

# REVISTA DE ESTUDOS DA LINGUAGEM

Faculdade de Letras da UFMG

ISSN

Impresso: 0104-0588

On-line: 2237-2083

V.29 - N° 3



# REVISTA DE ESTUDOS DA LINGUAGEM

**Universidade Federal de Minas Gerais**

REITORA: Sandra Regina Goulart Almeida

VICE-REITOR: Alessandro Fernandes Moreira

**Faculdade de Letras**

DIRETORA: Graciela Inés Ravetti de Gómez (*in memoriam*)

VICE-DIRETORA: Sueli Maria Coelho

## **Editor-chefe**

Gustavo Ximenes Cunha (UFMG)

## **Revisão e Normalização**

Alda Lopes Durães Ribeiro

Gustavo Ximenes Cunha

Jairo Venício Carvalhais Oliveira

## **Editoras-associadas**

Ana Larissa Adorno Maciotto Oliveira (UFMG)

Carla Viana Coscarelli (UFMG)

Helcira Maria Rodrigues de Lima (UFMG)

## **Secretaria**

Gustavo Ximenes Cunha (UFMG)

## **Revisão de Língua Inglesa**

Ana Larissa Adorno Marciotto Oliveira (UFMG)

Junia de Carvalho Fidelis Braga (UFMG)

Mara Passos Guimarães (UFMG)

Marisa Mendonça Carneiro (UFMG)

## **Editoração eletrônica**

Alda Lopes Durães Ribeiro

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

ISSN: Impresso: 0104-0588

On-line: 2237-2083

# REVISTA DE ESTUDOS DA LINGUAGEM

V. 29 - Nº 3 - jul.-set. 2021

## Indexadores

*Diadorim [Brazil]*

*DOAJ (Directory of Open Access Journals) [Sweden]*

*DRJI (Directory of Research Journals Indexing) [India]*

*EBSCO [USA]*

*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**

## **Editor-chefe**

Gustavo Ximenes Cunha (UFMG, Belo Horizonte/MG, Brasil)

## **Editoras-associadas**

Ana Larissa Adorno Maciotto Oliveira (UFMG, Belo Horizonte/MG, Brasil)

Carla Viana Coscarelli (UFMG, Belo Horizonte/MG, Brasil)

Helcira Maria Rodrigues de Lima (UFMG, Belo Horizonte/MG, Brasil)

## **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)

## **Comissão Científica**

Aderlande Pereira Ferraz (UFMG, Belo Horizonte/MG, Brasil)  
Alessandro Panunzi (Unifi, Florença, Itália)  
Alina M. S. M. Villalva (ULisboa, Lisboa, Portugal)  
Aline Alves Ferreira (UCSB, Santa Barbara/CA, Estados Unidos)  
Ana Lúcia de Paula Müller (USP, São Paulo/SP, Brasil)  
Ana Maria Carvalho (UA, Tucson/AZ, Estados Unidos)  
Ana Paula Scher (USP, São Paulo/SP, Brasil)  
Anabela Rato (U of T, Toronto/ON, Canadá)  
Aparecida de Araújo Oliveira (UFV, Viçosa/MG, Brasil)  
Aquiles Tescari Neto (UNICAMP, Campinas/SP, Brasil)  
Augusto Soares da Silva (UCP, Braga, Portugal)  
Beth Brait (PUC-SP/USP, São Paulo/SP, Brasil)  
Bruno Neves Rati de Melo Rocha (UFMG, Belo Horizonte/MG, Brasil)  
Carmen Lucia Barreto Matzenauer (UCPEL, Pelotas/RS, Brasil)  
Celso Ferrarezi (UNIFAL, Alfenas/MG, Brasil)  
César Nardelli Cambraia (UFMG, Belo Horizonte/MG, Brasil)  
Cristina Name (UFJF, Juiz de Fora/MG, Brasil)  
Charlotte C. Galves (UNICAMP, Campinas/SP, Brasil)  
Deise Prina Dutra (UFMG, Belo Horizonte/MG, Brasil)  
Diana Luz Pessoa de Barros (USP/UPM, São Paulo/SP, Brasil)  
Edwiges Morato (UNICAMP, Campinas/SP, Brasil)  
Emília Mendes Lopes (UFMG, Belo Horizonte/MG, Brasil)  
Esmeralda V. Negrão (USP, São Paulo/SP, Brasil)  
Flávia Azeredo Cerqueira (JHU, Baltimore/MD, Estados Unidos)  
Gabriel de Avila Othero (UFRGS, Porto Alegre/RS, Brasil)  
Gerardo Augusto Lorenzino (TU, Filadélfia/PA, Estados Unidos)  
Glaucia Muniz Proença de Lara (UFMG, Belo Horizonte/MG, Brasil)  
Hanna Batoréo (UAb, Lisboa, Portugal)  
Heliana Ribeiro de Mello (UFMG, Belo Horizonte/MG, Brasil)  
Heronides Moura (UFSC, Florianópolis/SC, Brasil)  
Hilario Bohn (UCPEL, Pelotas/RS, Brasil)  
Hugo Mari (PUC-Minas, Belo Horizonte/MG, Brasil)  
Ida Lucia Machado (UFMG, Belo Horizonte/MG, Brasil)  
Ieda Maria Alves (USP, São Paulo/SP, Brasil)  
Ivã Carlos Lopes (USP, São Paulo/SP, Brasil)  
Jairo Nunes (USP, São Paulo/SP, Brasil)

Jairo Venício Carvalhais Oliveira (UFMG, Belo Horizonte/MG, Brasil)  
Jean Cristtus Portela (UNESP-Araraquara, Araraquara/SP, Brasil)  
João Antônio de Moraes (UFRJ, Rio de Janeiro/ RJ, Brasil)  
João Miguel Marques da Costa (Universidade Nova da Lisboa, Lisboa, Portugal)  
João Queiroz (UFJF, Juiz de Fora/MG, Brasil)  
José Magalhaes (UFU, Uberlândia/MG, Brasil)  
João Saramago (Universidade de Lisboa, Lisboa, Portugal)  
José Borges Neto (UFPR, Curitiba/PR, Brasil)  
Kanavillil Rajagopalan (UNICAMP, Campinas/SP, Brasil)  
Laura Alvarez Lopez (Universidade de Estocolmo, Stockholm, Suécia)  
Leo Wetzels (Free Univ. of Amsterdam, Amsterdã, Holanda)  
Laurent Fillietaz (Université de Genève, Genebra, Suíça)  
Leonel Figueiredo de Alencar (UFC, Fortaleza/CE, Brasil)  
Livia Oushiro (UNICAMP, Campinas/SP, Brasil)  
Lodenir Becker Karnopp (UFRGS, Porto Alegre/RS, Brasil)  
Lorenzo Teixeira Vitral (UFMG, Belo Horizonte/MG, Brasil)  
Luiz Amaral (UMass Amherst, Amherst/MA, Estados Unidos)  
Luiz Carlos Cagliari (UNESP, São Paulo/SP, Brasil)  
Luiz Carlos Travaglia (UFU, Uberlândia/MG, Brasil)  
Marcelo Barra Ferreira (USP, São Paulo/SP, Brasil)  
Marcia Cançado (UFMG, Belo Horizonte/MG, Brasil)  
Márcio Leitão (UFPB, João Pessoa/PB, Brasil)  
Marcus Maia (UFRJ, Rio de Janeiro/RJ, Brasil)  
Maria Bernadete Marques Abaurre (UNICAMP, Campinas/SP, Brasil)  
Maria Cecília Camargo Magalhães (PUC-SP, São Paulo/SP, Brasil)  
Maria Cecília Magalhães Mollica (UFRJ, Rio de Janeiro/RJ, Brasil)  
Maria Cândida Trindade Costa de Seabra (UFMG, Belo Horizonte/MG, Brasil)  
Maria Cristina Figueiredo Silva (UFPR, Curitiba/PR, Brasil)  
Maria Luíza Braga (PUC/RJ, Rio de Janeiro/RJ, Brasil)  
Maria Marta P. Scherre (UNB, Brasília/DF, Brasil)  
Micheline Mattedi Tomazi (UFES, Vitória/ES, Brasil)  
Miguel Oliveira, Jr. (UFAL, Maceió, Alagoas, Brasil)  
Monica Santos de Souza Melo (UFV, Viçosa/MG, Brasil)  
Patricia Matos Amaral (UI, Bloomington/IN, Estados Unidos)  
Paulo Roberto Gonçalves Segundo (USP, São Paulo/SP, Brasil)  
Philippe Martin (Université Paris 7, Paris, França)  
Rafael Nonato (Museu Nacional-UFRJ, Rio de Janeiro/RJ, Brasil)  
Raquel Meister Ko. Freitag (UFS, Aracaju/SE, Brasil)

