da agentividade nos nichos domésticos: a interação em foco. *Cuadernos de la ALFAL*, v. 2, n. 14, p. 225-252, nov. 2022. DOI: 10.5935/2218-0761.2022012
- LEMO, C. T. G. de. Sobre Aquisição de Linguagem e seu dilema (pecado) original. *Boletim da Abralin*, Curitiba, v. 3, p. 97-126, 1982.
- LEMO, C. T. G. de. Specularity as a constitutive process in dialogue and language acquisition. In: CAMAIONI, L.; DE LEMO, C. T. G. (orgs.). *Questions on social explanation: piagetian themes reconsidered*. Amsterdam: John Benjamins, 1985. p. 23-31.

- LEMOS, C. T. G. de. Interacionismo e aquisição de linguagem. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, São Paulo, v. 2, n. 2, p. 231-248, 1986a. Disponível em: <https://revistas.pucsp.br/index.php/delta/article/view/31271>. Acesso em: 03 nov. 2025.
- LEMOS, C. T. G. de. A sintaxe no espelho. *Cadernos de Estudos Linguísticos*, Campinas, v. 10, p. 5-15, 1986b. DOI: 10.20396/cel.v10i0.8636714
- LEMOS, C. T. G. de. Os processos metafóricos e metonímicos como mecanismos de mudança. In: TEBEROSKY, A.; TOLCHINSKY, L. (orgs.). *Substratum: Temas fundamentais em psicologia e educação*, v. 1, n. 3. Porto Alegre: Artes Médicas, 1998a [1992]. p. 151-172.
- LEMOS, C. T. G. de. A função e o destino da palavra alheia: três momentos da reflexão de Bakhtin. In: BARROS, D. L. P. (org.). *Dialogismo, polifonia, intertextualidade*. São Paulo: EDUSP, 1994. p. 37-43.
- LEMOS, C. T. G. de. Língua e discurso na teorização sobre aquisição de linguagem. *Letras de Hoje*, Porto Alegre, v. 30, n. 4, p. 9-28, dez. 1995. Disponível em: <https://revistaseletronicas.pucrs.br/fale/article/view/15682>. Acesso em: 03 nov. 2025.
- LEMOS, C. T. G. de. Native speaker's intuitions and metalinguistic abilities: what do they have in common from the point of view of language acquisition? *Cadernos de Estudos Linguísticos*, Campinas, v. 33, p. 5-14, jul.-dez. 1997. DOI: 10.20396/cel.v33i0.8637091
- LEMOS, C. T. G. de. Sobre a Aquisição da Escrita: algumas questões. In: ROJO, R. (org.). *Alfabetização e Letramento*. Campinas: Mercado de Letras, 1998b, p. 13-32.
- LEMOS, C. T. G. de. Interrelações entre a Linguística e outras ciências. *Boletim da Abralin*, n. 22, p. 19-32, jun. 1998. Disponível em: <https://abralin.org/wp-content/uploads/2018/12/boletim22a.pdf>. Acesso em: 03 nov. 2025.
- LEMOS, C. T. G. de. Sobre o "Interacionismo". *Letras de Hoje*, Porto Alegre, v. 34, n. 3, p. 11-16, 1999a. Disponível em: <https://revistaseletronicas.pucrs.br/fale/article/view/14944>. Acesso em: 03 nov. 2025.
- LEMOS, C. T. G. de. A criança com(o) ponto de interrogação. In: Lamprecht, R. (org.). *Aquisição de Linguagem: questões e análise*. Porto Alegre: EDIPUCRS, 1999b. p. 39-50.
- LEMOS, C. T. G. de. Desenvolvimento da linguagem e processo de subjetivação. *Interações*, São Paulo, v. V, n. 10, p. 53-72, jul./dez., 2000. Disponível em: <https://www.redalyc.org/pdf/354/35401004.pdf>. Acesso em: 03 nov. 2025.
- LEMOS, C. T. G. de. Das vicissitudes da fala da criança e de sua investigação. *Cadernos de Estudos Linguísticos*, Campinas, n. 42, p. 41-69, jan./jun., 2002. DOI: 10.20396/cel.v42i0.8637140
- LEMOS, C. T. G. de. Uma crítica (radical) à noção de desenvolvimento na aquisição de linguagem. In: LIER-DEVITTO, M. F.; ARANTES, L. (orgs.). *Aquisição, patologias e clínica de linguagem*. São Paulo: EDUC, 2006 [2000]. p. 21-32.
- LEMOS, C. T. G. de. Entre o falante ideal e o sujeito falante: por onde se move a pesquisa lingüística e/ou por onde circula o lingüista. *Revista Conexão Letras*, Porto Alegre, v. 3, n. 3, 2008. DOI: <https://doi.org/10.22456/2594-8962.55632>
- LEMOS, C. T. G. de. A criança e o linguista: modos de habitar a língua? *Estudos Linguísticos*, São Paulo, v. 43, n. 2, p. 954-964, 2014. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/493>. Acesso em: 03 nov. 2025.

- LEMOS, C. T. G. de. *La língua: acontecimento e transmissão*. In: PORGE, É.; LEITE, N. V. de A. (Orgs.). *Savoir-faire avec la langue*. Campinas: Mercado de Letras, 2016, p. 39-49.
- LEMOS, C. T. G. de. Do encontro com a fala da criança. *YouTube*, 26 abr. 2018. 01h49min20s. Disponível em: <https://www.youtube.com/watch?v=TqKF5j7JkOY&t=124s>. Acesso em: 03 nov. 2025.
- LEMOS, M. T. G. de. *A língua que me falta: uma análise dos estudos em aquisição de linguagem*. Campinas: Mercado de Letras; São Paulo: FAPESP, 2002, 216 p.
- LIER-DEVITTO, M. F.; CARVALHO, G. M. M. O interacionismo: uma teorização sobre a Aquisição da Linguagem. *Revista da ABRALIN*, [S. l.], v. 23, n. 2, p. 462–490, 2024. DOI: 10.25189/rabralin.v23i2.2226
- LINAZA, J. L. Introducción. In: BRUNER, J. S. *Acción, pensamiento y lenguaje*. Compilación de José Luis Linaza. Alianza Editorial: Madrid, 2007 [1984]. p. 9-29.
- OLIVEIRA, G. F. *Do homo loquens ao homo loquens scriptor: por uma perspectiva semiológico-enunciativa da aquisição da escrita*. 2022. 428 f. Tese (Doutorado em Estudos da Linguagem) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2022. Disponível em: <https://www.lume.ufrgs.br/handle/10183/257970>. Acesso em: 17 nov. 2024.
- OLIVEIRA, G. F.; SILVA, C. L. da C. O que os estudos sobre a aquisição devem a Benveniste. *Eutomia*, Recife, v. 1, n. 33, p. 153-184, jun. 2023. DOI: <https://doi.org/10.51359/1982-6850.2023.258892>
- PEREIRA DE CASTRO, M. F. Sobre o Projeto de Aquisição de Linguagem e a obra de sua fundadora: uma homenagem a Cláudia de Lemos. *Cadernos de Estudos Linguísticos*, Campinas, v. 47, n. 1 e 2, p. 11-15, jan./dez. 2005. DOI: 10.20396/cel.v47i1/2.8637265
- SILVA, C. L. da C. *A criança na linguagem: enunciação e aquisição*. 1. ed. Campinas: Pontes Editores, 2009.

As interrogativas polares e a falha de pressuposição em português brasileiro: uma análise sob a perspectiva da semântica e da pragmática formais

Polar Interrogatives and Presupposition Failure in Brazilian Portuguese: An Analysis from the Perspective of Formal Semantics and Pragmatics

Fernanda Rosa da Silva
Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
fernandarosa@ufmg.br
<https://orcid.org/0000-0002-0599-8805>

Elisa de Oliveira Silva
Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
elisaoliversilva@gmail.com
<https://orcid.org/0009-0003-2589-6755>

Resumo: Esta pesquisa investiga a interação entre a dinâmica das sentenças interrogativas e o fenômeno de falha de pressuposição à luz da semântica e da pragmática formais. O estudo focaliza perguntas polares com descrições definidas, que desencadeiam pressuposições de existência e unicidade. Objetiva-se compreender como tais falhas influenciam a dinâmica contextual e em quais situações o interlocutor acomoda ou contesta a pressuposição. Também buscam-se descrever formalmente os efeitos pragmáticos e conversacionais decorrentes dessas respostas. A pesquisa justifica-se pela escassez de estudos formais sobre interrogativas no português brasileiro e pela lacuna na literatura sobre falhas de pressuposição e suas implicações na interação conversacional. Adotou-se o método introspectivo, em que se construíram contextos comunicativos com perguntas polares seguidas de respostas variadas, permitindo-se observar processos de acomodação, nos quais o ouvinte integra a informação pressuposta ao *common ground*, e de contestação, em que a pressuposição é negada e substituída por uma proposição alternativa. Exemplos como *O gato da pracinha foi adotado?* evidenciam que respostas como *Não sabia que tinha um gato na pracinha* configuram acomodação, enquanto respostas como *Não tem nenhum gato na pracinha* ou *Tem muitos gatos na pracinha* caracterizam contestação. Nos casos de falha de pressuposição de existência, a contestação impede a pergunta de ser respondida, retirando-a do



question set. Já na contestação da unicidade, a pergunta permanece, subdividindo-se para refletir múltiplos referentes possíveis. Em síntese, o estudo contribui para a compreensão dos mecanismos formais que regem o comportamento semântico-pragmático das interrogativas no PB, destacando como interlocutores interagem diante de falhas pressuposicionais e atualizam o contexto conversacional.

Palavras-chave: sentenças interrogativas polares; falha de pressuposição; acomodação e contestação.

Abstract: This research investigates the interaction between the dynamics of interrogative sentences and the phenomenon of presupposition failure within the frameworks of formal Semantics and Pragmatics. The research focuses on polar questions containing definite descriptions, which trigger existence and uniqueness presuppositions. It aims to understand how such failures influence contextual dynamics and under which circumstances the interlocutor accommodates or challenges the presupposition. Furthermore, it seeks to provide a formal description of the pragmatic and conversational effects resulting from these responses. The study is justified by the scarcity of formal analyses of interrogatives in Brazilian Portuguese and by the lack of literature addressing presupposition failures and their implications for conversational interaction. An introspective method was adopted, constructing communicative contexts with polar questions followed by varied responses, which allowed for the observation of accommodation processes, where the hearer integrates the presupposed information into the common ground, and contestation processes, where the presupposition is denied and replaced by an alternative proposition. Examples such as *Was the cat from the square adopted?* illustrate that responses like *I didn't know there was a cat in the square* represent accommodation, whereas responses such as *There is no cat in the square* or *There are many cats in the square* exemplify contestation. In cases of existence presupposition failure, contestation prevents the question from being answered, removing it from the question set. In contrast, in uniqueness presupposition failures, the question remains but subdivides to reflect multiple possible referents. In sum, this study

contributes to understanding the formal mechanisms governing the semantic-pragmatic behavior of interrogatives in Brazilian Portuguese, highlighting how interlocutors manage presupposition failures and update the conversational context.

Keywords: dynamics of interrogative sentences; presupposition failure; accommodation and contestation.

1 Introdução

A presente pesquisa busca explorar, sob a ótica da semântica formal (Chierchia, 2003; Chierchia; McConnell-Ginet, 1996; Heim, 1983; Heim; Kratzer, 1998; Partee, 1973) e da pragmática formal (Gazdar, 1979; Grice, 1975; Levinson, 1983; Stalnaker, 1999; entre outros) a interação entre a dinâmica das sentenças interrogativas (Groenendijk, 1999; Groenendijk; Stokhof, 1997) e o fenômeno de falha de pressuposição (Beaver, 2001). Mais especificamente, serão analisados contextos em que haja perguntas polares, e que essas apresentem pressuposições desencadeadas por descrições definidas singulares. Serão observadas tanto falhas de pressuposição de existência, quanto falhas de unicidade para o indivíduo referido.

A partir da análise teórica dessa interação, objetiva-se compreender como a falha de pressuposição impacta a dinâmica interrogativa. Além disso, busca-se identificar em quais situações o ouvinte acomoda a pressuposição, em quais situações o ouvinte contesta a pressuposição e como essas diferentes respostas impactam o processo conversacional. Esses processos de falha, acomodação e contestação da pressuposição serão descritos formalmente, a partir das ferramentas teóricas e metodológicas utilizadas pela semântica e pela pragmática formais.

Esta pesquisa se justifica, pois ainda há poucos estudos em português brasileiro que exploram a dinâmica das sentenças interrogativas, em especial, sob uma perspectiva formal do estudo do significado. Ainda, pouco se sabe sobre o fenômeno de falha de pressuposição e sobre como esse interage com a dinâmica das interrogativas.

Para a análise teórica da interação entre o fenômeno de falha de pressuposição e a dinâmica das sentenças interrogativas, foi utilizado o método introspectivo, em que as autoras elaboraram contextos específicos e, a partir das teorias semânticas e pragmáticas de base, analisaram os fenômenos ocorridos. Para isso, foram criadas perguntas polares. Para cada pergunta, foram desenvolvidas diferentes respostas em situações comunicativas diversas, todas com falha de pressuposição. Há contextos em que a falha é acomodada pelo ouvinte e contextos em que a falha é contestada. A seguir, apresentamos alguns exemplos de diálogos a serem investigados:

- 1) Ana: O gato da pracinha foi adotado?
Bruno: Não sabia que tinha um gato na pracinha.

- 2) Ana: O gato da pracinha foi adotado?
Bruno: Não tem nenhum gato na pracinha.
- 3) Ana: O gato da pracinha foi adotado?
Bruno: Só tem cachorro na pracinha.
- 4) Ana: O gato da pracinha foi adotado?
Bruno: Tem muitos gatos na pracinha.
- 5) Ana: O gato da pracinha foi adotado?
Bruno: O gato laranja foi, os outros não.

Nos diálogos de (1) a (5), há falha de pressuposição, já que ao fazer a pergunta, Ana acredita que Bruno saiba da existência do gato da pracinha. Entretanto, no contexto em (1), há um processo de acomodação (Lewis, 1979, p. 341), em que Bruno declara que não sabia da existência do gato, mas a partir de então toma essa informação como verdadeira. Nos diálogos de (2) a (5), por outro lado, não ocorre esse processo de acomodação, mas sim uma contestação por parte de Bruno, ou da existência do gato, como em (2) e (3); ou de que haja apenas um gato na pracinha, como em (4) e (5). Tanto a existência quanto a unicidade são pressuposições desencadeadas pelas descrições definidas singulares (Levinson, 1983, p. 171).

Diante desses dados a serem investigados, nossas perguntas são: (i) O que leva o ouvinte em alguns casos acomodar as falhas de pressuposição e em outros casos contestar?; (ii) O que acontece na dinâmica conversacional a partir dos processos de acomodação e contestação?; (iii) Quais as relações lógico-contextuais envolvidas?; (iv) As relações de contradição e contradição estão envolvidas nesses processos?

Para isso, este artigo está organizado da seguinte forma: na seção 2, serão apresentadas as bases teóricas que assumimos para a semântica e a pragmática das sentenças interrogativas. Em 3, será exposta a relação entre o conceito de dinâmica contextual e as sentenças interrogativas. A seção 4 é reservada para analisar as relações entre as sentenças interrogativas e o fenômeno da falha de pressuposição. Por fim, em 5, serão apresentadas as considerações finais da pesquisa.

2 Semântica e pragmática de sentenças interrogativas

Nesta seção, serão apresentadas as teorias e conceitos, a partir de uma perspectiva formal, pertinentes à interpretação semântica e pragmática das sentenças interrogativas. Essas são imprescindíveis para o presente estudo, visto que aqui se discute a interação entre interrogativas e o fenômeno de falha de pressuposição.

As sentenças interrogativas se dividem em, pelo menos, dois tipos básicos: as de constituinte e as polares. As interrogativas de constituinte são aquelas que apresentam um pronome interrogativo associado a um constituinte sintático (Borges-Neto, 2007, p. 2; Ferreira, 2023, p. 163), como:

6) Quem comeu o bolo de chocolate?

Em (6), o pronome interrogativo é 'quem' e o constituinte sintático ao qual ele se associa é o sujeito. A resposta a essa pergunta pode ser extensa, como em *Ana e Bruno comeram o bolo de chocolate*, ou abreviada, como em *Ana e Bruno*, em que a informação pressuposta, desencadeada pela pergunta, não é proferida. O outro tipo de interrogativa, as polares, são aquelas que apresentam apenas duas possibilidades de resposta: uma afirmativa e uma negativa:

7) Bruno comeu o bolo de chocolate?

Em (7), as respostas possíveis são: *Sim, Bruno comeu o bolo de chocolate* e *Não, Bruno não comeu o bolo de chocolate*, no caso das extensas; ou simplesmente *Sim* e *Não*, no caso das abreviadas. Tanto para a semântica como para a pragmática, as respostas extensas são equivalentes às suas contrapartes abreviadas, ou seja, não há diferença de sentido entre os dois tipos de resposta (Ferreira, 2023, p. 163).

Para a semântica formal, o significado das interrogativas está nas suas condições de resposta.¹ Hamblin (1973, p. 47) afirma que, diferentemente das sentenças declarativas, que possuem condições de verdade, as sentenças interrogativas apresentam condições de resposta. Essas são representadas formalmente pelo conjunto de respostas possíveis atribuídas a pergunta. Para uma pergunta de constituinte, o conjunto de respostas possíveis é aberto e será delimitado pelos indivíduos disponíveis no contexto, como podemos observar em (8). Já para a pergunta polar, o conjunto de respostas possui duas alternativas: uma afirmativa e outra negativa, conforme representado em (9):

8) $[[6]] = \{\text{Ana comeu o bolo de chocolate, Bruno comeu o bolo de chocolate, Paulo comeu o bolo de chocolate, Jane comeu o bolo de chocolate, ...}\}$

9) $[[7]] = \{\text{Bruno comeu o bolo de chocolate, Bruno não comeu o bolo de chocolate}\}$

A representação em (8) corresponde ao conjunto de alternativas da pergunta de constituinte em (6). E em (9), o conjunto de alternativas da polar em (7). Groenendijk; Stokhof (1997, p. 41) propõem que essa representação pode-se dar utilizando-se do aparato teórico da semântica de mundos possíveis (Kripke, 1980). Antes de apresentar a proposta de representação do significado de uma sentença interrogativa na semântica de mundos possíveis, é imprescindível que se discuta a proposta de representação para uma sentença declarativa.

Em semântica formal, saber o significado de uma sentença declarativa é saber as suas condições de verdade, ou seja, saber as condições necessárias e suficientes para que esta seja considerada verdadeira (Ferreira, 2019; Ferreira, 2022; Pires De Oliveira, 2010; Quadros-Gomes; Sanchez-Mendes, 2018). Por exemplo:

10) Faz 30° C em Belo Horizonte.

¹ Para saber mais sobre semântica das condições de verdade, consulte: Ferreira (2019; 2022); Pires de Oliveira (2010); Quadros-Gomes; Sanchez-Mendes (2018).

A condição necessária e suficiente para que a sentença (10) seja considerada verdadeira é que faça 30° C na cidade de Belo Horizonte no momento da fala. Então, essa é a condição de verdade da sentença e, por conseguinte, o seu significado semântico.

Em uma perspectiva da semântica dos mundos possíveis, igualmente relevante para o fenômeno investigado nesta pesquisa, é o uso do aparato de mundos possíveis, assume-se uma ontologia teórica que inclui um conjunto (possivelmente infinito) de mundos possíveis, entendidos como alternativas logicamente admissíveis em relação ao mundo atual (Kripke, 1980). Cada mundo possível é representado por w (da palavra em inglês *world*, “mundo”) com um número n subscrito, como em: w_1 , w_2 , w_3 , e etc. Há um conjunto que contém todos os mundos possíveis e este é representado por W . Sendo um mundo possível, o mundo real também pertence à W e convencionou-se que esse é representado por w_0 (Ferreira, 2022, p. 75).

Com a teoria dos mundos possíveis, é possível descrever as condições de verdade de sentenças não apenas a partir dos fatos e do estado de coisas do mundo real, mas também a partir dos fatos e do estado de coisas em outros mundos possíveis. Por exemplo:

11) A raposa falou.

Uma sentença como (11) seria considerada falsa no mundo real, visto que em w_0 , as raposas não possuem a capacidade de falar. Porém, no mundo fictício de *O Pequeno Príncipe*, (11) seria considerada verdadeira, já que na história há uma raposa que fala.

Outra grande utilidade da abordagem teórica de mundos possíveis é a formalização do conceito de proposição como um conjunto de mundos possíveis (Kripke, 1980). Retomamos o exemplo em (11). A proposição dessa sentença é o conjunto de todos os mundos possíveis em que é verdade que raposas falam. Em termos formais:

12) $[[[p_{11}]]] = \{w \in W \mid \text{a raposa falou em } w\}$

A representação em (12) pode ser parafraseada como: a proposição da sentença (11) corresponde ao conjunto de mundos possíveis w , pertencente ao conjunto de todos os mundos possíveis, em que a proposição *a raposa falou* é verdadeira.

Por outro lado, todos os mundos w que estão fora do conjunto delimitado pela proposição são mundos em que a sentença é falsa. No caso do exemplo em (11), o mundo real está fora do conjunto delimitado pela proposição, então a sentença é falsa em w_0 . Porém, o mundo fictício de *O Pequeno Príncipe* está dentro do conjunto delimitado pela proposição, então a sentença é verdadeira nesse mundo possível.

É importante ressaltar que os mundos possíveis nada mais são do que «[...] uma maneira de modelar formalmente a noção de possibilidade lógica, e que nos serão instrumentais na construção de uma teoria semântica.» (Ferreira, 2022, p. 75). Em outras palavras, esta é apenas uma ferramenta teórica e não propõe que os mundos possíveis existam de fato.

Na abordagem dos mundos possíveis, as sentenças interrogativas representam um conjunto de proposições. Se proposições representam um conjunto de mundos possíveis, então as sentenças interrogativas são um conjunto de conjuntos de mundos possíveis. Assim, a pergunta em (6) ‘Quem comeu o bolo de chocolate?’ será representada da seguinte forma:

13) $[[[Q_6]]] = \{w \in W \mid \exists x: x \text{ comeu o bolo de chocolate em } w\}$

Na representação em (13), temos um conjunto de questões Q , formado por proposições em diversos mundos possíveis em que é verdade que *existe um indivíduo x , tal que x comeu o bolo de chocolate*. Por outro lado, uma pergunta polar, como em (7), que compõe nosso objeto de estudo, é representada formalmente como segue:

- 14) $[[Q]] = \{w \mid \text{Bruno comeu o bolo de chocolate em } w\} \cup \{w' \mid \text{Bruno não comeu o bolo de chocolate em } w'\}$

Neste caso, a pergunta polar divide o conjunto de mundos possíveis em dois: um conjunto em que é verdade que o Bruno comeu o bolo de chocolate: w ; e um conjunto em que é falso que o Bruno comeu o bolo de chocolate: w' .

Segundo Groenendijk (1999, p. 111), enquanto uma sentença declarativa, que é representada por uma proposição, seleciona um conjunto de mundos em que aquela sentença é verdadeira, uma sentença interrogativa tem como função semântica realizar uma partição de conjuntos de mundos possíveis. No caso de uma interrogativa polar, em especial, o conjunto de todos os mundos possíveis W é repartido em dois conjuntos: um em que a proposição é verdadeira e um em que a proposição é falsa.

A semântica das sentenças interrogativas é representada, na semântica de mundos possíveis, como uma partição de conjuntos. Já as respostas, segundo Groenendijk; Stokhof (1997, p. 41) podem ser completas ou parciais. Uma resposta completa para uma pergunta apresenta todas as opções logicamente possíveis de resposta e seleciona todos os mundos possíveis.

No caso da interrogativa de constituinte (6), numa situação comunicativa em que apenas Ana e Bruno são suspeitos de ter comido o bolo de chocolate que estava na geladeira, têm-se o seguinte conjunto de possíveis respostas e suas respectivas proposições:²

- 15) R1: Ana e Bruno.
 $p_{R1}: \{w \in W \mid \text{Ana comeu o bolo de chocolate em } w\} \cap \{w' \in W \mid \text{Bruno comeu o bolo de chocolate em } w'\}$
 R2: Ana.
 $p_{R2}: \{w \in W \mid \text{Ana comeu o bolo de chocolate em } w\} \cap \{w' \in W \mid \text{Bruno não comeu o bolo de chocolate em } w'\}$
 R3: Bruno.
 $p_{R3}: \{w \in W \mid \text{Bruno comeu o bolo de chocolate em } w\} \cap \{w' \in W \mid \text{Ana não comeu o bolo de chocolate em } w'\}$
 R4: Ninguém.
 $p_{R4}: \{w \in W \mid \text{Bruno não comeu o bolo de chocolate em } w\} \cap \{w' \in W \mid \text{Ana não comeu o bolo de chocolate em } w'\}$

A partir do proferimento da interrogativa, o conjunto de todos os mundos possíveis W é repartido em subconjuntos que correspondem a cada uma das proposições apresentadas em (15): um conjunto de mundos em que Ana e Bruno comeram o bolo de chocolate; um conjunto de mundos em que apenas Ana comeu o bolo de chocolate; um conjunto de mundos

² Neste contexto, limitamos a partição de mundos possíveis a apenas dois indivíduos: Ana e Bruno. Entretanto, em um contexto real, a partição de mundos possíveis seria mais complexa e com mais indivíduos incluídos.

em que apenas Bruno comeu o bolo de chocolate; e um conjunto de mundos em que nenhum dos dois comeu o bolo de chocolate.

Em um exemplo como o da interrogativa polar em (7), as opções de resposta são sempre limitadas a uma afirmativa e uma negativa:

16) R1: Sim.

$p_{R1}: \{w \in W \mid \text{Bruno comeu o bolo de chocolate em } w\}$

R2: Não.

$p_{R2}: \{w' \in W \mid \text{Bruno não comeu o bolo de chocolate em } w'\}$

Neste caso, o conjunto W também é repartido em subconjuntos que correspondem às proposições apresentadas em (16): um conjunto de mundos em que Bruno comeu o bolo de chocolate; e um conjunto de mundos em que Bruno não comeu o bolo de chocolate. A resposta para a pergunta selecionará um desses conjuntos de mundos possíveis e eliminará outro.

Dessa forma, uma sentença interrogativa sempre partirá o conjunto W em um número de seções correspondente ao número de respostas logicamente possíveis para a interrogativa. Cada seção em que W foi repartido é um conjunto de mundos possíveis que corresponde a uma proposição de uma das respostas à interrogativa. Esse é o significado semântico de uma sentença interrogativa (Groenendijk, 1999).

Nesta seção, foram apresentadas as interpretações semântica e pragmática das sentenças interrogativas. Essas servirão de subsídio para as análises feitas na seção 4 deste artigo sobre a interação entre as sentenças interrogativas e o fenômeno de falha de pressuposição.

3 Dinâmica contextual e as sentenças interrogativas

Esta seção é reservada para apresentar a proposta da dinâmica contextual, outra importante teoria no estudo formal das sentenças interrogativas. Essa é uma proposta de descrição formal dos diferentes estágios do desenrolar de uma conversa. A dinâmica contextual é composta pelo *common ground* (conhecimento compartilhado)³; *context set* (conjunto contextual); e o *question set* (conjunto de questões) (Roberts, 2012, p. 65; Stalnaker, 1999, p. 100).

O conceito de *Common Ground*, doravante CG, baseia-se, em parte, no conceito de pressuposição do falante definido por Stalnaker (1999, p. 49). As pressuposições do falante, segundo o autor, são as proposições que um falante toma como verdadeiras e que servem de pano de fundo para uma conversa. Aqui, o termo pressuposição é tomado como uma atitude mental que relaciona uma proposição a um falante⁴. Por exemplo, numa situação em que Ana e Bruno conversam em uma cafeteria, se Ana sabe que Bruno faz aniversário no dia 18/4 e que Carlos, amigo de Bruno, vai fazer uma festa de aniversário surpresa para ele, então as proposições em (17) são pressuposições de Ana:

³ Tradução nossa.

⁴ Para a semântica, as pressuposições são pré-condições para que uma sentença possa ser avaliada quanto ao seu valor de verdade (Ferreira, 2023; Strawson, 1950). Esse conceito semântico de pressuposição será explorado com mais detalhes na seção seguinte (seção 2.4).

17) Conjunto de pressuposições de Ana:

$p_1: \{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\}$

$p_2: \{w \in W \mid \text{Carlos vai fazer uma festa de aniversário surpresa para Bruno em } w\}$

Diferentemente das pressuposições do falante, que dizem respeito a apenas um indivíduo, o CG é o conjunto de pressuposições (novamente, pressuposição enquanto atitude mental) compartilhadas entre todos os participantes de uma conversa. É um conhecimento mútuo, que inclui não apenas o que cada um sabe, mas também o que cada um sabe que o outro também sabe. Segue o exemplo de Ana e Bruno: Bruno, claro, sabe que seu aniversário é dia 18/4, Ana também sabe dessa informação e ambos estão cientes que o outro sabe dessa informação. Sendo assim, a proposição: $\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\}$ é uma informação compartilhada entre os participantes da conversa e, portanto, faz parte do CG.

Informações extralinguísticas também podem fazer parte do CG. Dessa forma, se chove do lado de fora da cafeteria em que Ana e Bruno conversam e ambos não apenas notaram a chuva como notaram também que o outro sabe que está chovendo, então a proposição em (18) fará parte do CG:

18) $p_3: \{w \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w\}$

Por outro lado, se Bruno não descobriu sobre a festa surpresa sendo planejada por Carlos, então a proposição: $\{w \in W \mid \text{Carlos vai fazer uma festa de aniversário surpresa para Bruno em } w\}$ não faz parte do CG.

Formalmente, o CG é modelado como um conjunto que contém todas as proposições tomadas como verdadeiras por todos os participantes da conversa em um dado momento da interação (Stalnaker, 1999, p. 67). Nesse exemplo de Ana, Bruno e Carlos, o CG seria o conjunto:

19) $CG_1 = \{\{w \in W \mid \text{O aniversário de Bruno é no dia 18/4 em } w\}, \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}\}$

O *context set*, doravante CS, por outro lado, é o conjunto de “[...] mundos compatíveis com tudo que se pressupõe em um dado instante da conversa” (Stalnaker, 1999, p. 68). Em outras palavras, é o conjunto de mundos em que todas as proposições compartilhadas entre os falantes de uma conversa são tomadas como verdadeiras. Formalmente, o CS é modelado pela intersecção entre todas as proposições de um CG. No exemplo de Ana e Bruno, tem-se o seguinte CS:

20) $CS_1 = \{\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\} \cap \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}\}$

A cada nova asserção aceita pelos participantes da conversa, o CG atualiza e se amplia, visto que este é a junção de todas as proposições de conhecimento compartilhado entre os participantes de uma conversa; e o CS diminui, visto que este é a intersecção de todas as proposições que fazem parte do CG. Se, durante a conversa na cafeteria, Bruno afirma:

21) Bruno: Fiquei sabendo que o *bubble tea* dessa cafeteria recebeu um prêmio.

- 22) $p_4: \{w \in W \mid \text{o } bubble\ tea \text{ da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w\}$

E Ana aceita essa nova informação, então a proposição de (21) é adicionada ao CG:

- 23) $CG_{II} = \{\{w \in W \mid \text{O aniversário de Bruno é no dia 18/4 em } w\}, \{w' \in W \mid \text{chove o lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}, \{w'' \in W \mid \text{o } bubble\ tea \text{ da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}\}$

Consequentemente, o novo CS passa a ser:

- 24) $CS_{II} = \{\{w \in W \mid \text{O aniversário de Bruno é no dia 18/4 em } w\} \cap \{w' \in W \mid \text{Chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\} \cap \{w'' \in W \mid \text{O } bubble\ tea \text{ da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}\}$

O *question set*, doravante QS, se baseia no pressuposto teórico de que toda conversação está voltada para responder perguntas. Roberts (2012, p.65) defende que todo processo conversacional está voltado para responder perguntas, ao que ela chama de conjunto de perguntas em discussão. Essas perguntas direcionam o discurso para indicar o que é relevante e o que não é em uma conversa. Elas podem ser implícitas ou explícitas e há uma relação de hierarquia entre questões mais amplas e questões mais específicas. Seguimos com a seguinte interação entre Ana e Bruno:

- 25) a. Ana: Que incrível! Você já experimentou o *bubble tea* daqui?
Bruno: Já! É muito gostoso mesmo.
b. Ana: Qual sabor você recomenda? Recomenda o de *matcha*?
Bruno: Recomendo sim, experimenta.

No diálogo (25)b, percebe-se essa relação hierárquica, em que a pergunta mais geral é *Qual sabor você recomenda?* e a pergunta mais específica é *Recomenda o de matcha?*. Vale ressaltar, porém, que todas as perguntas em (25) podem ser perguntas mais específicas da pergunta implícita 'Como é a cafeteria em que Ana e Bruno conversam?'.

O QS é um repositório composto por todas as perguntas em discussão. Ou seja, todas as sentenças interrogativas explícitas ou implícitas que estejam disponíveis para serem respondidas na situação comunicativa. Quando um falante introduz um assunto geral novo ou faz uma nova pergunta, essas questões são inseridas no QS. Assim, a cada pergunta proferida explicitamente e a cada pergunta implícita que guia o diálogo, o QS aumenta. Por outro lado, quando uma interrogativa é respondida completamente, ela é retirada do conjunto, que diminui (Ferreira, 2023; Roberts, 2012).

Formalmente, o QS é representado por um conjunto de questões, ou seja um conjunto contendo as proposições das partições feitas pela sentença interrogativa em W . No exemplo em (25), têm-se o QS inicial (QS_I), o QS intermediário (QS_{II}) e o QS final (QS_{III}), como apresentado em (26), (27) e (28) respectivamente:

- 26) $[[QS_I]] = [[\text{Você já experimentou o } bubble\ tea \text{ daqui?}]] = \{\{w \in W \mid \text{Bruno já experimentou o } bubble\ tea \text{ da cafeteria em que ele e Ana conversam em } w\}, \{w' \in W \mid \text{Bruno não experimentou o } bubble\ tea \text{ da cafeteria em que ele e Ana conversam em } w'\}\}, Q_2, Q_3, \dots Q_n\}$

Com a resposta completa de Bruno a pergunta em (25)a e as novas perguntas feitas por Ana em (25)b, têm-se o QS intermediário:

- 27) $[[QS_{II}]] = \{\{w \in W \mid \text{Bruno recomenda o sabor tradicional em } w\}, \{w' \in W \mid \text{Bruno recomenda todos os sabores em } w'\}, \{w'' \in W \mid \text{Bruno não recomenda nenhum sabor em } w''\}, \{w''' \in W \mid \text{Bruno recomenda o sabor de } matcha \text{ em } w'''\}, \{w'''' \in W \mid \text{Bruno não recomenda o sabor de } matcha \text{ em } w''''\}\}, Q_2, Q_3, \dots Q_n\}$

Como a resposta de Bruno em (25)b responde completamente a pergunta mais específica e é apenas uma resposta parcial à pergunta mais geral, têm-se o QS final, em que a proposição: *Bruno não recomenda o sabor de matcha em w* é eliminada do conjunto de questões.

- 28) QS_{III} : Qual sabor você recomenda? $\{\{w \in W \mid \text{Bruno recomenda o sabor tradicional em } w\}, \{w' \in W \mid \text{Bruno recomenda o sabor de } matcha \text{ em } w'\}, \{w'' \in W \mid \text{Bruno recomenda todos os sabores em } w''\}, \{w''' \in W \mid \text{Bruno não recomenda nenhum sabor em } w'''\}\}, Q_2, Q_3, \dots Q_n\}$

A seguir, são apresentados todos os passos da atualização contextual que ocorre no diálogo em (25), demonstrando como CG, CS e QS interagem na dinâmica contextual. No início da conversa em (25)a, o CG e o CS são os mesmos apresentados após a asserção (21):

- 29) $CG_{II} = \{\{w \in W \mid \text{O aniversário de Bruno é no dia 18/4 em } w\}, \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}, \{w'' \in W \mid \text{o } bubble \text{ tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}\}$
- 30) $CS_{II} = \{\{w \in W \mid \text{O aniversário de Bruno é no dia 18/4 em } w\} \cap \{w' \in W \mid \text{Chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\} \cap \{w'' \in W \mid \text{O } bubble \text{ tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}\}$

Com a pergunta de Ana em (25)a, *Você já experimentou o bubble tea daqui?*, o conjunto de todos os mundos possíveis W é dividido em $\{w \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}$, para a resposta afirmativa, e $\{w \in W \mid \text{Bruno não experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}$, para a resposta negativa. A pergunta é adicionada ao QS:

- 31) QS_{III} : $[[\text{Você já experimentou o } bubble \text{ tea daqui?}]] = \{\{w \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}, \{w' \in W \mid \text{Bruno não experimentou o } bubble \text{ tea da cafeteria em que ele e Ana conversam em } w'''\}\}, Q_2, Q_3, \dots Q_n\}$

A partir da resposta de Bruno, *Já! É muito gostoso mesmo*, os mundos em que $\{w \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}$ são selecionados e os mundos em que $\{w \in W \mid \text{Bruno não experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}$ são desconsiderados. A proposição da resposta de Bruno passa a fazer parte do CG e, consequentemente, do CS:

- 32) $CG_{III} = \{\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\}, \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}, \{w'' \in W \mid \text{o } bubble \text{ tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}, \{w''' \in W \mid \text{Bruno já experimentou o } bubble \text{ tea da cafeteria em que ele e Ana conversam em } w'''\}\}$

- 33) $CS_{III} = \{\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\} \cap \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\} \cap \{w'' \in W \mid \text{o bubble tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\} \cap \{w \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w\}\}$

Como a pergunta de Ana foi respondida completamente, ela é removida do QS. Em (25)b, a pergunta *Qual sabor você recomenda?* reparte o conjunto W em $\{w \in W \mid \text{Bruno recomenda o sabor tradicional em } w\}$, $\{w' \in W \mid \text{Bruno recomenda o sabor de matcha em } w'\}$, $\{w'' \in W \mid \text{Bruno recomenda todos os sabores em } w''\}$ e $\{w''' \in W \mid \text{Bruno não recomenda nenhum sabor em } w'''\}$. A segunda pergunta feita por Ana, 'Recomenda o de matcha?', reparte W em $\{w \in W \mid \text{Bruno recomenda o sabor de matcha em } w\}$ e $\{w' \in W \mid \text{Bruno não recomenda o sabor de matcha em } w'\}$. Ambas as perguntas são adicionadas ao QS, gerando o QS_{III} :

- 34) $QS_{III} : [[\text{Qual sabor você recomenda?}]] = \{\{w \in W \mid \text{Bruno recomenda o sabor tradicional em } w\}, \{w' \in W \mid \text{Bruno recomenda o sabor de matcha em } w'\}, \{w'' \in W \mid \text{Bruno recomenda todos os sabores em } w''\}, \{w''' \in W \mid \text{Bruno não recomenda nenhum sabor em } w'''\}\}$
 $[[\text{Recomenda o de matcha?}]] = \{\{w \in W \mid \text{Bruno recomenda o sabor de matcha em } w\}, \{w' \in W \mid \text{Bruno não recomenda o sabor de matcha em } w'\}, Q_2, Q_3, \dots Q_n\}$

A resposta de Bruno, *Recomendo sim, experimenta*, seleciona os mundos $\{w \in W \mid \text{Bruno recomenda o sabor de matcha em } w\}$ e desconsidera os mundos $\{w \in W \mid \text{Bruno não recomenda o sabor de matcha em } w\}$. A proposição é adicionada ao CG e ao CS:

- 35) $CG_{IV} = \{\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\}, \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\}, \{w'' \in W \mid \text{o bubble tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\}, \{w''' \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w'''\}, \{w'''' \in W \mid \text{Bruno recomenda o sabor de matcha em } w''''\}\}$
 36) $CS_{IV} = \{\{w \in W \mid \text{o aniversário de Bruno é no dia 18/4 em } w\} \cap \{w' \in W \mid \text{chove do lado de fora da cafeteria em que Ana e Bruno conversam em } w'\} \cap \{w'' \in W \mid \text{o bubble tea da cafeteria em que Ana e Bruno conversam recebeu um prêmio em } w''\} \cap \{w''' \in W \mid \text{Bruno já experimentou o bubble tea da cafeteria em que ele e Ana conversam em } w'''\} \cap \{w'''' \in W \mid \text{Bruno recomenda o sabor de matcha em } w''''\}\}$

Como essa é uma resposta completa à segunda pergunta em (25)b, a interrogativa é removida do QS. Porém, como nada é afirmado explicitamente sobre Bruno recomendar ou não os outros sabores que não *matcha*, a resposta é apenas parcial no caso da primeira pergunta. Essa se mantém no QS:

- 37) $QS_{IV} : [[\text{Qual sabor você recomenda?}]] = \{\{w' \in W \mid \text{Bruno recomenda o sabor tradicional em } w'\}, \{w'' \in W \mid \text{Bruno recomenda todos os sabores em } w''\}, \{w''' \in W \mid \text{Bruno não recomenda nenhum sabor em } w'''\}\}, Q_2, Q_3, \dots Q_n\}$

Dessa forma CG, CS e QS interagem entre si para compor e atualizar a dinâmica contextual. Esses conceitos serão importantes para a análise feita na seção 4, que abordará a dinâmica contextual presente em diálogos com sentenças interrogativas que apresentam falhas de pressuposição.

4 As sentenças interrogativas polares e fenômeno da falha de pressuposição

Nesta seção, serão analisados diálogos que apresentem falha de pressuposição na sentença interrogativa polar, a partir dos conceitos e teorias explorados nas seções 2 e 3. Nosso objetivo é identificar, na dinâmica contextual, o que ocorre quando há falha de pressuposição. Uma das perguntas a ser respondida é: Quais contextos favorecem a acomodação e em quais há a contestação da falha?

A fim de descrever como ocorre essa interação, foram criadas sentenças interrogativas polares em que há uma pressuposição desencadeada por uma descrição definida singular, como o exemplo em (38):

38) O gato da pracinha foi adotado?

Na interrogativa (38), a descrição definida singular é *o gato da pracinha*. Descrições definidas singulares disparam pressuposição de existência e de unicidade. Ou seja, uma descrição definida singular pressupõe que: a) existe um referente capaz de satisfazer o predicado e b) não existe mais do que um referente capaz de satisfazer o predicado (Ferreira, 2019, p. 79; Levinson, 1983, p. 171).

Para a análise do fenômeno da falha de pressuposição foram criadas respostas em situações que há: i) acomodação da informação pressuposta; ii) contestação explícita da pressuposição de existência; iii) contestação implícita da pressuposição de existência; iv) contestação explícita da pressuposição de unicidade; v) contestação implícita da pressuposição de unicidade.

Para fins de comparação, descreve-se primeiramente o que ocorre nas respostas em situações nas quais não há falha de pressuposição. Consideramos que a interação ocorre entre Ana e Bruno, moradores de uma mesma vizinhança. Nessa vizinhança, há uma praça pública conhecida como pracinha e tanto Ana como Bruno sabem que há um gato de rua vivendo nela. Então, nessa situação, têm-se os CG e CS simplificados:

39) $CG_1 = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há um gato de rua vivendo na pracinha em } w'\}\}$

40) $CS_1 = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há um gato de rua vivendo na pracinha em } w'\}\}$

Durante a interação, ocorre o diálogo em (41):

41) Ana: O gato da pracinha foi adotado?
Bruno: Foi sim! Alguém levou ele ontem.

Alternativamente, Bruno também pode responder algo como:

42) Ana: O gato da pracinha foi adotado?
Bruno: Não foi. Vi ele lá na pracinha hoje de manhã.

Tanto em (41) como em (42), a pergunta de Ana divide o conjunto W em $\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}$, para a resposta afirmativa, e $\{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}$, para a resposta negativa. A pergunta está disponível no QS:

- 43) $QS_I = \{[Q_I] = \{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}\}, Q_2, Q_3, \dots Q_n\}$

Em (41) a resposta de Bruno, 'Foi sim! Alguém levou ele ontem.', seleciona o conjunto $\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}$ e descarta o conjunto restante. Inversamente, em (43), a resposta de Bruno, 'Não foi. Vi ele lá na pracinha hoje de manhã.', seleciona o conjunto $\{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}$ e descarta o conjunto restante. Em ambos os casos, o CG aumenta com a nova proposição introduzida pela resposta de Bruno e o CS diminui. Explicando o diálogo em (41), têm-se:

- 44) $CG_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há um gato de rua vivendo na pracinha em } w'\}, \{w \in W \mid \text{o gato da pracinha foi adotado em } w'\}\}$
- 45) $CS_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há um gato de rua vivendo na pracinha em } w'\} \cap \{w \in W \mid \text{o gato da pracinha foi adotado em } w'\}\}$

Independente de qual a resposta de Bruno, a pergunta de Ana é respondida completamente e, portanto, removida do QS:

- 46) $QS_{II} = \{Q_2, Q_3, \dots Q_n\}$

Daqui em diante, representaremos o QS por variáveis como $Q_2, Q_3, \dots Q_n$ para que a apresentação fique mais enxuta. A pergunta Q_1 , que corresponde a *O gato da pracinha foi adotado?*, foi eliminada do QS, uma vez que foi totalmente respondida.

Em suma, nos casos em que não há falha de pressuposição de existência do referente: o falante faz a pergunta polar que contém a pressuposição de existência disparada por descrição definida singular; a pergunta divide o conjunto W entre o conjunto de mundos correspondentes à resposta afirmativa e o conjunto de mundos correspondentes à resposta negativa; a pergunta está disponível no QS; o ouvinte responde à pergunta e a sua resposta seleciona um dos conjuntos (afirmativo ou negativo). Ao final da interação, o conjunto selecionado pela resposta é inserido ao CG e ao CS e a pergunta é removida do QS.

Feita a descrição do que ocorre com a dinâmica do diálogo em (41), em que não há falha de pressuposição, exploraremos daqui para frente os casos em que há falha de pressuposição, que são objeto de estudo da presente pesquisa. Para essas situações, considera-se que Ana e Bruno moram numa mesma vizinhança e que existe nessa vizinhança uma praça pública, a pracinha; mas apenas Ana sabe ou acredita que há um gato vivendo nessa praça. Como a informação pressuposta não é compartilhada por todos os participantes da conversa, ela não faz parte do CG e nem do CS, o que constitui uma falha de pressuposição. Vale notar que quando há falha de pressuposição, a informação pressuposta não é acarretada pelo CS e, por isso, a pergunta é considerada pragmaticamente inadequada. Sendo assim, os novos CG e CS iniciais para os diálogos que serão analisados adiante são:

- 47) $CG_1 = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}\}$
- 48) $CS_1 = \{\{w' \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w'\}\}$

Essa nova situação comunicativa e os novos CG e CS serão considerados como ponto de partida para todos os exemplos que se seguem. Para dar continuidade à descrição, segue em (49) uma interação em que há acomodação da pressuposição presente na interrogativa:

- 49) Ana: O gato da pracinha foi adotado?
Bruno: Não sabia que tinha um gato na pracinha.

Independente da situação comunicativa considerada, a pergunta de Ana faz uma partição do conjunto W em $\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}$ e $\{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}$. Além disso, a pergunta fica disponível no QS:

- 50) $QS_1 = \text{O gato da pracinha foi adotado?}_1 \{\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}, Q_2, Q_3, \dots Q_n\}$

Em (49), Bruno não tem nenhuma informação sobre o gato da pracinha em suas pressuposições de falante. Então, ele não pode dar uma resposta afirmativa ou negativa, que selecione algum dos conjuntos em que W foi particionado. Como a pergunta de Ana não é e não pode ser respondida completamente por Bruno, ela se mantém em aberto no QS:

- 51) $QS_{II} = \{\text{O gato da pracinha foi adotado?}_1 \{\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}\}, Q_2, Q_3, \dots Q_n\}$

A acomodação é um processo de reparação contextual no qual o ouvinte trata a pressuposição como se fosse uma asserção, inserindo-a no CG e no CS (Birner, 2013, p. 165; Lewis, 1979, p. 341). Dessa forma, apesar de Bruno não conseguir responder à pergunta de Ana, ele aceita a existência de um gato que vive na praça e essa informação é inserida no CG e no CS:

- 52) $CG_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há um gato na pracinha em } w'\}\}$
- 53) $CS_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há um gato na pracinha em } w'\}\}$

A pergunta de Ana em (49), que era considerada inadequada por causa da falha de pressuposição, passa a ser considerada adequada com a inserção da pressuposição no CG e no CS. Porém, essa pergunta continua disponível no QS, como demonstrado em (54):

- 54) $QS_{III} = \text{O gato da pracinha foi adotado?}_1 \{\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}, Q_2, Q_3, \dots Q_n\}$

Resumidamente, quando há falha de pressuposição seguida de acomodação: o falante faz a pergunta polar que contém pressuposição de existência disparada por descrição definida

singular; a pergunta particiona W em dois, o conjunto de mundos que correspondem à resposta afirmativa e o conjunto de mundos que correspondem à resposta negativa; a falha de pressuposição torna a pergunta inadequada; o interlocutor acomoda a pressuposição; a pressuposição é inserida no CG e no CS; com a acomodação, a pergunta passa a ser adequada e a pergunta fica disponível no QS; o ouvinte não pode responder à pergunta, que é mantida em aberto no QS. Ao final da interação, a informação pressuposta é adicionada ao CG e ao CS, resolvendo a falha de pressuposição. A pergunta continua em aberto no QS, pois não foi respondida, mas se torna adequada à situação comunicativa. Enquanto nos contextos em que não há falha de pressuposição, a asserção (afirmativa ou negativa) é adicionada ao CG e ao CS; nos contextos em que há falha sucedida de acomodação, apenas a proposição correspondente à pressuposição é adicionada ao CG e ao CS. Essas considerações podem ser resumidas no quadro 1:

Quadro 1 – Quadro comparativa entre diálogos com e sem falha de pressuposição

TIPO DE FALHA	COMMON GROUND	CONTEXT SET	QUESTION SET
Sem falha	Uma proposição é adicionada (asserção) e o CG aumenta.	Uma proposição é adicionada (asserção) e o CS diminui.	A questão é respondida e eliminada do QS.
Com falha e acomodação	Uma proposição é adicionada (pressuposição) e o CG aumenta.	Uma proposição é adicionada (pressuposição) e o CS diminui.	A questão não é respondida e é mantida no QS.

Fonte: elaboração das autoras (2025)

A seguir, descrevemos o que ocorre nos casos em que há contestação da informação pressuposta. Mais especificamente, o exemplo (55) se refere a resposta em situação que há contestação explícita da pressuposição de existência. Aqui, o termo: *pressuposição de existência* é empregado para se referir especificamente à existência de um referente capaz de satisfazer o predicado. Na pergunta em (55), como já discutido, há uma pressuposição de que existe um gato na pracinha, desencadeada pela descrição definida singular: ‘o gato da pracinha’. Essa é contestada com a resposta de Bruno:

- 55) Ana: O gato da pracinha foi adotado?
Bruno: Não tem nenhum gato na pracinha.

A pergunta de Ana, como já vimos, faz uma partição do conjunto W em $\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}$ e $\{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}$. A pergunta é inserida no QS:

- 56) $QS_1 = \{\text{O gato da pracinha foi adotado?}, \{\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}, Q_2, Q_3, \dots, Q_n\}$

Diferentemente do que ocorre em casos de acomodação, nos casos de contestação, Bruno (o ouvinte) possui em suas pressuposições de falante uma informação que entra em conflito com a pressuposição presente na interrogativa. É essa informação que o leva a rejeitar a contestá-la. Para determinar a natureza do conflito gerado entre a pressuposição e a contestação, é necessário saber qual a relação entre a informação pressuposta e a proposição expressa pela contestação, se há uma relação de contrariedade ou de contraditoriedade:

- 57) Ana: O gato da pracinha foi adotado?
 Pressuposição: Há um gato na pracinha.
 Bruno: Não tem nenhum gato na pracinha.
 Contestação: Não há um gato na pracinha.

Considerando o quadrado das oposições originalmente apresentado por Aristóteles e apresentado em Ferreira (2022, p. 84) a relação de contraditoriedade é caracterizada por valores de verdade opostos entre si. As proposições não podem ser verdadeiras ao mesmo tempo (no mesmo mundo), nem falsas ao mesmo tempo, porque a verdade de uma implica na falsidade da outra e vice-versa. Como se observa em (57) a pressuposição da pergunta de Ana e a contestação de Bruno apresentam proposições contraditórias entre si: se é verdade que há um gato na pracinha, é necessariamente falso que não há nenhum gato na pracinha; e se é verdade que não há nenhum gato na pracinha, é necessariamente falso que há um gato na pracinha. Nesse caso, a proposição de que existe um gato na pracinha é verdadeira para Ana, mas é falsa para Bruno, por isso ele a contesta. Se Ana aceitar essa contestação, a pressuposição da pergunta é cancelada. Dessa forma, a informação pressuposta, que já não fazia parte do CG e do CS por causa da falha, não pode ser acomodada e inserida em nenhum dos dois. Em seu lugar, é adicionada ao CG e ao CS a proposição da contestação:

- 58) $CG_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{não há um gato na pracinha em } w'\}\}$
 59) $CS_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{não há um gato na pracinha em } w'\}\}$

Com o cancelamento da pressuposição contida na pergunta de Ana, essa não pode mais ser respondida. Não faz sentido dizer que o gato da pracinha foi adotado ou que o gato da pracinha não foi adotado quando não existe um gato da pracinha. Pragmaticamente, a pergunta permanece inadequada. Então, Q_i deixa de fazer parte do conjunto de perguntas em discussão e é removida do QS:

- 60) $QS_{II} = \{Q_2, Q_3, \dots, Q_n\}$

Vamos atualizar nosso quadro com descrição da contestação explícita da pressuposição de existência e a inclusão do tipo de oposição estabelecida:

Quadro 2 – Quadro comparativo entre diálogos com falha de pressuposição: acomodação e contestação

TIPO DE FALHA	COMMON GROUND	CONTEXT SET	QUESTION SET	TIPO DE OPOSIÇÃO
Sem falha	Uma proposição é adicionada (asserção) e o CG aumenta.	Uma proposição é adicionada (asserção) e o CS diminui.	A questão é respondida e eliminada do QS.	Não há.
Com acomodação	Uma proposição é adicionada (pressuposição) e o CG aumenta.	Uma proposição é adicionada (pressuposição) e o CS diminui.	A questão não é respondida e é mantida no QS.	Não há.
Com contestação explícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contraditoriedade

Fonte: elaboração das autoras (2025)

Em seguida, em (61), será analisado um contexto em que a mesma pergunta dos diálogos anteriores é feita e a resposta apresenta uma contestação implícita da pressuposição de existência:

- 61) Ana: O gato da pracinha foi adotado?
Bruno: Só tem cachorro na pracinha.

Como já vimos anteriormente, a pergunta de Ana faz uma partição do conjunto W em dois subconjuntos de mundos. A pergunta é inserida no QS:

- 62) $QS_1 = \{ \text{O gato da pracinha foi adotado?} \{ \{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\} \}, Q_2, Q_3, \dots, Q_n \}$

Em situações de contestação, o ouvinte possui em suas pressuposições de falante uma informação que entra em conflito com a pressuposição da pergunta. Para determinar a natureza desse conflito, é necessário saber qual relação entre a informação pressuposta com a proposição expressa pela contestação:

- 63) Ana: O gato da pracinha foi adotado?
Pressuposição: Há um gato na pracinha.
Bruno: Só tem cachorro na pracinha.
Contestação: Há apenas cachorro na pracinha.

Mais uma vez, retomando o quadrado de oposições aristotélico, Ferreira (2022, p. 84) afirma que a relação de contrariedade ocorre quando a verdade de uma proposição implica na falsidade da outra. Em outras palavras, duas proposições são contrárias entre si quando não podem ser verdadeiras ao mesmo tempo, mas podem ser falsas ao mesmo tempo (Ferreira, 2022, p. 85). Em (63), tem-se uma relação de contrariedade entre a informação pressuposta e a proposição expressa pela contestação: se é verdade que há um gato na pracinha, é necessariamente falso que há apenas cachorro na pracinha. Porém, há mundos possíveis em que

ambas as proposições podem ser falsas, nos casos em que não há gatos nem cachorros na pracinha. Se aceita por Ana, a contestação cancela a informação pressuposta, que não é inserida no CG e no CS. Em seu lugar entra a proposição da contestação:

- 64) $CG_{||} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há apenas cachorro na pracinha em } w'\}\}$
- 65) $CS_{||} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há apenas cachorro na pracinha em } w'\}\}$

Com o cancelamento da pressuposição, a pergunta permanece pragmaticamente inadequada. Como não se assume a existência do referente indicado pela pressuposição, a pergunta Q não pode ser respondida e é removida do QS:

- 66) $QS_{||} = \{Q_2, Q_3, \dots, Q_n\}$

Em síntese, o que ocorre nas situações em que há falha de pressuposição seguida de contestação da pressuposição de existência do referente: o falante faz a pergunta polar contendo a pressuposição de existência desencadeada por descrição definida singular; a pergunta divide o conjunto W em dois, o conjunto de mundos possíveis que correspondem a resposta afirmativa e o conjunto de mundos possíveis que correspondem a resposta negativa; a falha de pressuposição torna a pergunta inadequada; a pergunta é inicialmente está disponível no QS; o ouvinte contesta a informação pressuposta; se o falante aceita a contestação do ouvinte, a pressuposição na pergunta é cancelada. Ao final da interação, a proposição da contestação é inserida no CG e no CS e a pergunta é removida do QS.

É interessante notar que nos casos em (55) e (61), em que há a contestação da existência de um referente, há uma diferença entre a relação pressuposição-contestação nas respostas explícitas e nas respostas implícitas. Nas respostas explícitas, a relação é de contradição, que reparte o conjunto dos mundos possíveis W em duas partes: uma em que a pressuposição é verdadeira e a contestação falsa; e uma em que a contestação é verdadeira e a pressuposição falsa. Por outro lado, nas respostas implícitas a relação é de contrariedade, que reparte o conjunto dos mundos possíveis W em três partes: uma em que a pressuposição é verdadeira e a contestação falsa; uma em que a contestação é verdadeira e a pressuposição falsa; uma em que tanto a pressuposição como a contestação são falsas. Podemos observar essa distinção na sistematização a seguir:

Quadro 3 – Quadro comparativo entre diálogos com falha de pressuposição: acomodação, contestação da pressuposição de existência (explícita e implícita)

TIPO DE FALHA	COMMON GROUND	CONTEXT SET	QUESTION SET	TIPO DE OPOSIÇÃO
Sem falha	Uma proposição é adicionada (asserção) e o CG aumenta.	Uma proposição é adicionada (asserção) e o CS diminui.	A questão é respondida e eliminada do QS.	Não há.
Com acomodação	Uma proposição é adicionada (pressuposição) e o CG aumenta.	Uma proposição é adicionada (pressuposição) e o CS diminui.	A questão não é respondida e é mantida no QS.	Não há.
Com contestação explícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contraditoriedade
Com contestação implícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contrariedade

Fonte: elaboração das autoras (2025)

Na sequência, analisa-se a pergunta polar com resposta em situação em que há contestação da pressuposição de unicidade. Como já citado, uma descrição definida singular, além de disparar uma pressuposição de existência, também desencadeia a pressuposição de que existe um e somente um referente capaz de satisfazer o predicado. Em (67), a resposta a pergunta faz uma contestação explícita da pressuposição de unicidade:

- 67) Ana: O gato da pracinha foi adotado?
Bruno: Tem muitos gatos na pracinha.

Na contestação, o ouvinte possui em suas pressuposições de falante uma informação conflitante com a informação pressuposta:

- 68) Ana: O gato da pracinha foi adotado?
Pressuposição: Há apenas um gato na pracinha.
Bruno: Tem muitos gatos na pracinha.
Contestação: Há mais de um gato na pracinha.

Assim como em (63), no exemplo em (68) a relação também é de contrariedade. Ou seja, se é verdade que há apenas um gato na pracinha, então é necessariamente falso que há mais de um gato na pracinha. Porém, nos mundos possíveis em que não há nenhum gato na pracinha, ambas as proposições são falsas. Se aceita por Ana, a contestação de Bruno em (67) cancela a pressuposição, fazendo com que essa não possa ser adicionada ao CG e ao CS por acomodação. O que é inserido no CG e no CS é a proposição da contestação:

- 69) $CG_{II} = \{ \{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há mais de um gato na pracinha em } w'\} \}$

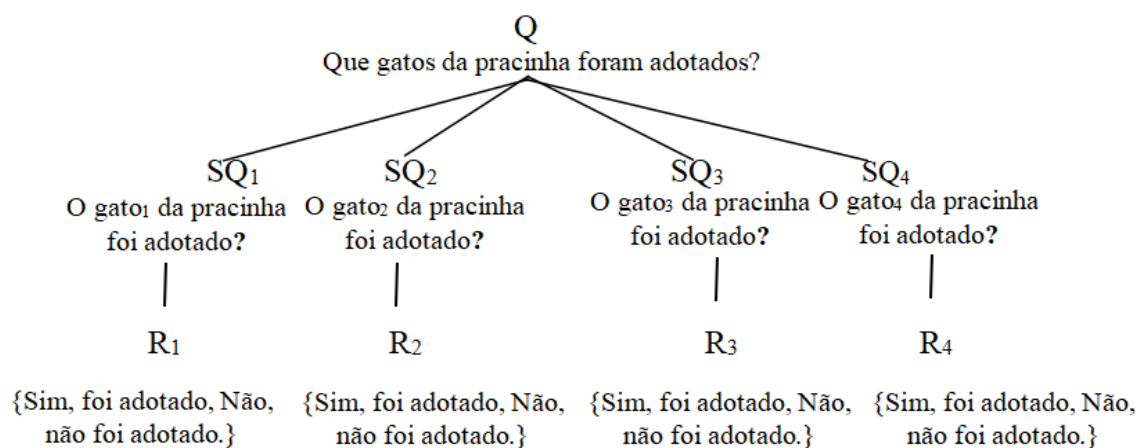
- 70) $CS_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há mais de um gato na pracinha em } w'\}\}$

Diferentemente dos casos em que há contestação da pressuposição da existência de um referente, o cancelamento da pressuposição da unicidade desse referente não inviabiliza a pergunta. Existem candidatos a referente para 'o gato da pracinha', por isso, a pergunta se mantém no QS, mas se subdivide:

- 71) $QS_{II} = \{\{w \in W \mid \text{o gato}_1 \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_1 \text{ da pracinha não foi adotado em } w'\}, \{w \in W \mid \text{o gato}_2 \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_2 \text{ da pracinha não foi adotado em } w'\}, \{w \in W \mid \text{o gato}_3 \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_3 \text{ da pracinha não foi adotado em } w'\} \dots \{w \in W \mid \text{o gato}_n \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_n \text{ da pracinha não foi adotado em } w'\}\}$

Essas relações podem ser representadas por uma *d-tree* ("árvore do discurso") (Büring, p. 516):

Imagem 1 – Árvore do discurso para (71)



Fonte: elaboração das autoras (2025)

Na imagem 1, temos uma questão geral: Q, que inclui todos os gatos da pracinha. Quando Bruno, em (67), contesta a pergunta de Ana, uma pergunta mais geral como Q, é adicionada ao QS e essa é subdividida em sub-questões, como as representadas por SQ₁, SQ₂, SQ₃ e SQ₄. Contudo, o uso da descrição definida singular é inadequado, na medida em que existe mais de um gato da pracinha e não um único gato. Para que a descrição definida singular e, conseqüentemente, a pergunta que a carrega sejam adequadas à situação comunicativa, seria necessário que o falante especificasse de qual gato está falando, como por exemplo *O gato preto da pracinha foi adotado?*. A seguir, apresentamos o quadro atualizado:

Quadro 4 – Quadro comparativo entre diálogos com falha de pressuposição: acomodação, contestação de existência (explícita e implícita), contestação de unicidade (explícita)

TIPO DE FALHA	COMMON GROUND	CONTEXT SET	QUESTION SET	TIPO DE OPOSIÇÃO
Sem falha	Uma proposição é adicionada (asserção) e o CG aumenta.	Uma proposição é adicionada (asserção) e o CS diminui.	A questão é respondida e eliminada do QS.	Não há.
Com acomodação	Uma proposição é adicionada (pressuposição) e o CG aumenta.	Uma proposição é adicionada (pressuposição) e o CS diminui.	A questão não é respondida e é mantida no QS.	Não há.
Com contestação explícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contraditoriedade
Com contestação implícita da pressuposição de existência	A proposição da contestação é adicionada ao CG. O CG aumenta.	A proposição da contestação é adicionada ao CS. O CS diminui.	A questão é inadequada e é eliminada do QS.	Contrariedade
Com contestação explícita da pressuposição de unicidade	A proposição da contestação é adicionada ao CG. O CG aumenta.	A proposição da contestação é adicionada ao CS. O CS diminui.	A questão se mantém no QS, mas é subdividida em perguntas mais específicas.	Contrariedade

Fonte: elaboração das autoras (2025)

Por fim, encerra-se a análise da interrogativa polar com pressuposição disparada por descrição definida singular com uma resposta em situação na qual há contestação implícita da pressuposição de unicidade, o exemplo (72):

- 72) Ana: O gato da pracinha foi adotado?
Bruno: O gato laranja foi, os outros não.

Em conformidade com todos os outros casos, a pergunta de Ana faz uma partição do conjunto W . Esta é incorporada ao QS:

- 73) $QS_1 = \{ \text{O gato da pracinha foi adotado?} \{ \{ w \in W \mid \text{o gato da pracinha foi adotado em } w \}, \{ w' \in W \mid \text{o gato da pracinha não foi adotado em } w' \}, Q_2, Q_3, \dots, Q_n \} \}$

Como ocorre nos casos em que há contestação, Bruno (o ouvinte) possui em suas pressuposições de falante uma informação que entra em conflito com a pressuposição. Então, faz-se necessário observar qual a natureza semântica entre a pressuposição e a contestação:

- 74) Ana: O gato da pracinha foi adotado?
Pressuposição: Há apenas um gato na pracinha.
Bruno: O gato laranja foi, os outros não.
Contestação: Há mais de um gato na pracinha.

Assim como em (61), na contestação implícita da pressuposição de unicidade em (72), a relação entre a informação pressuposta e a proposição da contestação é de contrariedade. Ou seja, se é verdade que há apenas um gato na pracinha, é necessariamente falso que há mais de um gato na pracinha e vice-versa. Porém, nas situações em que não há nenhum gato na pracinha, ambas as proposições são falsas. Se aceita por Ana, a contestação cancela a pressuposição, que não pode ser inserida no CG e no CS por acomodação. Em seu lugar, entra a proposição da contestação:

- 75) $CG_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\}, \{w' \in W \mid \text{há mais de um gato na pracinha em } w'\}\}$
- 76) $CS_{II} = \{\{w \in W \mid \text{na vizinhança de Ana e Bruno há uma praça pública conhecida como pracinha em } w\} \cap \{w' \in W \mid \text{há mais de um gato na pracinha em } w'\}\}$

O cancelamento da pressuposição de unicidade não exclui completamente as possibilidades de um referente para expressão *o gato da pracinha*. Isso faz com que ainda exista uma possibilidade de resposta para a pergunta, que se mantém no QS, mas se subdivide:

- 77) $QS_{II} = \{\text{O gato da pracinha foi adotado? } \{\{w \in W \mid \text{o gato da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato da pracinha não foi adotado em } w'\}\}, Q_2, Q_3, \dots, Q_n\}$
- 78) $QS_{III} = \{\{w \in W \mid \text{o gato}_1 \text{ foi adotado em } w\}, \{w' \in W \mid \text{o gato}_1 \text{ não foi adotado em } w'\}, \{\{w \in W \mid \text{o gato}_2 \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_2 \text{ da pracinha não foi adotado em } w'\}\} \dots \{\{w \in W \mid \text{o gato}_n \text{ da pracinha foi adotado em } w\}, \{w' \in W \mid \text{o gato}_n \text{ da pracinha não foi adotado em } w'\}\}\}$

Neste caso, a resposta de Bruno responde parcialmente a uma pergunta mais ampla do que a solicitada, como: *Quais gatos da pracinha foram adotados?* (Roberts, 2012, p. 5). Se o diálogo continua e Ana confirma que o gato ao qual ela se referia era o preto, então a resposta de Bruno seleciona o conjunto $\{w \in W \mid \text{o gato}_1 \text{ da pracinha foi adotado em } w\}$, por exemplo, e descarta o conjunto restante. Se o gato ao qual Ana se referia não era o preto, a resposta de Bruno seleciona o conjunto $\{w \in W \mid \text{o gato}_1 \text{ da pracinha não foi adotado em } w\}$ e descarta o conjunto restante. Com a resposta completa, a pergunta Q_1 é removida do QS.

- 79) $QS_{II} = \{Q_2, Q_3, \dots, Q_n\}$

Como observado nos diálogos em que há falha de pressuposição de existência, o cancelamento da pressuposição de existência do referente remove qualquer possibilidade de resposta para a pergunta, visto que não há um referente para o qual a descrição definida singular aponta. Em contrapartida, como observado em (61) e principalmente em (72), o cancelamento da pressuposição de unicidade do referente não remove todas as possibilidades de resposta para a pergunta. A pergunta ainda pode ser respondida e a descrição definida singular ainda pode apontar para um único referente, desde que este seja especificado.

Resumidamente, quando há falha de pressuposição seguida de contestação da pressuposição de unicidade do referente: o falante profere a pergunta polar com pressuposição de unicidade disparada por descrição definida singular; a pergunta parte o conjunto W em dois, o conjunto dos mundos que correspondem a resposta afirmativa e o conjunto dos mundos que correspondem a resposta negativa; a falha de pressuposição torna a pergunta inadequada; o

ouvinte contesta a informação pressuposta; se o falante aceita a contestação, a pressuposição é cancelada. Ao final da interação, a proposição da contestação é inserida no CG e no CS, mas a pergunta se mantém e é subdividida no QS. A sistematização completa desses resultados pode ser observada no quadro 5:

Quadro 5 – Quadro comparativo entre diálogos com falha de pressuposição: acomodação, contestação de existência (explícita e implícita), contestação de unicidade (explícita e implícita)

TIPO DE FALHA	COMMON GROUND	CONTEXT SET	QUESTION SET	TIPO DE OPOSIÇÃO
Sem falha	Uma proposição é adicionada (asserção) e o CG aumenta.	Uma proposição é adicionada (asserção) e o CS diminui.	A questão é respondida e eliminada do QS.	Não há.
Com acomodação	Uma proposição é adicionada (pressuposição) e o CG aumenta.	Uma proposição é adicionada (pressuposição) e o CS diminui.	A questão não é respondida e é mantida no QS.	Não há.
Com contestação explícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contraditoriedade
Com contestação implícita da pressuposição de existência	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão é inadequada e é eliminada do QS.	Contrariedade
Com contestação explícita da pressuposição de unicidade	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão se mantém no QS, mas é subdividida em perguntas mais específicas.	Contrariedade
Com contestação implícita da pressuposição de unicidade	A proposição da contestação é adicionada ao CG e o CG aumenta.	A proposição da contestação é adicionada ao CS e o CS diminui.	A questão se mantém no QS, mas é subdividida em perguntas mais específicas.	Contrariedade

Fonte: elaboração das autoras (2025)

5 Considerações finais

Como apresentado no presente artigo, uma sentença interrogativa polar com falha de pressuposição, assim como qualquer interrogativa polar, divide o conjunto *W* de todos os mundos possíveis em dois: um conjunto de mundos possíveis em que a proposição é verdadeira e um conjunto de mundos possíveis em que a proposição é falsa. Também como outras interrogativas polares comuns, a polar com falha de pressuposição está disponível no QS. Contudo, perguntas com falha de pressuposição são consideradas pragmaticamente inadequadas.

Diante desse cenário, o ouvinte pode reagir de duas formas diferentes: acomodar ou contestar a informação pressuposta presente na pergunta. Se o ouvinte não possuir em suas pressuposições de falante (Stalnaker, 1999, p. 49) uma informação que entre em conflito com a informação pressuposta, ele a acomoda. Por outro lado, se o ouvinte possuir em suas pressuposições de falante uma informação que entre em conflito com a informação pressuposta, ele a contesta.

Quando o ouvinte acomoda a pressuposição, esse processo age como uma ferramenta de reparo da falha de pressuposição. Em outras palavras, o ouvinte trata a informação pressuposta como uma informação nova, inserindo-a no CG e no CS, resolvendo, assim, essa falha. Dessa maneira, a sentença interrogativa, que era considerada pragmaticamente inadequada, passa a ser adequada. Todavia, como o ouvinte não tem nenhuma informação sobre o referente da pressuposição de existência, desencadeada por descrição definida singular, então ele não pode responder à pergunta, que permanece em aberto no QS.

Quando o ouvinte contesta a pressuposição, a contestação cancela a informação pressuposta, impedindo-a de entrar no CG e no CS. Em seu lugar, entra a proposição da contestação. Assim, a falha de pressuposição não é resolvida e o uso da interrogativa polar permanece sendo considerado pragmaticamente inadequado. Para as contestações da pressuposição de existência do referente, o cancelamento da pressuposição desencadeia a negação dessa pressuposição. Isso impede a pergunta de ser respondida pelo ouvinte, então, essa é removida do QS. Para as contestações da pressuposição de unicidade do referente, apenas que não se sabe qual o indivíduo específico que é pressuposto. Isso não impede a pergunta de ser respondida pelo ouvinte, então, ela é mantida no QS. Porém, diferente do que acontece nos casos de acomodação, no QS a pergunta se subdivide em outras mais específicas, para refletir a multiplicidade de possíveis referentes.

As interrogativas com pressuposição disparada por descrição definida singular apresentam diferenças entre contestações explícitas e implícitas da pressuposição de existência do referente: enquanto as contestações explícitas estabelecem relação de contraditoriedade com a informação pressuposta, as implícitas estabelecem relação de contrariedade. Já as contestações da pressuposição de unicidade do referente sempre estabelecem uma relação de contrariedade com a informação pressuposta, independente do tipo de contestação (explícita ou implícita).

Se considerarmos que a relação de contraditoriedade é mais forte do que a relação de contrariedade, então esses resultados podem indicar que a pressuposição de existência é mais forte e, conseqüentemente, mais dificilmente cancelável do que pressuposição de unicidade. Ainda, como houve diferença entre contestações explícitas e implícitas (uma vez que as contestações explícitas podem desencadear relação de contraditoriedade, enquanto as implícitas apenas contrariedade) pode-se afirmar que as respostas de contestação explícitas sejam mais fortes do que suas contrapartes implícitas.

Esperamos que essas reflexões e análises venham a contribuir com os estudos da pressuposição e da falha de pressuposição, sobretudo na abordagem mais formal de investigação do significado. Entretanto sabemos das limitações desta pesquisa, uma vez que focamos apenas nas interrogativas polares e na pressuposição desencadeada por descrição definida singular. Assim, ainda não se sabe o que ocorre com a falha de pressuposição em interrogativas de constituinte, nem o que ocorre quando a informação pressuposta é desencadeada por outros gatilhos. Ainda, é importante considerar que sejam desenvolvidas pesquisas empíricas futu-

ras, com aplicação de experimentos psicolinguísticos, a fim de testar se os processos comunicativos de acomodação e contestação da pressuposição ocorrem como descritos neste artigo ou não. Além disso, uma questão a ser investigada é se há diferenças de processamento em contextos de falha de pressuposição, comparando com contextos sem falha. Que este estudo seja um ponto de partida para pesquisas na interface entre a semântica e a pragmática formais, sobretudo na busca por compreender as relações lógico-contextuais que ocorrem na dinâmica conversacional do PB.

Declaração de autoria

Este artigo apresenta os resultados referentes à pesquisa de Iniciação Científica da segunda autora, sob orientação da primeira autora. No que diz respeito às contribuições de cada uma, a primeira autora realizou a revisão bibliográfica, auxiliou na criação de diálogos que foram objeto de análise, realizou a análise linguística conjuntamente com a segunda autora e escreveu a versão final do artigo. A segunda autora realizou a revisão bibliográfica, criou os diálogos que foram objeto de análise, realizou a análise linguística conjuntamente com a primeira autora, escreveu a primeira versão do artigo e fez a revisão final.

Agradecimentos

Agradecemos o apoio financeiro da Fundep – Fundação de Desenvolvimento da Pesquisa, por meio da concessão de bolsa de iniciação científica à segunda autora. Agradecemos igualmente à Universidade Federal de Minas Gerais, em especial à Faculdade de Letras, pelo suporte acadêmico, estrutural e administrativo oferecido ao longo do desenvolvimento desta pesquisa. Aproveitamos para agradecer a toda equipe editorial da RELIN, pelo suporte técnico e acadêmico. Por fim, registramos nosso agradecimento aos dois pareceristas anônimos, cujas leituras atentas e sugestões criteriosas foram fundamentais para o aprimoramento do artigo. Eventuais erros remanescentes são de nossa inteira responsabilidade.

Referências

- BEAVER, D. *Presupposition and Assertion in Dynamic Semantics*. Stanford: CSLI Publications, 2001.
- BIRNER, B. J. *Introduction to Pragmatics*. Oxford: Wiley-Blackwell, 2013.
- BORGES NETO, J. *A semântica das perguntas*. (Texto apresentado em simpósio durante o LV Seminário do GEL - Franca/SP, julho de 2007). Disponível em: https://docs.ufpr.br/~borges/publicacoes/para_download/Sem_Perguntas.pdf. Acesso em 31/10/2025.
- BÜRING, D. Contrastive Topic. In: FERY, C.; SHIN, I. (eds) *Handbook of Information Structure*. Oxford University Press, 2014.
- BÜRING, D. On D-trees, beans, and B-accent. *Linguistics & Philosophy*, Cidade, v. 26, n.5, p.511-545, 2003.

- CHIERCHIA, G. *Semântica* (Trad. Luis Arthur Pagani, *et al.*). Campinas, SP: Editora da UNICAMP; Londrina, PR: EDUEL, 2003.
- CHIERQUIA, G.; MCCONNELL-GINET, S. *Meaning and grammar: An introduction to semantics*. Cambridge: MIT Press, 1996.
- FERREIRA, M. *Curso de semântica formal*. Textbooks in Language Sciences 6.n: T Berlin: Language Science Press, 2019.
- FERREIRA, M. *Semântica: uma introdução ao estudo formal do significado*. São Paulo: Contexto, 2022.
- FERREIRA, M. *Pragmática: significado, comunicação e dinâmica contextual*. São Paulo: Contexto, 2023.
- GAZDAR, G. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press, 1979.
- GRICE, H. P. Logic and conversation. In: Cole, P.; Morgan, J. (eds.) *Syntax and Semantics*, vol. 3. New York: Academic Press, 1975. p. 41-58.
- GROENENDIJK, J.; STOKHOF, M. Questions. In: VAN BENTHEM, J.; TER MEULEN, A. (ed.). *Handbook of Logic and Language*. Amsterdam: Elsevier Science B.V., 1997.
- GROENENDIJK, J. The Logic of Interrogation: Classical Version. TIEKEN-BOON VAN OSTADE, I.; KOIER, G.; LOS, B.; VOS, M. (eds.). *Linguistics in the Netherlands 1999*. Amsterdam: John Benjamins, 1999. p. 13–24.
- HAMBLIN, C. L. Questions in Montague English. *Foundations of Language*, Dordrecht, v. 10, n. 1, p. 41–53, 1973.
- HEIM, I. File Change Semantics and the Familiarity Theory of Definiteness. In: R. BÄUERLE; C. SCHWARZE; A. VON STECHOW (orgs.). *Meaning, Use and the Interpretation of Language*. Berlin/New York: de Gruyter, 1983. p. 164–189.
- HEIM, H.; KRATZER, A. *Semantics in Generative Grammar*. Blackwell Textbooks in Linguistics, Oxford: Blackwell Publishers Ltd, 1998.
- KRIPKE, S. *Naming and Necessity*. Harvard University Press, 1980.
- LEWIS, D. Scorekeeping in a language games. *Journal of Philosophical Logic*, Dordrecht, v.8, n.1, p. 339-359, 1979. DOI: <https://doi.org/10.1007/BF00258436>
- LEVINSON, S. *Pragmatics*. Cambridge: Cambridge University Press, 1983.
- PARTEE, B. H. Some transformational extensions of Montague grammar. *Journal of Philosophical Logic*, Dordrecht, v.2, n.4, p. 509-534, 1973. DOI: <https://doi.org/10.1007/BF00262953>
- PIRES DE OLIVEIRA, R. *Semântica Formal: uma breve introdução*. Mercado das Letras – 2ª ed., 2010.
- ROBERTS, C. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, v. 5, n. 6, p. 1–69, 2012. DOI: <http://dx.doi.org/10.3765/sp.5.6>.
- QUADROS-GOMES, A. P.; SANCHEZ-MENDES, L. *Para conhecer: semântica*. São Paulo: Contexto, 2018.
- STALNAKER, R. C. *Context and Content*. New York: Oxford University Press, 1999.
- STRAWSON, P. F. On referring. *Mind, New Series*, v. 59, n. 235, p. 320-344, 1950. DOI: <https://doi.org/10.1093/mind/LIX.235.320>

Construções insubordinadas causais no português brasileiro: um estudo da fala espontânea

Insubordinated Constructions in Brazilian Portuguese Speech: A Study of Spontaneous Speech

**Alessandra Emanuelle
Macieira Silva**

Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
aems@ufmg.br
<https://orcid.org/0009-0005-9008-6665>

Luana Lopes Amaral

Universidade Federal de Minas Gerais
(UFMG) | Belo Horizonte | MG | BR
luanalopes@let.grad.ufmg.br
<https://orcid.org/0000-0002-4290-1208>

Resumo: Nesta pesquisa, analisamos as construções insubordinadas na fala espontânea do português brasileiro, com foco nas construções causais, como *eu acho que sim // porque ela trabalha o dia inteiro //*. Adotamos como base teórica a Gramática de Construções (Croft; 2001, 2022) e os trabalhos sobre insubordinação, desenvolvidos na literatura sobre o tema (Bossaglia; Mello e Raso, 2020; Evans, 2007; Mithun, 2008). De acordo com esse aparato teórico, a relação entre partes da construção é de tipo semântico. Conforme mostram os dados deste artigo, a insubordinada é prosodicamente independente, mas exerce funções discursivas próprias. Por se tratar de dados autênticos e representativos da dia-mesia falada, as 286 ocorrências foram extraídas do *corpus* C-ORAL BRASIL (Raso; Mello, 2012) e as construções analisadas foram observadas quanto à função pragmática e conteúdo semântico. Os dados mostraram que as construções insubordinadas causais analisadas, apesar de apresentarem sintaxe de subordinadas causais, estabelecem relações semânticas com as orações que as precedem, especificamente uma relação de justificação enunciativa, de modo que a sintaxe de subordinação fica em segundo plano. Logo, a construção é vista como um resultado do uso da língua. Há, ainda, casos em que as insubordinadas apresentam uma relação com outro ato de fala contíguo, no qual, diferentemente das insubordinadas, uma aparente oração principal é identificada e, formam, portanto, as construções semi-insubordinadas.

Palavras-chave: insubordinação; construções causais; Gramática de Construções; fala espontânea.



Abstract: In this research, we analyze the insubordinated constructions in spontaneous Brazilian Portuguese speech, focusing on causal constructions, based on Construction Grammar (Croft, 2001, 2022) and the studies on insubordination found in the literature on the subject (Bossaglia; Mello and Raso, 2020; Evans, 2007; Mithun, 2008). According to this theoretical framework, the relationships between parts of the construction are purely semantic, and there are no syntactic relationships, only roles played within the construction. Since the data is authentic and representative of spoken diamesis, the occurrences were extracted from the C-ORAL BRASIL *corpus* (Raso; Mello, 2012), with a focus on enunciative justification constructions regarding their pragmatic function and semantic content. The results showed that the constructions establish semantic relationships between the main and the subordinate clause, so that syntax takes a secondary role, and the construction is seen as a result of language use. There also cases in which insubordinate clauses show a relation to another contiguous speech act, in which an apparent main clause is produced and therefore forms the semi-insubordinate constructions.

Keywords: insubordination; causal constructions; Construction Grammar; spontaneous speech.

1 Introdução

Este trabalho se dedica a analisar a insubordinação em dados de fala espontânea do português brasileiro, com foco na realização de construções causais insubordinadas, com base na Gramática de Construções (Goldberg, 1995, 2006; Croft, 2001, 2022). As orações insubordinadas são aquelas que, segundo Evans (2007), apresentam características de orações subordinadas, mas são realizadas de forma independente e assumem diferentes funções discursivas como: controle interpessoal, modalização e sinalização de material pressuposto. Ou seja, são orações que apresentam marcas de subordinação, mas apresentam uso independente e convencionalizado como oração “principal” de uma oração complexa, mesmo que, à primeira vista, pareçam ser formalmente subordinadas.

Nesse ponto de vista, a insubordinação deve ser avaliada também no nível semântico, pragmático e discursivo; e não apenas sintático (Evans; Watanabe, 2016). No plano discursivo, conforme afirma Mithun (2008), é possível reconhecer funções metadiscursivas e dialógicas das construções. Prosodicamente, como observado por Schwenter (2016), em sua pesquisa

sobre insubordinadas condicionais no espanhol, as construções apresentam contorno entonacional específico, mais parecido com enunciados declarativos, em vez de apresentarem contorno prosódico não final. Logo, as orações condicionais independentes, conforme observado por Schwenter (2016), não apresentam entonação “inacabada” ou “não conclusiva”, e soam como completas, devido ao contorno entonacional mais conclusivo¹:

- 1) A: ¿Vamos a la playa?
 (“Vamos à praia?”)
 B1: *¿Si esté lloviendo!
 (“Se estiver chovendo!”)
 B2: *¿Si está lloviendo y si hace frío!
 (“Se está chovendo e está frio!”)

Em seu estudo, a autora observou que essas construções não aceitam formas verbais no subjuntivo, não permite a coordenação com outras orações e a entonação de “continuação ascendente”; ao contrário, pesquisas anteriores indicam que apresenta entonação exclamativa (Montolío, 1999; Schwenter, 2016). Vejamos mais um exemplo:

- 2) A: ¿A qué hora llegan los primos esta tarde?
 (“A que horas chegar:3PL os primos esta tarde?”)
 (“A que horas chegam nossos primos chegam esta tarde?”)
 B: ¿Si ya han llegado esta mañana!
 (“SI já haver:3PL chegar:PAPT esta manhã!”)
 (“Eles já chegaram esta manhã!”)
- 3) Si aún tenéis hambre... (falante aponta para uma panela cheia de comida)
 (“Se ainda ter:2PL fome”)
 (“Se você ainda tem fome...”)

A análise feita por Schwenter (2016, p. 76) sobre o estudo de orações *si* independentes refuta uma afirmação feita por outro falante em uma oração imediatamente anterior, e o movimento melódico final do enunciado se assemelha a contornos prosódicos finais, com tom de fronteira baixo, comum em enunciados declarativos. Por outro lado, em outros usos das insubordinadas, como em (3), o padrão melódico é de contorno entonacional não final, apresentando uma ideia de “suspensão” e “não conclusão” do enunciado, gerando uma inferência para interlocutor.

Nos dados observados nesta pesquisa, alguns padrões melódicos parecidos foram observados, nos quais a construção insubordinada, apresenta, prosodicamente, um perfil que orienta o interlocutor a fazer uma inferência a partir da construção realizada pelo falante. O exemplo abaixo, retirado do *corpus* C-ORAL BRASIL (Raso; Mello, 2012) e que faz parte dos dados desta pesquisa, mostra uma ocorrência de insubordinação:

¹ Traduções e adaptações de nossa autoria, retiradas de Schwenter (2016, p. 76).

4) [falantes conversam sobre a construção de um muro]²

*ROG³: eu vou &coloc [/3] eu vou suspender mais um pouquim aqui / vou pegar a linha / e vou colocar por cima //

*PAU: ah / **porque senão** //

As construções insubordinadas servem a propósitos comunicativos e discursivos entre falantes e ouvintes e as conjunções adverbiais veiculam significados diferentes, servindo ao contexto. No exemplo (4), a construção insubordinada *porque senão* não apresenta integração textual com a construção anterior *eu vou suspender mais um pouquim / vou pegar a linha e vou colocar para cima* no sentido de que cada uma compõe um enunciado diferente. Essa autonomia se dá nos níveis semântico-pragmático e prosódico. Funcionalmente, a construção é usada pelo falante para constatar um aspecto factual, uma obviedade, isto é, se a linha não for colocada, o muro pode desabar. Portanto, *PAU apresenta, em seu ato de fala, uma avaliação sobre o ato de fala de *ROG, constatando sua percepção em relação ao contexto situacional do ato de fala de seu interlocutor, e não do conteúdo locutivo.

Neste caso, as construções não apresentam nenhum material recuperável no contexto linguístico adjacente, sendo a estrutura, aparentemente dependente, totalmente solta do ponto de vista textual. Portanto, o uso do termo “integração textual” é uma referência ao fato de que as insubordinadas não se conectam textualmente à construção anterior como parte de uma estrutura lógica e coesa típica de uma oração subordinada e, portanto, não há integração textual com a construção adjacente visto que esta e a insubordinada são dois enunciados diferentes. Desse modo, são independentes sintática e pragmaticamente. Vejamos outro exemplo, também de nossos dados:

5) [falantes comentam a reforma do apartamento]

*CES: mas / só que a área de serviço ficou muito pequena //

*ANE: é / ué / isso é que [/3] é < isso que a > [/4] isso que a Alcione falou comigo / Anete / cê tem dúvida //⁴

*CES: < tendeu > //

*CES: é //

*CES: tem dúvida //

*CES: tendeu //

*CES: o' p' cê ver //

*CES: em relação / àquele apartamento //

*CES: < tendeu > //

*ANE: < **porque olha** > aqui //

*ANE: quer ver //

*ANE: esse aqui / tava aqui //

² Todos os áudios referentes aos dados presentes no artigo estão disponibilizados neste link: https://drive.google.com/drive/folders/1ImfTZDLp1d2qFouFlZSWOl7ViyzdJABi?usp=drive_link.

³ Segundo os critérios de transcrição do *corpus* C-ORAL BRASIL (Raso; Mello, 2012), os acrônimos precedidos por asterisco são utilizados para indicar os falantes.

⁴ A barra simples entre colchetes seguida de número marca o retracting, que ocorre quando o falante reformula seu conteúdo locutivo.

Neste caso, a construção *porque olha aqui* é uma estrutura insubordinada usada pela falante para justificar o seu próprio ato de fala, como defendemos neste artigo. Esse uso, inclusive, se torna mais evidente ao notarmos que a construção causal realizada por *ANE acontece após vários turnos de fala de *CES. Ou seja, a falante realiza uma afirmação para confirmar o que foi dito pela interlocutora (*isso que a Alcione falou comigo*) e em seguida, realiza uma pergunta para se certificar de que o conteúdo locutivo foi entendido (*cê tem dúvida*). Conforme acontece em uma interação espontânea entre falantes, há sobreposição e novos turnos de fala de *CES e, em seguida, *ANE veicula uma construção insubordinada causal, justificando o seu ato de fala anterior. Porém, a justificativa é “abstrata” dado que *ANE utiliza um elemento dêitico de lugar (*aqui*) para mostrar a área do apartamento e, portanto, referenciar o lugar da enunciação. Logo, dentro da interação comunicativa, a referência ao espaço é o que sinaliza a justificativa enunciativa, ou seja, *ANE localiza o espaço para justificar a sua opinião de que o apartamento ficou pequeno após as mudanças.

Por se tratar da diamesia falada, a ocorrência das orações insubordinadas apresenta relação no nível prosódico. Nesse sentido, segundo Raso (2012, p. 58), a fala espontânea é entendida como a fala que é planejada ao mesmo tempo em que é executada, sem planejamento “prévio”. Desse modo, as diamesias falada e escrita se diferem não apenas em relação a questões estruturais, mas também apresentam distinções na organização da informação devido a características próprias da fala. Durante as interações comunicativas, falante e ouvinte estão sujeitos a limitações de pouco tempo (de memória), isto é, de “pedaços” de informações que conseguem armazenar em um curto período de tempo. Isso se deve ao fato de que, como observado por Miller e Weinert (1988), esta diamesia apresenta características diferentes da escrita, uma vez que a fala espontânea envolve interação face a face e contato visual, por exemplo, de modo que não é possível “apagar” aquilo que foi dito.

As características próprias da fala, portanto, podem gerar interpretações e análises diferentes acerca das funções das construções, visto que essas são orientadas ao contexto e à comunicação com o interlocutor. Nesse sentido, a fala tem alguns parâmetros prosódicos como entonação, velocidade de fala, duração; além de elementos característicos como pausas, risadas e interrupções entre turnos de fala e tais fatores estão presentes nas interações comunicativas espontâneas e compõem a estrutura da fala.

É importante destacar que, além da dimensão prosódica, a fala é governada pelo nível pragmático, organizando o significado e o sentido no discurso. Por conseguinte, para capturar a existência de fenômenos próprios da fala, é importante registrar ocorrências que acontecem em interações não guiadas. No entanto, durante muito tempo a fala foi estudada com base apenas em textos transcritos, ou seja, categorias adequadas ao estudo da escrita. Considerando esse cenário, para que os dados reflitam, de fato, o uso da língua pelos falantes, neste trabalho, selecionamos o *corpus* C-ORAL BRASIL (Raso; Mello, 2012), por dispor de dados autênticos de fala, ser segmentado prosodicamente e apresentar variedades situacionais que permitam explorar, de forma mais ampla, o estudo das causais.

Diante disso, partindo da proposta teórica de Croft (2001, 2022), pretendemos investigar se as construções insubordinadas apresentam algum tipo de relação de subordinação ou dependência, não sintática, mas semântica, com as demais construções oracionais com que coocorrem na enunciação na fala espontânea. Para tanto, temos como hipótese que os eventos denotados pelas construções insubordinadas e suas adjacentes na enunciação na diamesia falada mantém uma relação semântica de justificativa do ato de fala, como apresentado no exemplo (5).

Além disso, neste trabalho, buscamos também entender o motivo pelo qual a forma causal subordinada, com a conjunção *porque*, acontece na fala espontânea com função de justificção do ato de fala na estrutura da insubordinada. Para isso, adotamos a hipótese de que a construçção causal usa a estratégia de recrutamento (Croft, 2022), que será explicada em detalhes mais adiante, por assumir uma forma canônica de uma oraçção formalmente subordinada, porém com outra função discursiva.

De modo geral, para resolver os problemas propostos e corroborar as hipóteses levantadas, temos como objetivo analisar o uso das construçções insubordinadas causais na fala espontânea do português brasileiro e identificar o motivo pelo qual os falantes usam a forma menos canônica, suas funções e se essas construçções são totalmente independentes ou se existe alguma motivação semântica ao optar por essa distribuiçção, dentro do plano prosódico-pragmático da construçção. Além disso, buscamos observar as ocorrências de construçções semi-insubordinadas, as quais apresentam uma relaçção com outro ato de fala contíguo, no qual uma aparente oraçção principal é realizada, estabelecendo uma relaçção textual com o conteúdo linguístico adjacente. Tais construçções, conforme analisado por Bossaglia, Mello e Raso (2020), são diferentes das insubordinadas uma vez que apresentam vínculo textual ou pragmático com outra construçção que contém um material de oraçção principal.

A seguir, apresentaremos o aporte teórico que orienta esta pesquisa, distribuído nas seções sobre insubordinaçção e Gramática de Construçções. Em seguida, abordaremos a metodologia que guiou este trabalho e, por fim, os resultados encontrados e as consideraçções finais.

2 Insubordinaçção

Como apresentado na seçção anterior, Evans (2007) destaca que as insubordinadas são oraçções que, mostram-se formalmente subordinadas, mas funcionam como principais. Nesse caso, as oraçções são usadas como independentes, mesmo que existam índices lexicais, como conjunçções, que prototipicamente introduzem construçções subordinadas. Nessa perspectiva, apresentam a estrutura formal de subordinaçção, mas são orientadas à função de uma oraçção principal, funcionando como construçções não prototípicas, como mostraremos em mais detalhes adiante. Portanto, representam certos desafios às análises tradicionais de construçções que apresentam relaçções de combinaçções entre as oraçções, como a subordinaçção e a coordenaçção. Apesar de ocorrerem em muitas línguas (Evans, 2007; Evans; Watanabe, 2016; Mithun, 2008), continuam sendo uma construçção é pouco estudada, em detrimento de outras construçções mais canônicas. Em outras abordagens, o fenômeno pode ser descrito como oraçções desgarradas (Decat, 2011) e *independent clauses* (Montolíó, 1999; Schwenter, 1999). Na abordagem de Traugott (2017, p. 3), por exemplo, em sua análise sobre oraçções independentes finitas introduzidas por subordinadores em inglês, foram identificadas algumas pistas formais de insubordinaçção, como a presença de: i) um subordinador, ii) modalidade/moço (subjuntivo em línguas em que isso é relevante), iii) ordem de palavras típica de oraçção subordinada (em línguas em que isso é relevante) e iv) uso sintático independente. Vejamos, em alguns dos nossos dados abaixo, a ocorrência das construçções insubordinadas:

- 6) [falante fala sobre o desmatamento na região]

*MAI: talvez agora já acabou porque já desmataram muito / né //

- 7) [falante pede ao seu interlocutor para pegar as pilhas]

*BAL: tira essas / sem &ca [2] essas não recarregáveis daí // **porque eu vou colocar elas em outro lugar** //

Nos exemplos, em (6), há uma construção subordinada e causa direta, pois a causa de ter acabado a vegetação é ter desmatado muito; em (7), a construção é insubordinada e não há a causa direta, visto que há uma justificação enunciativa: a falante dá uma ordem e produz uma construção insubordinada veiculada ao ato de fala, em uma relação do tipo “disse isso (*tire as pilhas*), por causa disso (*vou colocá-las em outro lugar*)”. Logo, a relação é de causa, de modo que *porque* justifica o ato de fala imperativo. Podemos considerar essa justificação como uma causa indireta. Sweetser (1990, p. 77) define que o uso como “causa indireta” corresponde a uma mudança da relação causal para o plano ou epistêmico (o falante quer justificar uma crença expressa em uma oração anterior ou marcar uma conclusão), ou pragmático (o falante querendo dar conta de um ato de fala que acaba de cumprir).

Portanto, neste trabalho usaremos “causa” para os casos em que há uma relação direta de causa-consequência e “justificação enunciativa” para quando essa relação canônica não existir⁵. Assim, nesse caso, o uso do item lexical que introduz as causais não parece estabelecer uma causa direta, isto é, causa/consequência, mas sim uma relação pragmática entre atos de fala, com uma justificativa. Logo, não há dependência entre a “principal” e a “insubordinada”, e sim um uso sintático independente, até pelo fato de estarem em níveis prosódicos diferentes.

Como postulado pela *Language into Act-Theory* (Cresti, 2000; Moneglia; Raso, 2014), teoria que orienta a segmentação do *corpus* C-ORAL BRASIL (Raso; Mello, 2012), a unidade de referência da fala é o enunciado e a marcação da fronteira dos enunciados é feita por quebras terminais (/). Nesse sentido, o enunciado é uma unidade com autonomia prosódica e pragmática. Por exemplo, *ham ham //* é considerado um enunciado uma vez que veicula uma ilocução de confirmação. Nessa abordagem teórica, o enunciado corresponde ao ato de fala (Austin, 1962), ao ato locutivo e, portanto, carrega uma força ilocucionária. Desse modo, como indicado por Raso (2012, p. 95), “a segmentação e a individualização de um enunciado não é então a sua estrutura morfossintática, mas a sua forma prosódica, que nos diz onde e como segmentar a fala”. Portanto, as insubordinadas ocorrem entre fronteiras de enunciados, separados pelas barras duplas (/). Logo, em (7) *tira essas não recarregáveis daí //* e *porque eu vou colocar elas em outro lugar //* estão em fronteiras prosódicas diferentes visto que estão separadas por barras duplas e, portanto, constituem enunciados diferentes e, sob a perspectiva de Croft (2022), apresentam construções diferentes – sendo, portanto, uma escolha do falante produzir as construções em unidades prosódicas separadas, o que reflete sua intenção comunicativa. Essa escolha também foi observada por Schwenter (2016) nos dados de orações iniciadas por *si* e realizadas de forma independente no espanhol, usadas pelos falantes para sinalizar uma inferência ao interlocutor. No exemplo do português, portanto, há duas construções oracionais independentes, de modo que construção insubordinada causal não exerce depen-

⁵ Por serem também independentes do ponto de vista sintático e semântico, as construções insubordinadas aqui estudadas podem ser confundidas com orações coordenadas explicativas. Porém, as insubordinadas apresentam funções diferentes e, portanto, não se assemelham às coordenadas visto que, para a insubordinada não é possível reconhecer uma oração principal no contexto linguístico. Portanto, a leitura como coordenada explicativa pressupõe a existência de duas construções independentes ligadas parataticamente, porém as insubordinadas não possuem uma principal reconhecida no contexto.

dência sintática em outra construção, uma vez que é autônoma e que o enunciado que é justificado pela insubordinada pode ser expresso por diferentes tipos de construções.

Uma vez delineado o fenômeno da insubordinação, é importante compreender o quadro teórico que permite tratar tais ocorrências como unidades significativas da língua. A **Gramática de Construções** (Croft, 2001, 2022) oferece um modelo adequado para analisar padrões formais e funcionais em uso, especialmente na língua falada. Assim, na seção seguinte, são apresentados os princípios fundamentais dessa abordagem e estudos que contemplam esse aporte teórico.

3 Gramática de Construções

A Gramática de Construções, como proposta em Goldberg (1995, 2006) e Croft (2001, 2022), além de outros autores, é considerada um conjunto de teorias que aderem aos princípios da Linguística Cognitiva e propõe um modelo uniforme de representação de todo o conhecimento gramatical (sintaxe, morfologia e léxico). Nesse sentido, a construção é um pareamento entre forma e função e é considerada uma unidade simbólica convencional (Langacker, 1987; Croft, 2005), ou seja, são associações parcialmente arbitrárias entre forma e significado compartilhadas entre os falantes e que podem ser mais ou menos frequentes no uso, de modo que o significado é constituído pelas propriedades semânticas, pragmáticas e discursivas da construção. Segundo a proposta de Croft (2022), a estrutura básica da construção é constituída pela forma, morfossintaxe e características fonológicas, inclusive a prosódia, e pela função, composta pelo conteúdo semântico e pela função discursiva. O conteúdo semântico, segundo Croft (2022), é o conteúdo informativo que é convencionalmente transmitido por uma construção: a palavra “quadrado”, por exemplo tem o significado de uma forma e figura específica. Já a função discursiva, também conhecida como empacotamento da informação, refere-se à forma como o significado ou o conteúdo semântico é organizado ou estruturado para fins de comunicação no discurso: o conceito de propriedade *enorme* pode ser expresso como uma predicação “aquela árvore é enorme”) ou pode ser empregado para modificar ou acrescentar informação a um objeto (“aquela árvore enorme ali”). Portanto, segundo Croft (2022), as construções expressam um conteúdo semântico, também chamado de significado, e um empacotamento desse conteúdo: o empacotamento de informação, também chamado de função discursiva.