Roberto de Almeida (Concordia University, Montreal/QC, Canadá)  
Ronice Müller de Quadros (UFSC, Florianópolis/SC, Brasil)  
Ronald Beline (USP, São Paulo/SP, Brasil)  
Rove Chishman (UNISINOS, São Leopoldo/RS, Brasil)  
Sanderléia Longhin-Thomazi (UNESP, São Paulo/SP, Brasil)  
Sergio de Moura Menuzzi (UFRGS, Porto Alegre/RS, Brasil)  
Seung- Hwa Lee (UFMG, Belo Horizonte/MG, Brasil)  
Sírío Possenti (UNICAMP, Campinas/SP, Brasil)  
Suzi Lima (U of T / UFRJ, Toronto/ON - Rio de Janeiro/RJ, Brasil)  
Thais Cristofaro Alves da Silva (UFMG, Belo Horizonte/MG, Brasil)  
Tommaso Raso (UFMG, Belo Horizonte/MG-Brasil)  
Tony Berber Sardinha (PUC-SP, São Paulo/SP, Brasil)  
Ubiratã Kickhöfel Alves (UFRGS, Porto Alegre/RS, Brasil)  
Vander Viana (University of Stirling, Stirling/Sld, Reino Unido)  
Vanise Gomes de Medeiros (UFF, Niterói/RJ, Brasil)  
Vera Lucia Lopes Cristovao (UEL, Londrina/PR, Brasil)  
Vera Menezes (UFMG, Belo Horizonte/MG, Brasil)  
Vilson José Leffa (UCPel, Pelotas/RS, Brasil)

## *Sumário / Contents*

---

Formalidade e pronomes de segunda pessoa do singular no português gaúcho: dados de interpretação

*Formality and second person singular pronouns in Gaúcho Portuguese: data from interpretation*

Ronan Pereira ..... 1651

Sentidos do discurso **coaching** financeiro no enunciado vídeo publicitário “Meu nome é Bettina” e possibilidade de cotejo

*Senses of the financial coaching discourse in the utterance advertising video “My name is Bettina” and possibilities of collation*

Grenissa Stafuzza

Maximiano Antonio Pereira ..... 1685

Uma gramática computacional de um fragmento do nheengatu

*A computational grammar for a fragment of Nheengatu*

Leonel Figueiredo de Alencar ..... 1717

Garimpando palavras: produção e circulação de conhecimento no século XIX a partir do gesto de autoria de Alencar

*Digging for Words: the production and circulation of knowledge in the 19th Century from the gestures of Alencar*

Vanise Medeiros

Thais Costa

Raphael Mendes ..... 1779

A família lexical de *usura*: um estudo etimológico e morfossemântico

*The lexical family of usura: an etymological and morphosemantic study*

Matheus Pinto

Mailson Lopes ..... 1813

<p>“Tô notando, sô!” A partícula vocativa <i>sô</i> e a interface sintático-pragmática</p> <p><i>“Tô notando, sô!” The vocative particle sô and the syntactic-pragmatic interface</i></p> <p>Juliana Costa Moreira .....</p>	1873
<p>Reeditar é preciso? As <i>cartas oficiais norte-rio-grandenses</i> e os <i>corpora</i> diacrônicos</p> <p><i>Is it necessary to reedit? The Rio Grande do Norte official letters and the diachronic corpora</i></p> <p>Felipe Morais de Melo .....</p>	1901
<p>Vivendo à margem da lei: histórias de brasileiros em situação irregular no contexto europeu</p> <p><i>Living on the fringes of the law: stories of Brazilians in an irregular situation in the European context</i></p> <p>Glaucia Muniz Proença Lara .....</p>	1943
<p>“Isso tudo me traz de novo a vida que eu tinha”: a coconstrução de uma narrativa autobiográfica na Doença de Alzheimer</p> <p><i>“It all brings me back to the life I had”: the co-construction of an autobiographical narrative in Alzheimer’s Disease</i></p> <p>Caio Mira</p> <p>Katiuscia Custodio .....</p>	1979
<p>Apagamento do rótico em coda no Português Santomense (PST): uma análise sociolinguística</p> <p><i>Rhotic deletion in coda in Santomean Portuguese (PST): an sociolinguistics analysis</i></p> <p>Nancy Mendes Torres Vieira</p> <p>Amanda Macedo Balduino .....</p>	2011

<p>Compilação, reciclagem e padronização de um <i>Corpus</i>  Colaborativo de Linguística: percursos metodológicos  <i>Compilation, recycling and standardization in a Collaborative</i>  <i>Corpus of Linguistics: methodological approaches</i>  Guilherme Fromm  Márcio Issamu Yamamoto .....</p>	2041
<p>Flavors of the progressive in the New Romania: the perfective  progressive periphrasis in Brazilian Portuguese and Argentinian  Spanish  <i>Sabores do progressivo na România Nova: a perífrase perfectiva</i>  <i>progressivo no português brasileiro e no espanhol argentino</i>  Romina Trebisacce  Victoria Ferrero  Renato Miguel Basso .....</p>	2079
<p>Modelos de procesamiento sintáctico y sus implicaciones para el  estudio del lenguaje  <i>Syntactic processing models and their implications for the study</i>  <i>of language</i>  Noelia Ayelén Stetie .....</p>	2117





## Formalidade e pronomes de segunda pessoa do singular no português gaúcho: dados de interpretação

### *Formality and second person singular pronouns in Gaúcho Portuguese: data from interpretation*

Ronan Pereira

Centro de Linguística da Universidade Nova de Lisboa (CLUNL), Lisboa / Portugal  
a57730@campus.fcsh.unl.pt  
<http://orcid.org/0000-0003-2675-6378>

**Resumo:** Na variedade do português falado no estado do Rio Grande do Sul, o pronome de segunda pessoa do singular “tu” ocorre com maior frequência do que o pronome inovador “você”, sendo o “tu” considerado marca da identidade local. No entanto, vem geralmente acompanhado de verbos com a flexão de terceira pessoa, tendo sido sugerido que a utilização da flexão canônica está reservada a momentos de maior formalidade. Este estudo propôs-se a obter dados empíricos acerca da percepção de formalidade dos diferentes pronomes de segunda pessoa do português gaúcho pelos seus falantes nativos, além de estabelecer se consideram o pronome “tu” parte do seu jeito de falar. Uma tarefa de seleção de vocábulos, além de uma tarefa de julgamento de formalidade com uma escala de cinco pontos foram realizadas com 233 participantes. Os resultados confirmam a visão de que o pronome “tu” faz parte da identidade dos falantes e sugerem um sistema triádico, composto, em ordem de formalidade por “tu”, “você” e “o senhor”, acompanhados de morfologia flexional de terceira pessoa, enquanto “tu” com a flexão canônica compete com “você” pela posição intermédia. Dentre os fatores sociolinguísticos que influenciam os julgamentos, somente a idade foi relevante, com os participantes acima de 50 anos de idade tendendo a considerar as frases com “tu” mais informais, independentemente da flexão a ele associada.

**Palavras-chave:** pronomes de segunda pessoa; formalidade; português brasileiro; português gaúcho; sociolinguística.

**Abstract:** In the variety of Portuguese spoken in the Brazilian state of Rio Grande do Sul, the second person singular pronoun “tu” occurs more frequently than the innovative pronoun “você”, being “tu” considered a mark of local identity. It is, however, usually followed by verbs conjugated in the third person, having been suggested that the usage of the canonical inflection is reserved to moments of a higher formality level. This study aimed to bring empirical data regarding how formal the different second person pronouns are interpreted by native speakers, as well as data confirming if they consider “tu” as part of their identity. A word selection task and a 5-point scale formality judgement task were conducted with 233 participants. The results confirm the idea that “tu” is part of the speakers’ identity and they also suggest a triadic system, composed, in order of formality by “tu”, “você”, and “o senhor”, all of them followed by a third person inflectional morphology, whereas “tu” followed by the canonical inflection competes with “você” for the intermediate position. Among the sociolinguistic factors that influence such judgments, only age was relevant, since participants who were 50 years old or older tended to consider sentences with “tu” more informal, regardless of the inflection associated to it.

**Keywords:** second person pronouns, formality, Brazilian Portuguese; Gaúcho Portuguese, sociolinguistics.

Recebido em 20 de janeiro de 2021

Aceito em 08 de fevereiro de 2021

## 1 Introdução

Quando se observa o paradigma dos pronomes pessoais no português brasileiro, nota-se que, em geral, o “tu” (pronome canônico de segunda pessoa do singular) foi substituído pelo pronome inovador “você”<sup>1</sup> (ou compete com ele) em diversas regiões do país, acabando este por ser a forma preferida em referência ao interlocutor. No entanto, na região mais ao sul do Brasil, observa-se o oposto: o “tu” é o pronome de eleição da maioria dos falantes. Nesse dialeto, o qual será referido como português gaúcho (PBRS) neste artigo, a flexão verbal que acompanha o pronome de segunda pessoa “tu” pode ocorrer ora associado à flexão verbal de 2ª pessoa do singular (flexão canônica), ora associado à flexão verbal de 3ª pessoa (BECHARA, 1999; CAVALHEIRO, 2016; CUNHA;

---

<sup>1</sup> Segue-se a posição de Rocha Lima (2011, p. 386), que define “você” como pronome pessoal e não como forma pronominal ou substantivada de tratamento.

CINTRA, 1984; MENON; LOREGIAN-PENKAL, 2002; SCHERRE *et al.*, 2015; dentre outros). Essa diferença parece ter levado a emergência de uma diferenciação na interpretação da utilização das flexões de 3<sup>a</sup> e 2<sup>a</sup> pessoa: esta incorpora um traço de formalidade (CAVALHEIRO, 2016; LOREGIAN-PENKAL, 2004; SCHERRE *et al.*, 2015).