De acordo com o aporte teórico da Gramática de Construções não há distinção rígida entre léxico e gramática, de forma que existe um *continuum* entre ambos e as construções se diferenciam de acordo com a função a que servem e ao tipo de entidade linguística que representam, de modo que a estrutura semântica é projetada diretamente na estrutura sintática. Além disso, do ponto de vista cognitivo, as construções são vistas como redes interconectadas, possibilitando que os falantes interpretem uma nova construção com base em algum padrão semelhante armazenado em seu conhecimento linguístico. Por conseguinte, as construções formam um inventário estruturado do conhecimento do falante sobre as convenções de sua língua, representado, normalmente, na forma de uma rede taxonômica, em que cada construção é um nó.

Nesse sentido, Croft (2022) assume a ausência de relações sintáticas entre elementos e a definição de categorias e relações por meio de papéis que os elementos apresentam em

uma construção complexa. Logo, no que tange às construções encontradas na fala espontânea, há uma relação semântica entre partes da construção sem, necessariamente, haver relação sintática entre elas. Portanto, por se tratar de uma diamesia governada pela dimensão pragmática, as relações entre as construções são pautadas pelo contexto e pelo uso na comunicação - e não por dependência sintática, como mostram os exemplos:

- 8) [a falante comenta o seu apelido]

*LUZ: ocês nũ acostumaram ainda / mas cês vão acostumar // **porque todo mundo me chama < de Lilisa > hhh⁶ //**

- 9) [falante comenta a construção de uma obra]

*PAU: e lá eu quero chegar com ele até a divisa com o Paulo // isso até a [/2] até aquele murinho lá / sabe como //

*ROC: sei //

*LUZ: **porque é capaz d'eu subir uma parede lá / cobrir //**

Os exemplos apresentam ocorrências de construções que não têm dependência sintática com as construções com as quais coocorrem. Nos dados, podemos observar que o item lexical que introduz a construção insubordinada é uma estratégia recrutada da construção prototípica (subordinada), de modo que assume uma nova função discursiva.

Antes de prosseguirmos, é importante esclarecer brevemente o conceito de construções prototípicas à luz da Gramática de Construções. Segundo Croft (2013, 2022), as construções prototípicas expressam combinações mais comuns ou “favorecidas” de função discursiva e conteúdo semântico. Já as construções não prototípicas expressam combinações menos frequentes e mais marcadas em relação às construções prototípicas. Portanto, como a forma das insubordinadas expressa funções diferentes, essas são consideradas não prototípicas. Neste caso, a estrutura morfossintática da construção causal expressa uma função discursiva diferente, ou seja, não há uma relação de causa-consequência veiculada pelo conteúdo semântico, e sim uma relação de justificação enunciativa.

De acordo com Croft (2022), a estratégia de recrutamento ocorre quando uma construção usada para uma função é usada para outra função. Essa estratégia tem origem diacrônica, uma vez que o recrutamento acontece no passado. No caso da insubordinação, há, como observado por Evans e Watanabe (2016), um recrutamento da forma das orações subordinadas pelas estruturas insubordinadas. De acordo com a proposta de Evans (2007, p. 370), a trajetória da insubordinação acontece da seguinte forma: (i) há o uso “normal” da oração subordinada; (ii), na segunda fase, ocorre o apagamento/elipse da oração principal, mas o ouvinte poderia reconstruí-la com base em processos de inferência conversacional sem restrições semânticas a princípio; (iii) porém, quando a reconstrução da oração principal começa a apresentar restrições quanto ao conteúdo, ou seja, quando há restrição à interpretação do material elíptico, surge o uso das orações insubordinadas, isto é, o uso convencionalizado como principal de uma oração formalmente subordinada. Seguem alguns exemplos dos nossos dados para analisarmos as ocorrências na fala espontânea:

⁶ Segundo os critérios de transcrição do *corpus* C-ORAL BRASIL (Raso; Mello, 2012), o símbolo hhh indica ruído paralinguístico (tosse, riso, choro ou ruídos feitos com a boca).

- 10) [homens conversam sobre reformas na estrutura das empresas]
 *JOR: *nũ tem interesse / porque todo mundo se defende //*
- 11) [mulher conversa sobre a filha adotiva]
 *CAR: *mas eu fico muito feliz porque ela tá / me dando / &he / bons retornos //*
- 12) [falante conta um caso que aconteceu em uma cidade]
 *ALO: *aí es / correram depressa / tamparam o caixão / < e > [/1] e racharam embora hhh //*
 *LUA: *< então vai > embora //*
 *ALO: *e foram embora com o seu Pedro //*
 *JOS: *Nossa / < porque aí era o > fim / < né > //*
 *ANA: *<é> //*
 *ALO: *<cabou> o trem //*

Nos exemplos, as construções são subordinadas, em seu uso canônico, com as orações principais *nũ tem interesse* em (10) e *eu fico muito feliz* em (11). Em (10), há uma quebra não terminal, representada pela barra simples (“/” na transcrição), entre a construção principal *nũ tem interesse* e a adverbial causal *porque todo mundo se defende*. Nesse caso, a construção subordinada ocorre dentro da mesma unidade sintática e fazem parte de um mesmo enunciado nos termos da *Language into Act-Theory* (Cresti, 2000; Moneglia; Raso, 2014), uma vez que estão dentro da mesma unidade prosódica, representada pelas barras duplas (//), e não são construções independentes.

Essas ocorrências apresentam construções prototípicas, com o uso tradicional de uma construção subordinada (principal + conjunção causal + subordinada). Entretanto, em alguns casos, por estratégia de recrutamento, o uso das construções insubordinadas pelos falantes acontece sem a principal e sem ser possível recuperá-la. Em (12), o interlocutor conta a história de uma senhora que foi traída pelo marido e, no velório dele, demonstrou “alívio” pela situação. Para expressar sua percepção sobre o acontecimento, a falante faz um comentário para demonstrar o quanto a situação retratada era intolerável, utilizando uma justificação enunciativa para apresentar a sua opinião perante o que foi dito por *ALO. Nesse caso, a construção insubordinada é usada para justificar o ato de fala, uma ilocução de pedido; e não apresenta uma relação de causa direta. Portanto, nos casos de insubordinação não é mais possível recuperar uma oração principal, pois as insubordinadas não apresentam significado relacionado ao significado denotativo da construção adjacente de modo que a relação não é entre conteúdos proposicionais das construções, e sim entre atos de fala.

Desse modo, a teoria assume uma estratégia definida em termos do uso da construção da língua para expressar a função em questão. Tal estratégia consiste no recrutamento, na qual a construção usada para uma função é recrutada para ser usada em uma função diferente, de modo que essa estratégia recrutada assume uma nova função, se adapta e é estendida a ela. De acordo com Croft (2001, 2022), nenhuma construção existe isoladamente, isto é, o sistema linguístico seria organizado em redes taxonômicas hierarquicamente organizadas, e a unidade básica da língua é a construção. Nessa perspectiva, a língua é um inventário de redes construcionais, com unidades simbólicas e complexas que compõem o conhecimento dos falantes sobre a língua.

No que tange à prosódia, a mesma estrutura morfossintática pode veicular diferentes funções discursivas a partir de diferentes prosódias, visto que o conteúdo sintático pode ser-

vir a funções discursivas diferentes. De acordo com Croft (2022, p. 1464), a prosódia “consiste em uma estratégia fonológica normalmente usada para expressar certas construções oracionais não predicativas e construções de atos de fala”. Nessa perspectiva, significados distintos podem ser veiculados por perfis prosódicos diferentes, mesmo que a estrutura morfossintática se mantenha a mesma. Isso ocorre, por exemplo, nas orações condicionais, como observado por Schwenter (2016), e em construções téticas, como apontam Kaltenböck e Heine (2014).

De acordo com quebras prosódicas específicas, terminais (//, que apresentam perfil conclusivo) e não terminais (/ , com perfil não conclusivo), as construções insubordinadas em análise assumem funções específicas orientadas ao nível pragmático, na interação com os falantes. Na visão de Bossaglia, Mello e Raso (2020), a força ilocucionária, veiculada pela prosódia, confere autonomia pragmática e o item lexical usado como subordinador estabelece uma relação pragmática em relação ao (con)texto, e não à sintaxe. Nessa perspectiva, “o estudo da insubordinação e, em geral, da sintaxe da fala, deve levar em conta a organização prosódica da fala” (Bossaglia; Mello e Raso, 2020, p. 248). Portanto, de acordo com características prosódicas específicas, as orações insubordinadas adquirem orientações discursivas específicas e funções pragmáticas (Bossaglia; Mello e Raso, 2020, p. 248). Por apresentar dados do mesmo *corpus* aqui utilizado e por critérios metodológicos adequados às análises da insubordinação sob o nível prosódico, a abordagem da dimensão prosódica adotada por Bossaglia; Mello e Raso (2020) será adotada para as observações feitas no presente trabalho. Desse modo, a força ilocucionária, veiculada por meios prosódicos, fornece construções com autonomia pragmática, de modo que o item lexical, formalmente subordinador, estabelece uma relação não sintática, mas pragmática e relacionada ao discurso.

Desse modo, após estabelecer o suporte teórico que orienta a análise, a próxima seção descreve a metodologia adotada para o estudo aqui apresentado e explicita o *corpus* utilizado, os critérios de seleção das ocorrências e os métodos quantitativos e qualitativos empregados na análise.

4 Metodologia

Os dados utilizados nesta pesquisa foram coletados do C-ORAL BRASIL⁷ (Raso; Mello, 2012), um *corpus* de fala espontânea do português do Brasil, principalmente da diatopia mineira. Como apontado por Raso (2012, p. 59), a fala espontânea varia de acordo com alguns parâmetros como: a tipologia comunicativa (monólogo, diálogo e conversação), canal comunicativo, contexto sociológico (privado, familiar ou público), variedades de registros ou gêneros possíveis, fatores sociolinguísticos, entre outros. A fim de respeitar essas condições, o *corpus* coleta dados em contexto natural e em contexto informal, mantendo uma estrutura com bom balanceamento e arquitetura que representa as variações que ocorrem em situações informais da fala espontânea, além de critérios de transcrição voltados para capturar fenômenos próprios dessa diamesia.

O *corpus* é composto por 139 textos (21:08:52 horas), divididos em dois contextos, familiar/privado (159.364 palavras) e público (48.766 palavras), e separado, de forma equilibrada, em três situações comunicativas: diálogos (dois participantes), conversação (mais de dois

⁷ Acessível em <http://www.c-oral-brasil.org>.

participantes) e monólogo (um participante), totalizando 208.130 palavras. O recurso apresenta transcrição em formato txt feita pelo sistema CHAT (Macwhinney, 2000), gravações que apresentam alta qualidade acústica, além de alinhamento texto-som feito pelo software WinPitch (Martin, 2010). O recurso integra o projeto C-ORAL ROM (Cresti; Moneglia, 2005), que dispõe de *corpora* comparáveis de outras línguas, como o espanhol, francês, italiano e português europeu.

Segundo a *Language into Act-Theory* (Cresti, 2000; Moneglia; Raso, 2014), uma teoria *corpus-driven* voltada para a análise pragmática e prosódica da fala espontânea, que orienta a segmentação dos *corpora* do projeto C-ORAL, a segmentação da fala é feita, a partir de percepções de variações prosódicas, conforme ilustram os exemplos:

13) se ficar alto demais / ele fica feio //

14) tendeu //

Conforme os exemplos, a segmentação é feita em quebras prosódicas que podem ser: conclusivas (“//” na transcrição); e não conclusivas (“/” na transcrição) que marcam unidades menores (as unidades tonais). Em (13), *se ficar alto demais* é uma construção que não tem independência, visto que o enunciado não tem autonomia prosódica nem pragmática, pois o falante ainda vai completar a sua ideia e o perfil prosódico é ascendente. Já o exemplo em (14) mostra uma ilocução autônoma, completa e independente, com perfil prosódico descendente e conclusivo, que veicula uma pergunta. Para garantir a representatividade de diamesia falada, em seu processo de compilação, as gravações dos áudios foram feitas com recursos sofisticados a fim de garantir a melhor qualidade acústica possível. Além disso, o *corpus* foi compilado para obter uma ampla variação diafásica, assim como uma arquitetura muito bem balanceada e equilíbrio quanto à idade e sexo dos falantes. Desse modo, o recurso apresenta dados autênticos de fala, registrados em diferentes situações comunicativas, e se mostra ideal para o estudo da fala espontânea, pois dispõe de critérios de compilação e segmentação próprios para análise dessa diamesia.

Nesse sentido, o recurso selecionado para a pesquisa obedece a critérios metodológicos bem definidos de compilação, arquitetura e transcrição e, por conseguinte, constituem recursos adequados e projetados para finalidade de estudo da fala espontânea. É importante destacar que, por razões metodológicas, não vamos nos ater às especificidades teóricas e categóricas da *Language into Act-Theory* (Cresti, 2000; Moneglia, Raso, 2014) e nos atentar apenas para a segmentação.

Em relação à etapa da coleta de dados, em primeiro lugar, foi feito o levantamento dos dados por meio do *software* AntConc (Anthony, 2014), buscando-se, nos textos, pelo subordinador *porque* em ocorrências nas quais o item lexical aparecia após quebras prosódicas terminais (/), uma vez que indica que se trata de um enunciado autônomo. A escolha do subordinador *porque* foi feita devido à sua alta produtividade nos registros orais e escritos, constituindo uma entrada para observar os padrões de insubordinação causal. De modo, outras conjunções causais subordinadas foram buscadas e não foi encontrada nenhuma ocorrência após quebras terminais para *porquanto*, *dado que* e *visto que*; uma para *uma vez que*; duas para *já que*; cinco para *que* e dezoito para *por isso que*. As ocorrências de *como* no início das construções eram com função de pronome interrogativo. Portanto, as construções menos frequentes não viabilizam uma análise contundente e, por esse motivo, apenas a observa-

ção das construções iniciadas por *porque* foi mantida, o que garante dados mais robustos para análises qualitativa e quantitativa.

Em seguida, as ocorrências de construções insubordinadas foram organizadas e classificadas quanto à relação causal (se há uma justificação enunciativa ou uma relação de causa-consequência). O processo de busca pelo item lexical nos textos foi feito sempre acompanhado pela oitiva dos arquivos, por meio de software WinPitch (Martin, 2010), que permite acesso simultâneo à transcrição, ao som e ao espectrograma.

Nesse sentido, a análise dos dados da presente pesquisa é de caráter qualitativo-quantitativo, utilizando como fonte o *corpus* C-ORAL BRASIL (Mello; Raso, 2012). Desse modo, a unidade de fala analisada foi a construção iniciada por *porque*, visto que outras conjunções não ocorrem, com frequência, em unidades tonais separadas. Após a identificação da construção, o critério para análise foi a função na fala, observando a entonação usada pelos falantes e a relação entre atos de fala e, por fim, os dados foram contabilizados. A análise permitiu quantificar a presença das construções e também compreender seu papel funcional na fala do português brasileiro.

Com a metodologia definida, serão apresentados os resultados obtidos a partir do levantamento e da análise das ocorrências no *corpus*. A próxima seção detalha os padrões identificados, tanto em termos de frequência quanto de função discursiva, oferecendo uma visão abrangente das construções observadas.

5 Resultados e análise dos dados

Com base nos preceitos da Gramática de Construções (Croft, 2022), até o momento as seções deste artigo apresentam os objetivos da pesquisa ao investigar o fenômeno da insubordinação no português brasileiro, sendo estes, em suma, identificar o motivo pelo qual os falantes usam a construção insubordinadas e suas funções. Para isso, buscamos investigar qual relação semântica as insubordinadas têm com as demais construções com que coocorrem, partindo da hipótese de que há uma relação de justificação enunciativa. Além disso, pretendemos entender porque a estrutura insubordinada apresenta forma de uma causal subordinada e, assumimos, a partir disso, que há uma estratégia de recrutamento. Na análise de Evans (2007, p. 370), insubordinação segue a seguinte trajetória:

Quadro 1 – Trajetória da insubordinação

Subordinação (1)	Elipse (2)	Elipse convencionalizada (3)	Reanálise como uma estrutura principal (4)
Construção subordinada	Elipse da oração principal	Restrição à interpretação do material elíptico	Uso convencionalizado como principal de oração originalmente subordinada

Fonte: Evans, 2007, p. 370, tradução de Coradini e Hirata-Vale, 2021, p. 322.

Na primeira fase da evolução diacrônica da insubordinação, há a construção subordinada, combinada com a oração principal, constituindo, portanto, a subordinação. Em seguida, ocorre a elipse da oração principal, a qual pode ser facilmente inferida pelo contexto e infe-

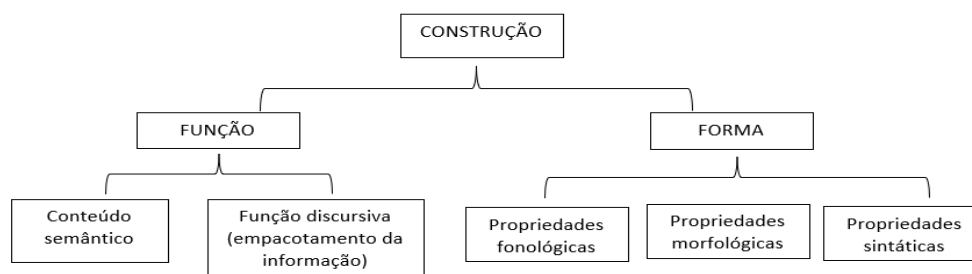
rências conversacionais, sem que existam restrições semântico-pragmáticas. No entanto, na etapa 3, a reconstrução da cláusula principal omitida se torna mais complexa e algumas estruturas sintáticas que poderiam ser consideradas como conteúdo elíptico são excluídas por convenção. Neste estágio, o contexto delimita algumas possibilidades de reconstituição do conteúdo da oração principal e surgem restrições quanto à sua reconstrução. Essa etapa, portanto, origina a fase final do processo, a convencionalização, na qual as construções subordinadas possuem significados próprios e são independentes, não sendo possível recuperar qualquer material elíptico.

Na próxima seção, apresentaremos de forma detalhada os dados e as análises referentes ao fenômeno das 286 ocorrências no português brasileiro falado, descrevendo não apenas sua distribuição e características estruturais, mas também os contextos discursivos em que aparecem e as possíveis interpretações que emergem a partir desses usos da língua. Na seção seguinte, por sua vez, serão apresentadas as considerações finais, nas quais retomaremos os principais resultados obtidos e refletiremos sobre suas implicações teóricas e metodológicas.

5.1 As construções causais insubordinadas

Segundo o aporte teórico da Gramática de Construções (Croft, 2001), as construções são associações parcialmente arbitrárias entre forma e significado (ou forma e função) compartilhadas entre os falantes e que podem ser mais ou menos frequentes no uso, de modo que no nível da forma, a construção é formada pelas propriedades fonológicas, morfológicas e sintáticas; e no nível da função, pelas propriedades semânticas, pragmáticas e discursivo-funcionais.

Imagem 1 – Modelo simbólico de uma construção



Fonte: Croft, 2022, p. 5, adaptado.

O modelo apresentado adaptado de Croft (2001, p. 5) demonstra a forma como a construção é concebida dentro da perspectiva construcional da linguagem (Traugott; Trousdale, 2013). Nessa abordagem teórica, forma e função se conectam por meio de relações de correspondência simbólica, de modo que orientam e são moldados por usos linguísticos, criando uma conexão entre ambos. A busca pelas orações causais no *corpus*, na pesquisa em andamento, levantou um total de 286 enunciados dos 27590 enunciados completos do *corpus*⁸:

⁸ Os dados completos podem ser encontrados neste link: <https://docs.google.com/spreadsheets/d/1h78CUH-vkHMPaX95qyw4VqzekOoNIt3N1/edit?usp=sharing&ouid=106176044600777471678&rtpof=true&sd=true>.

- 15) [falante comenta o seu visual para o casamento]
 *JAE: inda bem que cê vai tar co seu cabelo curto / pa nũ fazer aquela touca pra cima //
porque Deus me livre //
- 16) [amigos reunidos para jogar um jogo]
 *HEL: o outro grupo pode atrapalhar falando // < **porque quando** >⁹ eu jogava / < o
 povo ficava > / **repetindo a mesma < palavra / enchendo o saco > //**
- 17) [dois operários da construção civil planejando o trabalho]
 *ROG: eu vou &coloc [/3] eu vou suspender mais um pouquim aqui / vou pegar a linha /
 e vou colocar por cima //
 *PAU: ah / **porque senão //**

Conforme Traugott (2017), os sinais formais de insubordinação incluem, entre outros, a presença de um subordinador e o uso sintático independente. Nesse sentido, nos exemplos, é possível observar que as insubordinadas apresentam o item lexical *porque* e ocorrem em construções separadas da “principal”, mantendo sua autonomia sintática, como sinalizado por Croft (2001) em relação à natureza das construções.

Nos exemplos, é possível perceber que as estruturas parecem funcionar como justificativas, não no plano sintático, mas semântico. Em (15), *porque Deus me livre* é um ato de fala usado pela falante para expressar um certo “alívio” pelo fato de a amiga estar com cabelo curto e não precisar usar a touca. Em (16), a falante está justificando o porquê ela pergunta se o grupo pode atrapalhar o jogo uma vez que a equipe perturba ao ficar repetindo a mesma palavra. Um comportamento semelhante é observado retomando a construção (17), usada pelo falante para expressar que “se não usarem as linhas corretas na construção, algo pode dar errado”. Portanto, nos exemplos, não há dependência sintática entre a “principal” e a “insubordinada”, até pelo fato de estarem em níveis prosódicos diferentes. Porém, a relação de justificativa é mantida no plano pragmático, uma vez que a justificação ocorre no nível discursivo, ou seja, a construção insubordinada faz referência à situação discursiva, e não à “principal”. Logo, a relação de justificativa é preservada no plano pragmático, pois a conexão entre as partes não se dá pela estrutura sintática, mas pelo modo como o sentido é construído no discurso. No nível pragmático, a justificativa depende da intenção do falante, do contexto, das inferências que o ouvinte é capaz de fazer e da forma como a informação se articula com o fluxo discursivo. Assim, a construção insubordinada não se liga semântica ou sintaticamente à “principal”, mas sim ao contexto discursivo mais amplo, funcionando como um comentário ou justificativa para a situação comunicativa em curso. Desse modo, os exemplos apresentam eventos independentes, visto que, caso a insubordinada fosse removida, não haveria comprometimento de conteúdo no nível sintático, mas haveria uma lacuna semântico-pragmática. Vejamos alguns exemplos:

- 18) [vendedor conversando com a cliente em uma loja de sapatos]
 *EUG: por isso que cê tá sentindo // **porque esse aqui [sapato] tá todo forrado //** e
 aquele lá nũ é forrado //

⁹ A tag <> indica sobreposição de turno de fala.

- 19) [mulher fala sobre a cobertura da casa]
 *VER: a minha idéia é colocar uma telha colonial // **porque aí nũ preciso pôr laje** //
- 20) [pai comenta com o filho sobre ir embora de um local]
 *ONO: <uai> // <eu vou> / uai // eu tô querendo ir embora mesmo // eu nũ vou embora é por [/2] é por [/2] &he / é / por causa d'ocês / não // **porque ocês tá tudo bão / tá novo** // por causa da sua mãe e a <Maria Eduarda> //

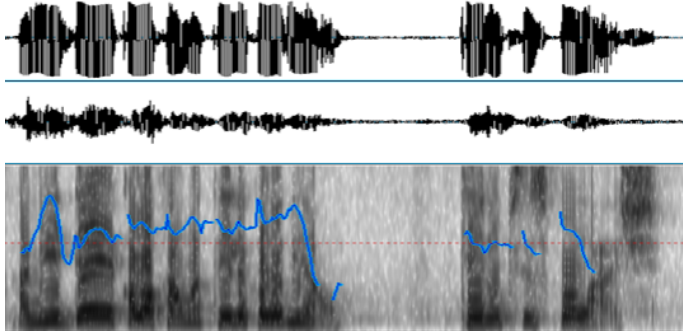
Nos exemplos, os falantes expressam uma crença: em (18), o vendedor deduz que a cliente está sentindo diferença entre os sapatos porque um deles é forrado, e o outro não; e em (19), a mulher diz ao seu interlocutor que acha que seria adequado colocar uma telha na casa. Por outro lado, em (20), o falante retoma o que ele disse antes justificando o seu ato de fala, reforçando o seu conteúdo locutivo. Nesse caso, a construção causal encontra-se “deslocada”, como se o falante interrompesse o próprio turno de fala para realizar essa construção insubordinada e, em seguida, retomar o que estava falando no início. Assim, o uso funciona como um “acréscimo” de informação durante o turno de fala. Segue mais um exemplo:

- 21) [falantes comentam o corte de cabelo]
 *MIC: eu queria trocar mas nũ dá // minha mãe vai yyyy //
 *LUC: acho bacana <hhh> //
 *FLL: eu < cortei> o meu / o' [cabelo]//
 *MIC: tô <vendo> //
 *FLL: <hoje> //
 *MIC: <tá ótimo> //
 *FLL: <porque eu cortei pro casamento> //

Na estratégia de recrutamento, uma construção é usada para uma função diferente, com a mesma forma. Em (21), a construção causal justifica o ato de fala e não há uma relação de causa e efeito. Em outros termos, a justificativa não é que a falante *FLL cortou o cabelo, que ela cortou o cabelo para o casamento; mas ela justifica o porquê ela disse que cortou o cabelo. Portanto, a construção causal recruta a estratégia da construção subordinada para assumir um novo significado no discurso.

De acordo com Croft (2022), a estrutura básica de uma construção é formada pela função, composta pela função discursiva e conteúdo semântico, e pela forma, representada pela estrutura morfossintática e fonológica/prosódica. Nesse sentido, as construções insubordinadas casuais são convencionalizadas como principais, visto que são independentes. Sintaticamente, dispõem do item lexical *porque* que, canonicamente, introduz uma oração causal, seguido do restante da construção. Prosodicamente, as construções insubordinadas têm sua própria força ilocucionária e são independentes, apresentando um contorno entonacional descendente (o que se assemelha ao ponto final na escrita). Nos exemplos em (18) e (21), portanto, a forma, aparentemente semelhante à de uma principal, apresenta uma nova função discursiva em um pareamento convencionalizado. Nos termos de Croft (2022), podemos considerar que o ouvinte reanalisa uma propriedade semântica/funcional da unidade sintática, no caso, a construção insubordinada, de modo que esta propriedade que passa a ser atribuída é inerente à unidade sintática e, portanto, a construção adquire uma nova função discursiva. Vejamos o quadro abaixo, que representa a forma e a função das construções insubordinadas:

Quadro 2 – Forma e função da construção insubordinada causal

Construção insubordinada causal	
Função	Conteúdo semântico: causa indireta Função discursiva: justificação enunciativa
Forma:	Morfossintaxe: porque oração Prosódia:  Figura 3: Imagem do espectrograma das orações principal “a minha ideia é colocar uma telha colonial” (à esquerda) e causal “porque aí nã preciso pôr laje” (à direita) feito pelo programa Praat (Boersma; Weenink, 2018).

Fonte: elaboração própria.

Para elucidar a forma como a informação é processada pelo usuário da língua, Chafe (1994) utiliza a ideia de unidade de informação e, posteriormente, a adapta para unidade de entonação, que ativa um novo “pedaço” de informação, um novo foco de atenção. Uma visão similar, com as devidas proporções, é apresentada pelo conceito de unidade tonal, apresentado pela *Language into Act-Theory* (Cresti, 2000; Moneglia; Raso, 2014), teoria que segmenta o *corpus* aqui utilizado, e corresponde a fronteiras prosódicas do enunciado que podem ser terminais “//” (delimitam o final do enunciado) ou não terminais “/” (apresentam um perfil não-conclusivo). De acordo com essa percepção no nível prosódico, as construções insubordinadas são segmentadas como independentes e constituem, por conseguinte, eventos autônomos sintaticamente.

Do ponto de vista da função, as construções se pautam nas interações entre falantes e ouvintes (Heine, Kaltenböck e Kuteva; 2016) e têm significados próprios e autônomos, carregando propriedades discursivas, pragmáticas e interacionais, apresentando justificativas dos atos de fala (como no exemplo 21), avaliações epistêmicas acerca do conteúdo locutivo (como nos exemplos 18 e 19), ou sinalizam um comentário ou acréscimo de informação dentro da situação discursiva, com caráter parentético, interrompendo o fluxo informacional (como no exemplo 20).

Além disso, segundo Croft (2022), as construções prototípicas são importantes para comparações linguísticas, definindo as estruturas morfossintáticas que caracterizam partes do discurso de uma língua e expressam combinações mais comuns ou “favoráveis” para o empacotamento da informação e conteúdo semântico. Seguem alguns exemplos:

22) [falante relata uma história de família]

*DFL: então / nessas férias / &q [1] a gente / quando veio pra Belo Horizonte / nas férias / ele trazia Maria Julieta / que era a única filha dele / **porque os pais da Dolores** / os sogros dele / passaram a morar nessa casa que ele morava //

- 23) [falantes comentam a organização do espaço]
 *LAU: não / tem que colocar uns espelhos aqui //
 *LAU: < nessas curvas > //
 *LUZ: < aqui também é > outra / aqui o' //
 *LUZ: **porque ocê < não vê > //**
 *LUZ: perigosíssima / né //
- 24) [falante conta sobre o dia do nascimento do seu filho]
 *REG: será que cê nũ tinha jeito de cê me arrumar um remédio / **_r porque a médica prescreveu //_r**¹⁰

Portanto, as subordinadas causais são consideradas prototípicas uma vez que são ocorrências mais frequentes e são formadas pela oração principal e pela oração dependente, expressando uma relação de causa e consequência, como em (22), no qual *DFL justifica que o pai precisava levar a filha para Belo Horizonte nas férias porque os sogros ocuparam a outra casa. Logo, o fato de os sogros morarem na antiga residência do pai é a causa de a viagem para Belo Horizonte ter sido necessária. Já as insubordinadas causais são menos prototípicas e expressam relações que estão no nível pragmático-discursivo, apresentando justificativas para os atos de fala, e não se reduzindo à natureza morfossintática. Em (23), a insubordinada apresenta uma relação pragmático-discursiva com o seu primeiro turno de fala (*não / tem que colocar uns espelhos aqui //*) de modo que *LUZ utiliza a construção causal para justificar a sua ilocução de instrução/ordem, mesmo após a interação com o turno de fala de seu interlocutor. O exemplo em (24), por sua vez, apresenta um discurso reportado, isto é, *REG usou o discurso direto para reproduzir exatamente a sua fala. Na ocorrência, a construção causal é usada para justificar o motivo pelo qual a paciente precisa do remédio, ou seja, porque a médica o prescreveu.

Após a descrição das construções totalmente insubordinadas, torna-se pertinente considerar casos que parecem desafiar a interpretação de orações formalmente insubordinadas. Na próxima seção, serão analisadas as construções causais semi-insubordinadas, que revelam ocorrências em que seria possível recuperar “orações principais” no contexto linguístico adjacente.

5.2 Semi-insubordinadas

Apesar de as construções que ocorrem em níveis prosódicos diferentes daquelas que seriam suas “principais” serem independentes sintaticamente, existem casos em que parece ser possível recuperar elementos das construções principais. Em uma abordagem da fala espontânea¹¹, segundo Bossaglia, Mello e Raso (2020, p. 246), as semi-insubordinadas são casos em

¹⁰ O símbolo _r indica o discurso reportado, isto é, quando o falante reproduz uma fala ou citação.

¹¹ Em outra perspectiva, Van Linden e Van De Velde (2014) destacam que as semi-insubordinadas apresentam um possível resquício da oração principal ausente e são construções precedidas de constituintes que aparentam funcionar, no nível da sentença, como fragmentos da oração principal. Ou seja, é possível recuperar algum material da “principal”. Conforme os trabalhos de Hirata-Vale (2015, 2017) sobre o português, há casos em que o significado da ocorrência é não restritivo, isto é, se refere à situação discursiva e não a uma oração principal. Na análise feita pela autora, foi observado que as condicionais insubordinadas apresentam usos metatextuais, com

que seria possível recuperar as “orações principais” no contexto linguístico adjacente. Nessa perspectiva, são autônomas e interpretáveis, prosódica e pragmaticamente, mas exibem vínculo com o ato de fala, recuperando uma possível “principal”, ou exibem algum tipo de coesão textual com ela no discurso. Vejamos alguns exemplos:

25) [homem comenta duas disciplinas do curso]

*BAL: <o problema> é / eu vou ter que &m [/1] estudar e me atualizar em duas coisas ao mesmo <tempo> //

*BEL: <ah> / tá // <ah / mas é yyyy> +¹²

*BAL: < não que isso me > [/4] não que isso fosse me confundir // **porque eu nunca confundo Letras com <informática> / nũ tem nem como //**

No exemplo, mais material sintático e semântico está presente no enunciado juntamente com a construção causal, de modo que, textualmente e discursivamente, parece haver uma “dependência/relação” entre a “principal” e a construção causal. Logo, é possível interpretar que o falante não confunde Letras com informática e, portanto, não seria capaz de fazer essa confusão entre as duas áreas. Nesse caso, nota-se uma inferência textual de uma possível principal para a construção “porque eu nunca confundo Letras com informática”, veiculada também pela repetição do verbo “confundir”.

26) [falante comenta um programa de televisão que assistia aos sábados]

*CLA: é &p [/2] eu assistia isso muito em Guarapari //

*CLA: tipo / na época das férias //

*CLA: **porque era <tipo / final da manhã> //**

*RAQ: <tá> //

*RAQ: <mas cê já tinha passado da idade de assistir isso> //

27) [mulheres falam sobre o prédio que estão visitando]

*CES: mas ele tinha que ter elevador // **porque ele é [/2] ele é alto / Anete //**

Em (26), a falante está relatando sobre um programa televisivo, que passava aos sábados, ao qual costumava assistir. Nesse sentido, a construção causal é usada como forma de acrescentar uma informação à construção anterior, isto é, ela assistia ao programa na época das férias, que era exibido no final das manhãs. Neste caso, podemos identificar uma relação textual entre as duas construções (*tipo na época das férias e porque era tipo final da manhã*) e inferir, pelo contexto, que ela assistia ao programa em Guarapari porque estava de férias e tinha tempo livre pelas manhãs. Em (27), a falante está constatando a informação de que o prédio é alto e, por isso, deveria ter um elevador. Nos dois exemplos, as construções apresentam relações textuais-discursivas entre si, de modo que é possível recuperar um fragmento da “principal”, e não ocorrem “soltas” do ponto de vista textual.

No entanto, é necessário destacar que, apesar de ser possível recuperar a construção principal no contexto linguístico adjacente, as construções causais semi-insubordinadas

um caráter de retomada, e se relacionam à organização textual que integra o discurso. Neste trabalho, vamos assumir as abordagens de Bossaglia, Mello e Raso (2020).

¹² A anotação yyyy indica um trecho ininteligível e o símbolo + indica interrupção do turno de fala.

ainda são independentes sintaticamente e prosodicamente. Ou seja, são autônomas e interpretáveis; mas mantêm coesão textual com o discurso produzido. Segue mais um exemplo:

28) [mulher relata o nascimento do seu filho]

*REG: assim / João nasceu dentro do carro // quê //_r menino / isso aí foi um acontecimento // **porque ninguém acreditava** / né //

Na ocorrência, a construção causal *porque ninguém acreditava* apresenta referência textual com a construção anterior *isso aí foi um acontecimento*. Na conversa, *REG está relatando o nascimento de seu filho e destaca que o parto aconteceu no carro e, por isso, foi um acontecimento inesperado - porque ninguém acreditava. Desse modo, no nível discursivo e textual, a causa expressa pela construção tem ligação com a construção adjacente e, por isso, apresenta o comportamento de uma semi-insubordinada. No nível discursivo, a semi-insubordinada apresenta a função comunicativa de acrescentar ou justificar informações no contexto da fala ou do texto. No plano textual, por sua vez, a construção apresenta uma ligação com o conteúdo linguístico adjacente, mesmo não apresentando dependência sintática plena. Portanto, dar luz às semi-insubordinadas se faz necessário, visto que essas construções, apesar de serem independentes sintática e prosodicamente, não possuem total autonomia pragmática, uma vez que elas dependem de conteúdo linguístico adjacente para que seu sentido seja plenamente interpretado. Desse modo, não possuem função semelhante às insubordinadas, as quais não apresentam ligação textual com o conteúdo linguístico adjacente.

Por fim, após a análise detalhada dos dados e a discussão das diferentes manifestações das construções causais, na próxima seção serão sintetizados os principais achados da pesquisa, retomando os objetivos propostos e destacando as contribuições do estudo.

6 Considerações finais

Este trabalho teve como objetivo analisar as ocorrências das construções insubordinadas no português brasileiro falado, com foco nas causais, a fim de trazer maior compreensão para seus usos e funções no discurso. Nesse sentido, buscamos observar, nos dados, se as construções apresentam algum tipo de relação, não sintática, mas semântica com as construções adjacentes. Para isso, partimos da hipótese de que existe uma relação de justificação enunciativa. Além disso, procuramos investigar a motivação e função para o uso de uma estrutura de subordinação causativa em uma construção insubordinada. Diante disso, assumimos a hipótese de que a construção causativa é uma estratégia de recrutamento por assumir a forma canônica de uma oração formalmente subordinada.

Tais propostas se fundamentaram sob o aporte teórico funcionalista e cognitivista de Croft (2022), segundo o qual a construção deve ser analisada em relação a como a sua forma expressa a função. Nessa perspectiva, observamos que as construções insubordinadas analisadas recrutam a forma das construções subordinadas, porém com uma nova função discursiva: a justificação enunciativa. Nessa nova função, a relação causal ocorre no plano epistêmico ou no nível do ato de fala, e não no plano sintático. No entanto, verificamos ocorrências em que as causais parecem manter uma relação textual e discursiva com a construção adjacente, sendo, consideradas, portanto, semi-insubordinadas.

Na análise apresentada neste trabalho, observamos casos diferentes, em que as insubordinadas ocorrem no mesmo turno de fala e em turnos diferentes. No entanto, analisamos as ocorrências do ponto de vista da função, sem avaliar as diferenças prosódicas entre as insubordinadas – o que será feito em análises mais aprofundadas que estão em desenvolvimento.

Desse modo, o estudo da realização das construções insubordinadas causais na fala espontânea do português brasileiro revela padrões de uso condicionados pelo contexto pragmático, no qual desempenham diferentes funções pragmático-discursivas, e a relação entre partes da construção é semântica, apresentando fenômenos próprios da diamesia falada.

Declaração de autoria

Alessandra Emanuelle Macieira Silva participou da conceptualização do estudo, da curadoria de dados, da realização da análise formal, da condução da investigação, da elaboração dos procedimentos metodológicos e da escrita do artigo. Luana Lopes Amaral contribuiu com a conceptualização, supervisionou o desenvolvimento do trabalho, colaborou na análise formal e participou da escrita do artigo.

Referências

- ANTHONY, L. *AntConc*. Versão 3.4.3. Tokyo, Japan: Waseda University, 2014. Disponível em: <http://www.laurenceanthony.net/>. Acesso em 08 fev. 2025.
- AUSTIN, L.J. *How to do things with words*. Oxford: Oxford University Press, 1962.
- BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* [Computer program], Version 6.0.37. Disponível em: <http://www.praat.org/>. Acesso em 08 fev. 2025.
- BOSSAGLIA, G.; MELLO, H.; RASO, T. Illocution as a unit of reference for spontaneous speech: An account of insubordinated adverbial clauses in Brazilian Portuguese. In: SHLOMO IZRE'EL, S.; MELLO, H.; PANUNZI, A.; RASO, T. (orgs.). *Search of Basic Units of Spoken Language*. John Benjamins Publishing Company, 2020. p. 221-256.
- CHAFE, W. L. *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press, 1994.
- CORADINI, M.; de M. HIRATA-VALE, BEZZERA, F. Os estágios de insubordinação em construções condicionais com a conjunção *se* no português: evidências históricas. *Working Papers em Linguística*, Florianópolis, v. 22, n. 2, p. 318-345, 2021. DOI: <http://doi.org/10.5007/1984-8420.2021.e75334>.
- CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000.
- CRESTI, E; MONEGLIA, M. (eds.). *C-ORAL-ROM*. Integrated reference corpora for spoken romance languages. Amsterdam/Philadelphia John Benjamins, 2005.
- CROFT, W. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: University Press, 2001.

- CROFT, W. Logical and typological arguments for Radical Construction Grammar. In: ÖSTMAN, J-O; FRIED, M. (eds.). *Construction grammars: cognitive grounding and theoretical extension*. Amsterdam: Benjamins, 2005. p. 273-314.
- CROFT, W. *Morphosyntax: constructions of the world's languages*. Cambridge University Press, 2022.
- EVANS, N. Insubordination and its uses. In: NIKOLAEVA, I. (ed.). *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press, 2007. p. 266-431.
- EVANS, N., WATANABE, H. *Insubordination*. Amsterdam: John Benjamins, 2016.
- DECAT, M. B. N. *Estruturas desgarradas em língua portuguesa*. Campinas: Pontes, 2011.
- DECAT, M. B. N. *Desgarramento, subordinação discursiva e insubordinação: abordagens funcionalistas*. Pontes Editores, 2021.
- GOLDBERG, A. *Constructions: a construction grammar approach to argument structure* (Cognitive theory of language and culture). Chicago: University of Chicago Press, 1995.
- GOLDBERG, A. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press, 2006.
- GORSKI, E. M.; GIBBON, A.; VALLE, C. R. M.; MAGO, D. Dal; TAVARES, M. A. Fenômenos discursivos: resultados de análises variacionistas como indícios de gramaticalização. In: RONCARATI, C.; J. ABRAÇADO (orgs.). *Português brasileiro: contato lingüístico, heterogeneidade e história*. Rio de Janeiro: 7Letras, 2003. p.106-122
- HEINE, B., KALTENBÖCK, G., KUTEVA, T. On insubordination and cooptation. In: EVANS, N., WATANABE, H. (eds.) *Insubordination*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2016. p. 39-63.
- HIRATA-VALE, F. B. M. *O processo de insubordinação nas construções condicionais do português do Brasil*. Relatório Científico Final. Bolsa de Pesquisa no Exterior. FAPESP: processo 13/24523-2. 2015.
- HIRATA-VALE, F. B. M. *Construções condicionais insubordinadas no português: usos metatextuais*. Estudos lingüísticos, São Paulo. 1978, v. 46, p. 83-97, 2017.
- HIRATA-VALE, F. B. M. Perspectivas teóricas para a análise do fenômeno da insubordinação: estado da arte e desafios futuros. In: DECAT, M.B.N; SÉ J. C. S.; HIRATA-VALE F.B. de M.; CAMPOS R. C. S e.; RODRIGUES, V. V. e CAIXETA, G. F. *Desgarramento, subordinação discursiva e insubordinação: perspectivas funcionalistas*. Campinas: Pontes, 2021. p. 75-113.
- KALTENBÖCK, G.; HEINE, B. Sentence grammar vs. thetical grammar: Two competing domains. In: MACWHINNEY, B.; MALCHUKOV, A.; MORAVCSIK, E. (ed.). *Competing motivations in grammar and usage*. New York: Oxford University Press, 2014. p. 348-363.
- LANGACKER, R. *Foundations of cognitive grammar: theoretical prerequisites*. Stanford: Stanford University Press, 1987. v. 1.
- MACWHINNEY, B. J. *The CHILDES project: tools for analyzing talk*. 3. ed. Mahwah: Lawrence Erlbaum Associates, 2000. 2 v.
- MARCUSCHI, L. A. Marcadores conversacionais do português brasileiro: formas, posições e funções. In: CASTILHO, A. (org). *Português culto falado no Brasil*. Campinas: Ed. UNICAMP, 1989. p. 281-318.

- MARTIN, P. *Winpitch10*, versão 1.00. Pitch Instruments. 2010. Disponível em: <<https://www.winpitch.com/>>. Acesso em 08 fev. 2025.
- MITHUN, M. The extension of dependency beyond the sentence. *Language*, Cambridge, v. 84, n.1, 2008, p. 69-119. DOI: <https://doi.org/10.1353/lan.2008.0054>.
- MILLER, J.; WEINERT, R. *Spontaneous spoken language: syntax and discourse*. Oxford: Oxford University Press, 1998. 457p.
- MONEGLIA, M.; RASO, T. Notes on Language into Act Theory (L-Act). In: RASO, T; MELLO, H. (ed.). *Spoken corpora and linguistic studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014. p. 468-494.
- MONTOLÍO, E. ¿Si nunca *he dicho que estuviera enamorada de él!* Sobre construcciones independientes introducidas por si con valor replicativo. *Oralia: Análisis del discurso oral*, Almería, n. 2, p. 37-70, 1999. DOI: <https://doi.org/10.25115/oralia.v2i.8529>.
- RASO, T. O Corpus C-ORAL-BRASIL. In: RASO, T; MELLO, H. (orgs). *C-ORAL BRASIL I: corpus de referência do português brasileiro falado informal*. Editora UFMG, Belo Horizonte, 2012. p. 55-90.
- RASO, T; MELLO, H. (eds.). *C-ORAL BRASIL I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: UFMG, 2012.
- SCHWENTER, S. A. Meaning and interaction in Spanish independent si-clauses. *Language Sciences*, Oxford, v. 58. p. 22-34, 2016. DOI: <https://doi.org/10.1016/j.langsci.2016.04.007>.
- SCHWENTER, S. A. Independent si-clauses in Spanish: Functions and consequences for insubordination. In: WATANABE, H.; EVANS, N. *Insubordination*. John Benjamins Publishing Company, 2016. p. 89-112.
- SCHWENTER, S. A. *Pragmatics of conditional marking: implicature, scalarity and exclusivity*. Nova York: Garland, 1999.
- SWEETSER, E. *From etymology to pragmatics*. Metaphorical and cultural aspects of semantic structure. Cambridge: Cambridge University Press, 1990.
- TRAUGOTT, E. C. 'Insubordination' in the Light of the Uniformitarian Principle. *English Language & Linguistics*, Cambridge, v. 21, n. 2, p. 289-310, 2017. DOI: <https://doi.org/10.1017/S1360674317000144>.
- TRAUGOTT, E. C., TROUSDALE, G. *Constructionalization and Constructional Change*. Oxford: Oxford University Press, 2013.
- VAN LINDEN, A.; VAN DE VELDE, F. (Semi-) autonomous subordination in Dutch: Structures and semantic-pragmatics values. *Journal of Pragmatics*, Amsterdam, v. 60, . p. 226-50, 2014. DOI: <https://doi.org/10.1016/j.pragma.2013.08.022>.

Sobre o papel das relações dialógicas em artigos científicos da área das Ciências Humanas

On the Role of Dialogic Relations in Scientific Articles in the Humanities

Fabício José da Silva

Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) | São José do Rio Preto | SP | BR
fabicio.jose@unesp.br
<https://orcid.org/0000-0001-7422-8989>

Resumo: Este artigo centra-se em aspectos relativos à análise da produção escrita de artigos científicos da área das Ciências Humanas. A reflexão ora apresentada advém de resultados de uma investigação mais ampla, uma Dissertação de Mestrado, que teve como objetivo central analisar e detalhar o papel das relações dialógicas na escrita de artigos científicos representativos das diferentes áreas do conhecimento. Com o objetivo de caracterizar como se configuram as relações dialógicas em artigos científicos da área das Ciências Humanas e discutir as implicações dessas relações para a construção da escrita científica, analisamos artigos científicos coletados em dois periódicos acadêmicos classificados no estrato A1 (Qualis), cujos fragmentos textuais ilustraram diferentes manifestações das relações dialógicas, examinadas com base nas classificações dos tipos e variantes de discurso discutidos em Bakhtin (2018) e em Volóchinov (2021). Metodologicamente, a análise empreendida baseou-se no estudo metodológico da língua exposto por Volóchinov (2021), conjugado com o paradigma indiciário (Ginzburg, 1991). Os resultados apontam como são diversas as diferentes formas de manifestação do dialogismo no gênero discursivo artigo científico e como os usos dos tipos e variantes de discurso orientam o processo de textualização dos artigos científicos na medida em que o discurso do outro é retomado, antecipado e citado.

Palavras-chave: relações dialógicas; tipos de discurso; artigo científico; escrita acadêmica.



Abstract: This article focuses on aspects related to the analysis of the written production of research articles in the field of the Humanities. The discussion presented here derives from the results of a broader investigation developed in a Master's thesis, whose main objective was to analyze and detail the role of dialogic relations in the writing of research articles representative of different areas of knowledge. In order to characterize how dialogic relations are configured in Humanities research articles and to discuss their implications for the construction of academic writing, we analyzed articles collected from two academic journals ranked A1 in the Brazilian Qualis system. Selected textual excerpts illustrate different manifestations of dialogic relations, examined on the basis of the classifications of discourse types and variants discussed in Bakhtin (2018) and Volóchinov (2021). Methodologically, the analysis was grounded in Volóchinov's methodological study of language (2021), combined with the evidential paradigm (Ginzburg, 1991). The results reveal the diversity of dialogism's manifestations in the research article as a discursive genre and demonstrate how the use of different discourse types and variants guides the textualization process, as the discourse of the other is taken up, anticipated, and cited.

Keywords: dialogic relations; types of discourse; scientific article; academic writing.

1 A título de contextualização

Escrever na esfera acadêmica é uma competência fundamental na formação e na atuação de pesquisadores nas diferentes áreas do conhecimento. Quer para cumprir requisitos enquanto bolsista, para promover a comunicação e a divulgação científica entre os pares, quer para pleitear bolsa de agência de fomento à pesquisa, etc., o desenvolvimento da escrita acadêmica é crucial nessas diferentes atuações do estudante ou do professor/pesquisador. Dentre os gêneros discursivos que circulam na esfera acadêmica, o *artigo científico* destaca-se tendo em vista sua ampla produção em todas as diferentes áreas do conhecimento como veículo privilegiado para publicização dos resultados das pesquisas empreendidas. Conforme apontam Motta-Roth e Hendges (2010, p. 66), “[...] o artigo acadêmico é o gênero textual mais conceituado na divulgação do saber especializado acadêmico”. Justamente por ocupar posição central na circulação e na legitimação do conhecimento científico, esse gênero constitui um espaço

privilegiado para observar como se configuram as relações dialógicas que sustentam a construção argumentativa e a produção de sentidos na escrita acadêmica. Assim, a escolha dos fragmentos dos artigos científicos como objeto de análise justifica-se por permitir examinar, em sua forma mais institucionalizada, os modos de interlocução que permeiam o processo de escrita nas Ciências Humanas.

Entendemos que o processo de escrita (e o ensino de escrita) é plasmado e (en)formado por relações dialógicas na universidade (Miranda, 2016) e complexamente organizados por essas relações (Maciel, 2014). Sob o olhar bakhtiniano, entendemos que a escrita – acadêmica ou não – é necessariamente um ato dialógico. Para discorrermos a esse respeito, discutimos neste texto como os modos de citação presentes nos artigos científicos são complexamente organizados e mostram nuances das diferentes manifestações das relações dialógicas que constituem o processo de escrita desse gênero discursivo.

Procuramos, destarte, responder neste artigo à seguinte questão: *como as relações dialógicas constitutivas da linguagem se expressam em termos dos modos de citação empregados em artigos científicos da área das Humanidades?* Objetivamos, desse modo, explicitar que as citações presentes em artigos científicos podem ser mais bem compreendidas quando se considera o papel das relações dialógicas estabelecidas pelo escrevente ao tecer vínculos dialógicos entre palavras próprias e palavras alheias.

Sublinhamos que o presente artigo decorre de uma pesquisa mais ampla, que contemplou o estudo e a análise das relações dialógicas constituintes do gênero acadêmico artigo científico nas três grandes áreas do conhecimento – Ciências Biológicas, Exatas e Humanas. Neste artigo, contudo, focalizaremos apenas as publicações que compreendem a grande área de Ciências Humanas, uma vez que a discussão referente às demais áreas do conhecimento renderia análises extensas, as quais não caberiam e não seriam contempladas a contento no espaço composicional deste trabalho¹.

Somadas a essa pergunta principal, anteriormente transcrita, entram em cena também outras questões norteadoras do presente trabalho, a saber: (i) como se dão as relações dialógicas entre o discurso autoral e as vozes outras (alheias) na escrita dos artigos científicos?; (ii) como os tipos e variantes de discurso expõem as interações dialógicas eventualmente encontradas nesses artigos?; e (iii) quais as possíveis relações dialógicas estabelecidas entre os artigos científicos representativos da grande área de Ciências Humanas e as vozes outras (alheias) que auxiliaram na construção desses artigos?

Tendo em consideração essas questões, o presente texto tem por objetivo central caracterizar o papel das relações dialógicas na constituição da escrita de *artigos científicos* representantes das publicações relativas à área de Ciências Humanas². Essa análise está ancorada nos tipos e variantes de discurso que se encontram nas obras do Círculo de Bakhtin – *Marxismo e filosofia da linguagem* (2021) e *Problemas da Poética de Dostoiévski* (2018).

A fim de alcançar esse objetivo, valemo-nos da metodologia de estudo da língua proposta por Volóchinov em *Marxismo e filosofia da linguagem* (2021) e que será detalhada em

¹ O trabalho completo de que resulta parte das discussões expostas neste artigo pode ser consultado em Silva (2024).

² Optamos pelo uso da expressão “área de Ciências Humanas” como um modo de uniformizar as muitas classificações que são feitas pela CAPES em relação às Humanidades. Além disso, partimos também do pressuposto de que as diferentes áreas do conhecimento são tradicional e historicamente divididas em três grandes áreas: Ciências Biológicas, Ciências Exatas e Ciências Humanas.

nosso texto mais à frente. Associado a essa abordagem, também nos valem do paradigma indiciário, proposto por Ginzburg (1991), dado que não analisaremos todos os casos de citação possíveis encontrados nos artigos científicos, mas apenas ocorrências que “saltaram a nossos olhos” (Ginzburg, 1991, p. 179), conforme a proposta do paradigma.

A escolha de alguns excertos exemplares dos modos de citação buscou ilustrar de maneira satisfatória como os tipos e variantes de discurso influenciaram, sobretudo, na construção composicional dos artigos da área de Ciências Humanas. Buscou, também, ilustrar como esses tipos de discurso orientaram a constituição da escrita desses artigos em termos de menção ao discurso do outro.

Este artigo delimita-se à análise de artigos científicos da área das Ciências Humanas, diferentemente da pesquisa mais ampla que lhe deu origem, a qual contemplou outras grandes áreas do conhecimento. O recorte ora assumido permite aprofundar a análise qualitativa dos modos de citação enquanto manifestações das relações dialógicas, concentrando-se nas especificidades axiológicas e discursivas que caracterizam esse domínio disciplinar em específico. Assim, a contribuição deste estudo reside na descrição detalhada do funcionamento dos tipos e variantes de discurso na constituição do artigo científico em Ciências Humanas.

O presente artigo está organizado em duas seções relativas à fundamentação teórica; nelas, discorreremos a respeito do gênero do discurso *artigo científico* no viés bakhtiniano e acerca da manifestação das relações dialógicas quanto ao uso e à apropriação dos tipos e variantes de discurso. Na sequência, apresentamos os procedimentos metodológicos com vistas à concretização dos objetivos propostos. Adiante, mostramos a análise de trechos selecionados de artigos científicos, a fim de ilustrar nossa discussão; por fim, tecemos as considerações finais, destacando as principais reflexões advindas deste estudo.

2 O gênero do discurso artigo científico como um complexo de relações dialógicas

Sob o olhar bakhtiniano e do Círculo, os gêneros são concebidos como tipos *relativamente estáveis* de enunciados, considerada a possibilidade de se alterarem no curso do desenvolvimento histórico do tempo. Os gêneros do discurso são enunciados por meio dos quais a linguagem se manifesta e, portanto, se realiza; os elementos constitutivos dos gêneros são: conteúdo (temático), construção composicional e estilo. No dizer de Bakhtin, todos esses três elementos estão “indissoluvelmente ligados *no conjunto* do enunciado e são igualmente determinados pela especificidade de um determinado campo da comunicação” (2016, p. 12, grifo do autor). Esses três elementos – conteúdo temático, construção composicional e estilo –, portanto, não estão separados do complexo do enunciado; ao contrário: formam um todo concreto.

Cada enunciado, também na concepção bakhtiniana e do Círculo, é concebido mediante as *particularidades* de determinada esfera da comunicação humana. No caso do artigo científico – objeto de análise do presente trabalho –, entram em cena determinações e implicações próprias a práticas letradas acadêmico-científicas, tais como as diretrizes de submissão e avaliação em determinados periódicos de divulgação científica, as normas da Associação Brasileira de Normas Técnicas (ABNT), *Vancouver*, *American Psychological Association* (APA), políticas científicas, dentre outros aspectos intrínsecos à seara da escrita acadêmica.

Bakhtin (2016) também divide os gêneros em primários e em secundários. Tal classificação é estabelecida observando o vínculo que esses gêneros podem apresentar em relação a uma dada realidade concreta. Os gêneros primários são aqueles que dispõem de um forte vínculo com a realidade imediata. Como exemplo, poderíamos citar a ata de uma reunião, dado que, se os participantes faltarem à reunião, não haverá possibilidade de registro das decisões tomadas nessa reunião. Os gêneros secundários, por seu turno, não apresentam esse vínculo tão fortemente ligado à realidade imediata. Como exemplo, poderíamos citar uma charge: ela pode ser lida a qualquer momento do tempo e em vários espaços. Desse ponto de vista, não estamos imersos em uma relação de dependência com o momento presente.

Isso posto, na perspectiva bakhtiniana e do Círculo, o *artigo científico* deve ser compreendido como um gênero discursivo secundário, uma vez que pertence a um contexto de produção cultural mais elaborado e demasiado complexo – o da seara da escrita acadêmica – e, desse modo, atende às especificidades/particularidades desse campo da atividade humana. Vale ressaltar, também, que o próprio Bakhtin comenta, em *Os gêneros do discurso* (2016), o fato de os textos advindos da pesquisa científica serem gêneros secundários, conforme observamos na citação a seguir:

Os gêneros discursivos secundários (complexos – romances, dramas, ‘pesquisas científicas de toda espécie’, os grandes gêneros publicísticos, etc.) surgem nas condições de um convívio cultural mais complexo e relativamente muito desenvolvido e organizado (predominantemente o escrito) – ficcional, científico, socio-político, etc. No processo de sua formação eles incorporam e reelaboram diversos gêneros primários (simples), que se formaram nas condições da comunicação discursiva imediata. (Bakhtin, 2016, p. 15, grifos nossos).

Sendo também um gênero do discurso, a produção do artigo científico passa pela construção composicional, pelo conteúdo temático e pelo estilo inerentes a esse gênero. Além disso, por se vincular a uma determinada área do conhecimento, o artigo científico refletirá e refratará as *particularidades* dessa esfera da atividade humana. Assim, entram em cena implicações específicas para a construção e para a formulação desses três elementos constitutivos do enunciado no todo do artigo científico.

Importante sublinhar também – e esse é o interesse particular do presente artigo – as diversas formas de retomada, antecipação e citação do discurso de outrem, isto é, as marcas evidentes de relações dialógicas que podem ser vislumbradas neste gênero. As relações dialógicas são interpretadas como as relações de diálogo (em sentido amplo) entre as vozes mobilizadas pelos autores nos enunciados por eles produzidos. Assim, no momento em que falamos ou escrevemos, esperamos sempre uma *posição responsiva* – uma resposta ativa – por parte do nosso interlocutor, ainda que não haja resposta em voz alta. Essa resposta ativa caracterizar-se-ia como a preocupação sentida no tocante à forma como o nosso texto ou fala poderá vir a ser recebido. Ou seja, preocupamo-nos com a recepção de nosso enunciado.

Ademais, todo enunciado se relaciona, necessariamente, com enunciados passados. Nas palavras de Bakhtin (2016, p. 57, grifo do autor), cada enunciado

[...] é pleno de ecos e ressonâncias de outros enunciados com os quais está ligado pela identidade da esfera de comunicação discursiva. Todo enunciado deve ser visto antes de tudo como uma *resposta* aos enunciados precedentes de um deter-

minado campo (aqui concebemos a palavra “resposta” no sentido mais amplo): ela os rejeita, confirma, completa, baseia-se neles, subentende-os como conhecidos, de certo modo os leva em conta. Porque o enunciado ocupa uma posição ‘definida’ em uma dada esfera da comunicação, em uma dada questão, em um dado assunto, etc. É impossível alguém definir sua posição sem correlacioná-la com outras posições.

Para Bakhtin, portanto, é impossível alguém definir sua posição sem correlacioná-la com outras posições; daí entram em cena os *elos antecedentes* e os *elos futuros*. Os primeiros são entendidos como as relações/conexões que se manifestam entre um enunciado com textos ou falas passadas; os segundos, como as possíveis relações/conexões que poderão ser suscitadas entre um enunciado e outros enunciados porvindouros. Toda a comunicação humana, sob o olhar bakhtiniano e do Círculo, tem por base esses elos cuja função é unir as palavras e as ideias de um indivíduo com outro indivíduo. A esses elos, atribui-se o nome de *relações dialógicas*. Podemos depreender, então, que o falante ou escrevente sempre leva em consideração enunciados passados para a construção dos próprios enunciados³ dele, os quais suscitarão outras respostas – não inéditas, pois partem de um já-dito – em momentos futuros. O gênero discursivo é, ao fim e ao cabo, interpretado como um complexo de relações dialógicas.

Na próxima seção, detalhamos os tipos e variantes de discurso, tendo por base *Marxismo e filosofia da linguagem* (2021) e *Problemas da poética de Dostoiévski* (2018). A seguir, em nossa análise, procuraremos discorrer acerca dos modos pelos quais tais tipos e variantes de discurso podem se manifestar na constituição da escrita de artigos científicos.

3 Relações dialógicas manifestadas pelos usos dos tipos e variantes de discurso

Em *Problemas da poética de Dostoiévski* (2018), Bakhtin realiza classificações quanto aos “tipos de discurso da prosa”. Ao fazer essas classificações, o autor lista formas de relacionamento possíveis entre a palavra própria e a palavra alheia. Por sua vez, Volóchinov, em *Marxismo e filosofia da linguagem* (2021), elenca e analisa vários tipos de discurso, diretos, indiretos e indireto livre.

De nossa parte, neste artigo, destacaremos apenas os tipos e variantes de discurso apresentados pelo Círculo, nas obras *Problemas da Poética de Dostoiévski* (2018) e *Marxismo e filosofia da linguagem* (2021), que se mostraram influentes no processo de escrita dos fragmentos textuais dos artigos científicos selecionados para análise.

Um primeiro tipo de discurso, apresentado por Bakhtin e que citaremos aqui, denomina-se *discurso direto orientado para o referente* (Bakhtin, 2018, p. 213). Esse tipo de discurso pode ser identificado no momento em que o discurso autoral se volta única e exclusivamente a um determinado objeto, a um único e exclusivo referente. A título de ilustração, conforme aponta Silva (2024), pensemos em um objeto material, como uma árvore. Caso alguém pronuncie a

³ Mais à frente, essa questão será observada com base na análise dos diferentes modos de citação em artigos científicos por meio dos quais poderá ser notada a réplica a enunciados passados, bem como a capacidade de suscitar respostas futuras, caracterizando, assim, um modo de olhar para as relações dialógicas atuantes na escrita de artigos científicos.

frase “esta árvore é linda!”, tem-se um exemplo de um discurso voltado exclusivamente a um objeto, a um único referente, a saber: a árvore.

Outro tipo de discurso apresentado por Bakhtin, em *Problemas da Poética de Dostoiévski* (2018), é o *discurso objetificado* (discurso da pessoa representada). Nesse tipo de discurso, um autor de prosa torna o discurso da personagem por ele criado o seu objeto.

Conforme Bakhtin (2018), o discurso da personagem é orientado única e exclusivamente para o seu objeto e esse discurso também é objeto de outra orientação – a orientação do autor. Nesse sentido, o discurso da personagem volta-se a um referente sobre o qual se fala; mas também esse mesmo discurso da personagem funciona como o objeto de discurso do autor. Ainda no que toca a esse tipo de discurso, a unidade de enunciação da personagem *subordina-se* à unidade de enunciação do autor, tendo em vista que é objeto dela. Logo, na proposta de Bakhtin, o *discurso objetificado* se refere ao discurso da personagem; neste trabalho, no entanto, esse conceito é expandido para a análise das formas de citação nos artigos científicos. Assim, em nossos dados, o *discurso objetificado* não compreende o discurso da personagem, mas o discurso de outros autores reais com os quais o(s) escrevente(s) dos artigos científicos dialoga(m).

Considerando o princípio bakhtiniano e do Círculo de que o enunciado sempre leva em conta enunciados alheios, entra em cena um terceiro tipo de discurso: o *discurso orientado para o discurso do outro*, a que Bakhtin atribui o nome de “discurso bivocal”. Bakhtin discrimina três variedades de *discurso bivocal* – em que há a presença de duas vozes imbricadas –, a saber: (i) discurso bivocal de orientação única; (ii) discurso bivocal de orientação vária; e (iii) discurso bivocal de tipo ativo.

Neste artigo, discorreremos apenas acerca de duas variedades – *discurso bivocal de orientação única* e *discurso bivocal de orientação vária* –, uma vez que apenas essas variedades se fizeram presentes de maneira sistemática no *corpus*, assumindo função significativa na estruturação do discurso autoral e na incorporação da palavra do outro. Quanto à primeira variedade do discurso bivocal, o *discurso bivocal de orientação única*, esclarece Bakhtin (2018, p. 221): “consiste em que o autor inclui no seu plano o discurso do outro voltado para as suas próprias intenções”. Nesse tipo de discurso, há duas vozes, mas essas vozes estão seguindo uma única orientação – a orientação prescrita pelo autor. Conforme cita Bakhtin (2018), um exemplo típico de *discurso bivocal de orientação única* é o discurso do narrador, dado que, nesse caso, a voz do narrador está a serviço da voz do autor. Nesse tipo de discurso, não há conflito entre ambas as vozes – a voz do discurso autoral e a voz do discurso alheio –, pois ambas convergem para uma mesma e única orientação: a orientação prescrita pelo autor do texto. Assim, a voz alheia está *a serviço* da voz do discurso autoral, uma vez que “serve”, por assim dizer, aos propósitos discursivos do discurso autoral.

Quanto à segunda variedade do discurso bivocal, *discurso bivocal de orientação vária*, temos que “o discurso se converte em palco de luta entre duas vozes com orientação diametralmente oposta” (Bakhtin, 2018, p. 221). O autor russo traz como exemplo o discurso parodístico, pois, nesse caso, o autor obriga a palavra do outro a servir a fins opostos àqueles originalmente empregados.

Por último e não menos importante, Bakhtin ainda cita o *diálogo velado* como uma das manifestações do discurso bivocal de tipo ativo. Assim considera Bakhtin (2018, p. 226):

Imaginemos um diálogo entre duas pessoas no qual foram suprimidas as réplicas do segundo interlocutor, mas de tal forma que o sentido geral não tenha sofrido qualquer perturbação. O segundo interlocutor é invisível, suas palavras estão ausentes, mas deixam profundos vestígios que determinam todas as palavras presentes no primeiro interlocutor.

No diálogo velado, o interlocutor não se manifesta de forma explicitamente representada no enunciado. Ainda assim, sua presença é pressuposta, na medida em que o discurso se organiza como réplica a enunciados anteriores ou antecipados. Bakhtin (2018) assinala que, na obra de Dostoiévski, esse tipo de dialogização é particularmente elaborado e assume posição de destaque na construção estética do romance. Tal procedimento discursivo comparece em diversas obras do autor, entre as quais *Gente pobre*, *O duplo*, *Memórias do subsolo* e *Os irmãos Karamázov*.

Assim, no *diálogo velado*, esse “outro” com o qual a personagem dialoga é ausente; porém, suas palavras deixam sinais as quais determinam que tudo o que será dito por ela – pela personagem⁴ – recebe influência desse outro. Bakhtin acrescenta que o *diálogo velado* é amplamente difundido nas narrativas de Dostoiévski. Verificaremos, adiante, se há, na escrita dos artigos científicos, exemplos de diálogo velado, por meio do qual o escrevente dialoga consigo mesmo tendo em conta as réplicas de um interlocutor ausente.

Passemos agora aos tipos de discurso abordados em *Marxismo e filosofia da linguagem* (2021), com destaque àqueles que se mostraram influentes nos fragmentos textuais selecionados para análise.

Um dos tipos de discurso examinado por Volóchinov (2021) é o *discurso antecipado*, *disperso*, *oculto*, cuja manifestação ocorre quando o discurso do outro passa a “contaminar” o discurso autoral. Desse ponto de vista, o discurso autoral recebe influências *diretas* do discurso alheio. Assim, há um embate entre duas entonações: a do autor e a das personagens, levando-nos, por consequência, ao fenômeno da *interferência discursiva* (Volóchinov, 2021, p. 284). Mostra-se esse embate no conflito entre o discurso narrativo e as particularidades linguísticas do discurso da personagem. Segundo Volóchinov (2021), as influências que o discurso autoral recebe das palavras outras (alheias) acabam por atenuar as fronteiras entre os aspectos afetivo-emocionais do autor e aqueles aspectos afetivo-emocionais do discurso do outro.

Outro tipo de discurso apresentado por Volóchinov (2021) é o *discurso direto substituído*. Nesse tipo de discurso, o autor passa a falar pelo outro, isto é, falar no lugar e em nome do outro com o qual dialoga. Nesse caso, o autor tenciona se antecipar ao discurso do outro, de modo a falar por ele aquilo que poderia ser dito pelo próprio personagem.

No *discurso direto em estilo monumental*, por sua vez, o enunciado alheio é interpretado como um todo compacto, indivisível, imutável e impenetrável. Esse tipo de discurso difere, em termos de sentido, do discurso indireto, em que há a transmissão analítica do discurso alheio.

Segundo Volóchinov (2021), uma das tendências do discurso indireto é realizar a análise do discurso alheio, daí que se nomeie essa tendência de analítica do discurso indireto. Nessa tendência do discurso indireto, os elementos afetivo-emocionais do discurso alheio são transferidos para o conteúdo do discurso indireto.

⁴ Mais uma vez, ressalte-se que esse conceito é expandido para a análise de formas de citação de outros autores reais.

Exemplifiquemos a título de esclarecimento. Quando alguém diz, por exemplo, “Que maravilha! Isso sim é um trabalho de qualidade!”, tem-se um discurso direto. Já no discurso indireto, teríamos algo como: “Ele disse que aquilo era uma maravilha e um verdadeiro trabalho de qualidade”. Temos em cena, portanto, a passagem dos aspectos afetivo-emocionais para o conteúdo do discurso indireto relato, o que não ocorre com o *discurso direto em estilo monumental*. Isso porque o discurso direto em estilo monumental é citado de modo compacto, sem interferências do discurso autoral em seu interior. Os aspectos afetivo-emocionais do discurso alheio permanecem inalterados, sem transmissão analítica por parte do discurso autoral.

Outro tipo de discurso apontado por Volóchinov (2021) e que se mostra pertinente para a análise empreendida a seguir é o *discurso direto preparado*, cuja manifestação ocorre quando o discurso autoral “prepara”, por assim dizer, a emergência do discurso alheio, de modo a sublinhar a sua diferença. Ocorre quando o autor antecipa, por exemplo, os temas do discurso alheio dos autores com os quais dialoga. O discurso indireto do autor “prepara” (antecipando) temas do discurso direto alheio que virão na sequência. O discurso direto alheio, assim, surge do discurso indireto autoral. Desse modo, o futuro discurso direto e os principais temas dele são antecipados e preparados pelo contexto e pelas entonações do discurso indireto do autor.

A mobilização dessas categorias teóricas não tem caráter meramente classificatório. Ao analisar o artigo científico sob a perspectiva das relações dialógicas, assume-se que cada modo de citação revela posicionamentos axiológicos e estratégias de inscrição do sujeito na cadeia discursiva da área. Desse modo, os tipos e variantes de discurso não são apenas formas estruturais, mas índices do modo como o pesquisador se situa frente à tradição teórica que convoca.

Importa reconhecer que a taxonomia dos tipos e variantes de discurso, tal como sistematizada por Bakhtin e Volóchinov, foi originalmente elaborada para a análise do romance polifônico, especialmente a partir da obra de Dostoiévski. Nesse contexto, categorias como *discurso bivocal*, *diálogo velado* e *discurso objetificado* visavam a descrever a coexistência de consciências plenivalentes em um universo ficcional estruturado pela não finalização e pelo confronto axiológico entre vozes.

A transposição dessas categorias para o domínio do gênero artigo científico não pode, portanto, ser realizada de modo acrítico. Há uma diferença fundamental de natureza axiológica entre os dois gêneros. No romance polifônico, o dialogismo radical constitui o próprio princípio estético organizador da obra, sustentando a multiplicidade de verdades não fundidas. Já no artigo científico, embora o diálogo com vozes anteriores seja constitutivo do gênero, a orientação argumentativa tende a organizar-se em direção à sustentação de uma tese específica, amiúde subordinando vozes alheias ao projeto enunciativo do autor.

Reconhecida essa diferença, sustentamos, entretanto, que as categorias bakhtinianas mantêm valor heurístico para a análise do discurso científico, desde que compreendidas não como descrição de uma polifonia plena, mas como instrumentos para examinar os modos de incorporação e reacentuação do discurso de outrem. O que se investiga, neste estudo, não é a existência de consciências plenivalentes em conflito aberto, mas as formas pelas quais o artigo científico articula, enquadra, legitima ou tensiona vozes precedentes na construção da argumentação.

Assim, a aplicação da taxonomia não implica equiparar o artigo científico ao romance polifônico presente nas obras de Dostoiévski, mas explorar a produtividade analítica das categorias para compreender como o discurso científico realiza, em seu próprio horizonte axiológico, processos de dialogização.

Na seção que se segue, apresentamos noções que permearam nossos procedimentos metodológicos.

4 Metodologia e *corpus* de pesquisa: artigos científicos representativos da área das Ciências Humanas

Em *Marxismo e filosofia da linguagem* (2021), Volóchinov aponta um caminho, uma ordem metodológica a partir da qual se pode estudar a língua. Tal ordem, conforme Volóchinov (2021, p. 220), passaria pelas:

- 1) formas e tipos de interação discursiva em sua relação com as condições concretas; 2) formas dos enunciados ou discursos verbais singulares em relação estreita com a interação da qual são parte, isto é, os gêneros dos discursos verbais determinados pela interação discursiva na vida e na criação ideológica; 3) partindo disso, revisão das formas da língua em sua concepção linguística habitual.

Desse modo, assumindo esse percurso metodológico e ajustando-o a nossos dados, procuramos seguir: (a) descrever o contexto de constituição dos artigos; (b) descrever como gênero discursivo o artigo científico por meio do qual se dão as interações discursivas analisadas; e (c) examinar as relações dialógicas constitutivas do gênero do discurso acadêmico artigo científico, privilegiando os tipos e variantes de discurso por meio dos quais se tecem as relações dialógicas nos artigos sob análise.

Para desenvolver essa exposição, optamos por determinados artigos e destacamos certos modos de citação. A escolha dos fragmentos a serem analisados a seguir se deu por intermédio da leitura e da comparação entre os três primeiros artigos científicos listados na edição de volume 44⁵ e número 3 do periódico “Cadernos Nietzsche” e de número 68 do periódico “Cadernos Pagu”.

A seleção dos fragmentos analisados foi orientada por três critérios: (i) presença explícita de modos de citação direta ou indireta; (ii) ocorrência de marcas linguístico-discursivas que evidenciam articulação entre discurso autoral e discurso alheio; (iii) potencial ilustrativo para exemplificar diferentes tipos e variantes de discurso descritos por Bakhtin e Volóchinov. O paradigma indiciário fundamenta a análise interpretativa, tendo em vista que se baseia em “elementos imponderáveis: faro, golpe de vista, intuição”, em “formas de saber tendencialmente *mudas* – no sentido de que [...] suas regras não se prestam a ser formalizadas nem ditas” (Ginzburg, 1991, p. 179, grifos do autor); contudo, não elimina a necessidade de critérios explícitos de seleção.

O paradigma indiciário, nesse contexto, nos interessa como procedimento metodológico não tanto por nos determos a indícios, mas, antes, por nos permitir selecionar, a partir de nossas vivências e conhecimentos, aquilo que nos “salta” aos olhos. Elegem-se, assim, fragmentos dos artigos científicos que puderam ilustrar as relações dialógicas em termos dos diferentes tipos e variantes de discurso. A metodologia empregada no presente trabalho, desse modo, conjuga os postulados bakhtinianos a elementos imponderáveis, conforme citado por Ginzburg (1991).

⁵ À época de realização da pesquisa, era a edição publicada mais recente.

Inicialmente, buscamos periódicos acadêmicos representativos da área das Ciências Humanas. No momento de nossa busca, segundo semestre de 2023 – setembro a novembro de 2023 –, na *Plataforma Sucupira*⁶, selecionamos as revistas “Cadernos Nietzsche” e “Cadernos Pagu”, as quais estavam inseridas na grande área de Ciências Humanas e classificadas como A1, o que nos possibilitou considerá-las representativas da área em que se encontravam inseridas e classificadas.

Optamos por esse estrato, dado que essa “nota” atribuída indica que os periódicos e os respectivos artigos científicos neles publicados alcançaram maior número de citações referentes à avaliação quadrienal 2017-2020 da CAPES. Como corolário dessa avaliação, os periódicos com maior número de citações tendem a gozar de mais prestígio na esfera acadêmica, sobretudo em termos de divulgação científica.

No que toca ao estrato⁷ Qualis, é oportuno lembrar como se efetiva essa classificação quadrienal dos periódicos científicos realizada pela CAPES. A classificação dos periódicos acadêmicos é utilizada historicamente como qualificação *indireta* da produção intelectual dos artigos científicos publicados nessas plataformas de divulgação científica – os periódicos (Capes, 2023). A qualificação é indireta tendo em vista que não se atribui um conceito *diretamente* aos artigos científicos vinculados a esses suportes, mas, por meio do Qualis, atribuem-se conceitos a esses veículos de divulgação nos quais se inscrevem os artigos.

Nesse contexto, os periódicos classificados no estrato A1 gozam de mais prestígio na seara da divulgação científica. Por extensão, os artigos científicos a eles vinculados também parecem receber mais prestígio acadêmico. Optamos por trabalhar com artigos científicos inscritos em periódicos cuja classificação esteja posicionada no estrato A1 na medida em que essa validação parece indicar que os textos presentes nesses periódicos são reconhecidos como bons representantes do gênero discursivo artigo acadêmico em termos temáticos, estilísticos e composicionais do gênero na esteira do que propõe Bakhtin (2016) no que concerne aos estudos dos gêneros do discurso. Consequentemente, os modos de citação empregados nos textos analisados são também (re)conhecidos como adequados a esse gênero discursivo e a essa esfera da atividade humana.

No quadro a seguir, listamos as revistas e os respectivos artigos que foram objeto de análise⁸.

Quadro 1 – Periódicos científicos que constituem o corpus deste artigo

Ciências Humanas
“Cadernos Nietzsche” Disponível em: https://www.scielo.br/j/cniet/i/2023.v44n3/
“Cadernos Pagu” Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/cadpagu/issue/view/2097

Fonte: Elaboração própria.

⁶ Disponível em: <https://sucupira.capes.gov.br/sucupira/>.

⁷ “Estrato” é preferível à “nota” em virtude de o Qualis ser mais propriamente uma espécie de ranqueamento, que enfeixa complexos cruzamentos de critérios.

⁸ Os artigos analisados concernem à edição mais recente do periódico à época de realização da pesquisa.

A seguir, transporemos em forma de imagem (captura de tela) como os três artigos estavam situados no momento de nossa coleta de dados, em novembro de 2023, das revistas escolhidas, a saber: “Cadernos Nietzsche” e “Cadernos Pagu”.

Figura 1 – Os três primeiros artigos listados na edição analisada da “Cadernos Nietzsche”



Fonte: <https://www.scielo.br/j/cniet/i/2023.v44n3/>

Figura 2 – Os três primeiros artigos listados na edição analisada da “Cadernos Pagu”

Artigos

Camila Fernandes	1-18
Desejo e acusação: a “novinha” e a sexualidade feminina como fonte de tensão	
PDF	
Diego Calmon	1-17
Bissexualidade e ambiguidade: relações metafóricas e processos metonímicos em produções discursivas sobre a bissexualidade	
PDF (English)	PDF (English)
Karla Garcia Luiz, Marivete Gesser	1-14
Mulheres com deficiência e dependência complexa: experiências de relações de cuidado para (sobre)viver	
PDF	PDF (English)

Fonte: <https://periodicos.sbu.unicamp.br/ojs/index.php/cadpagu/issue/view/2097>

À época de realização da pesquisa, optamos por selecionar apenas 3 (três) artigos de cada revista em função da necessidade de refinarmos os nossos dados, tendo em vista a impossibilidade de contemplar a contento todos os artigos publicados na edição analisada. Na sequência, passemos à análise e à discussão dos resultados.

5 Exame das relações dialógicas em artigos científicos da área das Ciências Humanas

Nesta seção, analisamos artigos científicos da área das Ciências Humanas, tencionando desvelar de que modo as relações dialógicas podem servir a diferentes projetos discursivos e a objetivos comunicacionais. Conforme já informado, optamos por focalizar determinados trechos dos três primeiros artigos científicos listados na edição analisada de ambos os periódicos escolhidos.

Dentre os três artigos listados primeiramente na edição de cada revista, analisamos, nesta seção, apenas excertos dos seguintes artigos: “Abrindo trilhas com leituras: reflexões sobre a gênese de *O nascimento da tragédia*”, “Entre a arte e a filosofia: Nietzsche e o lugar de Platão em *O nascimento da tragédia*”, publicados na Cadernos Nietzsche. Publicados na Cadernos Pagu, analisamos os artigos “Desejo e acusação: a ‘novinha’ e a sexualidade feminina como fonte de tensão”, “Bissexualidade e ambiguidade: relações metafóricas e processos metonímicos em produções discursivas sobre a bissexualidade”.

Esses artigos foram escolhidos para constituírem o *corpus* desta exposição, visto que os fragmentos selecionados ilustraram diferentes atuações dos tipos e variantes de discurso que influenciaram e orientaram o processo de escrita do(s) artigo(s) científico(s) analisado(s) em termos de menção ao discurso do outro. Assim, julgamos que os fragmentos textuais desses artigos analisados contemplam a contento nossos objetivos neste trabalho e podem ilustrar de maneira satisfatória os aspectos das relações dialógicas que nos propusemos a comentar.

Tendo em vista o nosso interesse em analisar os modos de citação como constitutivos das relações dialógicas, passemos a examinar alguns excertos dos artigos, para exemplificar as questões que pretendemos discutir.

O primeiro fragmento a ser analisado foi retirado do artigo “Abrindo trilhas com leituras: reflexões sobre a gênese de *O nascimento da tragédia*” e se apresenta da seguinte forma no todo composicional do artigo:

- (1) É nesse sentido que se deve compreender a afirmação de Nietzsche de que Dioniso é o “efetivo herói cênico” e que todas as figuras do palco grego são apenas “máscaras” do deus (GT/NT10, KSA1.69).

Nesse trecho, percebemos a atuação dos: a) *discurso bivocal de orientação única*, b) *discurso direto em estilo monumental*, c) *discurso direto preparado*, d) *discurso antecipado, disperso, oculto* e e) *discurso direto substituído*. As duas vozes presentes na constituição desse trecho não apresentam orientações opostas entre si; ao contrário: ambas convergem para uma mesma e única orientação. Orientação segundo a qual se deve compreender a afirmação de Nietzsche como “efetivo herói cênico” e que as outras figuras do palco seriam apenas “máscaras” do deus. Não há, portanto, conflitos/divergências entre ambas as vozes citadas, uma vez que elas seguem uma mesma e única orientação. Assim, a voz do discurso autoral e a voz do discurso alheio estão embasadas no uso do *discurso bivocal de orientação única*.

Em “efetivo herói cênico”, temos a atuação do *discurso direto em estilo monumental*. Por meio do uso desse tipo de discurso, o enunciado alheio é compreendido como um todo compacto, indivisível, imutável e impenetrável. Notamos que as palavras de outrem foram transpostas para o contexto do discurso autoral *ipsis litteris*, isto é, da mesma maneira com que originalmente foram empregadas. Assim, preserva-se o caráter “imutável” do enunciado do outro, uma vez que ele foi transposto para o contexto do discurso autoral sem alterações de qualquer ordem. Mediante o uso desse tipo de discurso, há também um modo de o autor⁹ se manter fiel às palavras do outro com o qual dialoga, tendo em vista que não houve alterações na passagem do discurso alheio para o contexto do discurso autoral.

Quanto aos usos do *discurso direto preparado* e do *discurso direto substituído*, eles se fazem presentes na constituição do excerto na medida em que o autor fala em nome do outro com quem dialoga. O uso das aspas, por exemplo, é um recurso gráfico que mostra textualmente a antecipação do discurso alheio para o contexto do discurso autoral. Elas – as aspas – antecipam/introduzem o discurso alheio no contexto do discurso autoral, visto que demarcam textualmente que o enunciado concerne ao outro, e não à voz autoral. Assim, temos em cena a atuação do *discurso direto preparado*. Com o uso da citação direta, o autor também fala no lugar do outro com o qual dialoga, já que a voz alheia apenas é trazida à baila por meio da voz autoral. Assim, temos em cena a atuação do *discurso direto substituído*.

O uso do *discurso antecipado, disperso, oculto* faz-se presente na medida em que notamos o fato de o discurso alheio passar a determinar o que foi dito pelo discurso autoral. Nesse ponto, podemos fazer um cruzamento, por assim dizer, com o discurso *bivocal de orientação única*, já que, se o fragmento dispõe de uma mesma e única orientação, é de se esperar que o discurso alheio tenha influenciado diretamente o que foi dito no contexto do discurso autoral. Assim, o discurso autoral recebeu influências *diretas* do discurso alheio; sem as quais o fragmento não teria sido escrito da maneira como foi.

O próximo fragmento sob análise ilustra a atuação do *diálogo velado*, uma variante do *discurso bivocal de tipo ativo*. Segue o trecho:

- (2) Escapa aos objetivos do presente artigo empreender uma análise detalhada de tal concepção de Schlegel, o que se pretende aqui é justamente estabelecer uma relação, a nosso ver fecunda, entre a leitura e o modelo entendido em sentido artístico.

Na seara da escrita acadêmica, como sabemos, é comum que se apresentem as limitações advindas de determinado trabalho realizado; algo que pode lançar luz, por exemplo, a sugestões para futuras pesquisas. É o que acontece com a construção dessa passagem sob análise: por meio do uso do *diálogo velado*, o escrevente considera as réplicas de um interlocutor ausente ou mesmo leva em conta a presença – não visível, mas sentida – de outrem. Assim, o autor dialoga consigo mesmo tendo em vista as possíveis réplicas de um interlocutor ausente – quer a comunidade acadêmica leitora, quer os pareceristas do artigo em voga. Apresentar as limitações de determinada pesquisa realizada é um movimento que leva em conta a presença do outro, pois esse movimento poderia ser uma réplica a possíveis indagações dos leitores, por exemplo, ou mesmo dos pareceristas em relação a objetivos não contemplados pelo presente artigo. Logo, o diálogo velado pode ser um modo pelo qual o autor

⁹ Sabemos que pode haver a presença de mais de um autor nos artigos científicos selecionados; optamos, no entanto, por “autor” e “escrevente”, no singular, por questões de uniformidade e de padronização.

pode evitar possíveis críticas no que tange à construção do seu artigo, contribuindo, também, em certa medida, para que o artigo seja aceito para publicação no periódico escolhido.

O próximo excerto a ser analisado ilustra a atuação do *discurso direto preparado*, do *discurso direto substituído* e do *discurso bivocal de orientação vária*. Vejamos como ele se apresenta no todo composicional do artigo:

- (3) Em sua Introdução, Machado afirma que, ao se referir aos gregos como “nossos luminosos guias”, Nietzsche dá continuidade ao projeto de Winckelmann, Goethe e Schiller de pensar a obra de arte moderna a partir de uma reflexão sobre a arte grega. Contudo, diferentemente destes autores, Nietzsche não toma os gregos como essencialmente ou exclusivamente apolíneos, e os relaciona ao que, para ele próprio, seria um aspecto mais profundo da cultura grega: o dionisíaco, que não havia sido pensado por aqueles.

Podemos notar que o escrevente narra, por assim dizer, as ideias do discurso do outro no momento em que introduz esse discurso no contexto do discurso autoral. Sentenças como “Machado afirma que” e “Nietzsche dá continuidade” são ilustrativas do modo pelo qual se dá a preparação do discurso do outro para fazer parte do discurso autoral. Essa antecipação do discurso alheio, por sua vez, é feita com base em uma narrativa, pois o discurso autoral “conta”, por assim dizer, sobre a introdução de Machado. Ao partir dessa introdução de Machado, o discurso autoral introduz as falas dos outros com os quais dialoga – Machado e Nietzsche. Desse modo, como houve a antecipação do discurso alheio de modo a fazer parte do discurso autoral e como o discurso autoral fala pelo e no lugar do outro com que dialoga, temos em cena a atuação do *discurso direto preparado* e do *discurso direto substituído*.

No que respeita ao uso do *discurso bivocal de orientação vária*, sua atuação é notada no momento em que as vozes presentes apresentam orientações diametralmente opostas. Ressalve-se, em tempo, que essas vozes não dizem respeito ao discurso autoral, isto é, não são aquelas vozes concernentes ao discurso autoral, mas sim ao discurso alheio. O papel do escrevente, nesse caso, é narrar, por assim dizer, a atuação do *discurso bivocal de orientação vária*. Essa narração se dá a partir do momento em que o escrevente cita a conjunção opositiva “contudo”, a qual ilustra a presença de um conflito. Tal conflito se dá entre a voz de Nietzsche e a de outros autores mencionados pelo escrevente. Assim, o discurso autoral narra a orientação oposta entre a voz de Nietzsche e a voz de outros autores com os quais dialoga. Essa orientação oposta reflete o pensamento diferente de Nietzsche em relação aos seus contemporâneos: ele não toma os gregos como única e exclusivamente apolíneos. Por isso, nesse cenário, o escrevente narra a atuação do *discurso bivocal de orientação vária* na composição desse excerto.

O fragmento a seguir foi retirado do todo composicional do artigo “Desejo e acusação: a ‘novinha’ e a sexualidade feminina como fonte de tensão”, publicado na “Cadernos Pagu”. O excerto ilustra a atuação do *discurso objetificado*, *discurso direto em estilo monumental*, *discurso direto preparado* e do *discurso direto substituído*. Vejamos como ele se apresenta:

- (4) Olha, você não vê? Essas novinhas não param de fazer filho!”, comenta Flor, uma mulher de 27 anos. Mulheres mais velhas, porém ainda jovens, circulam as mensagens com tom de advertência: “tá vendo, depois abandona aí, porque tem sempre gente pra cuidar”, diz Karen, uma mulher de 32 anos ao se referir às “novinhas”.

O *discurso objetificado* é influente na composição do trecho, especialmente porque o discurso alheio citado se torna o próprio objeto de discurso do escrevente. O discurso alheio citado é o tema sobre o qual se discorre no contexto do discurso autoral; e isso é feito de maneira evidente na composição do trecho, uma vez que as falas correspondentes ao discurso do outro são transpostas para o contexto do discurso autoral por meio do discurso direto. A voz alheia, ao ser objetificada, serve aos propósitos do projeto discursivo do autor, tendo em conta que é a respeito do discurso alheio que se discorre no contexto do discurso autoral. Por isso, o discurso alheio “serve” aos propósitos discursivos do autor, está *a serviço* do projeto discursivo do escrevente.

A influência do *discurso direto em estilo monumental* é percebida no momento em que a fala do discurso alheio é transposta para o contexto do discurso autoral *ipsis litteris*, isto é, da mesma forma com que originalmente foi empregada. Aqui, o uso do *discurso direto em estilo monumental* é também um modo de o autor se manter fiel às palavras de outrem, dado que elas são transpostas para o contexto do discurso autoral da mesma maneira com que originalmente foram empregadas. Por isso, há a interpretação de que o enunciado alheio seria um todo compacto, indivisível, imutável e impenetrável. Para o trabalho realizado, é importante que o autor transponha para o contexto do discurso autoral o discurso alheio da mesma forma com que originalmente foi empregado, pois é um modo de garantir credibilidade ao que está sendo discutido neste momento do artigo ao exemplificar o assunto tratado por meio da voz do discurso alheio. Ou seja, trata-se de uma estratégia argumentativa.

O uso do *discurso direto preparado* e do *discurso direto substituído* faz-se perceber na medida em que o autor antecipa/introduz o discurso do outro no contexto do discurso autoral e também no momento em que, a partir desse movimento, o escrevente fala por e no lugar do outro com o qual dialoga. Construções como “comenta Flor” e “diz Karen” são maneiras pelas quais o discurso autoral fala no lugar desse outro com o qual dialoga e também é uma maneira pela qual o discurso autoral introduz o discurso alheio na construção do trecho.

O próximo fragmento a ser analisado se apresenta da seguinte forma no todo composicional do artigo:

- (5) Cinthia emenda: “a culpa das novinha engravidar é das mães que não fica em cima, não fala de sexo, não fala de gravidez e às vezes fica feliz quando a filha fica grávida, posta no Facebook e diz que vai ser vovó”, sentença que localiza a natureza “errada” das “novinhas” num ponto nevrálgico para esta discussão, que diz respeito à relação entre mães e filhas, feita nas dobras do conflito e da amorosidade maternal.

Esse fragmento ilustra a atuação dos seguintes tipos de discurso: *discurso direto preparado*, do *discurso direto substituído* e do *discurso direto em estilo monumental*. O uso de “Cinthia emenda” ilustra de maneira evidente o modo pelo qual o discurso autoral “prepara”, por assim dizer, o discurso alheio para fazer parte do discurso autoral. Com esse movimento – o uso de “Cinthia emenda” –, o escrevente antecipa o discurso alheio no contexto do discurso autoral e também fala em nome desse outro (Cinthia) com o qual dialoga. Por isso, há a substituição do discurso do outro no momento em que o discurso autoral fala no lugar desse outro; sendo assim, a voz alheia apenas é trazida a lume por meio da voz concernente ao discurso autoral. Esses movimentos – antecipar o discurso alheio e falar no lugar desse discurso – caracterizam a atuação do *discurso direto preparado* e do *discurso direto substituído*.

Quanto à atuação do *discurso direto em estilo monumental*, esta é percebida no momento em que o discurso alheio é transposto para o contexto do discurso autoral *ipsis litteris*, isto é, da mesma forma com que originalmente foi empregado. Como sabemos, o uso desse tipo de discurso também pode ser um modo de o autor se manter fiel às palavras de outrem, já que, ao ser transposto para o contexto do discurso autoral, o discurso alheio não sofre alterações de qualquer ordem. Por isso, o enunciado alheio é entendido como um todo compacto, indivisível, imutável e impenetrável. Ao não sofrer alterações na transposição para o contexto do discurso autoral, preserva-se esse caráter “imutável” e “impenetrável” do enunciado.

O próximo fragmento textual a ser analisado foi retirado do artigo “Bissexualidade e ambiguidade: relações metafóricas e processos metonímicos em produções discursivas sobre a bissexualidade”. O excerto ilustra a atuação do *discurso direto preparado*, *discurso direto substituído*, *discurso direto em estilo monumental*, do *discurso antecipado*, *disperso*, *oculto* e do *discurso bivocal de orientação vária*. Vejamos como ele se apresenta:

- (6) Em entrevista a Gesser e Fietz (2019:141), Eva Kittay ressalta que é inevitável que pensem no cuidado, pois “um mundo em que ninguém se importasse com ninguém seria um mundo em que as necessidades daquelas(es) que não pudessem atender às suas próprias necessidades (e que somos todos nós em algum momento de nossas vidas) seriam negligenciadas”. Assim, a abordagem da temática do cuidado torna urgente e inevitável o rompimento com os ideais liberais de autonomia e independência, que estão fortemente presentes no contexto social em geral, inclusive nas teorias sobre justiça (Kittay, 2015).

O uso da expressão “Eva Kittay ressalta” é representativo da atuação do *discurso direto preparado* e do *discurso direto substituído*. Mediante o uso dessa expressão, o discurso autoral cumpre a função de antecipar/introduzir o discurso alheio no contexto do discurso autoral. Assim, a voz do discurso alheio apenas é trazida à cena por meio da voz do discurso autoral; a voz do discurso alheio apenas passa a fazer parte do discurso autoral porque é introduzida pela voz do discurso autoral, o que caracteriza a atuação do *discurso direto preparado*. Mediante o uso dessa expressão, o discurso autoral também cumpre a função de falar no lugar do outro com o qual dialoga; dessa maneira, a voz do discurso autoral apresenta o discurso alheio no momento em que fala por ela, fala no lugar dela, o que caracteriza a atuação do *discurso direto substituído*.

Temos também em cena a atuação de uma citação direta, o que significa dizer que temos em cena a atuação do discurso direto em estilo monumental. Por meio do uso desse tipo de discurso, o enunciado alheio é compreendido como um todo compacto, indivisível, imutável e impenetrável. O uso da citação direta preserva esse caráter “imutável” e “inalterável” do enunciado alheio, dado que, ao ser transposto para o contexto do discurso autoral, o enunciado alheio é deslocado *ipsis litteris*, isto é, da mesma maneira com que originalmente foi empregado. Assim, não houve alterações de qualquer ordem na transposição do enunciado alheio para o contexto do discurso autoral. O uso do *discurso direto em estilo monumental*, portanto, também pode ser um modo de o autor se manter fiel às palavras de outrem, visto que elas não sofrem alterações. Pode ser, também, uma maneira de o autor julgar que determinada ideia ou conceito pode ser mais bem explicado por meio das palavras do outro, considerando que a voz alheia é interpretada como uma voz de autoridade naquele assunto em voga, ao contrário de se valer de suas próprias palavras. Ou seja, trata-se novamente de estratégias argumentativas.

As inferências relacionadas à abordagem do cuidado apenas são tecidas com base na influência do discurso alheio. O discurso do outro passa, então, a determinar o que está sendo dito pelo discurso autoral. O discurso autoral se vale de uma paráfrase para citar as suas inferências que dizem respeito à abordagem da temática do cuidado. Essas inferências já foram ditas em algum contexto por outrem; o discurso autoral as reformula com as suas próprias palavras. Nesse caso, o autor julga que as suas próprias palavras podem explicar de maneira satisfatória a abordagem da temática do cuidado e não vê, portanto, a necessidade de recorrer às palavras alheias para discorrer acerca dessa temática. Essas influências do discurso alheio no contexto do discurso autoral caracterizam a atuação do *discurso antecipado, disperso, oculto*.

A voz do discurso autoral narra, por assim dizer, a atuação do *discurso bivocal de orientação vária*. O discurso autoral sugere que haja o rompimento com os ideais liberais de autonomia e independência. Esse rompimento ilustra o modo pelo qual o discurso se converteu em um palco de luta entre duas vozes – a voz do discurso autoral e a voz do discurso alheio – cujas orientações são diametralmente opostas. As orientações são opostas, pois, para o discurso autoral, deve haver o rompimento com os ideais liberais de autonomia e independência impregnados no contexto social em geral. Há um conflito, portanto, entre a voz do discurso autoral e a voz do discurso alheio (o que diz o contexto social mais amplo). Além de o discurso autoral não concordar com o que diz o contexto social em geral (o discurso alheio), ele também cita que deve haver esse rompimento, o que acaba por “narrar” a atuação do *discurso bivocal de orientação vária*.

Convém sublinhar neste momento algumas observações no que diz respeito aos resultados a que chegamos tendo por base as análises empreendidas e os tipos e variantes de discurso que recorrentemente atuaram na escrita desses artigos.

O uso do *discurso direto substituído* pode ocorrer simultaneamente ao uso do *discurso direto preparado*. Isso porque, como apontam as análises, no momento em que o discurso autoral introduz/antecipa o discurso do outro, o autor também faz o movimento de falar pelo outro, falar no lugar desse outro com o qual dialoga e em nome desse outro. Desse modo, a voz concernente ao discurso alheio apenas é trazida à baila por meio da voz do discurso autoral.

O uso do *discurso bivocal de orientação única* é oportuno no momento em que as vozes discursivas – a voz autoral e a voz alheia – convergem para uma mesma e única orientação. Assim, esse tipo de discurso nos indica que não há conflito entre ambas as vozes em determinado momento de composição do artigo científico e que há uma relação de conformidade/concordância entre ideias. São esperadas e muito comuns, portanto, nessas situações, conjunções conformativas, a exemplo: “conforme”, “segundo”, “de acordo”, etc.

A atuação do *discurso objetificado* faz com que o discurso do outro atenda aos propósitos discursivos do discurso autoral na medida em que o próprio discurso alheio funciona como o objeto de discurso do autor. Assim, a unidade de enunciação do discurso do outro está a serviço da unidade de enunciação do discurso autoral, de que é objeto.

O uso do *discurso antecipado, disperso, oculto* acompanha a presença das paráfrases realizadas pelo discurso autoral. As paráfrases do pensamento de outrem são um recurso por meio do qual o discurso autoral indica como as considerações alheias influenciaram as suas próprias considerações e, assim, com base na voz do outro, o discurso autoral apresenta um conteúdo (temático) sobre o qual possa discorrer. A atuação desse tipo de discurso, portanto, demonstra como o enunciado autoral recebeu influências do discurso alheio, pois o discurso do outro passa a fazer parte do discurso autoral.

O uso do *discurso direto em estilo monumental* acompanha a presença da citação direta. No momento em que o autor se vale dessa citação, o enunciado alheio, ao ser transposto para o contexto do discurso autoral, não sofre alterações de qualquer ordem, isto é, é transposto *ipsis litteris*. Assim, preserva-se, por meio do uso desse tipo de discurso, o enunciado alheio considerado como um todo compacto, indivisível, imutável e impenetrável. Desse modo, quando passam a integrar o contexto do discurso autoral, as palavras do outro não sofreram modificação. As palavras do outro, todavia, impactam o discurso autoral: o autor decidiu citá-las por alguma razão e isso já mostra a importância dessas palavras, traduzidas como posições axiológicas adotadas ou negadas pelo discurso autoral.