Tendo em vista essa questão, este estudo tentou elucidar se as diferentes formas de referência à segunda pessoa possuem níveis de formalidade distintos pelos seus falantes por meio de uma tarefa de juízo de formalidade. Além disso, obtiveram-se dados em relação à percepção dos falantes quanto ao pronome “tu” como parte da sua identidade linguística. Assim, este estudo traz um maior detalhamento da linguagem utilizada pelos falantes desse dialeto, sendo inovador no campo metodológico por não haver, até então, nenhum estudo que tenha tido uma abordagem de avaliar a questão da formalidade associada às diferentes formas de referência à segunda pessoa do singular no PBRS e nem a questão do pronome “tu” como parte da identidade gaúcha.

As próximas secções darão conta de apresentar a teoria por detrás desse fenómeno. A secção seguinte trará, resumidamente, alguns pontos relevantes da Teoria Variacionista dentro do escopo da Sociolinguística, nos moldes de Labov ([1972] 2008). Em seguida, abordar-se-ão algumas questões relativamente ao uso das formas “tu” e “você” no português brasileiro (PB) (secção 3) para então centrar-se na questão das diferentes formas de referência à segunda pessoa do singular na variedade em questão (secção 4). A partir disso, serão expostas as questões de investigação e hipóteses (secção 5), além da metodologia utilizada (secção 6). Por fim, são apresentados os resultados obtidos (secção 7) a discussão deles (secção 8) e as considerações finais (secção 9).

## **2 A Sociolinguística Variacionista**

Apesar da imprecisão que o termo Sociolinguística ainda possa trazer, Monteiro (2000) propõe duas perspectivas para o seu estudo. Uma primeira abordagem foca-se nas relações entre a sociedade e as línguas como uma unidade e volta-se a questões como políticas linguísticas em comunidades multilingues, por exemplo. A segunda abordagem, mais relevante para este trabalho, tem como objetivo relacionar os fatores sociais e as estruturas linguísticas, ou seja, o resultado proveniente da

interação dos dois: as formas como esta acaba por ser condicionada pelos fatores sociais em questão.

Essa segunda abordagem, comumente referida como Sociolinguística Variacionista (ou Teoria da Variação), preocupa-se em levantar dados que possam explicar a estrutura da língua a partir do seu uso (LABOV, [1972] 2008). De acordo com os sociolinguistas, a língua é composta por estruturas heterogêneas, as quais ocorrem sistematicamente de acordo com características sociais do falante (WEINREICH; LABOV; HERZOG, 2006, p. 101). Assume-se que não é possível compreender como a variação e a mudança linguística ocorrem sem que se considere a vida social da comunidade, visto que a linguagem sofre pressões sociais de modo contínuo (COAN; FREITAG, 2010). É nesse âmbito que Labov, o grande expoente desse quadro teórico, realiza os seus estudos iniciais relativamente a variáveis fonológicas, constatando uma forte correlação entre variáveis sociais e variáveis linguísticas.

O trabalho pioneiro de Labov foi realizado na comunidade da ilha de Martha's Vineyard. O autor constatou que os falantes nativos da ilha, incomodados com a presença dos turistas, “exageravam” a pronúncia de certos ditongos de forma a demarcar a sua pertença àquele território frente aos “estranhos” (nomeadamente turistas durante a época de verão). Dessa forma, uma pressão social externa era um fator condicionante à realização de uma pronúncia específica e que os demarcava socialmente (LABOV, [1972] 2008).

Weiner e Labov (1977) expandem os estudos para variáveis sintáticas e, segundo Paredes (1993, p. 885), quando o quadro teórico em questão passa a avaliar fenómenos sintático-discursivos, foi-se capaz de considerar que motivações fora da estrutura da língua, as quais decorrem de necessidades comunicativo-funcionais, eram a origem da variação. Destarte, é no processo de perceber melhor como que fatores (nomeadamente sociais) intervêm no uso de uma ou outra variante que a análise sociolinguística se insere. A interação entre a fala e a sociedade torna-se o seu objeto de estudo “na busca de estabelecer uma sistematização ao processo de variação linguística” (SALOMÃO, 2011).

Ainda, é importante chamar à atenção o facto de que a variabilidade é inerente ao processo de mudança linguística e, a partir dela, consegue-se extrair generalizações acerca da mudança (WEINREICH; LABOV; HERZOG, 2006). Salienta-se que a mudança possui estágios e não é a simples troca de um elemento por outro – envolve momentos de

concorrência, ou seja, em que as duas formas coexistem com o mesmo valor e há a possibilidade de que, eventualmente, uma delas se torne obsoleta (FARACO, 2005). Assim, como bem pontua Salomão (2011), os pressupostos teóricos dessa teoria “permitem ver regularidade e sistematicidade por trás do aparente caos da comunicação no dia-a-dia, procurando demonstrar como uma variante se implementa na língua ou desaparece”, ou seja, como mencionado anteriormente, é possível estabelecer que fatores condicionam o uso de uma variante ou outra.

### **3 Pronomes de Segunda Pessoa no PB**

Conforme a introdução, na variedade brasileira do português, o pronome “tu” tende a ser substituído pela forma inovadora “você” e diversos estudos têm abordado tal fenômeno. Sob uma perspectiva diacrônica, alguns estudos objetivaram analisar esse processo de “popularização” do pronome “você” no PB por meio de material escrito de diversos séculos. O trabalho de Barcia (2006) apoiou-se em cartas de jornais do século XIX e o de Lopes (2006) em cartas pessoais dos séculos XVIII e XIX, assim como o de Menon (2000), sendo, neste caso, de escritores brasileiros do fim do século XIX e início do século XX. Pelos dados obtidos, é a partir do fim do século XIX que a forma “você” passa a ocorrer em situações antes reservadas ao pronome “tu” e, no início da década de 1930, as ocorrências deste praticamente desaparecem. Entretanto, como bem aponta Menon (2000), o facto de uma forma surgir em registos escritos sugere que o seu uso oral já está presente há mais tempo no sistema.<sup>2</sup>

Assim, “você” acaba por substituir a forma canónica “tu”, passando a ser mais recorrente em tratamentos solidários e íntimos, ou seja, em atos diretivos simétricos (i.e., sem relações de poder entre os interlocutores) (YACONVENCO; SCARDUA, 2017), com a eventual sobrevivência de outras formas relacionadas à 2ª pessoa, como o pronome clítico “te” e o possessivo “teu”, mesmo quando o pronome sujeito

---

<sup>2</sup> Porém, uma reconstituição dos factos que levaram a uma generalização do “você” no PB é difícil, visto a inexistência de registos que atestem as formas linguísticas presentes no Brasil durante o início da sua ocupação pelos portugueses (FARACO, 1996). Ver o referido autor para uma reconstituição hipotética do período colonial e do impacto que poderá ter deixado na língua, além de Biderman (1972-1973).

utilizado é “você” (BIDERMAN, 1972-1973; LOPES, 2008; SILVA, 1982). Não obstante, Rumeu (2013), sobre o uso da forma “você” no século XX, considera que é possível um resquício de formalidade atrelado a essa forma. Lopes *et al.* (2009) apontam para a variação entre os dois pronomes no Rio de Janeiro, sendo “você” o mais frequente e uma forma não marcada, pois o “tu” é utilizado para expressar maior intimidade ou proximidade para com o interlocutor (ou seja, enquanto “tu” apresenta os traços referidos, “você” não, podendo ser utilizado em qualquer contexto).<sup>3</sup> Biderman (1972-1973) menciona que a forma “você” também se inseriu em contextos de uso da forma “o senhor” (no trato com superiores íntimos, i.e., pais e avós). Ainda assim, reconhece que, de facto, os pronomes de 2ª pessoa no PB são o “você” e “o senhor”, os quais representam a distinção T/V de Brown e Gilman (1960).<sup>4</sup>

No PB atual, a variação diatópica dos pronomes é clara: a depender da região do país, privilegia-se uma ou outra forma pronominal, mas ambas ocorrem em todas as regiões com proporções diferentes. Lopes e Cavalcante (2011) consideram a existência de três zonas dialetais em relação ao uso das duas formas: uma em que “você” é exclusivo, outra em que “tu” é exclusivo e uma terceira em que ambas ocorrem em variação.<sup>5</sup> O compilado de Cardoso *et al.* (2014) mostra a tendência à preferência por “você” em todas as capitais estaduais analisadas, com exceção de três: São Luís (Maranhão), Florianópolis (Santa Catarina) e Porto Alegre (RS), estando as duas últimas localizadas no extremo sul do Brasil.

Nas zonas em que a forma “tu” é utilizada, há outra questão a ter-se em conta: a morfologia flexional associada a ela. Segundo Scherre *et al.* (2015), ainda que o pronome “tu” tenha sido preservado, vem acompanhado na maioria das vezes pela flexão de 3ª pessoa. Menon (1995) considera que os falantes têm o morfema Ø como o morfema flexional de 2ª pessoa e a variação está na eleição entre “tu” e “você”,

<sup>3</sup> Lopes (2008) constata uma maior tendência à utilização de “você” por falantes de classes mais altas no Rio de Janeiro e variação dentre os de classes mais baixas.