O uso do *discurso direto preparado* foi um recurso bastante explorado pelo discurso autoral na formulação de seus textos. Com base nesse tipo de discurso, o autor consegue antecipar o discurso alheio, de modo a fazer com que ele seja introduzido no contexto do discurso autoral. O uso de citações diretas ou indiretas – sobretudo as diretas – mostra de maneira evidente a atuação desse tipo de discurso, pois, por meio delas, o discurso autoral introduz a fala alheia em momentos nos quais o discurso alheio contribui para a discussão promovida no contexto do discurso autoral.

O uso do *diálogo velado* é uma estratégia discursiva por meio da qual o autor considera a presença de um outro (interlocutor) e, dessa forma, antecipa possíveis indagações que esse outro possa fazer em relação ao que foi dito pelo autor. No caso dos artigos científicos, esse outro muitas vezes foi interpretado como o(s) parecerista(s) do artigo em questão e como os eventuais leitores interessados na discussão promovida por essas pesquisas realizadas. Assim, não há visivelmente a presença desse parecerista ou desses leitores, por exemplo, mas o autor “sente”, por assim dizer, essa presença do outro, ainda que de forma velada. Com essa presença – não visível, mas sentida –, o autor dialoga consigo mesmo tendo em vista as réplicas de (e a) um interlocutor ausente, que pode ser o parecerista do artigo em questão ou os leitores em potencial para a leitura do artigo científico produzido.

Por último e não menos importante, o uso do *discurso bivocal de orientação vária* indica que há um conflito entre a voz do discurso autoral e a voz do discurso alheio, pois ambas as vozes não convergem para uma mesma e única orientação; ao contrário: dispõem de uma orientação diametralmente oposta. Como tal, o uso de conjunções adversativas (“no entanto”, “contudo”, “todavia”, “entretanto”, “porém”, por exemplo) são bastante presentes nessas situações de discordância entre o discurso autoral e o discurso alheio.

A mobilização desses tipos e variantes de discurso pelos autores dos artigos científicos configura-se, a nosso ver, não apenas como estratégia argumentativa, mas como gesto enunciativo atravessado por posições axiológicas que organizam a relação com a palavra do outro. Essas escolhas não se limitam a procedimentos formais de citação; elas dizem respeito ao modo como o discurso alheio é incorporado à cadeia discursiva, reacentuado valorativamente e orientado responsivamente no interior do ato de enunciação. Ao convocar vozes precedentes, o escrevente não as reproduz de maneira neutra, mas as enquadra segundo seu projeto discursivo, atribuindo-lhes determinada entonação axiológica e inserindo-as em uma rede de relações dialógicas que as tensiona, confirma, reflete, refrata ou subordina. Nesse sentido, as formas de citação analisadas constituem modos específicos de realização do dialogismo na esfera científica, evidenciando que o relacionamento com a palavra do outro envolve sempre uma tomada de posição valorativa e uma orientação para o horizonte de compreensão responsiva da comunidade de pares.

6 Para concluir

O presente artigo teve por objetivo analisar modos de manifestação das relações dialógicas na constituição da escrita de *artigos científicos* da área das Ciências Humanas, tendo por base os tipos e variantes de discurso discutidos por Volóchinov e Bakhtin, respectivamente, nas obras *Marxismo e filosofia da linguagem* (2021) e *Problemas da Poética de Dostoiévski* (2018). Para isso, valemo-nos de uma metodologia de estudo da língua exposta por Volóchinov em *Marxismo e filosofia da linguagem* (2021) associada ao paradigma indiciário, como proposto por Ginzburg (1991).

A partir de nossa exposição, acreditamos ter respondido à pergunta de pesquisa principal que orientou a construção deste artigo: *como as relações dialógicas constitutivas da linguagem se expressam em termos dos modos de citação empregados em artigos científicos da área das Humanidades?* Esperamos, assim, ter exemplificado ao longo do presente trabalho que as relações dialógicas, sob o olhar bakhtiniano e do Círculo, sendo fenômeno constitutivo da linguagem, são também constitutivas da produção escrita do gênero artigo científico, conforme se observa nos excertos que foram analisados.

Dentre os tipos e variantes de discurso que se apresentaram na constituição da escrita dos exemplos analisados, podemos mencionar algumas considerações a respeito da atuação de cada um deles na construção dos artigos analisados: (a) o *discurso direto substituído* está presente em momentos nos quais o autor fala no lugar e em nome do pesquisador com o qual dialoga; (b) o *discurso bivocal de orientação única* é visível em momentos em que o discurso autoral e o discurso alheio citado estão em concordância; (c) o *discurso objetificado* é visível em momentos nos quais o discurso do outro está a serviço do projeto discursivo do autor; (d) o *discurso bivocal de orientação vária* faz-se presente em momentos em que o discurso autoral e o discurso alheio citado estão em discordância; (e) o *discurso antecipado, disperso, oculto* é notado no momento em que o discurso autoral deixou-se influenciar pelas palavras alheias; (f) o *discurso direto em estilo monumental* é visível quando o discurso autoral se vale de uma citação direta; (g) o *discurso direto preparado* pode ser notado quando o autor introduz o discurso do outro no contexto do discurso autoral e (h) o *diálogo velado* é influente em momentos nos quais o escrevente considera as réplicas de um interlocutor ausente.

Destacamos, assim, como as relações dialógicas são constitutivas do gênero do discurso *artigo científico*, sendo que, para o escopo do presente texto, optamos por analisar apenas artigos da área das Ciências Humanas. Da análise empreendida, constatamos que as relações dialógicas revelam modos peculiares de os autores dos artigos científicos se relacionarem com as palavras do outro em termos de estratégias argumentativas, discursivas e axiológicas adotadas (ou rejeitadas) em seu trajeto pela escrita do gênero.

Em vista do princípio bakhtiniano de que as esferas de comunicação humana condicionam a constituição dos enunciados e dos gêneros discursivos (Bakhtin, 2016), este estudo não pretendeu demonstrar esse fundamento teórico, mas observá-lo em funcionamento no interior de um *corpus* delimitado de artigos científicos da área das Ciências Humanas. As diferentes formas de citação analisadas vislumbram, assim, modos específicos de incorporação do discurso alheio que se articulam às finalidades argumentativas e às exigências axiológicas dessa esfera acadêmica. O que buscamos, portanto, não é generalizar resultados para a totalidade da área em voga, mas explicitar como, neste recorte empírico, o princípio dialogal preconizado pelo Círculo de Bakhtin se materializa em práticas concretas de escrita acadêmico-científica.

Entendemos, assim, que a discussão proposta aponta para a necessidade de se olhar para os modos de citação e de interpretação dos diferentes tipos de discurso (direto, indireto e indireto livre) para além de meras questões textuais ou de normalização da escrita acadêmica/científica. Ao trazer a voz do outro para o discurso autoral e ao fazê-lo de determinado modo, como atestam as análises empreendidas, o escrevente não se limita a “citar” palavras alheias; antes, estabelece com elas uma relação constitutivamente dialógica com o outro, articulando vozes que se entrecruzam e se refratam na tessitura de seu enunciado.

Agradecimentos

Agradecemos o apoio recebido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para a realização da investigação mais ampla, da qual provém parte dos resultados constantes neste trabalho.

Referências

AMERICAN PSYCHOLOGICAL ASSOCIATION. *Publication manual of the American psychological association*: The official guide to APA style (7th ed.). APA, 2020.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT 6022: *informação e documentação – artigo em publicação periódica científica impressa – apresentação*. Rio de Janeiro: ABNT, 2003.

BAKHTIN, M. M. *Os gêneros do discurso*. Organização, tradução, posfácio e notas de Paulo Bezerra; notas da edição russa de Serguei Botcharov. 1 ed. São Paulo: Editora 34, 2016.

BAKHTIN, M. M. *Problemas da poética de Dostoiévski*. Tradução direta do russo, notas e prefácio de Paulo Bezerra. 5 ed. Rio de Janeiro: Forense Universitária, 2018.

CADERNOS NIETZSCHE. v. 44. n. 3. 2023. Disponível em: <https://periodicos.unifesp.br/index.php/cniet> Acesso em: 12 jun. 2023.

CADERNOS PAGU. n. 68. 2023. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cadpagu> . Acesso em: 02 jul. 2023.

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Competências*. Disponível em: <http://www.capes.gov.br/acessoainformacao/80-conteudo-estatico/acesso-a-informacao/5418-competencias> . Acesso em: 21 nov. 2023.

GINZBURG, C. Sinais: raízes de um paradigma indiciário. In: GINZBURG, C. *Mitos, Emblemas Sinais: Morfologia e História*. Tradução Federico Carotti. São Paulo: Companhia das Letras, 1991. p. 143-179.

MACIEL, L. V. C. *Relações dialógicas em narrativas*. 2014. 360p. Tese (Doutorado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2014.

MIRANDA, F. D. S. *Letramentos (en)formados por relações dialógicas na universidade: (res)significações e refrações com tecnologias digitais*. 2016. 414p. Tese (Doutorado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2016.

MOTTA-ROTH, D.; HENDGES, G. H. *Produção textual na Universidade*. São Paulo: Parábola Editorial, 2010.

SILVA, F. J. *Relações dialógicas na constituição da escrita de artigos científicos*. 2024. 248p. Dissertação (Mestrado em Linguística) — Programa de Pós-Graduação em Linguística, Universidade Federal de São Carlos, São Carlos, 2024.

SUCUPIRA, *Plataforma*. Disponível em: <https://sucupira.capes.gov.br/sucupira/> Acesso em: 10 mai. 2023.

VOLÓCHINOV, V. *Marxismo e filosofia da linguagem*: problemas fundamentais do método sociológico na ciência da linguagem. Tradução, notas e glossário de Sheila Grillo e Ekaterina Vólkova Américo. 3. ed. São Paulo: Editora 34, 2021.

“Strikingly”, an Academic Writer’s Style Can Be Interactional: Towards Characterizing Metadiscourse Style in Applied Linguistics Research Articles

“Surpreendentemente”, o estilo de um autor acadêmico pode ser interacional: rumo à caracterização do estilo metadiscursivo em artigos de pesquisa da linguística aplicada

Janailton Mick Vitor da Silva
Universidade Federal da Paraíba (UFPB)
João Pessoa | PB | BR
janailtonm@gmail.com
<https://orcid.org/0000-0002-5137-5473>

Abstract: This paper aims at unveiling the interactional metadiscourse style of a prominent applied linguist, Douglas Biber. Comparisons are made between 25 of his research articles (RAs) to 126 RAs written by other applied linguists. In this work, a Corpus Linguistics (CL) methodological framework is proposed by using AntConc tools. The results show that the author’s interactional stylistic choices convey levels of certainty and uncertainty about his propositions; make his views explicit, comment on the message, and express evaluations (importance and suitability) and emotions (surprise); and directly involve readers in the ongoing discourse by overtly pointing them to other parts in the RAs. Strikingly, Biber’s interactional rhetorical practices are essentially evaluative and engaging, expressing solidarity, anticipating objections, and revealing the extent to which he works to jointly construct texts with his readers. As a result, the author’s tendency to pursue an interactional metadiscourse style reveals how an expert academic writer crafts an authorial style that may empower novice and other expert writers to craft a style that involves the audience in meaningful reading and writing experiences through academic texts.

Keywords: Style; metadiscourse; interaction; corpus linguistics; linguistic features.



Resumo: Este artigo objetiva desvendar o estilo meta-discursivo interativo de um importante linguista aplicado, Douglas Biber. São feitas comparações entre 25 de seus artigos de pesquisa e 126 artigos escritos por outros linguistas aplicados. Neste trabalho, uma proposta metodológica baseada na Linguística de Corpus (LC) é seguida usando ferramentas do AntConc. Os resultados desta pesquisa mostram que as escolhas estilísticas interacionais do autor transmitem níveis de certeza e incerteza sobre suas proposições; tornam suas visões explícitas, comentam a mensagem e expressam avaliações (importância e adequação) e emoções (surpresa); e envolvem diretamente os/as leitores/as no discurso em andamento, direcionando-os explicitamente para outras partes dos artigos. Surpreendentemente, as práticas retóricas interacionais de Biber são essencialmente avaliativas e envolventes, expressando solidariedade, antecipando objeções e revelando até que ponto o autor trabalha para construir conjuntamente textos com seus/suas leitores/as. Como consequência, a tendência do pesquisador em buscar um estilo metadiscursivo interacional revela como um escritor acadêmico experiente cria um estilo autoral que pode empoderar novos e experientes autores/as a criar um estilo que envolva o público em experiências significativas de leitura e escrita por meio de textos acadêmicos.

Palavras-chave: Estilo; metadiscorso; interação; linguística de *corpus*; recursos linguísticos.

1 Introduction

In our everyday life, we frequently associate *style* with someone's preference for particular songs, dishes, clothes, etc., but we may not promptly associate it with someone's choice of particular words or phrases in a given language. For example, the way an individual uses an adjective as predicative (*That house is beautiful*) and not attributive (*That is a beautiful house*) may characterize a personal linguistic preference. This can also be seen when an academic writer shows predilection for *moreover* over *furthermore*, or when the prepositional phrase *for instance* is preferred over *for example*. In sum, these and other frequent patterns of language use may subsidize the identification of an individual's language style. As proposed in this paper, a writer leaves their fingerprint in texts, revealing a style that may be promptly associated with them.

Implicit in the word *way* mentioned in the previous paragraph lies the notion of *style*. In this research, *style* is analyzed within the academic writing realm (Bacon, 2013), owing to the fact that most style studies have mainly focused on literary texts and literary authors (Jeffries; McIntyre, 2010). *Style* is hereby understood as a writer's fingerprint or personal form of expression (Baker, 2000; Garcia, 2006), marked by a preference for some linguistic items and structures across different texts and compared to other writers' language choices (Malmkjaer; Carter, 2002). *Style* is also defined as an individual's personality in linguistic terms and an individual's voice in the collectiveness, whereby authorial identity(ies) may be similarly revealed (Hyland, 2002a; 2002b; 2008b; 2010; Ivanič, 1998; Tardy, 2012).

The relevance of carrying out this research can be verified from different perspectives. Drawing on Leech and Short (2007), style reveals how language is used in a given context, by a given person, for a given purpose. Since no one writes in a vacuum, detached from a socio-cultural context and without pre-determined purposes and audiences, academic language is hereby examined insofar as it can create interpersonal bonds with readers, precisely through interactional linguistic choices. According to Hyland (2005), writers can be interactional when they comment on their message, make their views, evaluations, and opinions explicit to readers, and jointly construct the texts with them.

These interpersonal bonds materialized in texts have been investigated from different angles in literature (Hyland; Tse, 2004). For the sake of specificity, however, metadiscourse has been selected for this investigation, because it can refer both to scholar's patterns of discourse organization and writer's ways of expressing stance and engagement (Hyland, 2004). In this regard, metadiscourse is understood as part of author's style in academic writing, since scholars may present other types of styles, even potentially non-metadiscoursal, when writing in and/or out of academia. As can be seen, this definition adds to long-standing discussions which have surrounded the concept of style: "the elusiveness of style lies in the age-old dispute as to whether style is extrinsic or intrinsic in nature; whether it is the icing on the cake – in effect, an optional extra – or whether it is an inherent part of the cake itself." (Burke, 2014, p. 24).

Furthermore, additional reasons for carrying out this investigation stem from Crystal's (2013) attempt to define authors' stylistic profile. Once one's style is defined, pedagogical materials highlighting important stylistic features to be used in training courses, for instance, might be developed. Moreover, by doing this research, Bacon's (2013) proposal to challenge a classical dichotomy in the field of style in academic writing, particularly in the context of teaching style to college students, is supported. For the author, this dichotomy establishes that a scholar should either write with style or write for the academy, as if writers could not write stylistically within academia. Finally, the myth of a dry, impersonal discourse in academic writing is similarly refuted in this research (Hyland, 2002a).

Considering the reasons stated in this introductory section, it is important to understand how renowned applied linguists craft their interactional metadiscourse style and, by extension, express stance and establish engagement with readers. Stemming from Silva's (2024) PhD dissertation, the main goal of this paper is to unveil the *interactional* metadiscourse style of an applied linguist, Douglas Biber, based on frequent and consistent use of metadiscourse language features across research articles (RAs) and in comparison to other RAs. Biber has been chosen as protagonist in this style investigation because of his distinguishing contributions to corpus linguistics (CL) research in a variety of topics, especially in register studies. In this regard, this investigation is guided by the following research question: how does Biber express stance and establish engagement with readers in his RAs through interactional metadiscourse elements?

To answer this question and accomplish the main goal of this paper, a CL methodological framework has been proposed by using *AntConc*, version 4.2.0, designed by Lawrence Anthony (2022), specifically through *Key-Word-In-Context (KWIC)*, *Cluster*, *Collocate*, and *Keyword* tools, both for compilation and contextual analysis of *Biber Corpus (BC)* and *Corpus of Journal Articles in Applied Linguistics (CorJAL)*. These steps will be described and discussed in detail in the Methodology section. The linguistic data generated by the program has been submitted to corpus-based quantitative and qualitative linguistic analyses following the application of the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing*, which will be introduced in the upcoming sections.

2 Theoretical background

The elusiveness inherent to the meaning of style has often been attributed to its polysemic condition (Burke, 2014). Nevertheless, many scholars have contributed to defining it. For example, Malmkjaer and Carter (2002, p. 510) see style as “[...] a consistent occurrence in the [spoken and written] text of certain items and structures, or types of items and structures, among those offered by the language as a whole.” Furthermore, style is characterized as an individual’s or group’s “[...] personal form of expression [...]” (Garcia, 2006, p. 123)², or “[...] the definition of a personality in linguistic terms”³ (Câmara Junior, 1978, p. 13). In other words, writers select items and structures among many others similarly available in the language, resulting in the expression of their personality in linguistic terms.

Scholars’ self-expression may also be imprinted in written texts. As a matter of fact, Hyland (2005, p. 53) asserts that “all writing carries information about the writer [...]”, and that “writers cannot avoid projecting an impression of themselves and how they stand in relation to their arguments, their community and their readers.” From these quotes, it can be inferred that writing is a means by which writers can express authorial identities through different language resources, including personal pronouns, adjectives, lexical bundles, adverbs, among others (Hyland, 2002a; 2002b; 2008b; 2010).

Because all writing carries information about the writer, it can thus be acknowledged that “[...] all writing contains ‘voice’ [...]” (or style, as is proposed here) (Hyland, 2008b, p. 144). Although individuals have a voice of their own, everything they say or write is constrained by a social context and its elements, such as genre, register, audience, etc. (Tardy, 2012). As a result, a writer’s voice stems from a *mingling of voices* and discourses available in society, which resonates with Bakhtin’s (1981) *heteroglossia* notion: “language is not a neutral medium that passes freely and easily into the private property of the speaker’s intentions; it is populated – overpopulated – with the intentions of others” (Bakhtin, 1981, p. 294). In spite of this multi-voicedness feature of language, Tardy (2012) asserts that the way a particular writer mingles other individuals’ voices and discourses in a text is what characterizes their individual voice.

Style has traditionally been studied within the area of *stylistics*, which can be broadly “[...] defined as the (linguistic) study of style” (Leech; Short, 2007, p. 28), or the “[...] scientific

¹ “[...] forma pessoal de expressão [...]” (Garcia, 2006, p. 123).

² All translations provided here have been made by the author of this paper, unless otherwise stated.

³ “[...] a definição de uma personalidade em termos lingüísticos” (Câmara Junior, 1978, p. 13).

study of patterns of variation in written language. The object of study is the language of a single individual (idiolect), resulting in a description of his or her identifying linguistic characteristics.” (McMenamin, 2002, p. 176). Particularly in this paper, the focus is on the application of CL tools to help compile and analyze large portions of texts, including patterns of linguistic phenomena, which had earlier (in traditional stylistics) been unfeasible to undertake given the techniques available (McIntyre; Walker, 2019).

Taking these concepts presented so far into consideration, for the purposes of this investigation, style is defined within the context of academic writing as “[...] the particular way a scholar varies the use of language in academic genres and registers for specific purposes, resulting in the expression of voice(s), authorial identity(ies), stance(s), and potential effects in texts and audiences.” (Silva, 2024, p. 23). This definition of style encompasses some underlying elements which will be discussed next.

Firstly, when style is defined as the particular way an individual makes variation in the use of language in academia, it can be argued that scholars (synonymously regarded as authors or writers in this paper), show consistent preference for particular language items and/or structures over others, representing distinctive individual variation in the register used within the academic context in academic genres.

Moreover, this individual variation may contribute to the expression of academic writers’ voice(s), authorial identity(ies), and stance(s) in academic registers and genres, as they inevitably leave marks and traces on what they write. This expression, nevertheless, does not represent sole and individual voice(s), identity(ies), and stance(s) per se, because, as socially engaged subjects, scholars take part in discourse practices enacted in and out of academia, which is a social milieu consisting of specific principles and traditions.

Furthermore, authors’ personal form of expression in academia is not arbitrary, especially because it needs to acquiesce to academic and linguistic norms and conventions that regulate academic discourse and represent linguistic disciplinary variation (Hyland, 2009). As a result, language used by academics is imprinted in a myriad of text varieties, such as RAs, which reveals how style is also construed in and derived from the social discourse communities to which scholars, readers, and other agents belong in academia. For example, RAs undergo several publishing steps, and their content is affected by different literacy brokers, such as editors, translators, peer reviewers, academic peers, and friends (Lillis; Curry, 2010).

Finally, authors’ style is adopted with the intention of reaching particular purposes and causing potential effects in texts and audiences. Hence, in addition to expressing research-related facts, information, data, viewpoints, among other meanings, they may also end up creating interpersonal bonds with audiences by means of *interactive* and *interactional* meta-discourse features. For instance, previous research has shed light on the use of these features to shape authorial style in RAs (Silva, 2024, 2025).

In view of these concepts, a theoretical model can be applied to the analysis of academic writer’s texts and their language style. This model, entitled *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Silva, 2024, 2025), is a critical, blended up-to-date version of Hyland’s (2005) *interpersonal model of metadiscourse*, Hyland’s (2008a) *taxonomy for bundles functional analysis*, Hyland’s (2009) *key resources of academic interaction*, and Biber’s (2006) and Biber *et al.*’s (2021) *lexico-grammatical and semantic features used for stance analyses*. But prior to introducing the model, it is paramount to understand some key concepts: metadiscourse, stance, and formulaic language.

Similar to *style*, *metadiscourse* is also a fuzzy term. To start with, Crismore (1983, p. 2) defined *metadiscourse* as “the author’s intrusion into the discourse, either explicitly or non-explicitly, to direct rather than inform, showing readers how to understand what is said and meant in the primary discourse and how to ‘take’ the author”. Similarly, *metadiscourse* is deemed as some linguistic material which adds no propositional information to the text, but rather points to the author’s presence therein (Vande Kopple, 1985). In more recent studies, *metadiscourse* is characterized as aspects of a particular text, such as language features “[...] which explicitly refer to the organization of the discourse or the writer’s stance towards either its content or the reader” (Hyland, 2004, p. 109). However, Hyland also warns that metadiscourse has mistakenly been comprehended as either “discourse about discourse” or “talk about talk” (Hyland, 2005, p. 16), limiting its definition to textual over interpersonal aspects.

It is particularly on this interpersonal perspective that this investigation relies. Therefore, it is argued that the nature of metadiscourse is essentially dialogic, because “[...] all metadiscourse is interpersonal in that it takes account of the reader’s knowledge, textual experiences and processing needs and that it provides writers with an armoury of rhetorical appeals to achieve this” (Hyland; Tse, 2004, p. 161). It involves two main types of interaction, *interactive* and *interactional*, whereby Hyland (2005) pays homage to the works produced by Thompson and Thetela (1995) and Thompson (2001):

The interactive dimension. This concerns the writer’s awareness of a participating audience and the ways he or she seeks to accommodate its probable knowledge, interests, rhetorical expectations and processing abilities. The writer’s purpose here is to shape and constrain a text to meet the needs of particular readers, setting out arguments so that they [readers] will recover the writer’s preferred interpretations and goals. The use of resources in this category therefore addresses ways of organizing discourse, rather than experience, and reveals the extent to which the text is constructed with the readers’ needs in mind.

The interactional dimension. This concerns the ways writers conduct interaction by intruding and commenting on their message. The writer’s goal here is to make his or her views explicit and to involve readers by allowing them to respond to the unfolding text. This is the writer’s expression of a textual ‘voice’, or community-recognized personality, and includes the ways he or she conveys judgements and overtly aligns him- or herself with readers. Metadiscourse here is essentially evaluative and engaging, expressing solidarity, anticipating objections and responding to an imagined dialogue with others. It reveals the extent to which the writer works to jointly construct the text with readers. (Hyland, 2005, p. 50-51, emphasis in original).

In other words, the *interactive* dimension regards the extent to which writers have empathy with their readers by organizing the discourse so that they can fulfill the demands of the audience and make sure readers retrieve their objectives. The *interactional* dimension aims at writers’ direct, explicit intrusions along the discourse, making their epistemic and attitudinal stances clear to the audience. This is a more dialogic connection between reader and writer, engaging the audience in the discussion and inviting it to co-construct the texts.

In Hyland’s (2005) work, both of these dimensions present particular corresponding functional categories (*interactive* = *transitions, frame markers, endophoric markers, evidentials, code glosses*; *interactional* = *hedges, boosters, attitude markers, self-mentions, engagement markers*). In a more recent book, Hyland (2009) revisits his earlier model and splits the interactional

dimension into two, by making the differences between stance from engagement more explicit. Stance now encompasses all the four previous elements (*hedges, boosters markers, attitude mention, and self*) from the interactional dimension, whereas engagement consists of the following subcategories: *reader, directives, questions reference, shared knowledge, and asides*.

In addition to *metadiscourse*, another term has been used to refer to writer’s presence in texts: *stance*. In literature, *stance* figures as one among many other language features used to express writers’ personal feelings, attitudes, assessments, and value judgements, having been labeled differently over time: “‘evaluation’; ‘intensity’; ‘affect’; ‘evidentiality’; ‘hedging’; ‘modality’ and ‘stance’” (Biber, 2006, p. 87). Despite this varied terminology, the expression of stance is important in academia: “in claiming a right to be heard, and to have their work taken seriously, writers must simultaneously take a stance towards what they discuss and display competence as disciplinary insiders.” (Hyland; Jiang, 2019, p. 131).

Furthermore, another key term in this paper is *formulaic language*, which includes “[...] chunks, collocations, fixed expressions, formulaic language, idioms, lexical bundles, multiword units, prefabricated units, set phrases” (Frankenberg-Garcia, 2018, p. 93). Nevertheless, although these terms are interconnected, the term *lexical bundles* is hereby adopted as part of the large superordinate *formulaic language*, because bundles are considered “[...] sequences of word forms that commonly go together in natural discourse” (Biber *et al.*, 2021, p. 982).

Considering the theoretical notions laid out in this section, Table 01 summarizes the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Silva, 2024, 2025), which was elaborated based on a critical analysis of Hyland (2005; 2008a; 2009), Biber (2006), and Biber *et al.* (2021):

Table 01–Stylistic metadiscourse model of interpersonal interaction in academic writing (Silva, 2024, 2025)

General features of interaction	Categories		Subcategories
Interactive (helps guide readers through the text)	Discourse organizers (research-oriented and text-oriented)		Transitions
			Frame markers
			Evidentials
			Code glosses
Interactional (involves readers in the text)	Participant-oriented systems of interaction	Stance markers	Hedges (epistemic)
			Boosters (epistemic)
			Attitude markers (attitudinal)
			Self-mentions (epistemic)
	Engagement markers		Reader pronouns
			Personal asides
			Endophoric markers and directives
			Questions
			Appeals to shared knowledge

Source: elaborated by the author.

Based on the model introduced above, in the interactive dimension, *transitions* express additive, causative, and contrastive relations between main clauses (*moreover; furthermore; and*); *frame markers* refer to research/discourse acts, sequences, stages, or topic shifts (*first; then; a wide range of*); *evidentials* point to information from other texts, reporting on but not taking a stance towards other scholars' research and findings (*argue; assert; according to; claim*); and *code glosses* supply additional information by rephrasing, explaining, or elaborating on what has been stated (*for instance; for example*) (Hyland, 2005; 2008a; 2009; Biber, 2006a; Biber *et al.*, 2021).

In addition to these interactive elements, the model also presents interactional (sub) categories. To begin with, in the stance category, *hedges* withhold writer's commitment to a proposition, allowing information to be presented as an opinion rather than a fact and opening that position to negotiation (*believe; doubt; may; might*); *boosters* emphasize writer's certainty and confident voice, strengthening an argument but closing down possible alternatives to a proposition (*determine; discover; find; must; should*); *attitude markers* express writer's affective attitudes, emotions, evaluations to/on a proposition, conveying surprise, agreement, importance, obligation, frustration, etc. (*amazed; amazing; astonishingly*); and *self-mentions* make explicit reference to the writer (*me; we; I*). As for engagement markers, *reader pronouns* overtly build relationship with reader (*your; you; inclusive we*); *personal asides* interrupt the argument to offer a comment on what has been stated (through appositive phrases or clauses in between dashes or parentheses); *endophoric and directive markers* direct readers to another part of the text or to another text, inform them how to carry out some action in the real world, and instruct them how to interpret an argument (*as shown in figure; as noted above; as can be seen*); *questions* invite engagement, encourage curiosity, and bring interlocutors into the discourse arena (through interrogative and/or rhetorical questions); and *appeals to shared knowledge* ask readers to recognize something as familiar or accepted, identify themselves with particular views, and assign to them a role in creating the argument (*of course; by the way; obviously*) (Hyland, 2005; 2008a; 2009; Biber, 2006a; Biber *et al.*, 2021).

It should be noted that this model is quite comprehensive. However, considering the goal and research question laid out for this paper, only the features found in the research corpus and classified as *interactional* will be investigated, including *boosters, hedges, attitude markers, endophoric and directive markers*. These features, however, will only be introduced and analyzed in section 4. Before that, methodological procedures adopted for the current investigation are introduced next.

3 Methodology

In this paper, CL has been used as an approach (Tognini-Bonelli, 2001) to carry out stylistic analyses by comparing study and reference corpora. The methodological steps followed to carry out this investigation are presented and discussed in this section.

3.1 The study corpus

The first corpus compiled in this research was *BC*, the study corpus. *BC* consists of 25 single-authored RAs, mostly quantitative empirical (Gray, 2015), written by Douglas Biber and published in a myriad of peer-reviewed journals. These RAs span a time period of 35 years and were published from 1984 to 2019. Table 02 displays an overview of the texts, their time range, and the number of types and tokens obtained from *AntConc* (Anthony, 2022):

Table 02 –Overview of the study corpus

Biber Corpus - BC	
Sub-register/Sub-corpus	Journals RAs (JA)
Time range	1984-2019
Number of texts	25
Types	5,263
Tokens	144,719

Source: elaborated by the author

A note should be made regarding the time range informed on the table, for Biber published more than 25 texts throughout his career, including books, book chapters, reviews, monographs, and proceeding papers, especially in collaboration with other scholars, as listed on his personal website CV⁴. Nevertheless, in view of the style focus of this investigation, only 25 single-authored journal RAs were selected for the study corpus compilation. Because this collaborative practice is part of academic practice, Biber (2013, p. 375) has stated the following on his collaborative work and authorship attribution:

I like collaborating with a coauthor and working as part of a collaborative team. [...] But I somehow have ended up doing as much collaborative research as individual research. [...] The first author listed on my books and articles is the person who takes the lead in the analysis, and who actually does most of the writing. But all authors are usually fully involved in the analysis process and help with commenting and revising the paper.

Four are the major reasons for compiling only journal RAs in the present research. First, Biber’s greatest number of single-authored texts are, in fact, RAs. Hence, similar to Hyland’s (2008b; 2010) and McIntyre and Walker’s (2019) sample corpora, *BC* was considered an appropriate sample to investigate Biber’s style in comparison to his contemporaries’. Second, including only journal RAs in the current sample and excluding other academic genres (i.e. books or book chapters) would not pose a problem, especially because such genres share particular features in comparison to other genres (Biber; Gray, 2016) and are often produced in similar situations (Biber; Conrad, 2019). Third, a great amount of metadiscourse research focuses on RAs (Pearson; Abdollahzadeh, 2023), so it seemed feasible to carry on this research trend.

⁴ BIBER, D. Home. Arizona: Northern Arizona University, [201-?]. Available at: <https://doughbiber.weebly.com/>. Accessed on: 11 Nov. 2025.

Last, it is much easier to track Biber’s prominence as a scholar (and potentially as an academic writer) on the basis of journal RAs citation metrics.

3.2 The reference corpus

A reference corpus was compiled, *CorJAAL*, consisting of texts published by three peer-reviewed journals: *Applied Linguistics*, *System*, and *TESOL Quarterly*. In style research, it is important to have a comparison parameter to illustrate what is peculiar to a writer’s language repertoire and is probably not so to other academic writers’. For that, a methodological procedure suggested by Baker (2000) concerns the comparison of more than one author’s text to another writer’s set of works. A reference corpus works as a norm to which the study corpus will be compared, but as a relative norm, as “[...] an absolute norm for English cannot be relied on [...]” (Leech; Short, 2007, p. 41). Therefore, since Biber writes RAs in this area, the purpose is to investigate how he varies language within this expected relative norm as far as the use of interactional metadiscourse elements is concerned.

Despite the importance of having a reference corpus for style investigation, “the issue of reference corpus selection is far from decided” (Scott; Tribble, 2006, p. 64). Many scholars have struggled to establish the best layout for a reference corpus, without reaching much consensus. However, some researchers agree that size, genre, register, content, and research goals are some of the issues that should be borne in mind. In this regard, the reference corpus of this investigation considers various elements, including size (five times larger than the study corpus), register (academic writing), sub-register and genre (empirical and theoretical RAs), authors, journals, themes, topics, and fields within applied linguistics.

Regarding size, *CorJAAL* is five times larger than *BC*, following Berber Sardinha’s (2000) study. Table 03 presents an overview of the reference corpus and its time range, as well as the number of types and tokens obtained from *AntConc*:

Table 03 –Overview of the reference corpus

Corpus of Journal Articles in Applied Linguistics - CorJAAL	
Sub-register/Sub-corpus	Journal RAs (JA)
Time range	1984-2019
Number of texts	126
Types	21,311
Tokens	819,612

Source: elaborated by the author

Concerning genre and register, selecting only RAs published in journals represents a way to control variables that could otherwise impact the results. In addition, as to Geluso and Hirsch (2019), when a reference corpus contains a similar register⁵ to the study corpus,

⁵ The authors are referring to the following academic sub-register: academic writing in applied linguistics. In their research, they made different comparisons. To reach the conclusion that keywords characterize the study corpus and not the register, Geluso and Hirsch (2019) compared their study corpora (made of papers published by faculty members from two universities in the United States, including book chapters and journal publications) to a reference corpus of the same sub-register: research articles in the applied linguistics discipline.

the results from a keyword search are very particular to the study corpus and not to the register. Put differently, the words identified characterize the study corpus (and potentially the author's style, as will be shown) and not the register. However, this is not to say that register may not influence someone's individual style, since research like Grieve's (2023) has shown evidence on this issue.

Moreover, *CorJAAL* contains empirical (qualitative and quantitative) and theoretical RAs (Gray, 2015). As to empirical RAs, there is a balanced number of both qualitative and quantitative texts published in *Applied Linguistics*, but more qualitative RAs in *System* and *TESOL Quarterly*. The journals and RAs which compose *CorJAAL* were selected based on the following criteria:

- ◆ Criteria for selecting *CorJAAL* journals:
 - ◆ Be among top 20 with the highest impact factor (IF)⁶ throughout 2022 and across a 5-year period as published in Clarivate's (2023) Journal Citation Reports;
 - ◆ Publish RAs on a variety of topics, themes, and fields within applied linguistics.
- ◆ Criteria for selecting RAs from these journals:
 - ◆ Have been published over a 35-year period, the same time span of the RAs in *BC* (from 1984 to 2019);
 - ◆ Show the highest number of citations in the volumes and issues published in the same time span as the *BC* articles.

Furthermore, a plethora of authors and their texts compose *CorJAAL*, which serve a consistent diversity of authorship against which Biber's texts have been compared.

3.3 The compilation of the study and reference corpora

Regarding corpus compilation, all study and reference corpora texts have been downloaded from the web. The *.pdf* files were converted into *.txt* documents (UTF-8) using free online pro-

⁶ IF has figured as the dominant bibliometric for evaluating academic output for decades, and this was the major reason why it was chosen to select journals for the ongoing research. However, it raises many problems, especially when it is wrongly used by institutions to assess articles' quality and prevent scholars from getting jobs, tenure, or grants. In general, a common critique of IF is that "[...] citations are a shallow measure of research quality or impact" (Lillis; Curry, 2010, p. 9). In addition, Lillis and Curry (2010) notice that journals with higher IF receive more submissions; authors make immoderate self-citations and deliberate citations of other scholars publishing in the same journal; journals may pressure academics to cite papers from the same journal and limit scholars' citations to rival journals; university library subscriptions favor journals with higher IF. Furthermore, Waltman and Traag (2020) uphold that, under certain conditions, the number of citations an article receives is a more accurate indicator of the value of the article than IF. Hyland (2022, p. 18) also notes that "[...] a few highly cited papers can dramatically increase the impact of average or less significant papers."

grams, such as *I love Pdf*⁷, *Ocr 2 Edit*⁸, and *Online 2 PDF*⁹. By using basic automated *Microsoft Word* tools, the texts were edited to correct typos and other conversion-resulting errors, such as *distribu-tion* → *distribution*, so as not to compromise files processing on *AntConc*. Later, pieces of text from the original papers were withheld in brackets (< >), including the following cases: in-text short and long citations (only in the case of BC); titles of paper sections and sub-sections (e.g., <Conclusion>); information related to Biber and other writers (e.g., <Douglas Biber, Northern Arizona University, Flagstaff>). Because only the IMRD sections from BC and CorJAAL were used for the corpora compilation, such post-textual elements as authors' bios, notes, and acknowledgements sections, as well as images, tables, graphs, and lists of bibliographical works cited in the references section were deleted. For all these cases, the tag <DELETED EXCERPT> was inserted in place. Moreover, text-related identification tags were added at the beginning of each file, as displayed in Table 04:

Table 04– Example of text identification tags used in the corpora files

BC	CorJAAL
<text id = BC.JA.1993.04> <title = > <author = >	<text id = CorJAAL.AL.1993.04> <title = > <author = >
BC = Biber Corpus	CorJAAL = Corpus of Journal Articles in Applied Linguistics
JA = Journal RAs (subcorpus)	SYS = abbreviation for the journal System (AL = Applied linguistics; TQ = TESOL Quarterly)
Year of publication = 1993	Year of publication = 1993
Order of texts per year ¹⁰ = 04	Order of texts per year = 04
Title = the title of each research article (RA) is written	Title = the title of each RA is written
Author = the name of each author is given	Author = the name of each author is given

Source: elaborated by the author

At last, all the edited *.docx* files were saved into *.txt* (UTF-8) format. All these procedures were similarly followed in the reference corpus compilation.

3.4 Selecting keywords for analysis

The main *AntConc* tool used to identify interactional stylistic words was the *Keyword tool*. According to Scott (2022), the definition of keyness in CL is given below:

A word is said to be 'key' if
a) it occurs in the text at least as many times as the user has specified as a minimum frequency

⁷ I LOVE PDF. [S. l.: s. n], 2025. Available at: <https://www.ilovepdf.com/> Accessed on: 11 Nov. 2025.
⁸ OCR2EDIT. *OCR2Edit - Extract text from scans, images and more*. [S. l.: s. n], 2025. Available at: <https://www.ocr2edit.com/> Accessed on: 11 Nov. 2025.
⁹ ONLINE 2 PDF. *Conversor PDF online*. [S. l.: s. n], 2007. Available at: <https://online2pdf.com/> Accessed on: 11 Nov. 2025.
¹⁰ It was necessary to add this information in the tag as more than one paper was published in the same year.

- b) its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p value specified by the user
- c) in addition, the strength of keyness must be at least as great as the minimum log ratio set by the user. (Scott, 2022).

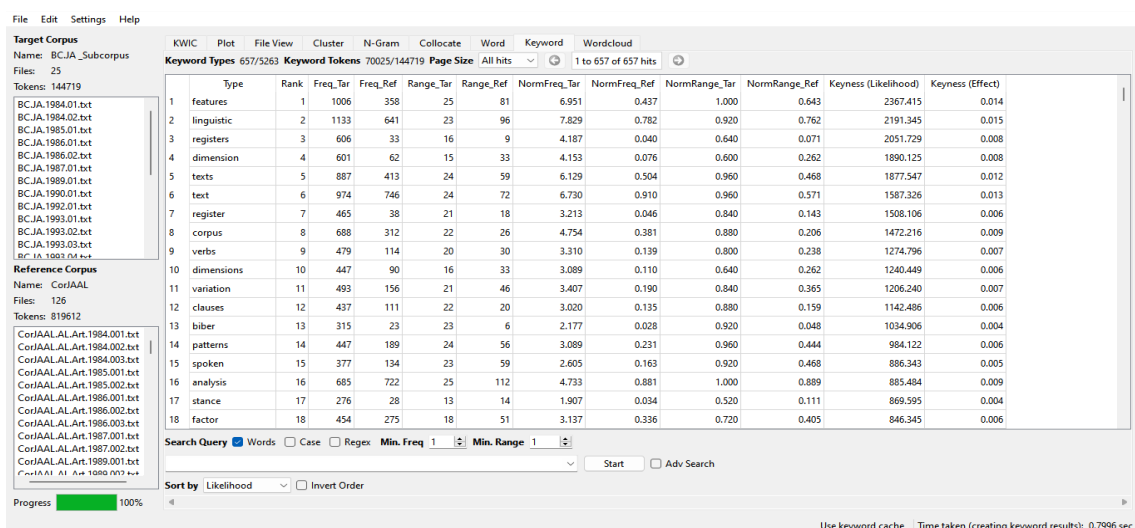
To obtain such keywords, the program compares two wordlists automatically generated by *AntConc Word* tool (in this investigation: each wordlist from a respective subcorpus of BC, which is the study corpus – or target corpus, as named by Anthony – and the second one from *CorJAAL*, or reference corpus). The results displayed by *Keyword Tool* are classified based on their statistical significance and labeled as either positive or negative:

a word which is *positively* key occurs *more often* than would be expected by chance in comparison with the reference corpus.

A word which is *negatively* key occurs *less often* than would be expected by chance in comparison with the reference corpus. (Scott, 2022, emphasis added).

For the purposes of the keyword analysis, BC was compared to *CorJAAL* following the default Keyness Values available in the program. In terms of data normalization, the value 1,000 words was set in the *Statistics global settings*, although this option may be altered at any time to suit the user's preference. This calculus is exemplified in Biber (2006a): the number of types is divided by the number of tokens and the result is multiplied by one thousand (1000/normed rate). For instance, take the case of the keyword *registers*: it occurs 606 times across BC (which has 144,719 tokens): $(606 / 144,719) * 1000 = 4.187$. Thus, the normed rate of the keyword *registers* in BC is 4.187 per one thousand words. After setting these values, both BC and *CorJAAL* were loaded in the program by using the *Corpus Manager* function. BC was used as the target corpus and *CorJAAL* as the reference. Upon selecting both corpora, the button *Start* in the *Keyword tool* was pressed, and a list of 657 keywords was shown on screen, as seen in Figure 01:

Figure 01– Screenshot of *AntConc* keyword tool



Source: elaborated by the author

Traditionally, researchers have limited themselves to analyzing only “top 10, 50 or 100 keywords” (Brezina, 2018, p. 85). Although dealing with long lists of keywords is time-consuming and may lead to errors, selecting some keywords over others may overlook important features of the corpus. In this regard, “a crucial next step when dealing with lists of key items [...] is working out which ones, if any, help most to answer your research questions.” (McIntyre; Walker, 2019, p. 283). Therefore, the keywords were grouped into three major categories: text-aboutness (keywords that regarded the topic/theme of the RAs); non-text-aboutness metadiscourse (interactional and interactive metadiscourse keywords); non-metadiscourse and non-text-aboutness (keywords that are neither text-aboutness nor metadiscourse). Despite these three major categories, only the non-text-aboutness metadiscourse keywords underwent contextual analysis in this paper. It should be mentioned that this entire selection included keywords that were not only in the top 100, but were also above this classical threshold (Brezina, 2018). For this reason, the issue of sorting by likelihood or effect size keyness values did not bear much importance, as all the keywords were eventually categorized, including those which had shown varying likelihood and effect size values (this is also the reason why normalized and percentage values are used for discussion in the next section). Since the main goal here was to focus on interactional keywords of the study corpus, *BC* keywords that fit this metadiscourse dimension were identified and used for the subsequent contextual analysis, as will be discussed in the next section.

During this selection process, some issue was raised and needed to be attended to. It was observed that some keywords collocated with other words (some even key), forming lexical bundles, illustrating Sinclair’s (2004) co-occurrence principle and that isolated tokens may be little informative (Berber Sardinha, 2004). To better locate these bundles, the *Cluster tool* was used and Biber *et al.*’s (2021, p. 985) cut-off points adopted. To be considered a bundle in this investigation, the expression should appear in at least five (05) texts (minimum range) and recur at least 10 times (minimum frequency). Another issue was raised even when adopting this cut-off point as the program returned many bundles: they needed to be classified as metadiscoursal according to the stylistic model presented in section 02, otherwise they would be deleted from the analysis. In sum, single and compound (bundles) interactional keywords were selected for analysis, as will be discussed next.

4 Results

The number of types and tokens of key interactional language items categorized according to the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* is presented in Table 05:

Table 05– Number of stylistic metadiscourse interactional types and tokens across BC

Interactional features	types	raw tokens	normed tokens
boosters (stance)	43	2,267	15.66
attitude markers (stance)	8	730	5.04
endophoric and directive markers (engagement)	12	226	1.56
hedges (stance)	1	45	0.31
total	64	3,268	22.58

Source: elaborated by the author

To express stance, Biber uses *boosters*, *attitude markers*, and *hedges*. To establish engagement with readers, he opts for *endophoric markers and directives*. Overall, these participant-oriented subcategories were realized by key language items through adverbs, adjectives, verbs, and bundles. In order to answer the research question posed in the introduction, this section will show how Biber expresses stance and establishes engagement with readers in his RAs and, as a consequence, favors an interactional metadiscourse style. More specifically, it will be explored how Biber's individual stylistic choices convey levels of certainty (through *boosters*) and uncertainty (through *hedges*) about his propositions; make his views explicit, comment on the message, and express evaluations and emotions through *attitude markers*; and directly involve readers in the ongoing discourse by overtly pointing them to other parts in the texts through *endophoric markers and directives*. As will be seen, these rhetorical practices are “[...] essentially evaluative and engaging, expressing solidarity, anticipating objections [...], reveal[ing] the extent to which [Biber] works to jointly construct the text with readers” (Hyland, 2005, p. 51).

It should be mentioned, however, that Biber's style could have been more interactional if other metadiscourse features had been identified as positively key in BC. In other words, the corpus was bereft of statistically significant results of *self-mentions*, *reader pronouns*, *personal asides*, *questions*, and *appeals to shared knowledge*. These are more self-expressing and engaging features, particularly two of them, which will be commented on in the next paragraph. This finding equally illustrates the recent trend of texts in applied linguistics concerning the decrease of engagement features, as well as “[...] an increase in more quantitative studies which restrict opportunities for overt stance-taking.” (Hyland; Jiang, 2019, p. 134). Biber's texts fit this quantitative paradigm and thus prove Hyland and Jiang's (2019) assumption.

Self-mentions are by far the most powerful means of self-representation in academic writing, revealing “[...] the degree of explicit author presence in the text measured by the frequency of first-person pronouns and possessive adjectives [...]” (Hyland, 2005, p. 53). This decision is not merely by chance, because “the presence or absence of explicit author reference is generally a conscious choice by writers to adopt a particular stance and a contextually situated authorial identity.” (Hyland, 2005, p. 53). Another feature absent in BC, *reader pronouns*, “[...] explicitly address readers, either to focus their attention or include them as discourse participants”, serving to “[...] highlight or downplay the presence of readers in the text.” (Hyland, 2005, p. 53). The style guidelines of BC journals publishing companies posed no restriction for the use of *self-mentions* or *reader pronouns*, suggesting Biber's decision to avoid using them altogether or using them less than expected.

Despite the absence of these features, Biber’s texts show significant use of other interactional elements, which will be explained as of the next subsection. For that, this section presents and discusses a few examples (at least one) from each participant-oriented subcategory as used in their respective context of occurrence across BC.

4.1 Boosters

Biber opts for verbs and a single adverb to stress his confidence in attesting claims in texts. Biber’s *boosters* stylistic choices recur in over 50% of all 25 texts and in almost all three time slots, revealing how he consistently opts for these markers across his academic production and time. *Boosters*, like *hedges*, convey epistemic modality, but there is a plausible difference between them, especially in terms of either opening (*hedges*) or closing (*boosters*) space for readers to reach specific conclusions while interpreting texts. Despite the fact that *boosters* tend to limit the variability of viewpoints on the part of readers, they still end up joining together writers and readers, as both agents are part of the same discourse community and, to a certain extent, agree with one another:

Hedges imply, then, that a statement is based on plausible reasoning rather than certain knowledge, and allow readers the freedom to dispute it. At first glance, boosters seem to contradict such conciliatory and defensive tactics. They emphasize the force of propositions and display commitment to statements, thereby asserting the writer’s conviction and restricting the negotiating space available to the reader. But while an apparently risky tactic, boosters nevertheless allow writers to strategically engage with colleagues, effecting interpersonal solidarity and membership of a disciplinary in-group. (Hyland, 1998b, p. 353).

According to Silva (2024), booster verbs state the writers’ stance towards their views and claims; booster verbs are endophoric in nature, pointing readers to some internal aspects of the authors’ research and building solidarity with the audience. These verbs were identified in BC, and their semantic classification and normed frequencies (all occurring 15,55 times per 1,000 words) following Biber *et al.* (2021) are displayed in Table 06:

Table 06 –Semantic classification of booster verbs and their normed frequencies in BC

Activity	Communi- cation	Mental	Existence	Occurrence	Causative
3.18	0.49	2.96	6.67	1.71	0.55
analyze; analyzed; analyzing; carried; combines; investi- gate; show; shows; shown	describe; describing	compares; compute; considered; considering; identify; identified; interpreted;	characterize; characterized; comprises; differ; mark; illustrate; illustrates; illustrated; include; included; presents; marks; reflect; reflects; reflecting; represents; representing; repre- sent; represented; uncover	occur; occurring; resulting	required

Source: elaborated by the author

As can be seen in Table 06, Biber favors existence verbs, followed by activity, mental, occurrence, causative, and communication verbs. Biber's stance expression for existence verbs serves to report states that exist between entities, representing the state of existence or a particular relationship between entities. In example 1 in BC, Biber explains what a comparison corpus is used for within stylistic studies. By opting for the verb *represent*, he expresses certainty about the state of *representing* that this type of corpus is submitted to because of the way it functions within a stylistic investigation. Moreover, the author inhibits the negotiating space available to readers by claiming that this corpus purpose is to *represent*, but nothing else, illustrating Biber's choice of one form out of the array of all available forms:

- (1) The comparison corpus is used to *represent* 'typical' patterns of use, making it possible to empirically identify distinctive linguistic patterns in the target corpus that depart from those typical patterns. (Biber, 2011, p. 16)

In the conclusion section of another text, Biber revisits and reiterates what has been his purpose throughout the RA: to show that corpora of different sizes can be useful (example 2). For that, he chooses the activity verb *show* to denote the action of *showing* while ascertaining the usefulness of such corpora. Furthermore, he prevents readers from identifying other research goals other than the one already laid out. At the same time, however, this limitation offers readers interpersonal solidarity, preventing them from going astray:

- (2) Rather, my purpose has been to *show* that existing corpora are adequate in many respects [...] (Biber, 1990, p. 269)

Elsewhere in BC (example 3), while Biber admits the systematicity and importance of lexico-grammatical association patterns, he nonetheless recognizes some limitation in interpreting these findings only this way, making explicit to the audience that additional reflections must be made. For that, he employs the mental verb *interpreted* in a passive voice structure to denote the interpreting action that must be executed by someone, either himself as researcher/writer or another scholar investigating the same subject. Consequently, he makes an apparent effort to engage with the disciplinary community and so displays a supportive authorial identity. At the same time, Biber restricts his readership to dispute other possibilities for this case besides the urgent need to interpret it:

- (3) Although patterns such as these must be *interpreted* much more fully, the present section has illustrated the systematicity and importance of these association patterns in describing the use of related grammatical features. (Biber, 1996, p. 185)

Biber's certainty and tendency to restrict disputing alternatives to his propositions can also be recovered through the occurrence verb *occur*. In example 4, the author reports on his research findings concerning fillers that complete the lexical frame *the * of the*. These results are even shown in an illustration followed by the statement, proving that Biber's claim is accurate:

- (4) In most cases, the high frequency patterns in academic writing are extremely productive. For example, there are numerous fillers that *occur* in the highest frequency frame the * of the, and as Table 4 shows, the most common of these fillers accounts for only 2% of the total occurrences of the pattern. (Biber, 2009, p. 298)

At the end of another publication (example 5), Biber urges for the need for additional research on use factors and relations between functions and association patterns. To make this demand, he opts for the causative verb *required*, especially in passive voice structures, also indicating that someone else (or perhaps himself) should carry on the research he started. This choice signals the author's certainty for his claim, displaying a confident authorial identity, as can also be seen by the attitude marker *obviously* at the beginning of the sentence. Although it is not obvious who is *required* to undertake additional research, this absence may also be seen as an attempt to engage with the disciplinary community by inviting it to follow on Biber's investigation:

- (5) Obviously, future research is *required* to investigate the relative importance of use factors and the ways in which particular functional considerations relate to particular kinds of association patterns. (Biber, 1996, p. 193)

Another way of ascertaining Biber's claims is through the communication verb *describe*. In example 6, the author lays out three features of a text-linguistic approach to linguistic variation, and limits how his readership may see such an approach. To address issues about the first feature, he admits that the research goal of this approach entails *describing* the characteristics of texts, and not *identifying* or *illustrating* them, actions which belong to other verb semantic categories:

- (6) The text-linguistic approach to linguistic variation differs in all three respects: the research goal is to *describe* the characteristics of texts, rather than the characteristics of a linguistic feature; each text constitutes an observation; and the quantitative findings represent the rates of occurrence of linguistic features in texts rather than the proportional preference for a linguistic variant in comparison to other variants. (Biber, 2012, p. 33)

Biber also uses a single adverb to stress his confident, assertive authorial identity in *BC*. In example 7, in order to highlight two important issues not considered by his analysis, he opts for the booster adverb *reliably* to refer to particular texts that could truthfully represent mean scores calculated on the basis of six linguistic features compared across 10-text samples:

- (7) Secondly, this analysis addressed how many texts were needed to *reliably* represent mean scores, but did not address the representation of linguistic diversity in registers. (Biber, 1993a, p. 253)

Given this overall analysis on the use of *boosters* in *BC*, Biber's individual boosting choices disclose an interactional metadiscourse style whereby the author stresses the strength of his propositions, displays commitment and declares his conviction to statements, even if this entails, on the one hand, restricting the negotiating space available to the participating audience and, on the other hand, offering it interpersonal solidarity occasionally. These choices are also responsive to the rhetorical conventions of the genre and illustrate how the author expresses stance in *BC* to reveal confident, assertive, and supportive authorial identities. Next in subsection 4.2, the focus will be given to the feature the author uses to convey uncertainty about his propositions, *hedges*, which is the counterpart of *boosters*.

4.2 Hedges

Biber opts for only one adverb to show “a lack of complete commitment to the truth value of an accompanying proposition [...] or a desire to express that commitment categorically” (Hyland, 1998a, p. 1). This stylistic choice recurs in almost all BC texts and in all three time slots, proving the author’s consistent decision for this marker across his academic production and time. As a mechanism of author’s self-effacing, epistemic modality, hedging is realized by only one key adverb in BC: *nearly*. This result resonates with Biber *et al.*’s (2021) claim that *hedges* are considered a sub-class of adverbs and that, like conversation, academic prose shows a heavy reliance on single adverbs as stance devices, especially those indicating epistemic stance. Although only one single hedging adverb was found key, this result still represents Biber *et al.*’s (2021) finding that authors in academic prose tend to favor this grammatical device for the expression of epistemic modality, like Biber himself across his RAs.

Hyland (1998a) advocates for the scientific investigation of *hedges* in academia because they are essential to building the rhetoric of academic discourse, especially when it comes to writer’s providing interpretative statements on paper’s theory, method, or data:

Hedging enables writers to express a perspective on their statements, to present unproven claims with caution and to enter into a dialogue with their audiences. It is therefore a substantial means by which scientists confirm their membership of the scientific community [...]. (Hyland, 1998a, p. 6).

In BC, the only key hedge is *nearly*, often collocating with such words as *all*, *identical*, and *every*, as well as with numbers and percentage. These patterns demonstrate that “hedges are also very common with numbers, measurements, and quantities. These forms are also called *approximators*, and typically function as modifiers of numerical or other quantifying expressions.” (Biber *et al.*, 2021, p. 554, emphasis in original). Biber’s choice of *nearly* over other available forms in the language also depicts an individual stylistic preference.

For instance, in example 8, Biber is unsure as to whether specialized dimensions arise from all MD analyses, and in doing so enters into a dialogue with his audience to dispute this claim. In the meantime, he reveals hesitant and dialogic authorial identities:

- (8) And to some extent, this expectation is met, with specialized dimensions emerging in *nearly* all MD analyses. (Biber, 2019, p. 52)

In example 9, Biber is cautious to determine that every discourse domain of English and other languages and cultures favors the presence of a narrative dimension, allowing his readership to refute whether this claim holds true:

- (9) Similar to the oral/literate dimensions, a narrative dimension emerges in the MD analysis of *nearly* every discourse domain of English, and in *nearly* every language/culture. (Biber, 2014, p. 25)

Moreover, when reporting on numerical results from two studies, Biber hedges his commitment to, respectively, providing absolute values for model improvement (example 10) and

two syntactic functions of prepositional phrases (example 11). It can be seen that the author opens possibilities for his readers to notice how approximate (and not absolute) the results are:

- (10) It represents a significant improvement over all other models and a relatively strong improvement in Delta over other models – *nearly* a 5% improvement over the 4-D model, and *nearly* a 10% improvement over the 1-D model. (Biber, 1992, p. 147)
- (11) For editorials, this table shows that there is *nearly* a 50/50 split for prepositional phrases attached as nominal versus verbal modifiers [...] (Biber, 1993b, p. 226)

Given this overall analysis, Biber's individual hedging choice of *nearly* uncovers an interactional metadiscourse style by which he refrains from committing himself to the truth values of his propositions. At the same time, however, he allows readers to respond to the unfolding text, opening up possibilities for (re)assessing the claims and propositions. This stylistic choice also reveals the author's hesitant and dialogic authorial identities. Next in subsection 4.3, the focus will be allocated to the last feature the author uses with a stance-marking purpose: *attitude markers*.

4.3 Attitude markers

Biber opts for adjectives and adverbs to overtly comment on the message and express evaluations and emotions in the meantime. This stylistic choice recurs in over 60% of BC texts and in almost all three time slots, proving the author's consistent decision for these markers across his academic production and time. Furthermore, there is one keyword that works as metadiscoursal both on its own and as part of a lexical bundle: *particular*.

Differently from *hedges* and *boosters*, which are epistemic in nature and convey degrees of (un)certainity and (un)likelihood, *attitude markers* are attitudinal. This is to say that they are used to express writers' affective attitudes, emotions, evaluations, and feelings to/on a proposition, conveying surprise, agreement, importance, obligation, frustration, etc. (Biber *et al.*, 2021; Biber, 2006; Hyland, 2005). As a result, this self-expression helps construct an interactional metadiscourse style in texts. Compared to frequently used epistemic stylistic choices, Biber's underuse of *attitude markers* appears to reflect academic prose register: "overall, attitudinal stance markers are less common than epistemic markers [...]" (Biber *et al.*, 2021, p. 966).

It is perceivable that Biber favors the expression of evaluations and emotions while overtly commenting on the messages, echoing his voice plainly to the readers. In the former case, he opts for the adjectives *important*, *particular*, *primary*, the adverb *especially*, and the bundle *in particular* to convey importance; and the adjective *adequate* to transmit a sense of suitability. In the latter, he chooses the adjective *surprising* and the adverb *strikingly* to express his surprise towards claims.

To begin with, Biber's expression of importance unveils his evaluative authorial identity. As seen in example 12, the author opts for the adjective *important* to highlight that textual co-occurrence patterns between frequent nouns, adjectives, and prepositional phrases are relevant for the identification of linguistic characteristics of registers and styles. It is thus possible to detect the way Biber comments on his message and leaves his fingerprint for readers to spot, also representing an individual stylistic choice:

- (12) For example, frequent nouns, adjectives, and prepositional phrases commonly co-occur in academic prose texts, working together to provide a dense integration of information. Textual co-occurrence patterns such as these are *important* in identifying the salient linguistic characteristics of registers and styles. (Biber, 1996, p. 175)

The expression of importance is also recognized by the adverb *especially*. In some cases, this adverb is even followed by the stance adjective *important*. In other cases, though, it collocates with other stance adjectives (*noteworthy, robust, conservative*) that did not return key by *AntConc*, but which can disclose Biber's evaluative authorial identity. In example 13, the author recognizes that, out of all dimensions of MD analysis often identified in many different studies, two bear especial importance because they help build basic human communication:

- (13) Two such dimensions are *especially important*, regardless of the discourse domain: a dimension associated with oral vs. literate discourse, and a dimension associated with narrative discourse. The robustness of narrative dimensions across languages and discourse domains indicates that this rhetorical mode is basic to human communication, whether in speech or in writing. (Biber, 2014, p. 31)

Another evaluative stance adjective used to transmit a sense of suitability in *BC* is *adequate*. In example 14, Biber refutes some criticisms and leads readers to grasp that existing computer-based corpora are suitable for many studies of linguistic variation:

- (14) However, the findings reported here argue against the claim that existing computer-based corpora are unsuited for the study of linguistic variation; instead, they indicate that, with respect to the above questions, the corpora are *adequate* for many studies of linguistic variation across texts. (Biber, 1990, p. 258)

The expression of emotion in *BC* is observed through the stance *attitude markers surprising* and *strikingly*, disclosing Biber's emotional authorial identity. In example 15, Biber is fully surprised with some research results because they may have been influenced by the layout of the corpus used. Again, it is possible to see how Biber comments on his message and emits his voice for readers to hear:

- (15) It is perhaps not *surprising* that Dimension 1 in the original 1988 MD analysis was strongly associated with the oral/literate opposition, given that the corpus in that study ranged from spoken conversational texts to written expository texts. (Biber, 2014, p. 20-21)

In another use, the adjective *surprising* is modified by the adverb *especially* presented earlier. Below, it can be perceived how Biber's evaluative and emotional authorial identities are revealed in the sentence, with particular stress on the former. As can be seen in example 16, a common assumption about syntactic complexity held by many linguists is weakened when research revealed unexpected findings. Once again, Biber's interactional choice evidently discloses how he feels about the results found:

- (16) The oral style relies on clauses to construct discourse – including a dense use of dependent clauses. In contrast, the complexity of the literate style is phrasal. This finding, replicated across languages and across discourse domains, is *especially surprising*

sing, because it runs counter to assumptions about syntactic complexity held by many linguists. (Biber, 2019, p. 53)

Furthermore, emotion is also conveyed through the adverb *strikingly*. As seen in example 17, Biber's stance choice of this word reveals his surprise towards particular lexico-grammatical patterns detected in some analyses:

- (17) These analyses show that *strikingly* different lexico-grammatical patterns are associated with each complement clause type and with each register, and that those associations can be explained in terms of the typical topics and communicative purposes of each register. (Biber, 1999, p. 136)

Given this overall analysis on the use of stance-marking *attitude markers* in BC, Biber's individual stylistic attitudinal choices unveil an interactional metadiscourse style whereby the author overtly comments on the messages, taking the participating audience to recognize his evaluative and emotional authorial identities. Lastly, in subsection 4.4, the focus will be allocated to the only feature the author uses with a more overtly engaging purpose: *endophoric and directive markers*.

4.4 Endophoric and directive markers

Biber opts for bundles to explicitly direct readers to particular points in his RAs. This stylistic choice recurs in over 50% of BC texts and in almost all three time slots, proving the author's consistent decision for these markers across his academic production and time. In the first assessment of the keyword list, however, *figure*, *shows*, *table*, *sample*, *illustrates*, *presents*, *above*, *section*, *shown*, and *summarizes* were identified as key, but when their collocates were searched for, it was observed that they formed bundles and could also function as metadiscoursal in BC.

Prior to illustrating some examples of *endophoric and directive markers* from BC, it is important to remember that, *endophoric markers* were originally classified as an interactive feature by Hyland (2005), but in a revisited version of the model, the author excluded it altogether and created *directives* (Hyland, 2009). However, in this paper, it is hereby understood that there is a connection between *endophoric markers* (originally interactive) and *directives* (originally interactional), for they have the same purpose: to show perceptible assessments of the material and the audience, direct readers to particular points in the text, and influence them to reach specific understandings and conclusions (Hyland, 2005). For this reason, in this paper, *endophoric markers* are a type of *interactional engagement markers*, together with *directives*, for both involve "[...] rhetorically positioning the audience, pulling readers into the discourse at critical points, predicting possible objections and guiding them to particular interpretations." (Hyland, 2005, p. 54). These features are also concerned with genre conventions (Swales, 1990). In sum:

Directives instruct the reader to perform an action or to see things in a way determined by the writer. [...] they are signalled mainly by imperatives (like *consider*, *note*, and *imagine*), modals of obligation addressed to the reader (e.g. *must*, *should*, *ought*) [...] Directives are the only engagement feature which occur more frequently in papers in the sciences and engineering. (Hyland, 2005, p. 154).

Endophorics [...] refer the reader to explanatory or related material elsewhere in the text. They represent the writer's assessment of both the material and the audience, relating the propositions to the reader's assumed ability to process, and accept, the ongoing argument. In the soft fields this largely means reinforcing an argument by ensuring the reader has immediate access to relevant data or arguments located elsewhere in the text [...] Endophorics, however, are overwhelmingly a feature of writing in the hard disciplines, where the referent is usually a nearby table or graph. (Hyland, 2005, p. 156-157).

In BC, Biber explicitly directs readers to other parts of the texts through textual acts by means of different bundles, such as *see table* and *see section*. It should be noted that the journal publishing companies' style guidelines encouraged authors to explicitly make cross-references along the texts, with two publishers suggesting the use of mental verbs like *see*. In this regard, Biber's choices seem to have undergone style guidelines influences, suggesting that he may have ended up abiding by academic conventions administered by literacy brokers and gatekeepers, including peer reviewers. Despite that, an interesting pattern across BC is that all uses of *see table* and *see section* come in parentheses and mostly at the end of sentences, implying that these directives behave similarly to non-integral citations, receiving minor focus and attention from the author. For instance, through textual acts, Biber discloses a supportive authorial identity by directing readers to *see*: a) a list of texts used to compare British and American writing, classified according to genre (example 18); and b) a section where it has been discussed how some dimensions characterized English and some discourse domains (example 19):

- (18) Each dimension will be discussed briefly with respect to both written and spoken genres, but the focus here is on the 796 British and American written texts taken from the LOB and Brown corpora, representing the same nine written genres in each dialect (*see table 3*). (Biber, 1987, p. 109)
- (19) Subsequent MD studies have shown that some of these dimensions turn out to be peculiar to English or to particular discourse domains (*see Section 5*). (Biber, 2014, p. 15)

Moreover, the keyword *shows* often collocates with *figure* and *table*. In example 20, it is possible to notice that Biber motivates readers to look at an illustration in the text, in which a marked decline with respect to the use of modals in different registers over a 5-decade period is depicted. The author's choice regards a textual act to be undertaken by readers:

- (20) The study here provides additional evidence that modals have undergone a recent decline in use. As *Figure 4 shows*, modals have undergone a marked decline in all registers, especially over the last 50 years. (Biber, 2004, p. 117)

As can be seen, the use of *endophoric and directive markers* demonstrates an involvement between different semiotic resources available in the writer's texts to which he makes explicit reference. This connection can also be identified in example 21, proving that "[...] writing is a constant switching between written and diagrammatic representation because scientific concepts are typically semiotic hybrids which combine verbal, mathematical and visual elements." (Hyland, 2005, p. 157). The way Biber refers to data in tables illustrates this hybrid combination of elements. Again, readers are encouraged to achieve a textual act by

looking at a specific table in the article which presents numerical data on mean score and standard deviation:

- (21) *Table 4 presents* the mean score and standard deviation of seven linguistic features in the pilot corpus, together with the computed tolerable error for each feature. It can be seen that the tolerable error ranges from 9.03 for nouns (which have a mean of 180.5) to 0.13 for conditional clauses (which have a mean of only 2.5) (Biber, 1993a, p. 253)

A further textual act is fostered in another excerpt in *BC*. In example 22, Biber leads readers to another part of the text, although not certainly where, encouraging them to note that text-linguistic register studies have now unveiled previously undiscovered language patterns. This choice also signals an individual stylistic decision:

- (22) Second, *as noted above*, text-linguistic register studies have also been important because they have uncovered patterns of use that were not even anticipated in previous research. (Biber, 2019, p. 49)

As seen in this subsection, Biber's preference for some *endophoric and directive markers* demonstrates how he engages with readers by leading them to particular text extracts, some even made of hybrid semiotic resources (*figures, tables*). The author's stylistic engaging choices disclose a supportive authorial identity and unveil an interactional metadiscourse style by which he explicitly bonds with the participating audience so that it can promptly recognize relevant information referred to in the RAs. In some cases, though, these choices may have been influenced by literacy brokers. Final Remarks are presented next as a way to summarize the findings in this research and illustrate how Biber has constructed a metadiscourse style through interactional features.

5 Conclusion

This research has attempted to unveil the interactional metadiscourse style of Biber by pursuing the following research question: how does Biber express stance and establish engagement with readers in his RAs through interactional metadiscourse elements? In view of the analyses carried out in this paper, it has been detected that Biber has used *boosters, attitude markers*, and *hedges* to express stance, and *endophoric markers and directives* to establish engagement with readers. Consequently, these patterns result in the creation of an interactional metadiscourse style through an individual preference for particular adverbs, adjectives, verbs, and bundles.

As seen across *BC*, the author's individual stylistic choices across texts and time convey levels of certainty (through *boosters*) and uncertainty (through *hedges*) about his propositions; make his views explicit, comment on the message, and express evaluations (importance and suitability) and emotions (surprise) through *attitude markers*; and directly involve readers in the ongoing discourse by overtly pointing them to other parts in the RAs through *endophoric markers and directives*.

More specifically, Biber's use of *boosters* has stressed the strength of his propositions and displayed his commitment and conviction to statements, revealing confident, assertive, and supportive authorial identities. Moreover, the author's *hedge* choice has demonstrated bivalent decisions on his part: he refrains from committing himself to some claims, but also allows the audience to respond to the unfolding text, opening up possibilities for (re)assessing the propositions. Together, these choices have disclosed the author's hesitant and dialogic authorial identities. Furthermore, Biber's adoption of *attitude markers* has allowed him to overtly comment on the discourse through the expression of evaluations and emotions, leading the participating audience to perceive his evaluative and emotional authorial identities. At last, the author's preference for some *endophoric and directive markers* have demonstrated how he engages with readers by leading them to promptly recognize relevant information located in figures, tables, sections, and other parts of the RAs, disclosing a supportive authorial identity.

Notwithstanding the accomplishment of this research, broader claims about academic writing may be substantiated through further comparative analyses of multiple authors or contexts. Therefore, some limitations have been noticed and may be resolved in future studies. *CorJAAL* may not have been the best reference corpus, but has nonetheless proven so on the basis of the research goals. Unarguably, the best layout of a reference corpus is an unsolved issue and has continued to foster additional discussions. Nevertheless, as previous research (Geluso; Hirsch, 2019) has shown, a different layout might affect the findings. Furthermore, because a style analysis is comprehensive, it can be argued that gatekeepers/literacy brokers may have influenced Biber's stylistic choices. However, the attention has hereby been directed to the texts resulting from such editing and publishing processes, since tracking other people's editions to Biber's texts might have been a daunting, if not an impossible task to carry out here. In terms of keywords chosen for analysis, like the research carried out by Pojanapunya and Watson Todd (2016), future studies might compare keyword lists generated separately by likelihood and effect size measures and see how they illustrate potentially different stylistic authorial choices. At last, the *Stylistic Metadiscourse Model of Interpersonal Interaction in Academic Writing* (Silva, 2024,2025) was devised for the purposes of this research, but if applied in additional investigations, it could be verified whether it continues to disclose academic writers' style.

In view of the data analyzed in this paper, "strikingly" (an interactional keyword used in BC), an academic writer's style can be interactional in applied linguistics research articles. Biber's interactional rhetorical practices are essentially evaluative and engaging, expressing solidarity, anticipating objections, and revealing the extent to which he works to jointly construct texts with his readers. By debunking the myth of a dry, impersonal discourse in academic writing and taking on Crystal's (2013) early proposal, it is possible to see that Biber's tendency to an interactional metadiscourse style shows potential implications for the understanding of how an expert academic writer crafts an authorial style that may empower novice and other expert writers to craft a style that involves the audience in meaningful reading and writing experiences through academic texts.

Acknowledgements

The author of this paper is thankful to Professor Deise Prina Dutra, PhD, from Federal University of Minas Gerais (UFMG), for advising the PhD dissertation from which this paper has originated.

References

- ANTHONY, L. *AntConc* (Version 4.2.0) [Computer Software]. Tokyo: Waseda University, 2022. Available at: <https://www.laurenceanthony.net/software/antconc/>. Accessed on: 11 Nov. 2025.
- BACON, N. Style in academic writing. In: DUNCAN, M.; VANGURI, S. M. (ed.). *The centrality of style*. Anderson: Parlor Press, 2013. p. 173-190.
- BAKER, M. Towards a methodology for investigating the style of a literary translator. *Target*, Amsterdam/Philadelphia, v. 12, n. 2, p. 241-266, 2000. DOI : <https://doi.org/10.1075/target.12.2.04bak>
- BAKHTIN, M. *The dialogic imagination: four essays by M. M. Bakhtin*. Austin: University of Texas Press, 1981.
- BERBER SARDINHA, A. P. Comparing Corpora with Wordsmith Tools: How Large Must the Reference Corpus Be?. In: Annual Meeting of the Association for Computational Linguistics, 38, 2000, Hong Kong. *Proceedings of the Workshop on Comparing Corpora...* Hong Kong: Hong Kong University of Science and Technology, 2000. p. 7-13.
- BERBER SARDINHA, A. P. *Lingüística de corpus*. Barueri: Manole, 2004.
- BIBER, D. et al. *Grammar of spoken and written English*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2021.
- BIBER, D. A textual comparison of British and American writing. *American Speech*, [S. l.], v. 62, n. 2, p. 99-119, 1987. DOI: <https://doi.org/10.2307/455273>
- BIBER, D. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, [S. l.], v. 5, n. 4, p. 257-269, 1990. DOI: <https://doi.org/10.1093/lc/5.4.257>
- BIBER, D. On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, [S. l.], v. 15, n. 2, p. 133-163, 1992. DOI: <https://doi.org/10.1080/01638539209544806>
- BIBER, D. Representativeness in Corpus Design. *Literary and Linguistic Computing*, [S. l.], v. 8, n. 4, p. 243-257, 1993a. DOI: https://doi.org/10.1007/978-0-585-35958-8_20
- BIBER, D. Using register-diversified corpora for general language studies. *Computational Linguistics*, [S. l.], v. 19, n. 2, p. 219-241, 1993b.
- BIBER, D. Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics*, [S. l.], v. 1, n. 2, p. 171-197, 1996. DOI: <https://doi.org/10.1075/ijcl.1.2.02bib>
- BIBER, D. A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies*, [S. l.], v. 1, n. 2, p. 131-150, 1999. DOI: <https://doi.org/10.1177/1461445699001002001>

- BIBER, D. Historical patterns for the grammatical marking of stance: A cross-register comparison. *Journal of Historical Pragmatics*, [S. l.], v. 5, n. 1, p. 107–136, 2004. DOI: <https://doi.org/10.1075/jhp.5.1.06bib>:
- BIBER, D. *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.
- BIBER, D. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, [S. l.], v. 14, n. 3, p. 275–311, 2009. DOI: <https://doi.org/10.1075/ijcl.14.3.08bib>
- BIBER, D. Corpus linguistics and the study of literature: Back to the future?. *Scientific Study of Literature*, [S. l.], v. 1, n. 1, p. 15–23, 2011. DOI: <https://doi.org/10.1075/ssol.1.1.02bib>
- BIBER, D. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, [S. l.], v. 8, n. 1, p. 9–37, 2012. DOI: <https://doi.org/10.1515/cllt-2012-0002>
- BIBER, D. Interview with Douglas Biber. [Interview given to] Bethany Gray. *Journal of English Linguistics*, Thousand Oaks, v. 41, n. 4, p. 359–379, 2013. DOI: <https://doi.org/10.1177/0075424213502237>
- BIBER, D. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast*, [S. l.], v. 14, n. 1, p. 7–34, 2014. DOI: <https://doi.org/10.1075/lic.14.1.02bib>
- BIBER, D. Text-linguistic approaches to register variation. *Register Studies*, [S. l.], v. 1, n. 1, p. 42–75, 2019. DOI: <https://doi.org/10.1075/rs.18007.bib>
- BIBER, D. *Home*. Arizona: Northern Arizona University, [201-?]. Available at: <https://dougiber.weebly.com/>. Accessed on: 11 Nov. 2025.
- BIBER, D.; CONRAD, S. *Register, genre, and style*. 2. ed. Cambridge: Cambridge University Press, 2019.
- BIBER, D.; GRAY, B. *Grammatical complexity in academic English: linguistic change in writing*. Cambridge: Cambridge University Press, 2016.
- BREZINA, V. *Statistics in corpus linguistics: a practical guide*. Cambridge/New York: Cambridge University Press, 2018.
- BURKE, M. Rhetoric and poetics: The classical heritage of stylistics. In: BURKE, M. (ed.). *The Routledge Handbook of Stylistics*. London/New York: Routledge, 2014. p. 11–30.
- CÂMARA JUNIOR, J. M. *Contribuição à estilística portuguesa*. 3. ed. Rio de Janeiro: Ao livro técnico, 1978.
- CLARIVATE. *Journal Citation Reports*. London: Clarivate, 2023.
- CRISMORE, A. *Metadiscourse: what is it and how is it used in school and non-school social science texts*. Urbana-Champaign: University of Illinois, 1983.
- GARCIA, O. M. *Comunicação em prosa moderna: aprenda a escrever, aprendendo a pensar*. 26. ed. Rio de Janeiro: Editora FGV, 2006.
- GELUSO, J.; HIRCH, R. The reference corpus matters: comparing the effect of different reference corpora on keyword analysis. *Register Studies*, Amsterdam/Birmingham, v. 1, n. 2, p. 209–242, 2019.
- GRAY, B. *Linguistic variation in research articles: when discipline tells only part of the story*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.

GRIEVE, J. Register variation explains stylometric authorship analysis. *Corpus Linguistics and Linguistic Theory*, Berlin, v. 19, n. 1, 1-31, 2023.

HYLAND, K. The scholarly publishing landscape. In: HANGANU-BRESCH, C.; ZERBE, M. J.; CUTRUFELLO, G.; MACI, S. M. (ed.). *The Routledge handbook of scientific communication*. London/New York: Routledge, 2022. p. 15-25.

HYLAND, K. Community and individuality: performing identity in applied linguistics. *Written Communication*, Thousand Oaks, v. 27, n. 2, p. 159-188, 2010.

HYLAND, K. *Academic Discourse: English in a global context*. London/New York: Continuum International Publishing Group, 2009.

HYLAND, K. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, Amsterdam, v. 27, n. 1, p. 4-21, 2008a.

HYLAND, K. Small bits of textual material: A discourse analysis of Swales' writing. *English for Specific Purposes*, Amsterdam, v. 27, n. 2, p. 143-160, 2008b.

HYLAND, K. *Metadiscourse: Exploring Interaction in writing*. London/New York: Continuum International Publishing Group, 2005.

HYLAND, K. *Disciplinary discourses: social interactions in academic writing*. Ann Arbor: The University of Michigan Press, 2004.

HYLAND, K. Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics*, Amsterdam, v. 34, p. 1091-1112, 2002b. DOI: [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8)

HYLAND, K. Options of identity in academic writing. *ELT Journal Volume*, Oxford, v. 56, n. 4, p. 351-358, 2002a. DOI: <https://doi.org/10.1093/elt/56.4.351>

HYLAND, K. *Hedging in scientific research articles*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1998a.

HYLAND, K. Boosting, hedging and the negotiation of academic knowledge. *Text*, Berlin, v. 18, n. 3, p. 349-382, 1998b.

HYLAND, K.; TSE, P. Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, Oxford, v. 25, n. 2, p. 156-77, 2004. DOI: <https://doi.org/10.1093/applin/25.2.156>

HYLAND, K.; JIANG, F. (K.). *Academic discourse and global publishing: disciplinary persuasion in changing times*. London/New York: Routledge, 2019.

I LOVE PDF. [S. l.: s. n], 2025. Available at: <https://www.ilovepdf.com/> Accessed on: 11 Nov. 2025.

IVANIČ, R. *Writing and identity: the discursual construction of identity in academic writing*. Amsterdam/Philadelphia: John Benjamins Publishing, 1998.

JEFFRIES, L.; MCINTYRE, D. *Stylistics*. Cambridge: Cambridge University Press, 2010.

LEECH, G.; SHORT, M. *Style in fiction: a linguistic introduction to English fictional prose*. Harlow: Pearson/Longman, 2007.

LILLIS, T.; CURRY, M. *Academic writing in a global context: the politics and practices of publishing in English*. London/New York: Routledge, 2010.

- MALMKJAER, K.; CARTER, R. Stylistics. In: MALMKJAER, K. (ed.) *The linguistics encyclopedia*. 2. ed. London/New York: Routledge, 2002. p. 510-520.
- MCINTYRE, D.; WALKER, B. *Corpus stylistics: theory and practice*. Edinburgh: Edinburgh University Press Ltd, 2019.
- MCMENAMIN, G. R. (ed.) *Forensic Linguistics: advances in forensic stylistics*. Boca Raton: CRC Press, 2002.
- ONLINE 2 PDF. *Conversor PDFonline*. [S. l.: s. n], 2007. Available at: <https://online2pdf.com/> Accessed on: 11 Nov. 2025.
- OCR2EDIT. *OCR2Edit - Extract text from scans, images and more*. [S. l.: s. n], 2025. Available at: <https://www.ocr2edit.com/> Accessed on: 11 Nov. 2025.
- PEARSON, W. S.; ABDOLLAHZADEH, E. Metadiscourse in academic writing: a systematic review. *Lingua*, Amsterdam, v. 263, p. 1-32, 2023. DOI: <https://doi.org/10.1016/j.lingua.2023.103561>
- POJANAPUNYA, P.; WATSON TODD, R. Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, Berlin, v. 14, n. 1, 133-167, 2016. DOI: 10.1515/cllt-2015-0030
- SCOTT, M. *WordSmith Tools manual*. Stroud: Lexical Analysis Software Ltd., 2022. Available at: https://lexically.net/downloads/version_64_8/HTML/index.html . Accessed on: 11 Nov. 2025.
- SCOTT, M.; TRIBBLE, C. *Textual patterns: key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.
- SILVA, J. M. V. *(Un)surprisingly, style can be found in academic writing: towards characterizing Douglas Biber's metadiscourse style*. 2024. 176f. Dissertation (Ph.D. in Linguistic Studies) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2024.
- SILVA, J. M. V. Being Interactive in Applied Linguistics Research Articles: Towards Characterizing an Academic Writer's Metadiscourse Style. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 25, n. 1, p. 1-29, 2025. DOI: <https://doi.org/10.1590/1984-6398202549625>
- SINCLAIR, J. *Trust the text: language corpus and discourse*. London: Routledge, 2004.
- SWALES, J. M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.
- TARDY, C. Current conceptions of voice. In: HYLAND, K.; GUINDA, C. (ed.). *Stance and voice in written academic genres*. London: Palgrave Macmillan, 2012. p. 34-47.
- THOMPSON, G. Interaction in academic writing: learning to argue with the reader. *Applied Linguistics*, Oxford, v. 22, n. 1, p. 58-78, 2001.
- THOMPSON, G.; THETELA, P. The sound of one hand clapping: the management of interaction in written discourse. *TEXT*, Berlin, v. 15, n. 1, 103-27, 1995. DOI: <https://doi.org/10.1515/text.1.1995.15.1.103>
- TOGNINI-BONELLI, E. *Corpus linguistics at work*. Amsterdam/Birmingham: John Benjamins Publishing Company, 2001.
- VANDE KOPPLE, W. Some exploratory discourse on metadiscourse. *College Composition and Communication*, New York/ Ann Arbor, v. 36, n. 1, p. 82-93, 1985. DOI: <https://doi.org/10.2307/357609>
- WALTMAN, L.; TRAAG, V. A. Use of the journal impact factor for assessing individual articles: Statistically flawed or not? *F1000Research*, London, v. 9, p. 1-29, 2000.

Desbravando *O Continente*: a Estilística de *Corpus* aplicada à literatura regionalista brasileira

Unveiling O Continente: Corpus Stylistics Applied to Brazilian Regionalist Literature

Geovani Henrique Santos de Souza

Universidade Federal do Rio Grande do Sul (UFRGS) | Porto Alegre | RS | BR
geovani.henrique@ifsc.edu.br
<https://orcid.org/0009-0005-8218-7159>

Rozane Rodrigues Rebechi

Universidade Federal do Rio Grande do Sul (UFRGS) | Porto Alegre | RS | BR
Rozane.rebechi@ufrgs.br
<https://orcid.org/0000-0002-1878-7548>

Resumo: Estudos em Estilística de *Corpus* (EC), área que aplica os pressupostos teóricos e metodológicos da Linguística de *Corpus* (LC) na análise de estilo literário, ainda são incipientes nos estudos de literatura em língua portuguesa. Uma das razões que justificam a resistência em se utilizar uma metodologia já consolidada em outros campos do conhecimento para o levantamento de padrões estilísticos textuais e literários pode estar diretamente ligada às limitações das ferramentas computacionais no processamento de textos em português brasileiro, especialmente no que tange à identificação automática de domínios semânticos. Este estudo piloto testa o etiquetador semântico multilíngue PyMUSAS, com interface no Wmatrix 7, primeira versão multilíngue do software de análise textual, na investigação de *O Continente*, de Erico Veríssimo. A partir da comparação com um *corpus* de referência composto pela antologia de escritores contemporâneos ao romancista gaúcho, a ferramenta evidenciou os campos semânticos mais representativos da obra regionalista gaúcha quando comparada às dos outros autores. Os padrões estilísticos levantados automaticamente foram, então, analisados manualmente. Os resultados obtidos corroboram dados da crítica literária tradicional, além de demonstrarem outras possibilidades de análise literária. Contudo, esta pesquisa também aponta que são necessários ajustes no etiquetador, a fim de elevar sua fiabilidade e impulsionar os estudos em EC no Brasil.

Palavras-chave: Estilística de *Corpus*; *O Continente*; Erico Veríssimo; etiquetagem semântica automática; literatura regionalista brasileira.



Abstract: Studies in *corpus* stylistics (CS), an area that applies the theoretical and methodological assumptions of *corpus* linguistics (CL) to the analysis of literary style, are still in their infancy in Portuguese literature studies. One of the reasons that explains the resistance to using a methodology already consolidated in other fields of knowledge for identifying textual and literary stylistic patterns may be directly linked to the limitations of computational tools in processing texts in Brazilian Portuguese, especially regarding the automatic identification of semantic fields. This pilot study tests the multilingual semantic tagger PyMUSAS, interfaced with Wmatrix 7, the first multilingual version of the text analysis software, to investigate *O Continente*, by Erico Veríssimo. By comparing it with a reference *corpus* composed of an anthology of writers contemporary to the southern novelist, the tool revealed the most representative semantic fields of the regionalist work when contrasted with the other authors. The stylistic patterns automatically identified were then manually analyzed. The results corroborate traditional literary criticism data and demonstrate other possibilities for literary analysis. However, this study also points out that adjustments to the tagger are necessary to increase its reliability and boost CS studies in Brazil.

Keywords: *corpus* stylistics; *O Continente*; Erico Veríssimo; automatic semantic annotation; Brazilian regionalist literature.

1 Introdução

A Linguística de *Corpus* (LC) é o campo de estudo que parte de uma coletânea de textos em formato eletrônico, organizada a partir de critérios definidos pelos objetivos de pesquisa, para levantamento de padrões textuais (Tagnin, 2013, p. 29-30). Nesse contexto, o avanço tecnológico e a informatização da ciência possibilitaram a construção de ferramentas que permitem a observação de fenômenos linguísticos textuais de forma (semi)automática. Essa revolução se deve à capacidade de compilação de imensas bases de textos e de consulta a esses bancos de dados com ferramentas computacionais, como AntConc (Anthony, 2021), WordSmith Tools (Scott, 2024) e Sketch Engine (Kilgarrieff, 2014).

Embora a LC seja tradicionalmente aplicada à Lexicografia, à Terminografia, ao Ensino de Línguas, à Análise de Variação de Registro e de Discurso, à Tradução, entre outras,

essa metodologia ainda está se estabelecendo nos estudos literários. Zyngier, Carneiro e Novodvorski (2023, p. 391) defendem o uso da LC para a análise literária baseada em evidências, diminuindo, portanto, o subjetivismo das interpretações hermenêuticas. Ressalta-se, contudo, que, após apresentarem algumas possibilidades de aplicação de recursos computacionais na análise literária e estimularem seu uso como forma de questionar paradigmas tradicionais, os autores enfatizam que a metodologia não prescinde da análise crítica.

Na última década, diversos estudos aplicaram recursos computacionais na análise de estilo, fundamentando a área que passou a ser conhecida como Estilística de *Corpus* (EC). McIntyre e Walker (2019) ocuparam-se de sistematizar os conceitos e as aplicações aos quais a EC se circunscreve. No livro, os autores revisam os estudos em língua inglesa que empregam métodos e ferramentas da LC na análise estilística literária e discutem aspectos que particularizam a EC como um campo de estudos.

Algumas aplicações de EC podem ser vistas em investigações de obras literárias em língua inglesa, como em Fischer-Starcke (2009), que analisa as especificidades de *Pride and Prejudice*, de Jane Austen (1813 [2002]), quando comparada aos outros romances da autora; Maturana (2012), cuja pesquisa investiga as vozes poéticas em *The Adoption Papers*, de Jackie Kay (1991); Wynne (2012), com um trabalho de exploração estilística do livro *The Magus*, de John Fowles (1965); e Mahlberg (2013), que realiza uma pesquisa minuciosa da obra digital de Charles Dickens, enfocando técnicas de caracterização e apresentação de linguagem corporal e dialógica a partir de padrões linguísticos. Mais recentes são as publicações de Čermákova (2018), cujo recorte explora uma abordagem de EC para o estudo da tradução literária voltada à literatura infantil; Ibrahim (2022), que avalia o uso da ferramenta Wmatrix na análise estilística de obras de Charles Dickens; e Can e Cangir (2022), que aplicam a EC na análise de poesias escritas por soldados que combateram na Primeira Guerra Mundial.

Já no contexto brasileiro, há alguns estudos que já recorrem à LC para analisar obras literárias, como Pimenta e Novodvorski (2018), Rangel et al. (2022) e Vital (2022), todavia, não foram encontrados trabalhos que mencionam ou empregam diretamente a metodologia de EC na análise de obras em português brasileiro em portais de periódico Capes e Scielo. Uma das razões da escassez de estudos em língua portuguesa que apliquem a metodologia de EC pode estar diretamente ligada às limitações das ferramentas de etiquetagem semântica dos textos em português.

Este estudo piloto tem por objetivo avaliar o desempenho da versão 7 do *software* Wmatrix¹ (Rayson, 2009), ferramenta on-line de análise de *corpus* integrada ao anotador semântico multilíngue PyMUSAS, e analisar o que a EC pode agregar à crítica literária já existente do romance *O Continente* (Veríssimo, 2013 [1949]).

Na próxima seção, será apresentada uma revisão bibliográfica de pesquisas que utilizaram a EC. Em seguida, a metodologia aplicada a esta pesquisa será explicitada, abrangendo a seleção do *corpus* de estudo e de referência e seu processamento com os recursos disponíveis na ferramenta Wmatrix. Adiante, serão discutidos os resultados obtidos a partir dos métodos selecionados, avaliando-se as vantagens e deficiências da ferramenta no processamento de textos em português. A Conclusão encerra este estudo piloto, apontando as limitações da pesquisa e perspectivas futuras.

¹ Disponível em: <https://ucl.ac.uk/wmatrix/>. A utilização da ferramenta está condicionada à licença paga ou gratuita, de acordo com requisitos de uso.

2 Fundamentação teórica

Nesta seção, são apresentados conceitos e ferramentas basilares da EC: estilística e anotadores semânticos automáticos, neste caso, o *framework* de etiquetagem semântica automática UCREL Python Multilingual Semantic Analysis System (PyMUSAS), disponível na ferramenta de análise textual Wmatrix. Essa ferramenta foi utilizada em diversas pesquisas de LC e EC, como, por exemplo, McIntyre e Walker (2019), McIntyre e Archer (2010), Culpeper, Archer e Rayson (2009), Balossi (2014, 2020), Ibrahim (2022) e Vathanalaoha (2022). Além disso, será feita uma breve apresentação da obra *O Continente*, de Erico Veríssimo.

2.1 Estilística de *Corpus*

A LC é uma abordagem que compreende um conjunto de procedimentos para o estudo de textos autênticos processáveis por programas computacionais e coletados a partir de critérios bem definidos (McEnery; Hardie, 2012; Tagnin, 2013). Trata-se de uma metodologia já consolidada em áreas como Tradução (Baker, 1993; Zanettin, 2012), Lexicografia (Teubert, 2001; Krishnamurthy, 2008), Discurso (Baker, 2023) e Ensino (Biber; Reppen, 2002; O’Keeffe; McCarthy; Carter, 2007), entre outras, mas ainda incipiente na análise de estilo literário, especialmente quando tomamos por base pesquisas em língua portuguesa no Brasil.

Já a Estilística é uma ramificação da Linguística Aplicada que possibilita o estudo de textos em geral e de literatura em particular. Esse campo de estudos dispõe de um conjunto de conceitos para a descrição, interpretação e avaliação textual (Short, 1996). De acordo com o autor, uma avaliação mais objetiva pode endossar ou refutar análises realizadas pela crítica literária tradicional, por se basear em fatos linguísticos observáveis, como aspectos textuais de nível discursivo, gramatical, grafológico, morfológico, lexical, semântico e fonológico (Short, 1996, p. 5).

Da combinação da LC com a Estilística adveio a Estilística de *Corpus* (EC), abordagem que propõe a integração entre dados levantados automaticamente e sua análise manual, que começou a se expandir nos últimos anos (McIntyre; Walker, 2019). Apesar de a EC não estar restrita ao levantamento de campos semânticos, mas também abranger a análise de palavras-chave, linhas de concordância e n-gramas para os estudos de estilo literário, a pesquisa que leva em conta o significado das palavras permite que se explorem padrões temáticos e ampara com dados estatísticos a investigação dos fatos linguísticos descritos por Short (1996).

McIntyre e Walker (2019, p. 14) salientam como vantagens da EC a praticidade na sistematização dos padrões presentes em obras literárias e a diminuição da subjetividade das críticas convencionais, aumentando, portanto, o rigor das análises. Outra vantagem da combinação de abordagens, conforme os autores, é a possibilidade de se alcançar a representatividade, diferentemente do que ocorre em relação à língua geral. Isso porque a EC, conforme McIntyre e Walker (2019, p. 5), permite que se analise, por exemplo, a totalidade das obras de determinado autor, ou de determinado período literário, enquanto que a língua geral jamais poderia ser analisada em sua completude, e qualquer tentativa de se construir um *corpus* representativo e balanceado de todas as possibilidades linguísticas será frustrado (Leech, 2007).

Vale ressaltar que a análise do estilo literário de autores e obras em língua inglesa por meio de ferramentas de LC já vem sendo realizada há algumas décadas. Fischer-Starcke (2009) combina métodos quantitativos e qualitativos no campo da estilística e constitui um marco nos estudos que mais tarde seriam consolidados sob a denominação de Estilística de *Corpus*. A pesquisadora analisou as palavras-chave e frases mais frequentes do romance *Pride and Prejudice* (Austen, [1813]2002), quando comparado aos outros romances da autora, e observou a predominância de palavras relacionadas a família, relações familiares e casamento, agrupando-as em campos semânticos. A análise revelou resultados que não só corroboraram o que já havia sido discutido pela crítica literária tradicional, mas também padrões dominantes relacionados a conceitos mentais e emoções, expressões de incerteza, expressões descrevendo atos comunicativos e predominância de prosódia semântica negativa observáveis no contexto de ocorrência das palavras-chave e frases estudadas.

Čermákova (2018) realiza um estudo em EC com os livros de Harry Potter e de Winnie the Pooh e suas traduções para o tcheco. O objetivo da autora é examinar e discutir a repetição nesses textos e suas traduções, uma vez que possuem públicos-alvo distintos, sendo a primeira coletânea voltada ao público infanto-juvenil, já alfabetizado, e a segunda a crianças ainda em processo de alfabetização. Do estudo com corpora paralelos, Čermákova (2018) conclui que palavras-chave e combinações-chave são pontos de partida adequados para identificar redes lexicais que construam o sentido e a coesão do texto literário e que precisam ser levadas em consideração no processo tradutório. A autora observa nos corpora analisados que muitas vezes essas redes lexicais e semânticas dos romances originais acabam alteradas ou mesmo perdidas devido à tentativa de compensação lexical com o uso de sinonímia. A análise estilística é apontada pela autora como uma aliada do tradutor de obras literárias (Čermákova, 2018, p. 130).

Braga (2020) demonstra como a EC pode auxiliar na construção de uma estratégia pré-tradutória ao se integrar a sua metodologia com o *framework* da análise de relevância tradutória de Nord (2009), a fim de se atenderem às especificidades da tradução literária. Isso porque, na tradução desse gênero, é proeminente o papel de certas funções textuais que devem ser consideradas durante o processo tradutório, como a função expressiva, que marca um texto como literário, segundo a autora. Para ilustrar seu argumento, Braga (2020) realiza um estudo estilístico do conto *Sickert at St Peter's*, de Denton Welch (1942). Para isso, construiu um *corpus* de referência com três romances do mesmo autor para verificar que itens lexicais eram salientes na narrativa selecionada, utilizando o *software* livre AntConc (Anthony, 2023) e a função *keyword* a partir da métrica estatística *log-likelihood*, que determina o índice de chavidade de uma palavra em relação ao *corpus* de referência. Ou seja, a ferramenta salienta as palavras estatisticamente mais recorrentes no *corpus* em análise do que no de referência. Após a análise qualitativa dos dados, a pesquisadora constrói um resumo instrucional com observações de estratégias linguísticas para a tradução de narrativas, considerando o *framework* de Nord (2009) e os resultados de EC que Braga (2020) encontrou na análise do conto de Welch (1942). Deste modo, o artigo atesta a efetividade dessa combinação metodológica e aponta essa abordagem como contributiva no desenvolvimento de estratégias pré-tradutórias.

Por sua vez, a pesquisa de Vathanalaotha (2022) se baseia em um *corpus* de dramas americanos e britânicos contemporâneos escritos após a Segunda Guerra Mundial e busca responder aos seguintes questionamentos: que campos semânticos são salientes, que tipos de palavras-chave podem ser vistos nesses campos semânticos e de que modo os campos

semânticos e as palavras-chave do *corpus* se inter-relacionam. Para essa análise, o pesquisador utiliza as ferramentas LancsBox (Brezina; Platt, 2024) e Wmatrix (Rayson, 2009), que realizam o processo de etiquetagem semântica por meio do sistema USAS. Vathanalaotha (2022, p. 59) depreende da pesquisa que o uso das ferramentas auxiliou na decodificação da linguagem literária e caracterização do *corpus* em uma perspectiva holística, permitindo encontrar intersecções entre os personagens dos romances selecionados. Essa análise do *corpus* possibilitou a identificação de um padrão de ênfase no delírio e em ambições dos protagonistas, marcado especialmente por uma prosódia negativa.

Vital (2022), embora não mencione explicitamente que utiliza a abordagem de EC, emprega a metodologia da LC para analisar a obra literária poética *Ave, Palavra* (Rosa, 1985), composta por poemas atribuídos a quatro pseudônimos anagramáticos de João Guimarães Rosa. O objetivo do estudo foi identificar diferenças de estilo entre os quatro pseudônimos, combinando abordagens da LC e do Processamento de Linguagem Natural (PLN). O pesquisador analisou quantidade de palavras, riqueza lexical, número de estrofes e palavras mais frequentes para cada um dos anagramáticos. Do estudo, Vital (2022, p. 11-12) identifica traços estilísticos de cada um desses pseudônimos e apresenta, por meio de nuvens de palavras, um comparativo entre eles. A pesquisa é concluída apontando a contribuição da LC para o estudo de obras literárias e a aproximação interdisciplinar entre Ciências Exatas e Humanas que, segundo o autor, é escassa na produção científica em língua portuguesa.

Como já apontado neste trabalho, uma das possíveis razões para essa escassez encontra-se nas limitações de recursos para a etiquetagem semântica automática do português. A seguir serão apresentados dois dos principais recursos com essa finalidade.

2.2 Etiquetagem semântica automática para a língua portuguesa

A necessidade de etiquetadores semânticos para o português tem impulsionado o desenvolvimento de alguns recursos para esse fim e também o aprimoramento de ferramentas pre-existentes. No contexto dos estudos de Lexicografia e de Semântica Cognitiva, Bick (2022) sugere o sistema de anotação automática Danish Framenet (Bick, 2011), centrado em papéis semânticos desempenhados sintaticamente, para aplicação em textos em língua portuguesa, obtendo resultados interessantes. A ferramenta, denominada PFN-PT, realiza essa etiquetagem por meio de uma rede que organiza o significado das palavras a partir de *frames*, que são estruturas conceituais que representam eventos, situações ou conceitos abstratos, cujo conjunto se denomina FrameNet. O objetivo do PFN-PT é auxiliar no preenchimento da lacuna existente nos anotadores automáticos baseados em regras contextuais, como o PALAVRAS (Bick, 2014), além de prover recursos para auxiliar na desambiguação do sentido de palavras (WSD, na sigla em inglês) e para tarefas de Inteligência Artificial, como tradução automática. O etiquetador adaptado por Bick (2022) alcançou um desempenho satisfatório para a desambiguação do sentido de *frames* verbais, com 92,2% de acerto.