<sup>4</sup> Os autores consideram que em muitas línguas há duas formas de referência à 2ª pessoa, as quais dependem da dimensão de poder, assimétrica (não permitindo o uso da mesma forma de tratamento entre os interlocutores), e da dimensão de solidariedade, simétrica (permitindo o uso da mesma forma).

<sup>5</sup> Na verdade, em nenhuma área o uso é exclusivo, sendo preferível utilizar o termo “predominante”.

a qual é determinada pela sua variedade dialetal. No mesmo caminho, Jensen (1981) diz que a forma “tu” é simplesmente outro item lexical, e não um paradigma inteiro. No entanto, ainda que em taxas muitíssimo mais baixas, a flexão de 2ª pessoa também ocorre.

Assim, o que alguns autores têm proposto é que o sistema pronominal do PB é diádico, composto por uma forma [+formal] (“o senhor”) e por uma [-formal] (“você”, a qual pode alternar com “tu” a depender dos fatores anteriores) (cf. BIDERMAN, 1972-1973; HEAD, 1976). Todavia, há indícios de que, pelo menos para variedades em que o “tu” é a forma predominante, o sistema seria triádico – ou até mesmo quadriádico (cf. LOREGIAN-PENKAL, 2004; RAMOS, 1989), mas os contextos em que uma forma ou outra ocorre (ou pode ocorrer), ainda devem ser mais bem investigadas.

#### 4 Os pronomes de segunda pessoa do singular no PBRs

Como descrito anteriormente, no território brasileiro, a eleição entre “tu” e “você” como a forma de referência ao interlocutor é variável. Há zonas em que o uso de “tu” ocorre em taxas muito altas, nomeadamente no extremo sul do Brasil, zona que compreende o PBRs, podendo até mesmo ser quase categórico como demonstrado por Amaral (2003). De acordo com Scherre *et al.* (2015, p. 140),

(...) tudo indica que a forma “você” tende a ser sentida pelo gaúcho como uma expressão pouco característica da cultura do Rio Grande do Sul, embora seja registrada em percentuais que variam de 5% a 15% na pesquisa de Loregian-Penkal (2004, p. 138). Na comunidade gaúcha ou de cultura gaúcha, o uso do “tu” é percebido e avaliado como mais natural.

Segundo o Mapa 5 do Atlas Linguístico-Etnográfico da Região Sul do Brasil (ALTENHOFEN, 2002, p. 144), é possível observar-se a clara predominância do pronome “tu” no Rio Grande do Sul (75% de ocorrências). Os excertos abaixo retirados de Loregian-Penkal (2004),<sup>6</sup> a partir de dados orais de falantes do PBRs, exemplificam tal fenômeno:

---

<sup>6</sup> Mantiveram-se os grifos da autora e a ortografia utilizada.

- (1) aí convidou: «Olha, Cláudio, quem sabe *tu* aceita, né? sai com ele”. (p. 119)
- (2) Bom, é isso que *tu* querias sobre vinho? (p. 98)
- (3) a irmã Maria Isabel era a minha professora de português e na época ela disse: “Não, *você* vai entrar porque *você* tem facilidade”. (p. 87)

Consoante os outros estados da Região Sul, um uso mais equilibrado de “tu” é observado em Santa Catarina. Esta zona, aparentemente, funciona como uma área de transição, pois, no Paraná, o pronome de eleição é o “você”,<sup>7</sup> havendo maior ocorrência do “tu” no litoral e no oeste catarinenses, sendo esta uma zona de forte migração de gaúchos (p. 125).

A possibilidade de variação entre as duas formas no PB parte do pressuposto de que elas não existem em distribuição complementar, ou seja, possuem o mesmo valor, não havendo “distinção de grau de cortesia ou de maior ou menor familiaridade nas localidades em que ambas as formas se encontram em variação” (NASCIMENTO; MENDES; DUARTE, 2018).<sup>8</sup> Esta afirmação surge da comparação do uso das formas em português europeu (PE), variedade na qual o pronome “tu” é utilizado em situações que demonstram uma relação de proximidade e intimidade entre os interlocutores (i.e., entre amigos ou colegas). Como bem expressam Nascimento, Mendes e Duarte (2018, p. 249),

[o] tratamento por tu tende a ser recíproco, isto é, simétrico, entre interlocutores com idade e posição hierárquica semelhante. [...] Depende ainda de hábitos familiares e sociais: embora esteja hoje em dia generalizado o uso de tu entre pais e filhos, em certas famílias as pessoas mais velhas tratam por tu os mais jovens.

Quanto ao uso de “você” em PE, observa-se uma maior complexidade. Segundo Cunha e Cintra (1984), na língua padrão, em uma

---

<sup>7</sup> Loregian (1996), por exemplo, não encontra nenhuma ocorrência de “tu” em Curitiba (PR).

<sup>8</sup> As autoras mencionam que uma alternativa ao uso dos pronomes é o sujeito nulo com a forma verbal em 3ª pessoa do singular quando os falantes não sabem que forma de tratamento utilizar para referir-se ao interlocutor (“tu/você” ou uma forma de tratamento mais formal como “o(a) senhor(a)”).

interação entre pessoas cultas, o uso de tal forma é possível quando existe uma relação social de igualdade ou de superioridade (i.e., a interação entre um patrão e um dos seus empregados). Os autores continuam o raciocínio dizendo que também pode ser interpretado como uma forma de tratamento pouco cortês, podendo ser até mesmo ofensiva, pelo que se prefere o uso de outras formas nominais como “o senhor” ou o nome próprio do ouvinte, por exemplo.<sup>9</sup> Destarte, por ter uma conotação ofensiva para muitos falantes, tal pronome é evitado.

Como referido anteriormente, o pronome “você” ocorre, ainda que em proporções baixas, no PBRS. Loregian (1996) comenta que, de acordo com os resultados obtidos na sua pesquisa, feita com habitantes da capital do estado, Porto Alegre, no âmbito do Projeto VARSUL,<sup>10</sup> a variação existe na comunidade, mas não nos indivíduos (18 sujeitos utilizaram somente “tu”, um utilizou somente “você” e cinco utilizaram ambas as formas nos seus enunciados). Os resultados de Menon e Loregian-Penkal (2002) (também com falantes de Porto Alegre) foram similares, tendo 14 sujeitos utilizado o pronome “tu”, um somente o pronome “você” e nove fizeram uso de ambas as formas. Já Loregian-Penkal (2004) constatou 34 sujeitos que utilizaram somente “tu”, um somente “você” e 30 ambas as formas. Este estudo foi mais amplo, com dados coletados em quatro cidades diferentes. Mais recentemente, Fleck e Simioni (2017), numa forma de trazer à sala de aula questões de sociolinguística, mais especificamente a variação das formas de segunda pessoa do singular, realizou um projeto com a ajuda de alunos de uma escola de Itaquí, na fronteira com a Argentina, observando o uso de “tu”

---

<sup>9</sup> É importante ressaltar que, também em Portugal, há zonas em que o uso do pronome “você” tem vindo a ser alargado, nomeadamente entre membros de classes menos cultas. Nascimento, Mendes e Duarte (2018) sobre isto dizem que seja possível que a utilização desse pronome corresponda a uma alternativa para o tratamento na 2ª pessoa do singular em conversas que, apesar de informais, não apresentam intimidade entre os interlocutores, um reflexo das mudanças da sociedade “no sentido da perda de formalidade na comunicação corrente” (p. 251). Também atribuem este fenómeno à influência do PB, nomeadamente por meio das telenovelas brasileiras e ao grande número de brasileiros em Portugal.

<sup>10</sup> Este projeto é constituído por gravações efetuadas nas três capitais dos estados sulistas, além de mais três cidades representativas das etnias mais importantes na sua colonização. É concebido e desenvolvido por pesquisadores desses estados, vinculados a universidades situadas nos mesmos.

por 18 sujeitos da pesquisa e o uso de “você” por 17. Todos os trabalhos referidos neste parágrafo fizeram uso de dados orais.

O estudo pioneiro na variação “tu”/“você” no PBRS foi conduzido por Guimarães (1979). A autora fez uso de um *corpus* composto por redações de estudantes de diversas idades, compreendendo três grupos: um grupo de estudantes no fim do primário, outro no início do secundário e um terceiro no início dos estudos universitários. Tais redações continham diálogos entre amigos (podendo o redator incluir-se como um dos amigos). Dos 120 participantes, 59 preferiram o pronome “tu”, 60 o pronome “você” e somente um utilizou ambos na mesma redação.

Outro estudo similar já data de 2016. Durante um curso de escrita<sup>11</sup> realizado na FEEVALE, universidade localizada em Novo Hamburgo, os 50 alunos tiveram de redigir uma carta pessoal, a qual serviu de amostra para o *corpus* (MARTINS; MAURER; SEVERO, 2016). Diferentemente dos resultados de Guimarães (1979) e muito mais distantes dos dados orais supracitados (FLECK; SIMIONI, 2017; LOREGIAN, 1996; LOREGIAN-PENKAL, 2004; MENON; LOREGIAN-PENKAL, 2002), as autoras constataram somente oito redações que fizeram uso exclusivo do pronome “tu”, 35 utilizaram exclusivamente o “você” e outras sete alternaram entre “tu” e “você”.