Uma ferramenta que permite anotações semânticas e já está integrada a alguns softwares de LC é a biblioteca PyMUSAS (Rayson; Berridge; Francis, 2004), desenvolvida em Python pela Universidade de Lancaster, que combina estratégias de Aprendizagem de Máquina (AM) para atribuir uma categoria semântica para cada palavra do *corpus* a ser etiquetado. Essa ferramenta foi integrada à interface web Wmatrix (Rayson, 2009), utilizada neste estudo, e será apresentada na próxima subseção e discutida em detalhes na Metodologia.

2.3 Wmatrix na EC

Wmatrix (Rayson, 2009) é uma ferramenta que oferece recursos semelhantes aos de *softwares* comumente utilizados para pesquisas com *corpus*, como Antconc (Anthony, 2023), Wordsmith Tools (Scott, 2024) e Sketch Engine (Kilgarriff et al., 2014), mas com o diferencial de possibilitar funções relacionadas à anotação semântica. Até a versão 5, a ferramenta permitia exclusivamente a anotação semântica de textos em inglês, por meio da interface com o sistema USAS. Wmatrix 6 foi lançado como uma versão beta de testagem fechada, que passou a integrar o etiquetador semântico multilíngue PyMUSAS, que processa e etiqueta textos em outros oito idiomas, entre eles o português². Atualmente, a ferramenta está em sua versão 7, que manteve a anotação semântica multilíngue, com algumas melhorias de interface.

O sistema USAS (Rayson; Berridge; Francis, 2004) foi criado a partir de um esquema de etiquetas semânticas contendo 21 Domínios Semânticos de Nível Superior (DSNS) e 232 etiquetas de subcategorias de campos semânticos específicos. Por exemplo, o DSNS “O corpo e o indivíduo” (B) contém como subcategorias o Domínio Semântico (DS) “Anatomia e fisiologia” (B1) e “Doença” (B2-), entre outras. Essa lista de DS é baseada no *Longman Lexicon of Contemporary English* (McArthur, 1981 apud Rayson; Berridge; Francis, 2004), sendo ampliada com as subcategorias, conforme mencionado. A ferramenta, desenvolvida por linguistas da Universidade de Lancaster, combina diferentes recursos para PLN, como o anotador automático de classe gramatical (POS tagger) CLAWS, dicionários semânticos, listas de modelos e regras contextuais para a desambiguação de sentido.

Esse sistema vem sendo utilizado em diversos estudos das ciências da linguagem e foi aprimorado para uma versão de suporte multilíngue (Piao *et al.*, 2015), o PyMUSAS. Assim como o FrameNet, o PyMUSAS também possui suas limitações em relação à necessidade de bancos de dados robustos para treinamento de AM, resultando em palavras não categorizadas ou etiquetadas erroneamente.

Cabe salientar que os desenvolvedores optaram por criar categorias semânticas nas quais são agrupados itens gramaticais, lexicais e discursivos. Portanto, nem todo DS apresenta necessariamente uma temática, como pode-se observar em subcategorias como “Frequência”, que agrupa advérbios que desempenham essa função (às vezes, frequentemente), “Velocidade: Rápido” (imediatamente, subitamente, de repente) e “Números” (numerais em geral)³. A relação de DS de nível superior completa pode ser observada no Quadro 1, com as categorias traduzidas para o português:

Quadro 1 – USAS - Domínios semânticos de nível superior

A	Termos Abstratos e Gerais	N	Números e Medidas
B	Corpo e o Indivíduo	O	Substâncias, Materiais, Objetos e Equipamentos
C	Artes e Ofícios	P	Educação
E	Ações Emocionais, Estados e Processos	Q	Ações Verbais, Estados e Processos
F	Alimentação e Agropecuária	S	Ações Sociais, Estados e Processos

² Para detalhes sobre o sistema PyMUSAS, ver <https://pypi.org/project/pymusas/>.

³ Para o detalhamento dos domínios semânticos do Wmatrix, recomendamos a leitura de Archer, Wilson e Rayson (2002).

G	Governo e Domínio Público	T	Tempo
H	Arquitetura, Construção, Casas e o Lar	W	Mundo e Nosso Ambiente
I	Dinheiro e Comércio	X	Ações Psicológicas, Estados e Processos
K	Entretenimento, Esportes e Jogos	Y	Ciência e Tecnologia
L	Vida e Seres Vivos	Z	Nomes e Palavras Gramaticais
M	Movimento, Localidades, Viagem e Transporte		

Fonte: Rayson et al. (2004) [tradução própria].

Por fim, outro aspecto da EC que McIntyre e Walker (2019) destacam é a abordagem definida como *corpus-informed stylistics* (estilística informada por *corpus*, em tradução livre), que consiste em usar grandes corpora, como o British National Corpus (BNC), para apoiar a análise estilística de textos ou trechos textuais. Segundo os precursores desse campo de estudos, esses corpora atuam como bancos de dados de linguagem, oferecendo a possibilidade de confirmar, com base em evidências empíricas, uma percepção ou interpretação intuitiva sobre as características estilísticas de um texto ou conjunto de textos, ou de questionar, incentivando uma reavaliação dessa intuição inicial (McIntyre; Walker, 2019, p. 26).

Para que a investigação do *corpus* seja eficaz, mediante os objetivos da pesquisa, é preciso selecionar a ferramenta adequada, em meio a tantas disponíveis. Caso, por exemplo, o pesquisador tenha como objetivo contrapor uma lista de palavras de determinado *corpus* de estudo com a lista extraída de um *corpus* de referência, ele terá à disposição diversas ferramentas, inclusive de acesso gratuito, que cumprem com esse propósito. Contudo, se a intenção for partir de um levantamento de campos semânticos de obras em português, como é o caso deste estudo, o pesquisador não encontrará a mesma disponibilidade, conforme apontado anteriormente.

A ferramenta Wmatrix, bem como o etiquetador USAS, já foram utilizados em diversas pesquisas em EC. Como exemplo, aponta-se o estudo de McIntyre e Walker (2010), que utiliza o Wmatrix para a análise comparativa de duas obras de William Blake: *Songs of Innocence* (Blake, 1789) e *Songs of Experience* (Blake, 1794). Por meio da investigação das diferenças lexicais e semânticas manifestadas em ambas as obras do mesmo autor, o objetivo dos pesquisadores foi compreender de que modo elas diferem.

No mesmo artigo, McIntyre e Walker (2010) também demonstram a aplicabilidade da pesquisa baseada em *corpus* para o estudo de roteiros de filmes *blockbusters*, utilizando metodologicamente a EC para validar algumas observações da crítica especializada audiovisual sobre os papéis de gênero nesses filmes, concluindo que a EC pode cumprir com êxito o objetivo de amparar diferentes tipos de análises em obras literárias e ficcionais a partir de suas ferramentas informatizadas.

Ibrahim (2022) realizou uma pesquisa a fim de explorar os recursos da EC, particularmente o uso da ferramenta Wmatrix (Rayson, 2009), para analisar DS em um *corpus* da obra de Charles Dickens, relatando alguns dos limites e desafios da EC, incluindo a dificuldade de se obterem cópias digitais de textos literários ou o direito de usar os textos disponíveis. Da análise dos resultados, o autor também conclui que a sinergia entre as abordagens quantitativa e qualitativa possibilitadas pela metodologia empregada evidenciam a efetividade das ferramentas de análise de *corpus* para serem incorporadas à estilística tradicional.

A partir do exposto, pode-se verificar a eficácia da ferramenta Wmatrix, bem como de seu etiquetador USAS, para a análise de textos em inglês. Neste estudo, testamos as funcionalidades da ferramenta, bem como do etiquetador semântico PyMUSAS, em um *corpus* literário composto pela obra *O Continente*, de Erico Veríssimo.

2.4 Objeto de estudo: *O Continente* de Erico Veríssimo

Erico Veríssimo (1905-1975) é um escritor consagrado da literatura brasileira e sua obra possui fundamental importância para a cultura gaúcha, especialmente considerando-se o sucesso da trilogia *O Tempo e o Vento* (Veríssimo, 2004), composta por três romances: *O Continente*, *O Retrato* e *O arquipélago*, lançados ao longo da carreira do escritor. Erico Veríssimo não fora bem acolhido pela crítica com seus romances anteriores à trilogia, mas a receptividade do público e a maturidade literária do escritor ficaram nítidas com o lançamento do primeiro livro da saga, *O Continente* (Veríssimo, [1949]2013). concedendo ao autor um grande sucesso e fazendo abrir os olhos dos críticos literários que, até então, consideravam Erico Veríssimo um escritor de literatura inferior, como bem revisa Santos (2019):

Na atualidade, Erico Veríssimo é considerado um dos mais importantes escritores da Literatura Brasileira, tido como um dos grandes nomes do Modernismo. Suas principais obras são continuamente reeditadas e algumas delas foram adaptadas nos últimos 30 anos para a televisão e para o cinema [...]. Mas, se em relação ao público o autor sempre teve muito prestígio, em relação à crítica literária nem sempre foi assim. No início de sua carreira literária, nas décadas de 30 e 40, o escritor sofreu avaliações muito depreciativas de suas obras por parte dos críticos literários de sua época. Essas apreciações críticas negativas do passado deixaram sequelas que por muito tempo acompanharam o romancista (Santos, 2019, p. 100).

O romance *O Continente*, objeto desta pesquisa, desenrola-se como uma vasta saga multigeneracional, concentrando-se na formação e nos destinos entrelaçados das famílias Terra e Cambará na fronteira do Rio Grande do Sul com o atual Uruguai. O início da narrativa remonta aos tempos coloniais, apresentando os jesuítas e a sua missão de catequização, personificada na figura do Padre Alonzo e do índio Pedro, e o choque de culturas entre colonizadores e povos nativos. Em paralelo, a história mergulha na vida árdua da família Terra, a aparição de Pedro Missioneiro na vida da resiliente Ana Terra, cujo romance é brutalmente interrompido quando o pai descobre que sua filha está grávida de Pedro. Além das adversidades do ambiente inóspito, Ana Terra se vê em meio aos conflitos territoriais entre portugueses e castelhanos, acaba sendo estuprada por castelhanos que invadem as terras da família, assassinando todos os homens adultos da casa.

O enredo ganha novos rumos com a mudança de Ana Terra, sua cunhada, sobrinha e o filho para Santa Fé, onde as vidas dos Terra e dos Cambará se cruzam com a chegada na cidade do capitão Rodrigo Cambará, que desposa Bibiana Terra, neta de Ana, após um duelo de facas com Bento Amaral, filho do coronel e neto do fundador da cidade. Os capítulos do romance se desenvolvem com a alternância entre a narrativa do passado de Ana Terra e Pedro Missioneiro, a vida conjugal de Rodrigo Cambará e Bibiana Terra na cidade de Santa Fé, a ida dele à Revolução Farroupilha, sua morte após regressar e atacar o casarão da família Amaral

e o cenário de guerra da revolução federalista já em 1895, que marca o tempo presente da narrativa e em que há o fechamento do primeiro romance da trilogia.

Esta pesquisa elegeu *O Continente* como objeto de estudo devido ao seu papel fundamental no reconhecimento do escritor pela crítica literária da época, ao seu amplo alcance junto ao público, resultando, inclusive, em duas adaptações para a televisão. Por essa razão, muitas vezes o romance primogênito da trilogia é confundido pelo público leigo com o todo da obra. Esse impacto de *O Continente* sobre a crítica também influenciou a quantidade de estudos direcionados a esse livro, possibilitando uma revisão mais ampla sobre as interpretações literárias e sua correlação com os resultados encontrados com a EC.

Nesse contexto, destaca-se, aqui, a análise crítica empenhada por Regina Zilberman (2004a), que escrutina a trilogia supracitada, a partir da qual tece análises contundentes do caráter estilístico da obra, esmiuçando sua estrutura narrativa e revisitando, inclusive, declarações do autor sobre o processo de escrita dos volumes que a compõem. Dessa análise, salientam-se alguns pontos em relação ao primeiro volume, *O Continente* (Veríssimo, 2013 [1949]). Primeiramente, seu caráter de narrativa fechada, mas que, posteriormente, obteve continuidade com a publicação de *O Retrato* (Veríssimo, 1951) e *O Arquipélago* (Veríssimo, 1961). Em segundo lugar, a circularidade narrativa, construída após a finalização da trilogia, que revela o narrador, até então anônimo, conectando a última página de *O Arquipélago* à primeira de *O Continente*. Por fim, um dos recursos estilísticos que permeia a narrativa dos três livros, conforme destacado por Zilberman (2004a, p. 31), é o papel que os objetos desempenham no preenchimento das lacunas entre os episódios da trama, como a tesoura de Ana Terra, que acompanha as gerações de mulheres parteiras, que a utilizam para cortar o cordão umbilical dos recém-nascidos, e o punhal de Pedro Missioneiro, passado de geração para geração aos homens da família Terra-Cambará.

Entre as características da obra de Veríssimo, destaca-se seu caráter histórico e o planejamento do romance pelo autor, que buscou em fontes históricas primárias – notícias de jornais, depoimentos pessoais etc. – a matéria prima de sua feitura (Bordini, 2004, p. 76). A pesquisadora também aponta o dualismo da obra, que marca as contradições da trajetória da família Terra-Cambará e a presença da família Caré como representativa da miséria camponesa e da invisibilidade social. A trajetória de decadência das personagens femininas, conforme os anos avançam no romance, também é observada por Bordini (2004), tanto quanto a tipificação das personagens, marcada pela debilidade e hesitação conforme as gerações avançam em *O Continente* e o livro se encaminha para a conclusão.

Bordini (2004) afirma que o espaço de *O Continente* vem pouco caracterizado em termos físicos e muito mais em termos daqueles que o habitam: “As personagens são descritas com poucos traços, os mais significativos para suscitar as ações que praticam, do mesmo modo que o espaço diegético é apenas sugerido, acentuando seu potencial simbólico (Bordini, 2004, p. 82)

Santos (2005) analisa como *O Continente* desenvolve a temática da revolução federalista e de que modo o autor recorre ao espaço diegético do sobrado para retratar as implicações da guerra civil numa perspectiva familiar. O pesquisador traça um paralelo entre a ação das forças legalistas, vitoriosas do confronto, que fazem o cerco do sobrado da família Cambará, delineando o cenário de guerra, e o papel do patriarca e chefe político Licurgo, que, resistindo à derrota iminente, vê sua família em condições precárias e precisa lidar com os conflitos internos e a insatisfação dos entes aquartelados.

Já Santos (2009) revisita a fortuna crítica literária de *O Tempo e o Vento* para analisar o tratamento dado à natureza da obra. Parte dos críticos pesquisados pelo autor considera *O Continente* um romance histórico, enquanto se refere às sequências (*O Retrato* e *O Arquipélago*) como romances majoritariamente políticos. Por outro lado, também há críticos que se opõem a essa separação e apontam a totalidade dos textos como um romance histórico, independentemente de Erico Veríssimo ter vivido durante a época retratada na continuidade do primeiro volume. Em resumo, após uma revisão da crítica literária favorável e contrária à natureza histórica do romance analisado, o artigo de Santos (2009, p. 103) considera a trilogia *O Tempo e o Vento* em sua unidade um romance histórico, mas, se analisada individualmente, a condição de romance histórico seria atribuída ao primeiro livro e de romance político aos outros dois.

Contextualiza-se, a seguir, um ponto complexo da obra de Erico Veríssimo: a temática do machismo, que, inclusive, levou o autor a precisar se defender na década de 1970 de acusações de que sua obra glorificaria o comportamento que promove a superioridade do homem.

Garcia e Lisboa Filho (2013) realizaram um estudo da obra *O Continente* a fim de identificar as intersecções entre identidade gaúcha e violência no romance. As autoras constataram que a figura masculina na obra, especialmente em personagens como o protagonista Capitão Rodrigo Cambará, é forjada em valores como virilidade, honra e belicosidade, paradoxalmente coexistindo com comportamentos de machismo e irresponsabilidade. Paralelamente, elas destacam a força e a resiliência das personagens femininas, como Ana Terra e Bibiana, que, embora apareçam frequentemente em contextos de submissão, demonstram sua capacidade de superação diante das adversidades. O artigo também sublinha a naturalização e a transmissão transgeracional da violência – inclusive de abusos sexuais, agressões físicas e psicológicas e violência conjugal pelas quais as personagens são afligidas no enredo.

Bisol e Porto (2015), similarmente, fazem uma análise de como a violência é representada no romance *O Continente*, especulam se existe um posicionamento crítico da obra em relação a esse tema e discutem a relação da obra com a construção de uma memória da história do Rio Grande do Sul. Baseando-se em Ginzburg (2013) e Halbwachs (2006), as autoras inferem que a violência e as guerras são realidades naturalizadas entre as personagens. Embora estas, em alguns momentos, reflitam sobre tal condição – especialmente as mulheres e os personagens religiosos do romance –, não há, contudo, uma ação concreta em direção ao enfrentamento ou à mudança desse paradigma. Bisol e Porto (2015, p. 154-155) explicam que há uma visão positiva das personagens, especialmente masculinas, em relação à violência, e a narrativa enfatiza as consequências trágicas das decisões dessas personagens e seu impacto sobre as famílias retratadas no romance, bem como a aceitação da guerra e da violência. Por fim, concluem que o romance contribui para uma reflexão crítica do papel da violência na construção da formação do Rio Grande do Sul, recompondo uma memória coletiva da história que *O Continente* representa.

Alves (2005), por sua vez, analisa em sua dissertação a presença de passagens de *O Continente* em que se destacam os aspectos religiosos do contexto da obra. Nela, um capítulo inteiro é dedicado a contar a história dos Sete Povos das Missões, de onde surge um personagem chave para a trama, Pedro Missioneiro, em torno do qual o misticismo e a religiosidade se configuram no romance.

Rosa (2019) se vale da análise do discurso para avaliar as temáticas salientes do romance a partir da escolha de três temáticas: “o telurismo (a influência da natureza/solo/local no caráter, costumes ou tradição de determinado povo), as relações étnico-raciais (ques-

tões indígenas e negras apresentadas ou não na obra) e de gênero (o papel das mulheres no enredo)” (Rosa, 2019, p. 81). O pesquisador coleta manualmente as passagens da obra que abordam essas temáticas, e, a partir dessa coletânea de excertos relacionados aos temas apontados, realiza sua análise. De seu estudo, conclui-se que *O Continente* oferece vários caminhos de análise, como a representatividade dos negros, o racismo, a questão das mulheres e seu silenciamento no romance.

Um trabalho recente sobre a obra de Erico Veríssimo é o de Jacobi (2023), que aborda a presença da melancolia em *O Continente*, apontando as guerras apresentadas durante o romance como precursora de uma visão fatalista de mundo pelas personagens. O autor aponta que a perspectiva das personagens femininas no romance permite a compreensão da guerra como destruidora e fonte de perdas e decepções, bem como a resignação das mulheres ante a sede dos homens por conflitos e poder.

Dando continuidade ao estudo sobre o impacto dessa obra, e em particular de seu primeiro volume (*O Continente*), busca-se demonstrar como a EC pode contribuir para a análise do estilo presente, avaliando em que medida essa abordagem pode confirmar ou não as interpretações da crítica especializada, além de oferecer novas perspectivas.

3 Metodologia

Este estudo piloto se baseia na pesquisa realizada por Ibrahim (2022), que utilizou a versão 5 da ferramenta Wmatrix, bem como seu etiquetador USAS, para avaliar a eficácia da ferramenta na detecção de temas predominantes em quatro romances de Charles Dickens. O autor conclui que a significância estatística dos domínios semânticos (DS) identificados corroborou os temas relacionados às obras de Dickens pela crítica literária.

Conforme salientado anteriormente, o etiquetador desenvolvido para a língua portuguesa, o PyMUSAS, está disponível apenas a partir da versão 6, que foi beta e já descontinuada e, por fim, a partir de 2025 com a versão definitiva, Wmatrix 7. Para este estudo foi realizado um teste preliminar com um *corpus* de referência (CR) parcial no Wmatrix 6 e a pesquisa foi concluída com os corpora definitivos processados no Wmatrix 7. Além de contribuir com a avaliação das capacidades do Wmatrix para a EC, esta pesquisa também aponta algumas limitações da interface em relação ao processamento de textos em português brasileiro. Abaixo, elencam-se os procedimentos metodológicos utilizados neste estudo.

3.1 O *corpus* de estudo e o *corpus* de referência

O *corpus* de estudo (CE) é a obra que passará por análise de EC. Nesta pesquisa ele é composto pela íntegra da 4ª edição da obra *O Continente* (Veríssimo, [1949]2013). Para possibilitar o processamento automático por meio de software de análise linguística, o texto passou pelos processos de digitalização e limpeza - excluindo-se cabeçalhos, elementos pré-textuais e pós-textuais -, preservando-se apenas o conteúdo literário - e foi convertido para o formato Texto sem Formatação (TXT), codificação UTF-8, mais adequada para a análise de textos em línguas abundantes em diacríticos, como é o caso do português. O CE totaliza 247.326 palavras (*tokens*).

Vale enfatizar que, ao selecionar como *corpus* de estudo apenas *O Continente*, não temos como objetivo traçar o estilo autoral de Erico Veríssimo, uma vez que um único livro não permitiria generalizações dessa natureza (Cf. Evert, 2006). O foco da pesquisa é avaliar o potencial da ferramenta Wmatrix para a análise estilística a partir de campos semânticos, investigando características específicas dessa obra, possibilitando, assim, corroborar ou refutar interpretações oriundas da estilística tradicional.

A comparabilidade entre corpora é um princípio metodológico essencial, e está diretamente ligada à representatividade e à adequação da construção dos corpora aos objetivos específicos da pesquisa (Kilgarriff, 2001). Para salientar o que é característico do CE, é necessário contrastá-lo com um CR. De acordo com Tagnin (2013), o CR, também chamado *corpus* de comparação ou *corpus* de contraste, possibilita o levantamento de uma lista de palavras-chave, ou seja, das palavras estatisticamente mais frequentes no CE do que no CR.

Berber Sardinha (2004) recomenda que o *corpus* de referência (CR) tenha de três a cinco vezes o tamanho do *corpus* de estudo (CE), e que a composição dos textos varie de acordo com o tipo de pesquisa a ser realizada. Contudo, a relação de tamanho entre CE e CR permanece uma questão controversa. Pojanapunya e Todd (2018) ressaltam que o resultado de uma análise de *keywords* depende não apenas da proporção entre os corpora, mas também da escolha do índice estatístico utilizado para calcular a chavicidade.

No caso da interface Wmatrix, o CR evidencia, a partir da etiquetagem com o PyMUSAS, quais DS são salientes no CE em comparação com o CR. Para esta pesquisa, elegeu-se como CR uma compilação de escritores contemporâneos de Erico Veríssimo, pertencentes à mesma época de produção literária. Como critérios de seleção dos autores que comporiam o CR foi considerada sua menção no levantamento sugerido por Fischer (2008, p. 113): escritores do período histórico descrito como “República Pós-30 e modernização econômica; o romance neo-realista dos anos 1930-1950”.

Outro critério adotado foi a busca por representantes de diferentes regiões do país e com publicações acessíveis em meio digital, uma vez que a necessidade de digitalização de todas as obras poderia inviabilizar a construção do CR. O CR coletado possui um total de 117 romances e coletâneas, que passaram pelo mesmo procedimento de limpeza e armazenamento descrito acima para o CE e totaliza 7.146.884 palavras (*tokens*). No Quadro 2 é possível ver a relação completa dos autores e obras que compõem o CR:

Quadro 2 – Relação de obras do *corpus* de referência

Autor	Obras	Tokens
Bernardo Élis	<i>Veranico de Janeiro, Apenas um Violão, Os melhores Contos de Bernardo Elis, O Tronco</i>	237.280
Cornélio Pena	<i>Fronteira, Repouso, A Menina Morta</i>	305.662
Cyro dos Anjos	<i>Romances: Abdias, Montanha, O Amanuense Belmiro.</i>	196.165
Dyonellio Machado	<i>Os Ratos</i>	40.088
Graciliano Ramos	<i>A Terra dos Meninos Pelados, Angústia, Histórias de Alexandre, Infância, Memórias do Cárcere, Os Filhos da Coruja, Vidas Secas</i>	355.917

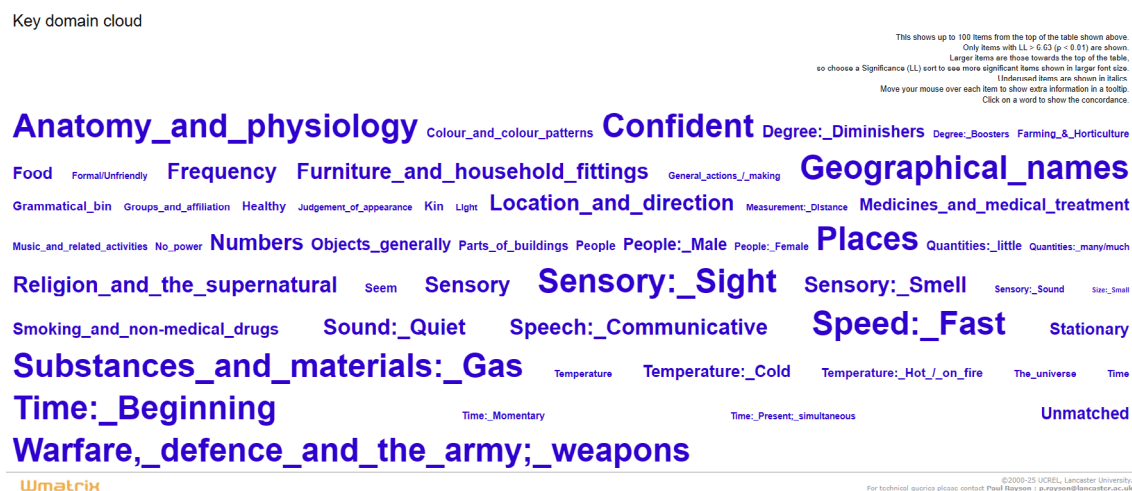
João Guimarães Rosa	<i>Ficção completa: Sagarana, Manuelzão e Miguelim, No Urubuquaquá, no Pinhém, Noites do Sertão, Grande Sertão: Veredas, Primeiras Estórias, Tutameia, Estas Estórias, Ave, palavra</i>	808.796
Jorge Amado	<i>Obra Completa: A Descoberta da América Pelos Turcos, A Morte e a Morte de Quincas Berro D'Água, Cacau, Capitães de Areia, Farda Fardão Camisola de Dormir, Gabriela Cravo e Canela, Jubiabá, Mar Morto, Navegação de Cabotagem, O Capitão de Longo Curso, O Menino Grapiúna, O País do Carnaval, A Luz no Túnel, Agonia da Noite, Os Ásperos Tempos, Seara Vermelha, Suor, Tenda dos Milagres, Tereza Batista Cansada de Guerra, Terras do Sem-fim, Tieta do Agreste, Tocaia Grande</i>	2.024.959
José Américo de Almeida	<i>Obra Completa: A Bagaceira, O Boqueirão</i>	54.876
José Lins do Rego	<i>Obra Completa: Água Mãe, Caminho de Pedras, Cangaceiros, Menino de Engenho, Doidinho, Banguê, Usina, Fogo Morto, Meus Verdes Anos, Pedra Bonita, Pureza, Riacho Doce</i>	920.542
Mário Palmério	<i>Vila dos Confins, Seleta, Chapadão do Bugre</i>	278.218
Murilo Rubião	<i>Obra Completa: O pirotécnico Zacarias, O ex-mágico da Taberna Minhota, Bárbara, A cidade, Ofélia, meu cachimbo e o mar, A flor de vidro, Os dragões, Teleco, o coelhinho, O edifício, O lodo, A fila, A Casa do Girassol Vermelho, Alfredo, Marina, a Intangível, Os Três Nomes de Godofredo, Memórias do Contabilista Pedro Inácio, Bruma (a estrela vermelha), D. José Não Era, A Lua, A Armadilha, O Bloqueio, A Diáspora, O Homem do Boné Cinzento, Mariazinha, Elisa, A Noiva da Casa Azul, O Bom Amigo Batista, Epidólia, Petúnia, Aglaia, O Convidado, Botão-de-rosa, Os Comensais</i>	53.943
Octávio de Faria	<i>Mundos Mortos, O Anjo de Pedra, O Lodo das Ruas, O Retrato da Morte, O Senhor do Mundo, Os Caminhos da Vida, Os Loucos, Os Renegados</i>	1.420.062
Rachel de Queiroz	<i>Obra Completa: As três Marias, Dôra Doralina, João Miguel, Memorial de Maria Moura, O Galo de Ouro, O Quinze</i>	450.376

Fonte: Elaborado pelos autores.

3.2 Processamento do corpus pela ferramenta

Como mencionado anteriormente, a ferramenta Wmatrix 7 (Rayson, 2009) permite a geração de nuvens de DS cujas etiquetas circunscrevem-se na lista de domínios do etiquetador PyMUSAS, apresentada na revisão de literatura. Esse utilitário permite que se ajuste o tamanho da fonte de acordo com a sua significância no CE. Além disso, o usuário pode utilizar um filtro que permite ajustar a visualização da nuvem e da tabela de acordo com seus objetivos. Na Imagem 1, observa-se a nuvem de palavras com o filtro de visualização “incluir itens usados em excesso” e “excluir itens subutilizados”. Esses filtros são selecionados para que a interface não exiba a lista completa de todos os campos semânticos, mas apenas os que são salientes no CE, o que facilita a visualização para a análise de dados do CE:

Imagem 1 – Nuvem de DSs do CE na interface Wmatrix



Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

Juntamente a essa nuvem, é apresentada uma tabela que lista a chavicidade, isto é, o grau de significância estatística (SE) de cada DS no CE em relação ao CR, com os dados estatísticos medidos a partir do índice *log likelihood* (LL). A tabela também disponibiliza os hiperlinks para o acesso às listas de palavras do CE (List1) e do CR (List2), bem como para as linhas de concordância dos itens (Concordance), o código da etiqueta semântica atribuída ao item (Item), a frequência observada no CE (O1), a frequência observada no CR (O2), a frequência relativa do DS em CE e CR (%1 e %2, respectivamente), o valor de SE - *log likelihood* (LL) - e o tamanho de efeito (TE) (%DIFF). De acordo com Gabrielatos (2018), a SE testa a confiabilidade da diferença de frequência entre os corpora comparados, enquanto o TE permite organizar os resultados de acordo com o tamanho da diferença de frequência entre CE e CR.

Na configuração padrão do Wmatrix, são apresentados todos os campos semânticos proeminentes tanto no CE quanto no CR, com a indicação de saliência ou pouco uso nesse contraste por meio dos sinais + e - precedendo o valor de SE (LL) na tabela, informando que determinado DS é mais ou menos saliente no CE do que no CR. A interface permite diversos ajustes, entre eles a possibilidade de ocultar os DS que não são salientes no CE. Esses elementos podem ser observados na Imagem 2, que apresenta os 10 campos semânticos mais representativos do CE, em ordem decrescente de chavicidade (LL):

Imagem 2 – Captura de tela da tabela de DS do CE na interface Wmatrix

			Item	O1	%1	O2	%2	LL	NOIFF		
1	List1	List2	Concordance	01	6091	1.94	93209	1.18 +	1211.70	63.90	Anatomy and physiology
2	List1	List2	Concordance	E6+	304	0.10	1104	0.01 +	600.07	590.62	Confidant
3	List1	List2	Concordance	G5	1615	0.51	22614	0.29 +	422.68	77.54	Marfare, defence and the army; weapons
4	List1	List2	Concordance	M7	1077	0.34	13561	0.17 +	391.51	99.19	Places
5	List1	List2	Concordance	T2+	606	0.19	6813	0.09 +	288.24	123.08	Time: Beginning
6	List1	List2	Concordance	X3.4	622	0.20	7345	0.09 +	264.78	112.39	Sensory: Sight
7	List1	List2	Concordance	H3.0+	506	0.16	6163	0.08 +	209.00	105.92	Speed: Fast
8	List1	List2	Concordance	Z2	731	0.23	18288	0.13 +	193.57	78.21	Geographical names
9	List1	List2	Concordance	O1.3	322	0.10	3273	0.04 +	188.07	146.74	Substances and materials: Gas
10	List1	List2	Concordance	H1	7966	2.54	172084	2.19 +	152.53	15.56	Numbers
11	List1	List2	Concordance	X3.2-	367	0.12	4406	0.05 +	159.23	108.91	Sound: Quiet
12	List1	List2	Concordance	Q2.1	2603	0.83	51014	0.65 +	140.02	27.97	Speech: Communicative
13	List1	List2	Concordance	H6	304	0.10	3512	0.04 +	136.11	117.10	Frequency
14	List1	List2	Concordance	S9	1278	0.41	23571	0.30 +	104.62	35.98	Religion and the supernatural
15	List1	List2	Concordance	H5	647	0.21	18554	0.13 +	99.63	53.75	Furniture and household fittings
16	List1	List2	Concordance	X3	68	0.02	382	0.00 +	91.22	346.46	Sensory
17	List1	List2	Concordance	M6	6998	2.23	158540	1.99 +	84.06	12.12	Location and direction
18	List1	List2	Concordance	X3.5	115	0.04	1027	0.01 +	84.38	180.84	Sensory: Smell
19	List1	List2	Concordance	O2	1548	0.49	38962	0.39 +	82.76	27.87	Objects generally
20	List1	List2	Concordance	F1	1047	0.33	19941	0.25 +	69.46	31.68	Food
21	List1	List2	Concordance	M8	123	0.04	1439	0.02 +	68.63	131.81	Stationary
22	List1	List2	Concordance	O4.6-	27	0.01	68	0.00 +	68.01	895.84	Temperature: Cold
23	List1	List2	Concordance	Z99	65312	20.79	1584837	20.11 +	67.67	3.36	Unmatched
24	List1	List2	Concordance	A13.6	1836	0.58	37730	0.48 +	65.47	22.05	Degree: Dindnshers
25	List1	List2	Concordance	F3	123	0.04	1389	0.02 +	57.94	122.09	Smoking and non-medical drugs
26	List1	List2	Concordance	B3	504	0.16	8766	0.11 +	57.39	44.20	Medicines and medical treatment
27	List1	List2	Concordance	S2.2	1489	0.47	30421	0.39 +	56.14	22.76	People: Male
28	List1	List2	Concordance	H2	1438	0.46	29370	0.37 +	54.37	22.80	Parts of buildings
29	List1	List2	Concordance	O4.6+	293	0.09	4612	0.05 +	52.30	59.34	Temperature: Hot / on fire
30	List1	List2	Concordance	Z5	50194	17.25	1316790	16.71 +	51.70	3.22	Grammatical bin

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).⁴

Durante a análise qualitativa apresentada nos resultados, avaliamos quais campos semânticos são salientes no CE e, de modo geral, representam o romance *O Continente* do ponto de vista temático. Também discutimos se esses resultados corroboram ou não as observações já feitas pela crítica literária. Em seguida, analisamos alguns dos DS encontrados na etapa anterior quanto às palavras-chave que os representam no romance, discutindo possíveis discrepâncias de classificação e interpretando os dados com base no que a crítica literária já identificou ou não em relação ao tema. Por fim, fazemos observações sobre as possíveis limitações da ferramenta, a partir dos dados analisados durante as etapas anteriores.

4 Resultados

4.1 Sobre o que, afinal, é *O Continente*?

Antes de proceder à análise dos dados obtidos com o Wmatrix, reitera-se, como explicado na fundamentação teórica deste estudo, que o etiquetador PyMUSAS atribui DS para itens lexicais e gramaticais. Portanto, nem todas as categorias se constituem como temáticas do CE. A depender dos objetivos da análise estilística, essas categorias podem ter maior ou menor relevância durante a etapa qualitativa da pesquisa. Em nosso estudo piloto, pretendemos nos ater aos DS que apresentam temáticas salientes do romance, quando comparados a outros de autores contemporâneos a Erico Veríssimo. Para gerar os resultados, foi utilizado, como linha de corte, o valor de LL 10.83, que corresponde a um nível de significância de 0.001 (0,1%). Rayson, Berridge e Francis (2004) esclarecem que esse limite mínimo já emite resultados precisos e confiáveis em

⁴ O quadro integral com todos os DS está disponível no Apêndice A deste artigo.

análises estatísticas que utilizam a métrica LL, indicando uma menor probabilidade de que os resultados se devam ao acaso⁵. A seguir, apresentam-se no Quadro 3 os DS salientes do CE comparado ao CR, em tradução literal para facilitar a compreensão dos resultados:

Quadro 3 – dados exportados de subcategorias de DS salientes no CE

Class.	Etiqueta	LL	Subcategorias de DS
1	B1	1211.70	Anatomia e fisiologia
2	E6+	600.07	Confiante
3	G3	422.68	Guerra, defesa e o exército; armas
4	M7	391.51	Lugares
5	T2+	288.24	Tempo: Início
6	X3.4	264.78	Sensorial: Visão
7	N3.8+	200.00	Velocidade: Rápido
8	Z2	193.57	Nomes Geográficos
9	O1.3	188.07	Substâncias e materiais: Gás
10	N1	152.53	Números

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

A partir dos resultados apresentados acima, foram removidos os DS referentes à categoria A (Termos Abstratos e Gerais), que contém, principalmente, palavras gramaticais, como advérbios e preposições, à categoria N, que contém números e medidas, à categoria Q - Ações Verbais, Estados e Processos, que concentra a maior parte dos verbos de elocução, à categoria Z (itens não classificados), e os falsos positivos de outras categorias (que serão discutidos na seção 4.3), de modo a facilitar a apresentação das temáticas salientes no romance analisado. O Quadro 4 apresenta uma compilação dos itens mantidos, cujos DS mostram indícios da temática que constrói a narrativa de *O Continente*:

Quadro 4 – Panorama dos DS temáticos do CE

Domínios Semânticos analisados					
1	B1	Anatomia e fisiologia	17	F3	Fumo e drogas não medicinais
2	G3	Guerra, defesa e o exército; armas	18	B3	Medicina e tratamentos medicinais
3	M7	Lugares	19	S2.2	Pessoas: Masculino
4	T2+	Tempo: início	20	H2	Partes de construções
5	X3.4	Sensorial: Visão	21	O4.6+	Temperatura: quente / em chamas
6	Z2	Nomes Geográficos	22	S2	Pessoas
7	O1.3	Substâncias e materiais: gás	23	O4.3	Cor e padrões de cores
8	X3.2-	Som: Silencioso	24	S4	Parentesco
9	Q2.1	Fala: Comunicativo	25	B2+	Saudável
10	S9	Religião e o Sobrenatural	26	O4.6	Temperatura

5 Para maiores detalhes sobre os limites de corte e sua efetividade nos resultados, recomendamos a leitura de Rayson, Berridge e Francis (2004).

11	H5	Mobília e utensílios domésticos	27	K2	Música e atividades relacionadas
12	X3	Sensorial	28	T1.2	Tempo: Momentâneo
13	X3.5	Sensorial: Olfato	29	S2.1	Pessoas: Feminino
14	O2	Objetos em geral	30	T1	Tempo
15	F1	Comida	31	S5	Grupos e Afiliações
16	O4.6-	Temperatura: frio	32	F4	Agricultura e Horticultura

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

4.2 Domínios Semânticos de *O Continente*: análise

Para realizar uma análise detalhada dos resultados, optou-se por agrupar os DS por suas categorias de nível superior (DSNS), conforme apresentado no Quadro 1 desta pesquisa e cuja organização pode ser visualizada no Apêndice B. Para cada DSNS, foi realizada uma análise dos dados encontrados e suas implicações no entendimento das temáticas de *O Continente*. Os resultados com discrepâncias devido a erros de etiquetagem serão comentados na seção que trata das limitações da ferramenta.

Reitera-se que não foram analisados os DSNS “A - Termos Gerais e Abstratos”, “N - Números e Medidas”, “Q - Ações Verbais, Estados e Processos” e “Z - Nomes e Palavras Gramaticais”. Deste último, a exceção é para a subcategoria “Z2 - Nomes Geográficos”, dada sua relevância para a compreensão da perspectiva histórica do romance, que será discutida junto à categoria “M - Movimento, Localidades, Viagem e Transporte”. As categorias excluídas podem ser investigadas futuramente em um estudo que vise às estratégias estilísticas de natureza gramatical e discursiva, que não são o foco da presente pesquisa.

4.2.1 DSNS: B - O Corpo e o Indivíduo

“Anatomia e fisiologia” (B1) aponta dados interessantes sobre o romance e tem destaque como o DS mais saliente de todo o *corpus*, com LL de 1211.70. “Olhos” ocupa a primeira posição, com 654 ocorrências, seguido por “cabeça” (486), “rosto” (266), “mão” (246), “mãos” (205) e “corpo” (199). Observando-se essas palavras em contexto, é possível constatar que a comunicação gestual e interação física das personagens de Erico Veríssimo é intensa e, de modo geral, elas estão constantemente direcionando sua atenção ou respondendo aos interlocutores com gestos enfáticos e reagindo emocionalmente por meio desses elementos, como se observa nas linhas de concordância “Alice sacode a cabeça dum lado para outro [...]”, “Alonzo meteu a cabeça no regaço materno e desatou o pranto [...]” e “Desdobrou-o com a mão trêmula e apresentou-o ao dono da casa [...]”.

Ainda sobre a categoria “Anatomia e fisiologia”, cabe uma ponderação em relação ao romance de (Veríssimo, 2013 [1949]).: há um total de 261 termos com uma frequência relativa, isto é, a proporção de ocorrências de uma palavra em relação ao total de palavras de um *corpus*, de 1,94%, e observa-se uma diversidade de palavras maior do que aquelas relacionadas ao espaço físico. Isso fica perceptível ao se analisarem as categorias M7 - Lugares (77; 0,34%), H2 - Partes de construções (66; 0,46%), H5 - Móveis e utensílios domésticos (46; 0,21%), corroborando a observação de Bordini e Zilberman (2004) de que Veríssimo trabalha pouco o espaço

diegético na trama, centrando seu estilo literário no desenvolvimento das personagens. Essa estratégia estilística, porém, é pouco explorada diretamente na crítica literária sobre o autor.

A segunda subcategoria mais saliente do grupo é “Medicina e tratamentos medicinais” (B3) e possui na primeira posição a palavra “doutor”, com 187 ocorrências, seguida por “médico” (157). A proeminência desse DS se justifica pela presença de um personagem secundário relevante para a trama, o médico alemão Dr. Winter, sendo que a maior parte das ocorrências das palavras coocorre com seu nome ou a ele se refere. A palavra “remédio” (30) ocorre na sequência, seguida por “cinta” (24) e “hospital” (10), que surge no início do romance nas reduções jesuíticas e na passagem pelo dr. Winter por Hamburgo, antes de sua chegada em Santa Fé, onde se tornou médico familiar e amigo dos Terra Cambará, como se observa em “Faço as minhas visitas quase diárias, como médico que sou da casa” e “Estou aqui não só como seu amigo, mas principalmente como médico da casa”. Ao se analisarem as coocorrências de “cinta” (24), observa-se um erro de classificação, pois todas as linhas de concordância estão relacionadas a um acessório em que os homens prendem suas armas, como se observa na Imagem 3:

Imagem 3 – Captura de tela das primeiras dez linhas de concordância do termo “cinta”

Line	Left context	Key	Right context	File
1	– Sim , eu o levava à	cinta	. O cura deu uma palmada na	CTN1-1
2	, experimentava-lhe a ponta , punha-o na	cinta	e imaginava-se um guerreiro como o corregedor	CTN1-1
3	penacho azul e amarelo , espadim à	cinta	e pés descalços . Os famosos Dragões	CTN1-1
4	de couro na cabeça , facão na	cinta	, veem os açorianos suando ao sol	CTN1-1
5	Deus que me mandou . Tira da	cinta	a faca , aproxima-a do pescoço de	CTN1-1
6	com um penacho , e trazia à	cinta	um espadagão e duas pistolas . E	CTN1-2
7	Examinou-a por alguns instantes , pô-la à	cinta	, ergueu-se e , sem dizer palavra	CTN1-2
8	tiracolo e o facão de mato à	cinta	. Decerto ele acampara ali numa noite	CTN1-2
9	o punhal de prata que trazia à	cinta	. Ana voltou para casa com a	CTN1-2
10	, segurando o punhal que trazia à	cinta	. Eulália perdera a fala , e	CTN1-2

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

4.2.2 DSNS: E - Ações Emocionais, Estados e Processos

A subcategoria “Confiante” (E6+) constitui-se como falso positivo, com a palavra “fé” ocorrendo 300 vezes, seguida apenas por “confidências” (2), “mente” (1) e “crenças” (1). Trata-se, todavia, de um erro de etiquetagem, pois a cidade fictícia em que se passa boa parte da trama chama-se Santa Fé, e, pelo fato de o anotador semântico não reconhecer nomes compostos, tratou todas as ocorrências da palavra como E6+.

4.2.3 DSNS: F - Alimentação e Agropecuária

A subcategoria “Comida” (F1) reflete uma diversidade de alimentos, frutas e pratos típicos do Rio Grande do Sul, como “churrasco”, “guisado”, “massa”, “bergamota”, “compota”, “bolacha” e “salame”, além de nomes de frutas, verduras e outros termos relacionados à alimentação. “Fumar e drogas não medicinais” (F3) revela um comportamento habitual das personagens masculinas no romance: o tabagismo. O termo “cigarro”, que figura em primeiro lugar, traz 80 ocorrências, todas envolvendo personagens masculinas. Há apenas uma ocorrência de “fumar” (14) e de “fumava” (8) em que é uma mulher o sujeito da ação : “— gritou para a Paraguaia, que fumava a um canto seu cachimbo de barro” e “A Paraguaia continuou a fumar, ouvindo agora os ruídos que vinham [...]”. Por fim, a categoria “Agricultura e horticultura”

(F4) apresenta o aspecto rural do romance, sendo o termo de maior ocorrência “terra” (213), seguido por “campos” (73), “campo” (71), “gado” (62), “lavoura” (37) e “lavouras” (18).

4.2.4 DSNS: G - Governo e Domínio Público

A saliência de “Guerra, defesa e o exército; armas” (G3) corrobora a perspectiva da crítica literária sobre o papel das guerras e seu impacto no romance, bem como sua responsabilidade na caracterização de *O Continente* como um romance histórico (Santos, 2009). Destacam-se aqui também as considerações de Santos (2005) sobre essa temática, que retrata as condições precárias causadas pelo cerco militar ao sobrado dos Cambará, ocasionando uma mistura entre espaço doméstico e palco de guerra. O DS conta com 111 termos e uma significância de 0.51%, ocupando a terceira posição na classificação geral dos DS salientes do CE. A palavra “guerra” lidera a categoria com 259 ocorrências, seguida de termos que designam as patentes das personagens - “capitão” (201), “coronel” (82) e “major” (76) -, seguidos de uma grande diversidade de termos que remetem ao confronto bélico, como “soldados”, “tropas”, “fortes” e palavras em geral que designam armamentos e agentes do conflito.

4.2.5 DSNS: H - Arquitetura, Construção, Casas e o Lar

Abarcam termos relacionados à construção do cenário as categorias “Móveis e utensílios domésticos” (H5), com 46 termos e “Partes de Edifícios” (H2), com 67. Observa-se em H5 o termo “cama” (172) na primeira posição, seguido de “mesa” (151) e “cadeira” (110). Analogamente, em H2 aparecem os principais ambientes onde se passa a trama: “quarto” (151), “sala” (122) e “janela(s)” (166). Considerando que a maior parte da trama se desenvolve no sobrado em que as personagens estão confinadas, a limitação do espaço diegético da trama por esses ambientes é compreensível.

4.2.6 DSNS: K - Entretenimento, Esportes e Jogos

“Música e atividades relacionadas” (K2) é a subcategoria saliente deste DS, com LL de 28.18 e aponta o papel que a música desempenha na vida das personagens de *O Continente*, especialmente na primeira parte do romance, que narra os tempos das reduções jesuíticas, cujo uso da música se mostra uma indispensável ponte entre o povo indígena e os colonizadores no romance. Isso pode ser observado nas linhas de concordância de “música” (84 ocorrências), palavra mais saliente do DS: “Aqueles índios amavam a música[...]”, “No princípio a música fora a linguagem por meio da qual padres e índios se entendiam.” e “Por meio da música os jesuítas induziam os índios ao estudo [...]”. Também é pelo uso da música que Pedro Missioneiro se aproxima de Ana Terra e aos poucos a conquista: “De repente Ana Terra descobriu que aquela música estava exprimindo toda a tristeza que lhe vinha” e “Aquele música saía do corpo de Pedro e entrava no corpo dela... Oh!”.

Assim como mencionado no encontro de Pedro Missioneiro com Ana Terra, observa-se que a música é um recurso estilístico recorrente de Erico Veríssimo para conduzir o estado emocional das personagens. Analisando todas as ocorrências de “música” também é possível observar que a sua presença estabelece um elo geracional entre vários personagens da história dos Terra Cambará, pois é com um violão nas costas que chega à Santa Fé o capitão Rodrigo e

com o instrumento ele se aproxima dos cidadãos: “E aquela noite as gentes de Santa Fé ouviram música de violão na casa de Nicolau”, também é tocando cítara que Luzia inquire seu filho Licurgo quanto ao tamanho de seu amor por ela: “A música doce envolvia Licurgo, que se imaginava no Angico olhando o pôr-do-sol.” e a banda marcial criada na cidade após os acontecimentos da revolução farroupilha também exerce um poder quase hipnótico sobre Licurgo: “Aos compassos vibrantes da música Curgo teve, mau grado seu, um estremecimento de entusiasmo”.

Além do exposto, a música também é utilizada para caracterizar os homens gaúchos, seu estilo de vida e relação com a arte: “[...] olhavam com repugnada desconfiança para os que se preocupavam com poesia, pintura ou certo tipo de música que não fossem as toadas monótonas de seus gaiteiros e violeiros.”, “Foi também na guerra que Chiru pela primeira vez na vida ouviu uma banda de música.”, “A fala deles tem música e é doce como laranja madura e meio parecida com a nossa.” e “As gaitas enchiam os salões com sua música rasgada e chorona.”.

4.2.7 DSNS: M - Movimento, Localidades, Viagem e Transporte e Subcategoria: Z2 - Nomes Geográficos

A subcategoria “Lugares” (M7) ocupa a quarta posição na classificação geral dos DS e possui 77 termos, tendo como palavra de maior impacto “praça” (130), local de grande importância para o desenvolvimento da trama de *O Continente*, uma vez que fica de frente para o sobrado e é onde se encontram as personagens antagonistas da família cercada. Na sequência aparecem os termos “província” (121), “vila” (106), “cidade” (81), “cemitério” (80), “rancho” (63) e “povoado” (61). A subcategoria “Nomes geográficos” (Z2) tem grande saliência no *corpus* devido à alta frequência das palavras “índios” (92), “índio” (85), “alemão” (43) e “Brasil” (40). Essas duas categorias são adequadas para demonstrar como se constrói o cenário em que a trama se passa e reflete cada um dos aspectos históricos do romance, que narra desde as missões jesuíticas, a vida no campo e os conflitos da cidade marcada pelo confronto político e a chegada de imigrantes no *O Continente* de São Pedro. M6 (Localização) tem impacto nos resultados devido à alta ocorrência da preposição “a” (3007), “sem” (439), “onde” (368) e “lá” (322), cujo análise foge ao escopo desta pesquisa por sua característica morfosintática.