Fleck e Simioni (2017) também realizaram recolhas de dados escritos no âmbito do projeto elaborado com alunos de Itaqui mencionado anteriormente. Com um *corpus* de poucos sujeitos, os autores apontam cinco textos com a segunda pessoa canônica “tu” e nove com o pronome “você”. Por fim, analisando 13 postagens na página do restaurante universitário da Universidade Federal de Santa Maria em uma rede social, Keller e Fontana (2019) constataram o uso de “tu” sete vezes e de “você” seis vezes.

Consoante os fatores que estimulam a produção de “tu” ou de “você” em dados orais, Loregian-Penkall (2004) realizou uma extensa análise estatística das variáveis linguísticas e sociolinguísticas que podiam interferir na eleição do pronome. A autora constatou que indivíduos com mais escolaridade (pelo menos em Porto Alegre), além de mulheres e os mais jovens tendiam a utilizar mais o pronome “tu”. Outro ponto que influencia o seu uso é o gênero do discurso: receitas, argumentações e,

---

<sup>11</sup> Como pontuar um texto? – Projeto Social Lavili – Laboratório de Vivências em Linguagem, da Universidade Feevale-RS.

em menor grau, explicações estimulam-no. Por fim, a determinação do referente é outro fator estimulante. Keller e Fontana (2019), a partir dos dados escritos das postagens referidas no parágrafo anterior, consideraram o fator interlocutor, observando que a forma “tu” ocorre majoritariamente nas interações entre amigos, enquanto funcionários da universidade, ao se dirigirem aos alunos, utilizam a forma “você”. No entanto, esta forma também ocorreu na interação entre alunos, o que as autoras consideram ter que ver com o facto de estarem os indivíduos a postar em uma página da universidade. Assim, alguns podem tê-la considerado um ambiente que exigiria mais formalidade (o que explicaria os resultados).

Se se considerarem os dados orais, vemos que o pronome “tu” é a forma mais comum no PBRs, apesar de estudos mais recentes demonstrarem uma perda de espaço para o “você”. De facto, a incidência de uso tem aumentado com o tempo, principalmente em fontes de dados escritos, em comparação ao estudo de Guimarães (1979). Contudo, a comparação com os dados orais vai de encontro ao mencionado por Menon (2000), em que uma forma ao surgir em registos escritos já deveria estar presente em proporções maiores na fala há mais tempo.

Devido à forma como os dados foram obtidos, parece haver, nos participantes, uma influência do paradoxo do observador (LABOV, [1972] 2008). Ou seja, por mais que se tomem os cuidados necessários para que um ambiente informal seja criado e as produções orais sejam as mais espontâneas possíveis, o gravador e o entrevistador (provavelmente, um desconhecido) estarão sempre presentes, influenciando as escolhas linguísticas do participante (cf. RAMOS, 1989). Isto explicaria o porquê de alguns participantes optarem pelo pronome “você” nessas situações de entrevista, supondo-se que esta forma esteja associada a uma maior formalidade.

No mesmo sentido, as normas da língua também podem servir para influenciar o uso dos pronomes. Segundo Scherre *et al.* (2015), o PBRs está inserido em uma zona em que, apesar da prevalência de “tu”, este vem acompanhado da flexão de 3ª pessoa. Essa associação é, segundo a gramática normativa, agramatical. Em situações mais monitorizadas (i.e., entrevistas, redações, etc.), as estratégias para evitá-la passam, portanto, pelo uso ou de “tu” com a flexão canónica, ou de “você” (novamente, uma possível influência do paradoxo do observador). Além disso, pelo facto de os estudos com maior incidência de “você” serem aqueles que recolheram dados escritos, podem sofrer outra influência do

ensino formal: segundo Martins, Maurer e Severo (2016), os materiais didáticos privilegiam o uso de “você” – as autoras comentam não ter ciência de material que utilize a segunda pessoa do singular canônica em enunciados ou em outros momentos direcionados aos alunos.

Assim, é comum a utilização de verbos sem a desinência -s (ou -ste no pretérito perfeito do indicativo) característica das formas verbais canônicas de segunda pessoa (cf. CUNHA; CINTRA, 1984; dentre outros). Nesse sentido, Guimarães (1979) analisou a taxa de concordância verbal dos alunos que participaram na sua pesquisa, e constatou que, dentre aqueles que utilizaram o pronome “tu” nas suas redações, a concordância ocorreu em cerca de 70% dos casos, com maiores taxas dentre os participantes que estavam no ensino superior (76,79%). No entanto, nos dados orais obtidos por Loregian (1996), a concordância ocorreu em somente 4% das enunciações dos falantes de Porto Alegre. A autora chama à atenção o facto de que, no seu estudo, os participantes com maior escolaridade utilizaram menos a flexão verbal canônica do que aqueles com menor escolaridade.

Amaral (2003), que já tinha realizado um estudo semelhante na cidade de Tavares, obtendo uma taxa de concordância de 17% (AMARAL *et al.*, 1999<sup>12</sup> *apud* AMARAL, 2003), realizou outro em Pelotas. Nesta cidade, a taxa de concordância foi de 7,4%. O autor concluiu que o uso da flexão canônica<sup>13</sup> com o pronome “tu” possui prestígio na comunidade, ainda que a flexão não canônica não traga nenhum tipo de estigma social. Por fim, em Martins, Maurer e Severo (2016), no seu *corpus* escrito, nas vezes em que a forma “tu” foi utilizada, observou-se o uso da flexão canônica praticamente em todas: apenas uma das cartas analisadas não a continha. No estudo de Keller e Fontana (2019), das sete ocorrências de “tu”, duas vieram acompanhadas da flexão canônica, uma em uma interação entre amigos e outra entre um funcionário e um aluno.

---

<sup>12</sup> “Aplicação de desinência número-pessoal na 2ª pessoa do singular em Tavares, RS”. Trabalho apresentado por Luis Amaral *et al.* no 3º Encontro do CELSUL, em Porto Alegre, RS, em 1999.

<sup>13</sup> O autor considera como flexão canônica também a variante -sse para a desinência -ste do pretérito perfeito, resultante da assimilação de [t] pela fricativa [s]. Garcia (2018) analisa (também em Pelotas) se ela ocorre na escrita, e obtém quase um quarto de formas verbais com tal variante.

Loregian-Penkall (2004), por outro lado, analisou o fenômeno em mais três cidades, as quais possuem populações representativas das principais etnias do estado (italiana, alemã e espanhola), além de Porto Alegre (luso-açoriana). Os resultados, no entanto, não foram muito diferentes dos obtidos por Loregian (1996): as percentagens de uso foram de 2%, 3%, 5% e 7%, respectivamente. A autora lista os fatores que podem ser relevantes para o uso da flexão de 2ª pessoa, nomeadamente a interlocução com o entrevistador, além de discursos relatados. Quanto às características do verbo, conjugações no indicativo, principalmente no pretérito perfeito, estimulam a produção do morfema em questão. Quanto às variáveis sociolinguísticas, uma menor escolaridade parece ser um fator importante e, em menor grau, a maior idade do falante.

Assim, vemos que o pronome “tu” no PBRs possui um estatuto próprio. É visto como parte da identidade gaúcha (cf. CAVALHEIRO, 2016; LOREGIAN-PENKEL, 2004; SCHERRE *et al.*, 2015;), mas a forma “você” também ocorre (sem que se tenha estabelecido exatamente o que a motiva, apesar de os dados apontarem para a sua preferência na modalidade escrita). Além disso, o pronome “tu” do PBRs tem um funcionamento mais amplo do que o “tu” do PE (CAVALHEIRO, 2016; GOUVEIA, 2008), sendo parte de um suposto sistema diádico.

Não obstante, para alguns falantes, uma estratégia de maior formalidade é justamente a utilização da flexão de segunda pessoa canônica (cf. AMARAL, 2003; CAVALHEIRO, 2016; FLECK; SIMIONI, 2017; LOREGIAN, 1996; LOREGIAN-PENKAL, 2004; MENON, 2000). Como bem aponta Scherre (2007, p. 204) sobre a região em que se insere o PBRs,

o pronome *tu* [...] é a forma não-marcada: transita também por contextos discursivos diversos; neste caso, o uso do pronome *você* pode estar sujeito a forte monitoração. Nas situações formais, aumenta-se, muitas vezes, a concordância com o *tu*.

Ou seja, aparentemente, pode-se postular uma espécie de sistema triádico composto por “tu” com flexão de 3ª pessoa, “tu” com flexão canônica e “o(a) senhor(a)”. Segundo Amaral (2003), a maior frequência de utilização da flexão canônica ocorre durante a interlocução com o entrevistador (dado também observado por Loregian-Penkall (2004)). É possível que no PBRs o uso do “tu” com a flexão canônica funcione como uma espécie de registo intermédio entre o informal, que seria

caracterizado pelo uso do “tu” sem a flexão de 2ª pessoa do singular, e o formal, caracterizado pelo uso de outras formas de tratamento, tais como “o(a) senhor(a)”. Já a forma “você” substituiria tanto o “tu” sem flexão, quanto o “tu” com flexão, sendo o seu uso condicionado mais por fatores estilísticos (i.e., os mencionados anteriormente).

Consoante os fatores sociais, Loregian-Penkal (2004) atribui à maior formalidade dos falantes mais velhos o facto de que utilizam mais a forma “você” do que os mais jovens (ou seja, são sensíveis ao registo criado pelo uso das diferentes formas). No entanto, ao mesmo tempo, utilizam a flexão verbal canónica em maiores taxas, o que parece dar suporte à ideia de que a flexão canónica está associada a um maior grau de formalidade. Cavalheiro (2016) menciona a escolaridade como um dos fatores para a interpretação supramencionada. Para ela, quanto mais escolarizado for o falante, maior é a relação que ele faz com o registo mais formal imposto pela utilização da flexão canónica. Não obstante, Loregian-Penkal (2004) também observou a relevância da escolaridade na questão do pronome “tu” no PBRs e constatou que uma maior escolaridade promove o uso de “tu” (em contraponto a “você”), mas é uma menor escolaridade que afeta o uso da flexão canónica.

Note-se que o exposto até aqui considera dados (e conclusões a partir deles) de produção (escrita ou oral). Não há, aparentemente, estudos que tenham observado a questão da interpretação dessas formas. Este estudo prestou-se justamente a isso. Os seus objetivos e hipóteses serão detalhados a seguir.

## 5 Questões de investigação e hipóteses

A partir do exposto nas secções anteriores, estabeleceram-se as seguintes questões de investigação:

*Questão 1:* Os falantes do PBRs consideram o pronome “tu” como traço característico do seu dialeto?

*Questão 2:* Os falantes do PBRs têm percepções distintas de formalidade consoante as formas de referência à 2ª pessoa do singular?

*Questão 3:* Os falantes do PBRs têm percepções distintas de formalidade consoante a forma verbal associada ao pronome “tu”?

*Questão 4:* Que fatores sociais influenciam na interpretação de formalidade das formas, havendo diferenças entre elas?

As hipóteses são, portanto, as seguintes:

*Hipótese 1:* O pronome “tu” é visto como traço característico do seu dialeto e, por extensão, da identidade gaúcha.

*Hipótese 2:* Os pronomes observam a seguinte hierarquia de formalidade: “tu” + 3ª pessoa > “você” > “tu” + 2ª pessoa > o senhor.

*Hipótese 3:* Os fatores sociais “escolaridade”, “renda”, “gênero” e “idade” influenciam nos julgamentos de formalidade das formas.

## 6 Metodologia

No total, 283 participantes dispuseram-se a participar no estudo. Somente foram considerados aqueles com o seguinte perfil: indivíduos que moravam no RS e que lá tinham vivido durante a primeira infância e durante a maior parte da sua infância e adolescência; indivíduos com pelo menos um dos pais com vivência no RS durante a maior parte da sua infância e adolescência; indivíduos maiores de 18 anos de idade no dia de realização do teste; indivíduos que aceitaram fazer parte deste estudo ao concordar com o Termo de Consentimento Livre e Esclarecido. Foram excluídos os indivíduos que, apesar de preencherem os critérios acima, tivessem morado em outro estado brasileiro ou país por mais de três meses nos dois anos que antecederam a realização do inquérito.

Ao todo, 50 participantes foram excluídos e os dados de 233 foram analisados. A Tabela 1 apresenta o perfil dos participantes.

TABELA 1 – Perfil sociolinguístico dos participantes. RS = Rio Grande do Sul.

Faixa etária (em anos)	18 – 29 = 50 (21,5%) 30 – 39 = 99 (42,5%) 40 – 49 = 43 (18,5%) 50 ou Mais = 41 (17,6%)	
Gênero	Feminino = 179 (76,8%) Masculino = 54 (23,2%)	
Escolaridade	Ensino Médio = 45 (19,3%) Ensino Superior = 90 (38,6%) Ensino Pós-Graduado = 98 (42,1%)	
Renda (média familiar)	Até 4 Salários = 57 (24,4%) 4 a 7 Salários = 65 (27,9%) 7 a 10 Salários = 45 (19,3%) 10 Salários ou Mais = 51 (21,9%) Não Respondeu = 15 (6,5%)	
Vivência em outro estado/país por até 3 meses durante os últimos dois anos	Não = 229 (98,3%) Sim = 4 (1,7%)	
Local de criação dos pais	Somente no RS = 219 (94%) Parcialmente no RS = 14 (6%)	
Outras L1 além do português*	Não = 193 (82,8%) Alemão = 19 (8,2%) Italiano = 13 (5,3%) Espanhol = 6 (2,6%) Guarani = 1 (0,4%) Inglês = 1 (0,4%) Ucraniano = 1 (0,4%)	
L2 (independentemente do nível de proficiência alcançado)	Não = 45 (19,3%) Inglês = 173 (74,2%) Espanhol = 87 (31,3%) Alemão = 29 (12,4%) Francês = 19 (8,2%) Italiano = 17 (7,3%)	Mandarin = 2 (0,9%) Russo = 2 (0,9%) Japonês = 1 (0,4%) Latim = 1 (0,4%) Grego = 1 (0,4%)

\*Uma pessoa indicou duas L1 além do português (inglês e alemão)

Fonte: elaboração própria.

Esta pesquisa apoiou-se em um formulário produzido na ferramenta Google Formulários (acedido por um *link* fornecido pelo investigador a quem se dispusesse a participar por meio eletrônico; também se o disponibilizou em redes sociais com a devida explicação da pesquisa). O *link* ficou disponível do dia 15 de novembro ao dia 15 de dezembro de 2020.

Todos os passos para a realização das tarefas estavam detalhados no formulário. A primeira parte continha o Termo de Consentimento Livre e Esclarecido, o qual devia ser aceite para que se pudesse dar

prosseguimento à secção seguinte. Se o participante se recusasse a participar, veria uma página de agradecimento. O participante podia também decidir fechar a janela do navegador a qualquer momento, o que acarretava na cessação da sua participação – os dados só eram efetivamente gravados aquando da submissão do formulário. Ao aceitar, era levado à página com o questionário sociolinguístico, cujos dados recolhidos já foram apresentados na Tabela 1.

A primeira tarefa a ser realizada consistia na seleção de vocábulos que os participantes considerassem como marcadores do jeito de falar do gaúcho. Dentre esses vocábulos estavam o “tu” e o “você”, de modo que se percebesse se tais formas são vistas como parte do dialeto.

Em um primeiro momento, praticavam com cinco vocábulos, somente para que entendessem como a tarefa era realizada. Ao passarem à próxima secção, viam as 14 opções para seleção, dentre os quais estavam “*tu (falando com outra pessoa)*” e “*você (falando com outra pessoa)*”, além da opção “*nenhuma das anteriores*”. Após a conclusão desta etapa, eram direcionados à Tarefa 2.

Para a tarefa de juízo de formalidade que compunha a segunda tarefa, reuniram-se 70 frases distribuídas nos seguintes contextos (seguidos de um exemplo de item): 10 frases com o pronome “tu” associado à flexão de 2ª pessoa do singular: “*Tu entendes a minha letra?*”; 10 frases com o pronome “tu” associado à flexão de 3ª pessoa do singular: “*Tu entende muito disso.*”; 10 frases com o pronome “você”: “*Você achou as minhas chaves.*”; e 10 frases com o pronome “o senhor”: “*O senhor fala francês muito bem.*”, além de 30 distratores com estruturas diferentes das anteriores: “*A Vera não pode vir hoje.*”

Todos os itens de teste possuíam o sujeito expreso e incluíam tanto frases declarativas, quanto interrogativas. Tomou-se o cuidado de utilizar vocabulário diverso e, nos itens avaliados, tomou-se também o cuidado de não utilizar léxico que pudesse ter uma leitura mais ou menos formal para que a leitura da formalidade fosse feita somente por meio do pronome utilizado. Antes do início da execução da tarefa, os participantes receberam as instruções e, em seguida, julgaram duas frases de 1 a 5 (sendo 1 informal e 5 formal) como treino, as quais não continham nenhum dos pronomes estudados.

Após o seu completamento, iniciava-se o teste. As 70 frases estavam divididas em cinco páginas com 14 frases em cada uma (1/5 do total de cada contexto). Todas as perguntas ocorriam de modo aleatório

cada vez que o formulário era acessado e a escala era também de 1 a 5. Além disso, todos os itens eram de resposta obrigatória.

## 7 Resultados

Relativamente à primeira tarefa, vê-se que todos os participantes selecionaram a opção “*Tu (falando com outra pessoa)*” como parte do jeito gaúcho de falar,<sup>14</sup> o qual se pode ver nos seguintes exemplos<sup>15</sup> de uso do pronome (além dos exemplos (1) e (2) mencionados na secção 3) retirados de Loregian-Penkal (2004) (4) e de Keller e Fontana (2019) (5), a partir de dados orais e escritos, respetivamente:

- 4) antes *tu* poderia encontrá isso num bairro assim bem mais pobre assim, né?
- 5) *tu* podes transferir aqueles créditos que *tu* não usas, para mim, por exemplo.

Somente cinco deles selecionaram a opção “*Você (falando com outra pessoa)*”, cujo uso a partir dos dados escritos de Keller e Fontana (2019) é exemplificado abaixo (e também pelo exemplo (3) mencionado na secção 3):

- 6) Se *você* possui a carteira antiga do RU *você* precisa registrar perda para depois solicitar a carteira nova.