Já “Estacionário (M8)” tem como palavras mais salientes “sentado” (57), “imóvel” (30) e “parado” (10). A palavra “sentado” é utilizada como recurso cênico em diversos momentos da narrativa a fim de indicar o estado emocional das personagens, como se observa nas linhas de concordância a seguir: “Sentado no seu canto, o velho Florêncio Terra está imóvel, de cabeça baixa”, “[...] Pedro, que continuava sentado, imóvel [...]”, “Ao entrar encontrou-o sentado, encurvado sobre a mesa, com a cabeça metida nos braços, soluçando como uma criança”, “Os homens estavam sentados em silêncio” e “Sentado numa cadeira, pitando tranquilamente seu cigarro de palha [...]”.

4.2.8 DSNS: O - Substâncias, Materiais, Objetos e Equipamentos

“Substâncias e materiais: Gás” (O1.3) é o DS de maior saliência do grupo e ocupa a nona posição geral dos DS, tendo “ar” (236) como a palavra mais recorrente, seguida de “fumo” (40) e “fumaça” (19). Há uma diversidade polissêmica notável no uso de “ar”, que não aparece apenas em seu sentido literal de “gás” ou “atmosfera”. A palavra em questão é utilizada como elemento descritivo para cenários, estados emocionais e também ações. Um dos usos mais

proeminentes está na descrição do ambiente físico (23 ocorrências), que surge relacionado à temperatura (“[...] o ar estava frio e úmido”, “[...] ar morno ao sol e fresco à sombra”) e à qualidade ambiental (“[...] ar límpido”, “[...] penumbra leitosa azulava o ar”). Além disso, também está presente em muitas descrições sensoriais (50) (“[...] o ar cheirava a sereno”, “[...] o ar cheirava a incenso”, “[...] um ar enfumaçado”) e sonoras (“o ar enchia-se de sinos e das vozes”, “[...] sinos subiam no ar”). Também constam nas linhas de concordância o uso da palavra em ações dinâmicas (79) (“[...] ergueu-o dramaticamente no ar [...]”, “[...] moscas voavam no ar pesado.”) e no sentido figurado (23) (“[...] palavras morreram no ar [...]” e “[...] de pernas pro ar.”). Por fim, é importante observar também o uso de “ar” para expressão de caráter e comportamento (53 ocorrências) das personagens, que se observa na incidência recorrente da construção sintática “com ar + adjetivo/substantivo”, delineando uma gama de emoções humanas: “ar palerma”, “ar céptico”, “ar indeciso”, “ar fúnebre”, “ar sentencioso”, “ar cansado”, “ar desamparado”, “ar agressivo”, “ar sonolento”, “ar trocista” etc.

“Temperatura” (O4.6) e “Temperatura: frio” (O4.6-) surgem entre os DS salientes com pouca diversidade de palavras, sendo as mais relevantes “morno” (22) para o primeiro e “geada” (20) para o segundo. “Temperatura: quente / em chamas” (O4.6+) tem “fogo” (109), seguido por “quente” (48) e “calor” (31). O uso de “fogo” traz a maioria das ocorrências no sentido literal: “[...] melhor é ir para junto do fogo [...]” e “[...] acender o fogo para aquecer a água do chimarrão.” são exemplos desse uso. Também há o uso como metáfora/ sentido figurado, observável em “aquecido pelo fogo numa raiva intensa.”, “[...] orelhas em fogo [...]” e “Um fogo ardia no peito de Rodrigo.”. As demais ocorrências da palavra surgem relacionadas a “arma de fogo/ disparo” e seriam classificadas de forma mais adequada como G3 (Guerra, defesa e o exército; armas): “[...] com a arma de fogo [...]”, “[...] faça fogo sem piedade.” e “Bolívar fez fogo e Dentinho de Ouro tombou de joelhos.”.

A subcategoria “Cor e padrões de cores” (O4.3) é um domínio semântico sobre o qual cabem algumas observações relevantes. O termo “Pardo” (94 ocorrências), de maior frequência da categoria, na maioria das entradas se refere a nome próprio, que deveria constar na categoria Z1. Na sequência temos o termo “vermelho” (51), extremamente significativo no contexto histórico do romance, uma vez que representa uma das facções políticas (os maragatos) envolvidas no cerne do conflito da trama e coocorre principalmente com a palavra “lenço” ou aparece em contextos associados a esse acessório - na cultura dos gaúchos, o lenço é parte integrante de sua indumentária, e, no contexto da revolução federalista, sua cor reflete a ideologia política do usuário.

Outro ponto relevante nessa mesma categoria é em relação às palavras “branca” (47) e “branco” (47), que são pouco utilizadas para a descrição étnica das personagens. A primeira vem, predominantemente, acompanhada de “bandeira”, “arma” ou “marmelada” e, das 47 ocorrências de “branco”, em apenas três a palavra se refere à etnia. Já os termos “negro(s)”, “negra(s)” e “preto(s)” surgem em grande maioria para designar personagens secundários. O termo “negro”, por exemplo, designa etnia em 75 das 91 ocorrências. Na maior parte das situações, se refere a personagens negras anônimas e que passam por situações ora de violência, ora de discriminação, ora de objetificação. Isso também é corroborado pela saliência do DS “Sem poder”, cujas palavras mais proeminentes são “escravos” (64) e “escravo” (31). O estudo desses DS corrobora a análise do discurso que Rosa (2019) realiza em sua dissertação sobre presença da temática do racismo no romance.

4.2.9 DSNS: S - Ações Sociais, Estados e Processos

“Religião e o sobrenatural” (S9) é a subcategoria com maior proeminência no grupo S e revela, em seus 109 termos, aspectos do romance *O Continente* relacionados ao seu contexto histórico religioso, objeto de estudo da crítica literária de Alves (2005). Alguns dos capítulos do romance ocorrem no período das missões jesuíticas no Rio Grande do Sul, e indícios dessa temática podem ser inferidos pelos termos “deus” (291), “igreja” (135), “céu” (100), “diabo” (76) e “alma” (65), que ocupam as cinco primeiras posições de S9. Os resultados dessa subcategoria também indicam a prevalência do cristianismo na obra.

“Objetos em geral” (O2) ocupa a segunda posição dentro da categoria; porém, levanta dúvidas sobre a prevalência de seus três primeiros termos: “coisas” (280), “amigo” (136) e “amigos” (69), isso porque a primeira palavra possui apenas 26 ocorrências relacionadas a objetos em geral, tal qual em “[...] as coisas do rancho [...]” e “[...] as coisas de comer[...]”, todas as demais são usos abstratos: “[...] aquelas coisas tristes [...]” e “[...] as coisas que sentiam antes”. A presença de “amigo” em “Objetos em geral” também é incoerente, já que existe o DS “Grupos e afiliações” (S5), que, a propósito, também é saliente no *corpus*. Porém, apresenta inconsistências de classificação e só contém as palavras contíguo (5) e contígua (4), ambas designando cômodos de uma casa: “quarto contíguo” e “sala contígua”.

A subcategoria Pessoas (S2) possui como palavra com maior frequência “filho” (317). Todavia, as demais palavras relacionadas à família foram etiquetadas na categoria Parentesco (S4) (“filha” e “filhos”, por exemplo, estão nessa subcategoria). Não foi possível compreender por que o anotador separou a palavra “filho” das demais que se referem a laços familiares com etiqueta S4.

A subcategoria “Pessoas: masculino” (S2.2) demonstra a prevalência das personagens masculinas na obra, com “homem” em primeiro lugar (579) e “homens” (261) na sequência. Comparativamente, observa-se uma menor quantidade de ocorrências das palavras “mulher” (328) e “mulheres” (165), ambas da subcategoria “Pessoas: feminino” (S2.1), que também é saliente no *corpus*. Além disso, uma análise das linhas de concordância de S2.2 corrobora a impressão da crítica especializada em relação à caracterização das personagens masculinas, como a de Santos (2005) e Rosa (2019), que comentam a temática do machismo que perpassa os capítulos de *O Continente*.

A obra traz para o primeiro plano a imposição da sociedade sobre o comportamento das personagens masculinas e da idealização do homem gaúcho, destemido, bruto e violento. Isso pode ser observado nas linhas de concordância dos termos de S2.2, como em “Um homem bem macho não chora nunca [...]”.

Ao se analisarem as linhas de concordância dos termos “menino” (143) e “rapaz” (95), é possível notar que esses não surgem em contextos de afirmação da masculinidade, como observado nos exemplos acima. O comportamento das personagens referidas por esses termos demonstra traços como pureza, educação e imaturidade. Como se observa em “[...] o rosto do menino tinha uma pureza de imagem.”, “Ele era um menino bom que não fazia mal pra ninguém.”, “Talvez o rapaz houvesse falsificado a firma do pai [...]” e “O rapaz respondeu com um sorriso meio constrangido”.

Por fim, é importante observar que o termo “macho” (35) revela nas linhas de concordância seu uso para afirmação da masculinidade das personagens e idealização do que é “ser

homem”, como se observa na Imagem 4. Esses resultados da ferramenta já dão indícios sobre a construção de um cenário de masculinidades exacerbadas:

Imagem 4 – captura de tela das linhas de concordância de “macho”

Line	Left context	Key	Right context	File
1	o último momento . “ Lírio é	macho	” , murmurou Liroca para si mesmo	CTN1-1
2	para si mesmo . “ Lírio é	macho	. ” Sempre que ia entrar num	CTN1-1
3	repetia estas palavras : “ Lírio é	macho	” . Levantou-se devagarinho , apertando a	CTN1-1
4	passos . Cinco segundos . Lírio é	macho	, Lírio é macho . José Lírio	CTN1-1
5	. Lírio é macho , Lírio é	macho	. José Lírio continuava imóvel , olhando	CTN1-1
6	. Pontaria , Liroca . Lírio é	macho	. Vamos . É Ámete bala .	CTN1-1
7	é negócio de mulher . É de	macho	. Maria Valéria abranda um pouco a	CTN1-1
8	reprimir as lágrimas . Um homem bem	macho	não chora nunca , haja o que	CTN1-1
9	vergonha ! O que ela queria era	macho	. E pensava em Pedro só porque	CTN1-2
10	ou para negro . Um homem bem	macho	devia saber manejar a espada , a	CTN1-2
11	a nuca , a bela cabeça de	macho	altivamente erguida , e aquele seu olhar	CTN1-2
12	– Por aqui hai também muito homem	macho	. Houve um silêncio desconfado . Juvenal	CTN1-2
13	podia jurar que nunca vira cara de	macho	mais insinuante . Os cabelos do capitão	CTN1-2
14	. Há um ditado : “ Cambará	macho	não morre na cama” . E ao	CTN1-2
15	lhe digo . – Nunca nenhum Cambará	macho	conseguiu passar dos cinquenta anos . Para	CTN1-2

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

A subcategoria Parentesco (S4) revela um traço característico da própria concepção de *O Continente* e da trilogia “O Tempo e o Vento”: a importância da família, já que toda a trama do romance se circunscreve em torno das linhagens dos Terra e dos Cambará e de seus descendentes e ascendentes, conflitos e dramas familiares em meio ao contexto histórico, algo bem observado na crítica literária de Bordini e Zilberman (2004) e Santos (2005). Ressalta-se que essa subcategoria teria um tamanho de efeito muito maior se as ocorrências do termo “filho” também tivessem a etiqueta S4, devido à alta incidência no romance.

4.2.10 T - Tempo

O DSNS T apresenta palavras relacionadas à passagem do tempo de maneira geral e à ideia abstrata de tempo como conceito. Em *O Continente*, são predominantes as subcategorias “Tempo” (T1), “Tempo: início” (T2+), “Tempo: momentâneo” (T1.2) e “Tempo: presente; simultâneo” (T1.1.2). A presença das subcategorias T1.1.2, T1.2 e T2+ entre os DS mais salientes está relacionada à marcação e passagem temporal na narrativa por meio de advérbios, locuções adverbiais e verbos, não sendo objeto do presente estudo, que é a análise temática dos campos semânticos.

A subcategoria “Tempo” (T1) possui entre o maior número de ocorrências os advérbios “nunca” (316) e “vezes” (158), ambos também utilizados para marcação temporal no romance. O maior número de ocorrências dessa subcategoria, porém, vai para o substantivo “tempo” (358). Dessas ocorrências destacam-se algumas considerações: ordenando as ocorrências na ferramenta pela palavra imediatamente à esquerda de “tempo” constatou-se que a locução adverbial “(ao) mesmo tempo” é o uso predominante da palavra com 73 recorrências expressando simultaneidade, seguido pelo uso nominal “o tempo” (58) e “algum tempo” (45).

Enquanto “mesmo tempo” e “algum tempo” desempenham exclusivamente a função acessória de adjunto adverbial nas linhas de concordância, como em “[...] cansada da corrida e ao mesmo tempo surpreendida de ter vindo” e em “Por algum tempo avistou as ruínas do rancho, [...]” é possível observar nas ocorrências nominais de “o tempo” um uso bastante diversificado do ponto de vista semântico, que será explorado a seguir.

O uso predominante de “o tempo” é o de sua principal acepção dicionarizada: “Aquilo que é medido em horas, dias, meses ou anos; período; duração” (Aulete, 2024), totalizando 47 ocorrências, como em “Como o tempo custa a passar quando a gente espera [...]”, sendo 12 desses resultados relacionados ao instante em que algo ocorre, como em “[...] durante todo o tempo das danças [...]” e “ficava o tempo todo como que sobre brasas [...]”. Cabe ressaltar também que, dentre o total mencionado, há 10 linhas de concordância que chamam a atenção por conterem prosopopeia, um recurso da função poética da linguagem comumente utilizado na literatura em que um objeto ou conceito abstrato é personificado. São exemplos: “Parece que o vento maneia o tempo”; “ele tinha a impressão de ver o tempo parado sobre os telhados [...]” e “E o tempo continuava a andar num tranco lento [...]” (nossa ênfase).

Há três registros como “Época, lapso de tempo futuro ou passado” (Aulete, 2024): “Recordou o tempo em que ela era menina [...]”. Além desses, um dos usos tem a acepção de “Estação” – “[...] pois o inverno era o tempo que mais custava [...]” e mais quatro como “Estado meteorológico”: “[...] outras maneiras de um campeiro prever o tempo sem precisar olhar naquelas geringonças [...]”. Por fim, há três ocorrências em que é utilizado o sentido figurado: “[...] puxou a adaga. ‘Fechou o tempo’, contava Fandango” e “como o tempo é remédio que cura tudo”.

A saliência desse DSNS ilustra a relevância que o tempo possui na narrativa, pois nesses resultados é possível visualizar de que modo esse conceito é explorado pelo romancista, como ele é utilizado linguisticamente e também analisar como as personagens da trama são impactadas por esse tema: percebido ora veloz e incontrolável – “Como o tempo voa – refletiu Rodrigo” –, ora irremediavelmente lento – “Na venda o tempo se arrastava como lesma”; em alguns momentos fonte de angústias – “Mas por que será que o tempo custa tanto a passar quando há guerra”; em outros, a solução de todos os males – “Mas, como o tempo é remédio que cura tudo [...]”.

4.2.11 DSNS: X - Ações Psicológicas, Estados e Processos

Esse DSNS demonstra algumas das estratégias estilísticas do autor para o desenvolvimento das personagens, uma vez que é a partir desse domínio que são expressas as percepções sensoriais e ações mentais. Agrupam-se nesse domínio as subcategorias a seguir, com seus respectivos resultados: “Som: Silêncio” (X3.2-) - com o termo “silêncio” (219) como o de maior ocorrência, seguido por “calado” (33); “Sensorial: Visão” (X3.4), com “olhando” (140) em primeiro lugar, seguido de “olhou” (131), “olhava” (103) e “olhar” (100); em “Objeto mental: Objeto conceitual” (X4.1), há o termo “ideia” (78), seguido por “pensamentos” (57) e “pensamento” (47); “Sensorial” (X3) tem apenas “sensação” (67) e “sensações” (1) como termos; por fim, “Sensorial: Olfato” (X3.5) traz “cheiro” (104) como termo mais saliente e “Sensorial: sonoro” (X3.2) apresenta “som” (29), “sons” (13) e “batidas” (11), enquanto a subcategoria “Ações psicológicas, estados e processos” (X1) apresenta “mente” (34), seguida por instinto (4) e mentes (1).

Conclui-se, a partir de uma visão geral a respeito da utilização desse DSNS, que é relevante para a caracterização das personagens do romance a apresentação das suas percepções sensoriais, impressões, pensamentos e ideias. Uma análise aprofundada das linhas de concordância de alguns dos termos sinalizados acima pode também corroborar as considerações de Jacobi (2023) a respeito da melancolia que permeia a obra, uma vez que, na categoria X3, por exemplo, observa-se a prevalência de uma prosódia semântica negativa em torno do termo “sensação”, como em “[...] uma sensação de desalento gelado a invadiu [...]”.

5 Discussão

Uma análise dos resultados permite inferir algumas características do estilo literário de Erico Veríssimo: pode-se observar o uso dos sentidos (olfato, visão, audição, tato) como recursos estilísticos que realçam diferentes percepções das personagens e também estabelecem o tom e a ambientação da narrativa, como se conclui pela saliência desses domínios semânticos na obra. Esse resultado apresenta uma nova perspectiva sobre o estilo literário presente no romance, uma vez que esses elementos não foram destacados na crítica de *O Continente*.

Outro fator que não teve proeminência na crítica literária é o papel não casual desempenhado pela música em diferentes momentos do romance, como observado ao se analisar o DS K2 “Música e atividades relacionadas” e suas linhas de concordância. O uso polissêmico como estratégia estilística que o autor faz das palavras, como observado nas análises de “ar” e “fogo” também se revela um aspecto pouco explorado nos estudos de *O Continente*.

A conformação do romance do ponto de vista temático, histórico e regional está presente na saliência dos DS “Tempo”, “Lugares”, “Objetos em geral”, “Nomes geográficos”, “Religião e o sobrenatural”, “Cores e padrões de cor”, “Móveis e acessórios domésticos”, além, é claro, da proeminência dos DS “Guerra, defesa e o exército; armas” e “Formal/não amigável” e endossa os estudos do romance como Zilberman (2004b). A análise dos DS relativos aos conflitos históricos retratados na trama, além do estudo das palavras caracterizadoras da violência, confronto e inimizades em *O Continente* endossam a visão de Garcia e Lisboa Filho (2013) e de Bisol e Porto (2015) sobre a naturalização da violência no romance.

Além disso, a prevalência de “Ações, estados e processos psicológicos” e “Ações, estados e processos emocionais gerais” corrobora a percepção de Bordini (2004) de que o romance enfoca a caracterização psicológica das personagens em detrimento de suas descrições físicas e do espaço diegético.

Há destaques, além disso, para o DS “Pessoas: Masculino”, indicando uma prevalência desse gênero na trama, e “Estacionário”, que demonstra a postura e linguagem corporal dos personagens ao longo das cenas, bem como “Fumar e drogas não medicinais” revela o comportamento predominante desses personagens gaúchos, enquanto os DS “Pessoas” e “Parentesco” sinalizam as relações familiares do romance.

Os resultados encontrados com o Wmatrix permitem uma exploração detalhada de cada um dos DS indicados como salientes, uma vez que é possível verificar que palavras pertencem a cada DS e também analisá-las em seu contexto por meio das linhas de concordância e conferir falsos positivos e possíveis inconsistências dos resultados. Na seção a seguir, pretende-se explorar alguns dos DS em relação às suas palavras-chave e seu papel na construção do estilo literário de *O Continente* e também tecer possíveis correlações com a crítica estabelecida da obra.

5.1 Wmatrix: Limitações da ferramenta

O primeiro fato a se ponderar, e esse diz respeito a uma das atuais limitações do sistema PyMUSAS na anotação semântica do português, é que ele depende do léxico (anotado manualmente) com o qual é treinado para uma etiquetagem efetiva. No caso do léxico para o portu-

guês, que foi feito a partir de tradução automática do léxico da versão inglesa, foi verificado na fonte do etiquetador que há anotações inconsistentes, que não correspondem ao modo como a língua portuguesa funciona, especialmente quanto às anotações morfossintáticas, que fazem parte da etapa automatizada, necessária para que o código do anotador analise o sentido das palavras e determine sua classificação. Esse fato causa dificuldades ao anotador para desambiguar palavras polissêmicas, lidar com nomes próprios e palavras compostas, por exemplo.

Observou-se, nesta pesquisa, que o anotador falhou com a desambiguação de algumas palavras, classificando-as equivocadamente, desconsiderando seus contextos de uso. Como exemplo, é possível observar a subcategoria “Saudável” (B2+), que revelou um falso positivo decorrente do erro de desambiguação da palavra “cura”, que ocorre 45 vezes, sendo apenas 3 com relação à saúde. Em 40 contextos, a palavra é utilizada pelo narrador para se referir a uma personagem religiosa nos primeiros capítulos do romance, conforme demonstrado nas linhas de concordância apresentadas na amostra da Imagem 5:

Imagem 5 – Captura de tela das primeiras dez linhas de concordância do termo “cura”

Line	Left context	Key	Right context	File
1	acesas . Preciso contar meu sonho ao	cura	– decidiu ele . E entrou no	CTN1-1
2	tempo em meditação . Por fim o	cura	ergueu-se , e Alonzo fez o mesmo	CTN1-1
3	velas e das lamparinas o rosto do	cura	tinha um tom alaranjado . Era uma	CTN1-1
4	nos sentar ali ... Sentaram-se . O	cura	respirava fundo . Era um homem corpulento	CTN1-1
5	à missão para servir de companheiro ao	cura	, que pouco sabia de sua vida	CTN1-1
6	instante . – Vamos – disse o	cura	– , conta tudo . – Nesse	CTN1-1
7	não ousou mencionar neste templo . O	cura	fez com a cabeça um grave sinal	CTN1-1
8	– É só ? – perguntou o	cura	– É só . Pelo menos	CTN1-1
9	mais tarde , quando se confessasse ao	cura	. Mas era preciso contar agora .	CTN1-1
10	de minha vida . De novo o	cura	estava de cabeça baixa e olhos cerrados	CTN1-1

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

Também é possível notar inconsistências de anotação dentro dos DS, tal como na subcategoria “Doença” (B2-), que apresenta variações de anotação ilustradas aqui com algumas linhas de concordância da palavra frio (92 ocorrências): “[...] envolvido pelo ar frio da noite. [...]”, “[...] tocam um objeto frio. [...]” e “[...] Estava exausta, com um frio de morte no corpo [...]”. Essa discrepância se dá porque há um DS específico para a palavra em questão: “Temperatura: frio” (O4.6-). Outros termos com discrepâncias também classificados como B2- são “mal” (25), em diversas das ocorrências utilizado como advérbio de modo, por exemplo, “[...] a luz da lamparina mal alumia. [...]” e “raiva” (30), relacionada à emoção sentida e não à doença, esse último caso observável na Imagem 6.

Imagem 6 – Captura de tela das linhas de concordância do termo “raiva”

Line	Left context	Key	Right context	File
1	esperando , subitamente aquecido pelo fogo duma	raiva	nascente . “ Dou-lhe um pontapé no	CTN1-1
2	em artes mágicas , ficaram loucos de	raiva	quando foram vencidos nas Cruzadas pelos cristãos	CTN1-2
3	coisas como uma grande aranha-caranguejeira . Numa	raiva	Anaagarrou com fúria os cabelos de Pedro	CTN1-2
4	mundo . Mas queria viver também de	raiva	, de birra . A sorte andava	CTN1-2
5	em cima da mesa num gesto de	raiva	e ao mesmo tempo de alegria .	CTN1-2

Fonte: Elaborado pelos autores com dados extraídos do Wmatrix 7 (Rayson, 2009).

Justamente pelo exposto acima, um dos primeiros passos da análise com a ferramenta diz respeito à análise manual criteriosa das listas de palavras indicadas do campo semântico que se quer analisar, a fim de se observarem possíveis discrepâncias, antes de chegar a conclusões baseadas apenas no levantamento automático. Também sugere-se uma análise atenta das

palavras classificadas na categoria Z99 - “Não Classificado” e que carecem de anotação manual a fim de serem incluídas no léxico do sistema PyMUSAS para aumentar sua efetividade. Isso porque o anotador não reconheceu palavras de variedades diatópicas, como gaúchos, fandangos, chimarrão, bombachas, nem diacrônicas, como vosmecê e vossuncê, e até em alguns casos, palavras simples como filha, lamparina e mamãe, etiquetando todas como Z99. Além disso, o sistema ainda não lida com unidades multi-palavras, classificando-as individualmente.

Outra limitação está nos equívocos de etiquetagem relacionados à sensibilidade de maiúsculas e minúsculas, como observado na etiquetagem de “fé”, que recebeu a etiqueta E6+ - “Confiante”, inclusive nos casos em que se referia ao nome de cidade - Santa Fé -, em vez de ser incluído na categoria Z, destinada a nomes próprios em geral. O mesmo ocorre com diversos nomes, como Ana Terra, Maneco Terra e outras pessoas dessa descendência, cujo sobrenome foi classificado como “Ações Gerais/ fazer” (A1.1.1), quando deveriam ser classificados como Z99.

6 Considerações finais

A EC é um campo de estudos promissor gestado no encontro de duas grandes metodologias de pesquisa: de um lado, as técnicas e recursos de PLN de que dispõe a LC, que já possui diversas ferramentas bem estabelecidas e estáveis à disposição dos estudos linguísticos e, de outro, um campo já consagrado no âmbito da análise literária, que é a Estilística tradicional. Com essa combinação, se torna mais fácil escrutinar uma ou mais obras com alguns cliques e revelar padrões e proeminências que enriqueçam uma crítica literária de forma quantitativa e qualitativa. Não se pretendeu aqui, todavia, esgotar as possibilidades de análise de cada DS relevante em *O Continente*, uma vez que esta pesquisa visou, sobretudo, avaliar a eficácia da versão 7 do Wmatrix (Rayson, 2009) e, complementarmente, investigar e demonstrar de que maneira a EC pode enriquecer e contribuir à crítica já existente da obra.

Considera-se que este trabalho alcançou seus objetivos de modo satisfatório, uma vez que foi possível vislumbrar aplicações da LC para a análise de uma obra literária, aproximando-se do que se espera ser a abordagem metodológica da EC. Ter a oportunidade de avaliar uma ferramenta promissora como o Wmatrix 7 (Rayson, 2009), que está implementando em sua interface o sistema PyMUSAS com uma série de novas funcionalidades, é uma experiência enriquecedora. Constatar, também, suas limitações e desafios em relação à classificação e desambiguação de palavras na língua portuguesa aponta caminhos possíveis de se contribuir nos estudos em EC, uma vez que os resultados podem ser aproveitados no aprimoramento da ferramenta.

Reitera-se, também, o caráter experimental desta pesquisa com o romance *O Continente*, uma vez que resultados mais precisos da análise de domínios semânticos é influenciado pela eficácia que a ferramenta testada pode entregar. Avalia-se que alguns dos resultados podem ser alterados significativamente quando a interface for capaz de classificar de forma apropriada os termos que recaem inadequadamente na categoria Z99, por exemplo, junto a outras palavras que o etiquetador PyMUSAS ainda não é capaz de reconhecer.

Por fim, esta pesquisa abre um leque de possibilidades futuras, uma vez que estudos preliminares com a trilogia integral de *O Tempo e o Vento* apontaram diferenças nos resultados obtidos, como o surgimento de outros DS, como “Política” e “Roupas e pertences pessoais”. Pesquisas futuras com essa ferramenta podem contribuir tanto em relação a estudos da obra

consagrada de Erico Veríssimo, assim como em relação à ampliação e ao refinamento de um léxico para PLN que proveja recursos mais robustos para que os etiquetadores semânticos do PyMUSAS sejam mais efetivos.

Referências

- ALVES, M. de B. *Tratado das gentes d'O Continente por uma definição da identidade gaúcha*. 2005. 123f. Dissertação (Mestrado em Literatura Comparada) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005. Disponível em: <https://lume.ufrgs.br/handle/10183/4432> Acesso em: 04 mai. 2023
- ANTHONY, L. *AntConc* (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University, 2021. Disponível em: <https://www.laurenceanthony.net/software/antconc/> Acesso em: 21 out. 2023
- ARCHER, D.; WILSON, A.; RAYSON, P. *Introduction to the USAS Category System*. Outubro 2002. Disponível em: https://ucrel.lancs.ac.uk/usas/usas_guide.pdf. Acesso em: 04 mai. 2023
- AULETE, C. *Aulete Digital – Dicionário contemporâneo da língua portuguesa*: Dicionário Caldas Aulete, vs online. Disponível em: <https://www.aulete.com.br/> Acesso em 14 jul. 2024.
- AUSTEN, J. *Pride and prejudice*. [S.l.]: [1813] 2002. Disponível em: <https://www.gutenberg.org/ebooks/1342> . Acesso em: 23 fev. 2024.
- BAKER, M. *Corpus linguistics and translation studies: Implications and applications*. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 1993. p. 233-250.
- BAKER, P. *Using Corpora in Discourse Analysis*. Bloomsbury Publishing, 2023.
- BALOSSI, G. *A Corpus Linguistic Approach to Literary Language and Characterization: Virginia Woolf's The Waves*. John Benjamins, 2014.
- BALOSSI, G. Key Pronouns through Wmatrix in a Novel of Formation: Conrad's The Shadow-Line. *Umanistica Digitale*, Bologna, v. 5, n. 9, 79-96, 2020. DOI: <https://doi.org/10.6092/issn.2532-8816/10542>
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.
- BIBER, D.; REPPEN, R. What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, Cambridge, v.24, n.2, p. 199-208, 2002. DOI: <https://doi.org/10.1017/S0272263102002048>
- BICK, E. A FrameNet for Danish. In: NORDIC CONFERENCE OF COMPUTATIONAL LINGUISTICS (NODALIDA 2011), 18, Northern European Association for Language Technology (NEALT), 2011, Tartu. *Proceedings Series*, Tartu: Tartu University Library, 2011. p. 34-41.
- BICK, E. PALAVRAS, a Constraint Grammar-based Parsing System for Portuguese. In: BERBER SARDINHA, T.; FERREIRA, T. de L. S. B. (ed.). *Working with Portuguese Corpora*. London/New York: Bloomsbury Academic, 2014. p. 279-302.
- BICK, E. PFN-PT: Anotação semântica automática: um novo Framenet para o português. *Domínios de Linguagem*, Uberlândia, v.16, n.4, p. 1401-1435, 2022. DOI: <https://doi.org/10.14393/DL52-v16n4a2022-7>

- BISOL, L. V.; PORTO, L. T. Violência e memória: uma leitura do romance *O Continente*, de Erico Verissimo. *Navegações*, [S. l.], v. 8, n. 2, p. 146–155, 2015. DOI: 10.15448/1983-4276.2015.2.20532.
- BLAKE, W. *Songs of Innocence*. Londres: Edição do Autor, 1789.
- BLAKE, W. *Songs of Innocence and of Experience*: shewing the two contrary states of the human soul. Londres: Edição do Autor, 1794. Disponível em: <https://www.blakearchive.org/copy/songsie.b?descld=songsie.b.illbk.01>. Acesso em: 8 ago. 2024.
- BORDINI, M. D. G. *O Continente*: um romance de formação? Pós-colonialismo e identidade política. In: BORDINI, M. D. G.; ZILBERMAN, R. (ed.) *O tempo e o vento: história, invenção e metamorfose*. Porto Alegre: Edipucrs, 2004. p.65-86.
- BORDINI, M. D. G.; ZILBERMAN, R. (ed.) *O tempo e o vento: história, invenção e metamorfose*. Porto Alegre: Edipucrs, 2004.
- BRAGA, G. da S. *Corpus* stylistics in translation-oriented text analysis: approaching the work of Denton Welch from a functionalist perspective. *Diacrítica*, Coimbra, v. 32, n. 3, p. 227-248, 2020. DOI: <https://doi.org/10.21814/diacritica.5163>
- BREZINA, V. & PLATT, W. *#LancsBox X* [software], Lancaster University, 2024. Disponível em: <http://lancsbox.lancs.ac.uk>.
- CAN, T.; CANGIR, H. A warring style: A *corpus* stylistic analysis of the First World War poetry. *Digital Scholarship in the Humanities*, Cidade, v. 37, n. 3, p. 660-680, 2022.
- ČERMÁKOVÁ, A. Traduzindo literatura infantil: algumas ideias de estilística de *corpus*. *Ilha do Desterro*, Florianópolis, v. 71, n. 1, p. 117-134, jan/abr 2018. DOI: <https://doi.org/10.5007/2175-8026.2018v71n1p117>
- CULPEPER, J.; ARCHER, D.; RAYSON, P. Love - 'A familiar of a devil'? An exploration of key domains in Shakespeare's comedies and tragedies. In: ARCHER, D. (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction*. Digital research in the Arts and Humanities. Ashgate, Farnham, 2009, pp. 136-157. Disponível em: https://eprints.lancs.ac.uk/id/eprint/12671/1/ACR_MethNet06.pdf Acesso em 01 out. 2024
- EVERT, S. How Random is a *Corpus*? The library metaphor. *Zeitschrift für Anglistik und Amerikanistik*, Berlin, v. 54, n. 2, 2006, pp. 177-190. DOI: 10.1515/zaa-2006-0208
- FISCHER, L. A. Sobre a vigência do Regionalismo no Brasil. *Terceira Margem*. Rio de Janeiro, v. 12, n. 19, p. 103-117, ago./dez. 2008. DOI: <https://doi.org/10.55702/3m.v12i19.11058>
- FISCHER-STARCKE, B. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A *corpus*-stylistic analysis. *International Journal of Corpus Linguistics*, Amsterdam, v. 14, n. 4, p. 492-523, 2009. DOI: <https://doi.org/10.1075/ijcl.14.4.03fis>
- FOWLES, J. *The Magus*. 1 ed. London: Jonathan Cape, 1965.
- GABRIELATOS, C. Keyness analysis: Nature, metrics and techniques. In: TAYLOR, C.; MARCHI, A. (eds.) *Corpus approaches to discourse: A critical review*. New York: Routledge, 2018. p. 225–258.
- GARCIA, C.; LISBOA FILHO, F. Representações da identidade do gaúcho e da violência: uma leitura de “O Tempo e o Vento”. *Violência intrafamiliar: discutindo facetas e possibilidades*. Jundiaí: Paco Editorial, 2013. p. 171-182, Disponível em: <https://www.ufsm.br/app/uploads/sites/513/2020/08/REPRESENTA%C3%87%C3%95ES-DA-IDENTIDADE-DO-GA%C3%A9CHO-E-DA-VIOL%C3%8ANCIA-UMA-LEITURA-DE-O-TEMPO-E-O-VENTO.pdf>. Acesso em: 22 Ago. 2025

- GINZBURG, J. *Literatura, violência e melancolia*. Campinas, SP: Autores Associados, 2013.
- HALBWACHS, M. *A memória coletiva*. Trad. Beatriz Sidou. São Paulo: Centauro, 2006.
- IBRAHIM, W. M. A. Utilizing *corpus* stylistics to facilitate literary analysis: An assessment of the effectiveness of semantic domains in identifying major literary themes in a selection of Charles Dickens novels. *AJELP: Asian Journal of English Language and Pedagogy*, Kedah, v. 10, n. 1, p. 114-138, 2022. DOI: <https://doi.org/10.37134/ajelp.vol10.1.9.2022>
- JACOBI, L. dos S. *A construção da melancolia em O continente, de Erico Veríssimo: a dor renitente do povo gaúcho*. 2023. 95 f. Dissertação (Mestrado em Letras) — Escola de Humanidades, Pontifícia Universidade Católica do Rio Grande do Sul, 2023. Disponível em: <https://tede2.pucrs.br/tede2/handle/tede/11068> Acesso em 22 Ago. 2025
- KAY, J. *The adoption papers*. 1 ed. Newcastle: Bloodaxe Books, 1991.
- KILGARRIFF, A. Comparing corpora. *International Journal of Corpus Linguistics*, Amsterdam, v. 6, n. 1, p. 97-133, 2001. DOI: <https://doi.org/10.1075/ijcl.6.1.05kil>
- KILGARRIFF, A. et al. The Sketch Engine: ten years on. *Lexicography*, Toronto, v. 1, p. 7-36, 2014. DOI: <https://doi.org/10.1007/s40607-014-0009-9>
- KRISHNAMURTHY, R. Corpus-driven lexicography. *International Journal of Lexicography*, Oxford, v. 21, n. 3, p. 231-242, 2008. DOI: <https://doi.org/10.1093/ijl/ecn028>
- LEECH, G. New resources, or just better old ones? The Holy Grail of representativeness. In: Hundt, M.; Nesselhauf, N.; Biewer, C. (eds.) *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 2007. p. 133-149. DOI: https://doi.org/10.1163/9789401203791_009
- MAHLBERG, M., SMITH, C. & PRESTON, S. Phrases in literary contexts: patterns and distributions of suspensions in Dickens's novels. *International Journal of Corpus Linguistics*, Amsterdam, v. 18, n. 1, p. 35-56, 2013. DOI: <https://doi.org/10.1075/ijcl.18.1.05mah>
- MATURANA, M. del C. C. *Maternidad y voces poéticas en 'The Adoption Papers' de Jackie Kay*. 2012. Tese (Doutorado em Literatura e Linguística Inglesas) - Universidad de Granada, 2012. Disponível em: https://digibug.ugr.es/bitstream/handle/10481/70114/Tesis_Coral%20Calvo%20Maturana_16Oct.pdf?sequence=4&isAllowed=y Acesso em: 22 Ago. 2024
- MCARTHUR, Tom. *Longman lexicon of contemporary English*. Harlow: Longman, 1981.
- MCENERY, T.; HARDIE, A. *Corpus Linguistics*. Cambridge: Cambridge University Press, 2012.
- MCINTYRE, D.; ARCHER, D. A corpus-based approach to mind style. *Journal of Literary Semantics*, Berlin, v. 39, n. 2, p. 167-182, 2010. DOI: <https://doi.org/10.1515/jls-2021-2045>
- MCINTYRE, D.; WALKER, B. How can corpora be used to explore the language of poetry and drama? In: O'KEEFFE, A.; MCCARTHY, M. (Eds.). *The Routledge Handbook of Corpus Linguistics*. 1 ed. London: Routledge, 2010. p. 516-530.
- MCINTYRE, D.; WALKER, B. *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press, 2019.
- NORD, C. *Textanalyse und Übersetzen*. 4 ed. Tübingen: Julius Groos, 2009.
- O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press, 2007.

PIAO, S.; BIANCHI, F.; DAYRELL, C.; D'EGÍDIO, A.; RAYSON, P. Development of the Multilingual Semantic Annotation System. In: *Proceedings of the 2015 Conference of The North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, Denver, Colorado. Association for Computational Linguistics, 2015. p. 1268- 1274. Disponível em: <https://aclanthology.org/N15-1137.pdf> Acesso em 25 Ago. 2025

PIMENTA, A.; NOVODVORSKI, A. Linguística de *Corpus* e Fraseologia: um estudo das colocações com “Feito” em Grande Sertão: Veredas. *Linguagem: Estudos e Pesquisas*, Goiânia, v. 21, n. 1, 2018. DOI: <https://doi.org/10.5216/lep.v21i1.52224>.

POJANAPUNYA, P.; TODD, R. W. Log-likelihood and odds ratio: keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, Berlin, v. 14, n. 1, p. 133-167, 2018. DOI: <https://doi.org/10.1515/cllt-2015-0030>

RANGEL, M. A. *Proposta metodológica para análises de obras literárias através da Linguística de Corpus: o caso de Paulo Coelho*. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Uberlândia, 2022. DOI: <http://doi.org/10.14393/ufu.di.2022.148>

RAYSON, P. *Wmatrix: a web-based corpus processing environment*. Computing Department, Lancaster University, 2009.

RAYSON, P.; ARCHER, D. E.; PIAO, S.; MCENERY, T. THE UCREL SEMANTIC ANALYSIS SYSTEM. In: FOURTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2004, Lisbon. *Proceedings...* Paris: European Language Resources Association (ELRA), 2004. p. 1813-1816. Disponível em: http://www.lancaster.ac.uk/staff/rayson/publications/usas_lreco4ws.pdf Acesso em: 12 Set. 2025

RAYSON, P.; BERRIDGE, D.; FRANCIS, B. Extending the Cochran rule for the comparison of word frequencies between corpora. In: INTERNATIONAL CONFERENCE ON STATISTICAL ANALYSIS OF TEXTUAL DATA (JADT), 7, 2004, *Proceedings...*, v. 2, Louvain-la-Neuve, Belgium, March 10-12, 2004, Presses universitaires de Louvain, pp. 926 - 936. ISBN 2-930344-50-4. Disponível em: http://ucrel.lancs.ac.uk/people/paul/publications/rbfo4_jadt.pdf

RAYSON, P.; GARSIDE, R. Comparing corpora using frequency profiling. In: WORKSHOP ON COMPARING CORPORA, 2000, Hong Kong. *Proceedings...* Hong Kong: [s.n.], 2000. p. 1-6. Disponível em: <https://aclanthology.org/W00-0901.pdf> Acesso em 12 set. 2025

ROSA, J. G. *Ave, Palavra*. Rio de Janeiro, RJ: Nova Fronteira, 1985.

ROSA, L. M. da. *Literatura de ficção e Educação Geográfica: Tradição inventada, modelação social e discurso na obra O Continente*, de Erico Verissimo. 2019. 148f. Dissertação (Mestrado em Geografia) – Programa de Pós-Graduação em Geografia, Instituto de Ciências Humanas, Universidade Federal de Pelotas, Pelotas, 2019. Disponível em: <https://guaiaca.ufpel.edu.br/handle/prefix/4733> Acesso em: 12 set. 2025

SANTOS, P. B. Aspectos do romance histórico em Erico Verissimo. *O Eixo e a Roda: Revista de Literatura Brasileira*, [S.l.], v. 11, p. 53-59, dez. 2005. DOI: <https://doi.org/10.17851/2358-9787.11..53-59>

SANTOS, D. A. dos. O tempo e o vento: romance histórico e romance político. *Cadernos do IL*, Porto Alegre, n. 39, p. 96-104, 2009. DOI: <https://doi.org/10.22456/2236-6385.25205>

SANTOS, D. A. dos. A crítica literária dos anos 40 e a autocrítica de Erico Verissimo. *Literatura em Debate*, Frederico Westphalen, v. 13, n. 24, p. 100-115, 2019.

- SANTOS, P. B. Aspectos do romance histórico em Erico Veríssimo. *O Eixo e a Roda: Revista de Literatura Brasileira*, [S.l.], v. 11, p. 53-59, dez. 2005. DOI: <https://doi.org/10.17851/2358-9787.11..53-59>
- SCOTT, M., *WordSmith Tools version 9* (64 bit version) Stroud: Lexical Analysis, 2024.
- SHORT, M. *Exploring the language of poems, plays and prose*. London: Longman, 1996.
- TAGNIN, S. E. O. *O jeito que a gente diz: combinações consagradas em inglês e português*. Barueri: DISAL, 2013.
- TEUBERT, W. *Corpus Linguistics and Lexicography. International Journal of Corpus Linguistics*, Amsterdam, v. 6, n. 1, p. 125-153, December 2001. DOI: <https://doi.org/10.1075/IJCL.6.SI.11TEU>
- VATHANALAOHA, K. *Corpus Stylistics in Contemporary English Dramas: Keywords and Semantic Fields of Delusions. GEMA Online Journal of Language Studies*, Bangi, v. 22, n. 2, p. 43-62, 2022. DOI: <https://doi.org/10.17576/gema-2022-2202-03>
- VERÍSSIMO, E. *O Continente*. Porto Alegre: Globo, 2013 [1949].
- VERÍSSIMO, E. *O Retrato*. Porto Alegre: Globo, 1951.
- VERÍSSIMO, E. *O Arquipélago*. Porto Alegre: Globo, 1961.
- VERÍSSIMO, E. *O Tempo e o Vento*. 5 v. São Paulo: Companhia das Letras, 2004.
- VITAL, Á. A. S. Do contato entre a Literatura, a Linguística de *Corpus* e o Processamento de Língua Natural: o caso dos anagramáticos de Guimarães Rosa. *Texto Livre*, Belo Horizonte, v. 15, p. e39316, 2022. DOI: <https://doi.org/10.35699/1983-3652.2022.39316>
- WELCH, D. *Sickert at St. Peter's*. Horizon, 1942.
- WYNNE, M. *Corpus Stylistics in Principles and Practice. A Stylistic Exploration of John Fowles' The Magus*. Yufang Ho. *Literary and Linguistic Computing*, Oxford, v. 27, n. 4, p. 474-476, December 2012. DOI: <https://doi.org/10.1093/lc/fqso23>
- ZANETTIN, F. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. New York: Routledge, 2012.
- ZILBERMAN, R. O Tempo e Vento: História, mito, literatura. In: BORDINI, M. D. G.; ZILBERMAN, R. (ed.) *O tempo e o vento: história, invenção e metamorfose*. Porto Alegre: Edipucrs, 2004a. p. 21-48
- ZILBERMAN, R. Luiza Silva Cambará – revendo a tradição do mito. In: BORDINI, M. D. G.; ZILBERMAN, R. (ed.) *O tempo e o vento: história, invenção e metamorfose*. Porto Alegre: Edipucrs, 2004b. p. 87-102.
- ZYNGIER, S.; CARNEIRO, R. M. O.; NOVODVORSKI, A. Reflexões sobre a estilística e o ensino de literatura: uma entrevista com Sonia Zyngier. *Trabalhos em Linguística Aplicada*, Campinas, v. 62, n. 2, 388-399, 2023. DOI: <https://doi.org/10.1590/01031813v62220238667192>

Apêndice A

Versão integral do Quadro 2: dados exportados de DS salientes no CE

Class.	Etiqueta	LL	Domínio Semântico
1	B1	1211.70	Anatomia e fisiologia
2	E6+	600.07	Confiante
3	G3	422.68	Guerra, defesa e o exército; armas
4	M7	391.51	Lugares
5	T2+	288.24	Tempo: Início
6	X3.4	264.78	Sensorial: Visão
7	N3.8+	200.00	Velocidade: Rápido
8	Z2	193.57	Nomes Geográficos
9	O1.3	188.07	Substâncias e materiais: Gás
10	N1	152.53	Números
11	X3.2-	150.23	Som: Silencioso
12	Q2.1	140.02	Fala: Comunicativo
13	N6	136.11	Frequência
14	S9	104.62	Religião e o sobrenatural
15	H5	99.63	Mobília e utensílios domésticos
16	X3	91.22	Sensorial
17	M6	84.66	Localização
18	X3.5	84.38	Ssensorial: Olfato
19	O2	82.76	Objetos em geral
20	F1	69.46	Comida
21	M8	68.63	Estacionário
22	O4.6-	68.01	Temperatura: frio
23	Z99	67.67	Não Classificável
24	A13.6	65.47	Grau: Diminutivos
25	F3	57.94	Fumo e drogas não medicinais
26	B3	57.39	Medicina e tratamentos medicinais
27	S2.2	56.14	Pessoas: Masculino
28	H2	54.37	Partes de construções
29	O4.6+	52.30	Temperatura: quente / em chamas
30	Z5	51.70	Caixa Gramatical
31	A8	50.92	Parecer
32	S2	45.35	Pessoas
33	N5-	44.61	Quantidades: pouco
34	O4.3	42.79	Cor e padrões de cores
35	S4	39.36	Família
36	B2+	38.93	Saudável
37	O4.6	37.75	Temperatura: quente / em chamas
38	K2	28.18	Música e atividades relacionadas

39	T1.2	26.45	Tempo: Momentâneo
40	S2.1	25.75	Pessoas: Feminino
41	T1	23.79	Tempo
42	S5	21.91	Grupos e afiliações
43	F4	20.72	Agricultura e horticultura
44	W1	19.66	O Universo
45	S7.1-	18.61	Sem Poder
46	A1.1.1	17.78	Ações em geral / fazer
47	A13.3	17.62	Grau: Aumentativo
48	S1.2.1-	16.81	Formal/ Não Amigável
49	X3.2	15.99	Sensorial: Som
50	W2	14.79	Luz
51	O4.2	14.26	Julgamento de aparência
52	N5++	13.07	Quantidades: bastante/muito
53	N3.3	13.01	Medidas: Distância
54	T1.1.2	12.76	Tempo: Presente, simultâneo
55	N3.2---	10.99	Tamanho: Pequeno

Fonte: próprios autores [tradução própria]

Apêndice B

Quadro de Domínios Semânticos agrupado por Categoria Geral

Domínio Semântico - Categoria geral	Etiqueta	Log Likelihood	DS - Subcategoria
A - Termos abstratos e gerais	A13.6	65.47	Grau: Diminutivos
	A8	50.92	Parecer
	A1.1.1	17.78	Ações em geral / fazer
	A13.3	17.62	Grau: Aumentativo
B - O Corpo e o Indivíduo	B1	1211.70	Anatomia e fisiologia
	B3	57.39	Medicina e tratamentos medicinais
	B2+	38.93	Saudável
E - Ações Emocionais, Estados e Processos	E6+	600.07	Confiante
F - Alimentação e Agropecuária	F1	69.46	Comida
	F3	57.94	Fumo e drogas não medicinais
	F4	20.72	Agricultura e horticultura
G - Governo e Domínio Público	G3	422.68	Guerra, defesa e o exército; armas
H - Arquitetura, Construção, Casas e o Lar	H5	99.63	Mobília e utensílios domésticos
	H2	54.37	Partes de construções
K – Entretenimento, Esportes e Jogos	K2	28.18	Música e atividades relacionadas

M - Movimento, Localidades, Viagem e Transporte	M7	391.51	Lugares
	M6	84.66	Localização
	M8	68.63	Estacionário
N - Números e Medidas	N3.8+	200.00	Velocidade: Rápido
	N1	152.53	Números
	N6	136.11	Frequência
	N5-	44.61	Quantidades: pouco
	N5++	13.07	Quantidades: bastante/muito
	N3.3	13.01	Medidas: Distância
	N3.2---	10.99	Tamanho: Pequeno
O - Substâncias, Materiais, Objetos e Equipamentos	O1.3	188.07	Substâncias e materiais: Gás
	O2	82.76	Objetos em geral
	O4.6-	68.01	Temperatura: frio
	O4.6+	52.30	Temperatura: quente / em chamas
	O4.3	42.79	Cor e padrões de cores
	O4.6	37.75	Temperatura
	O4.2	14.26	Julgamento de aparência
Q - Ações Verbais, Estados e Processos	Q2.1	140.02	Fala: Comunicativo
S - Ações Sociais, Estados e Processos	S9	104.62	Religião e o sobrenatural
	S2.2	56.14	Pessoas: Masculino
	S2	45.35	Pessoas
	S4	39.36	Parentesco
	S2.1	25.75	Pessoas: Feminino
	S5	21.91	Grupos e afiliações
	S7.1-	18.61	Sem Poder
	S1.2.1-	16.81	Formal/ Não Amigável
T - Tempo	T2+	288.24	Tempo: Início
	T1.2	26.45	Tempo: Momentâneo
	T1	23.79	Tempo
	T1.1.2	12.76	Tempo: Presente, simultâneo
W - O Mundo e Nosso Ambiente	W1	19.66	O Universo
	W2	14.79	Luz
X - Ações Psicológicas, Estados e Processos	X3.4	264.78	Sensorial: Visão
	X3.2-	150.23	Som: Silencioso
	X3	91.22	Sensorial
	X3.5	84.38	Ssensorial: Olfato
	X3.2	15.99	Sensorial: Som
Z - Nomes e Palavras Gramaticais	Z2	193.57	Nomes Geográficos
	Z99	67.67	Não Classificável
	Z5	51.70	Caixa Gramatical

Fonte: próprios autores [tradução própria]