Os resultados da Tarefa 1, além daqueles obtidos com os outros vocábulos<sup>16</sup> estão expressos na Figura 1.

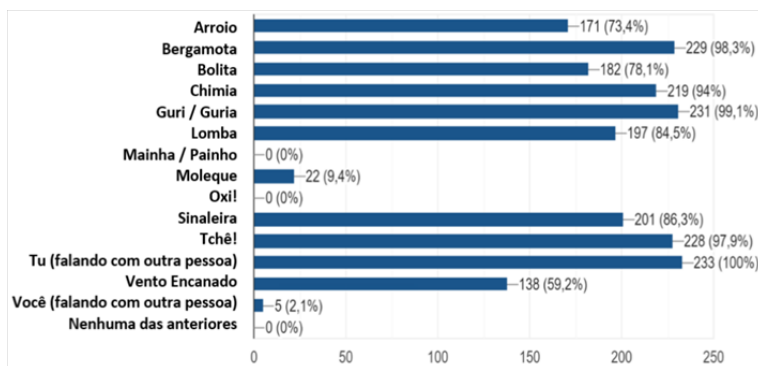
---

<sup>14</sup> Dois dos participantes excluídos, por falha da plataforma, procederam à execução das tarefas e não selecionaram a opção “*Tu (falando com outra pessoa)*”. Destes, um selecionou apenas uma opção (“*Tché!*”), o que pode sugerir que não compreendeu que podia selecionar mais de uma opção.

<sup>15</sup> Mantiveram-se os grifos das autoras somente nos pronomes. A ortografia utilizada é a original.

<sup>16</sup> Os dados dos outros vocábulos não serão abordados, pois não fazem parte do escopo deste trabalho, mas são divulgados, já que podem servir para estudos futuros. Não obstante, vale ressaltar-se que os dois que obtiveram 0% são tidos como parte dos dialetos do nordeste brasileiro. O vocábulo “*moleque*” é mais comum no sudeste. Os outros ocorrem no dialeto gaúcho com menor ou maior frequência de acordo com a região do estado, e alguns podem ocorrer nos outros estados do sul.

FIGURA 1 – Léxico considerado como parte do PBRS pelos falantes  
(em número total e em percentagem)



Fonte: elaboração própria.

Consoante a Tarefa 2, a média dos juízos foi, da menor à maior, de 2,28 (DP = 1,12) para as frases com o pronome “tu” associado à flexão de 3ª pessoa (T3), de 3,39 (DP = 1,00) para frases com o pronome “você” (V), de 3,42 (DP = 1,24) para frases com o pronome “tu” associado à flexão de 2ª pessoa (T2) e de 3,89 (DP = 1,01) para as frases com “o senhor” (S).

Quando se analisam os dados de acordo com os critérios sociolinguísticos alvo, vê-se que, em geral, o “T3” é considerado o mais informal (com médias que variam de 1,71 dentre os participantes com mais de 50 anos de idade até 2,52 dentre os participantes mais jovens) e “S” o mais formal (com médias que variam de 3,70 dentre os participantes com mais de 50 anos de idade até 4,05 dentre os participantes com renda entre 7 e 10 salários mínimos). Em nenhum contexto essas formas deixaram de aparecer com as médias mais baixas e altas, respetivamente.

As médias de “T2” e de “V” situam-se em uma posição intermédia na escala. As duas médias variaram, respetivamente, de 2,79 e 3,27 (ambas dentre indivíduos com mais de 50 anos) até 3,83 e 3,53 (ambas dentre os mais jovens). No entanto, não se observou uma tendência nas médias, com alternância entre qual possuía a maior a depender do contexto. A maior diferença entre elas encontrou-se dentre os indivíduos com mais de 50 anos (diferença de 0,48) e a menor dentre aqueles com maior renda (diferença de 0,03). Todos esses resultados estão expostos na Tabela 2.

TABELA 2 – Média de julgamentos por tipo de pronome e por contexto social

	<b>Tu + 3ª Pessoa</b>	<b>Você</b>	<b>Tu + 2ª Pessoa</b>	<b>O Senhor</b>
Média Geral	2,28 (DP = 1,12)	3,39 (DP = 1,00)	3,42 (DP = 1,24)	3,89 (DP = 1,01)
<b>Gênero</b>	<b>Tu + 3ª Pessoa</b>	<b>Você</b>	<b>Tu + 2ª Pessoa</b>	<b>O Senhor</b>
<i>Feminino</i>	2,27 (DP = 0,97)	3,36 (DP = 0,80)	3,46 (DP = 0,98)	3,86 (DP = 0,73)
<i>Masculino</i>	2,28 (DP = 0,81)	3,50 (DP = 0,63)	3,30 (DP = 1,07)	4,01 (DP = 0,74)
<b>Idade</b>	<b>Tu + 3ª Pessoa</b>	<b>Você</b>	<b>Tu + 2ª Pessoa</b>	<b>O Senhor</b>
<i>18 – 29 Anos</i>	2,52 (DP = 1,08)	3,53 (DP = 0,85)	3,83 (DP = 0,90)	4,05 (DP = 0,73)
<i>30 – 39 Anos</i>	2,34 (DP = 0,86)	3,41 (DP = 0,67)	3,61 (DP = 0,91)	3,89 (DP = 0,72)
<i>40 – 49 Anos</i>	2,38 (DP = 0,69)	3,31 (DP = 0,67)	3,13 (DP = 0,77)	3,91 (DP = 0,60)
<i>Mais de 50 Anos</i>	1,71 (DP = 0,79)	3,27 (DP = 0,95)	2,79 (DP = 1,18)	3,70 (DP = 0,89)
<b>Escolaridade</b>	<b>Tu + 3ª Pessoa</b>	<b>Você</b>	<b>Tu + 2ª Pessoa</b>	<b>O Senhor</b>
<i>Ensino Médio</i>	2,50 (DP = 1,13)	3,49 (DP = 0,88)	3,16 (DP = 1,17)	3,90 (DP = 0,87)
<i>Ensino Superior</i>	2,17 (DP = 0,83)	3,36 (DP = 0,78)	3,50 (DP = 0,98)	3,85 (DP = 0,77)
<i>Ensino Pós-Graduado</i>	2,27 (DP = 0,85)	3,37 (DP = 0,70)	3,48 (DP = 0,93)	3,93 (DP = 0,64)
<b>Renda<sup>17</sup></b>	<b>Tu + 3ª Pessoa</b>	<b>Você</b>	<b>Tu + 2ª Pessoa</b>	<b>O Senhor</b>
<i>Até 4 Salários</i>	2,50 (DP = 1,17)	3,39 (DP = 0,94)	3,32 (DP = 1,20)	3,87 (DP = 0,90)
<i>De 4 a 7 Salários</i>	2,34 (DP = 0,72)	3,40 (DP = 0,65)	3,71 (DP = 0,83)	3,83 (DP = 0,66)
<i>De 7 a 10 Salários</i>	2,23 (DP = 1,00)	3,48 (DP = 0,77)	3,22 (DP = 1,09)	4,05 (DP = 0,67)
<i>Mais de 10 Salários</i>	2,06 (DP = 0,72)	3,35 (DP = 0,77)	3,38 (DP = 0,88)	3,92 (DP = 0,75)

Fonte: elaboração própria.

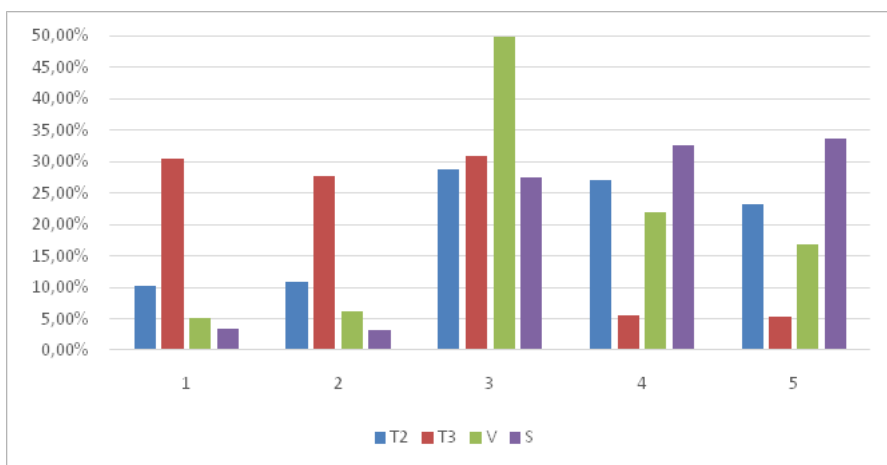
Um ponto importante a ser ressaltado é a questão do desvio padrão das médias. Como se pode ver na Tabela 2, os desvios foram altos, o que indica uma maior dispersão da amostra. Em outras palavras, o valor atribuído a cada pronome não se concentrou à volta de um valor específico.<sup>18</sup> De facto, ao serem analisadas o número de julgamentos para

<sup>17</sup> Sem os dados dos participantes que selecionaram a opção “Prefiro não responder” ( $n = 15$ ).

<sup>18</sup> Ainda assim, houve participantes que foram categóricos nas suas avaliações, totalizando 102 casos de julgamentos categóricos, distribuídos da seguinte forma: 38 com “S”, 46 com “T3”, 51 com “V” e 52 com “T2”. Note-se que houve participantes que foram categóricos em mais de um contexto.

cada valor por tipo de pronome, vê-se que todos foram julgados como 1, 2, 3, 4, ou 5 pelo menos uma vez. Contudo, observam-se tendências. Por exemplo, as frases com “T3” tenderam a ser julgadas com 1, 2 ou 3 (30,52%, 27,64% e 30,94%, respectivamente, totalizando 89,10% dos julgamentos nestes valores); as frases com “V” concentraram-se em 3, com 49,91% dos julgamentos neste valor, e em menor grau em 4 e 5 (21,93% e 16,87%, respectivamente); frases com “T2” tenderam a ser julgadas com 3, 4, ou 5 (28,67%, 27,04% e 23,26%, respectivamente); e frases com “S” tiveram a mesma tendência que a de “T2”, mas com a ordem invertida: 27,51% de julgamento no valor 3, 32,62% no valor 4 e 33,65% no valor 5. A Figura 2 exemplifica os achados.

FIGURA 2 – Percentagem de indivíduos que julgaram cada tipo de pronome em valores de 1 a 5



Fonte: elaboração própria.

Em escalas com valores ímpares, é costumeiro que haja uma maior concentração de juízos no seu centro, pois, nesses casos, não só representa uma situação neutra, como também é o valor tendencialmente selecionado quando não se consegue produzir um juízo ou se está em dúvida. Como se pode ver, à exceção de “S”, todas as outras formas foram julgadas na sua maioria como 3. Sem os julgamentos centrais, apenas a forma “T3” tem uma diminuição na média, que totaliza 1,95 (DP = 1,21). A forma “T2” obtém 3,59 (DP = 1,43), a qual é menor que a de “V”: 3,78 (DP = 1,31). “S”, por sua vez, tem média 4,23 (DP =

1,00). A Tabela 3 compara as médias originais e as médias obtidas após a eliminação dos julgamentos em 3.

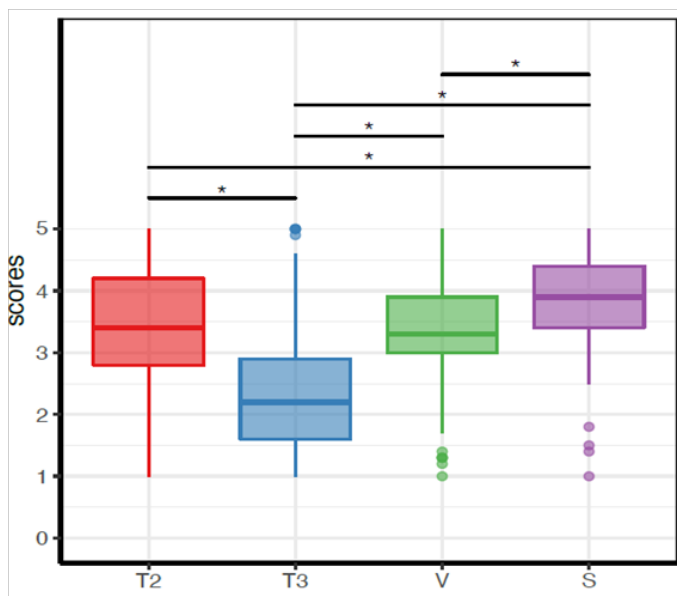
TABELA 3 – Média de julgamentos com e sem os julgamentos no valor central da escala

	Tu + 3ª Pessoa	Você	Tu + 2ª Pessoa	O Senhor
Média Geral	2,28 (DP = 1,12)	3,39 (DP = 1,00)	3,42 (DP = 1,24)	3,89 (DP = 1,01)
Média Sem Valor 3	1,95 (DP = 1,21)	3,78 (DP = 1,31)	3,59 (DP = 1,43)	4,23 (DP = 1,00)

Fonte: Elaboração própria.

Foram feitas duas análises estatísticas com os dados obtidos. A primeira verificou se cada contexto se comportava de maneira diferente dos outros. Constatou-se que todos os grupos eram diferentes entre si, exceto “V” quando comparado a “T2” (teste de Wilcoxon com valores de  $p$  ajustados pela correção de Bonferroni) (FIGURA 3).

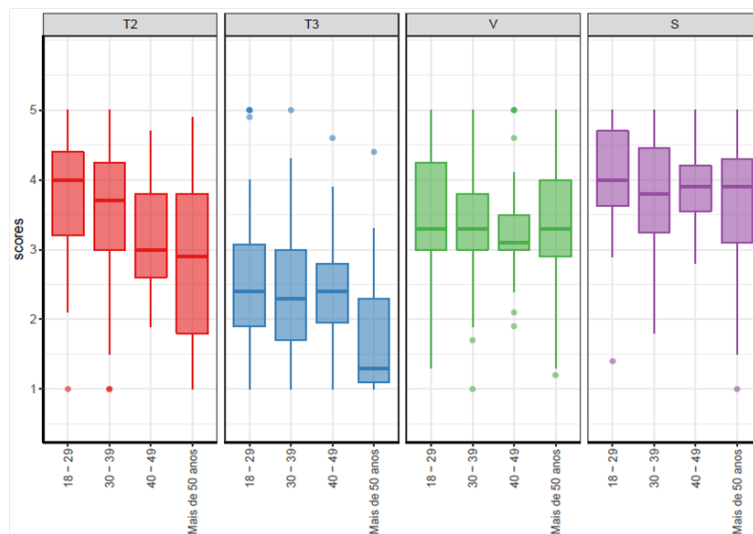
FIGURA 3 – Resultados dos testes de Wilcoxon pareados com correção múltipla de Bonferroni. \* =  $p < 0,001$



Fonte: elaboração própria.

A segunda análise procurou correlacionar os resultados com as variáveis sociais consideradas neste estudo (idade, gênero, renda e escolaridade). O teste de Wilcoxon não apontou efeitos do gênero sobre os resultados, e nem o de Kruskal-Wallis para a renda e a escolaridade. No entanto, o teste de Kruskal-Wallis encontrou efeitos da idade sobre os julgamentos das formas “T2” ( $H(3) = 29.9, p < .001$ ) e “T3” ( $H(3) = 21.4, p < .001$ ). A análise pareada mostrou que os dois grupos mais velhos se comportam de maneira diferente dos mais jovens com o pronome “tu” associado à flexão de 2ª pessoa ( $p < .01$ ), enquanto somente o grupo dos maiores de 50 anos se comporta diferente do resto consoante o pronome “tu” associado à flexão de 3ª pessoa ( $p < .01$ ) (FIGURA 4).

FIGURA 4 – Resultados do teste de Kruskal-Wallis



Fonte: elaboração própria.

A próxima seção discorrerá sobre os resultados obtidos.

## 8 Discussão

Os resultados das duas tarefas trazem importantes dados acerca do estatuto das diferentes formas de referência à 2ª pessoa no PBRs. A primeira tarefa, que consistiu na seleção de vocábulos considerados pelos

falantes como parte do “jeito gaúcho de falar”, trouxe dados empíricos acerca de uma ideia a qual muitos autores já consideravam: o pronome “tu” é visto como marca identitária da população (cf. CAVALHEIRO, 2016; MENON, 2000; dentre outros), confirmando a *Hipótese 1*. Destarte, não surpreende que tal vocábulo tenha sido selecionado por todos os participantes cujos dados foram analisados. Ao mesmo tempo, o pronome “você” só foi selecionado por cinco participantes, o que pode indicar que, para estes falantes, tal pronome já se encontra internalizado como parte característica do dialeto. Não obstante, não se pode descartar erros na compreensão da tarefa ou erros na hora da seleção dos vocábulos (afirmação válida também para a seleção do pronome “tu”).

A segunda tarefa mostrou que os falantes têm percepções distintas relativamente ao nível de formalidade dos pronomes estudados, sendo que o contexto “T3” é o menos formal (média = 2,28) e “S” é o mais formal (média = 3,89). Esse resultado tampouco surpreende. Historicamente, tanto a forma “tu” quando a forma “o senhor” têm sido atreladas a situações informais e formais, respetivamente (BARCIA, 2006; BIDERMAN, 1972-1973; LOPES, 2006; MENON, 2000; dentre outros) e, à partida, não há sinais de que isso se tenha modificado. A análise estatística não encontrou diferenças, no entanto, nas duas formas de formalidade intermédia: “T2” (média = 3,42) e “V” (média = 3,39). Logo, a *Hipótese 2* não se confirmou, porque considerava que estas duas formas seriam diferentes, com “T2” exibindo maior formalidade do que “V”.

Os resultados apontam para um sistema pronominal, pelo menos na percepção dos falantes, triádico. Do mais informal ao mais formal, “tu”, “você” e “o senhor” compõem-no, todos associados à forma verbal de 3ª pessoa do singular. A forma “T2” surge em competição com “V” pela forma intermédia. Devido ao facto de que “tu” é considerado parte da identidade gaúcha (cf. resultados da primeira tarefa), esperar-se-ia que “T2” então surgisse nos estudos de produção revistos anteriormente como a forma preferida para a situação de formalidade intermédia. Não obstante, o que se observa é o oposto.

É importante ressaltar que os dados cá obtidos são dados de interpretação, ao passo que os dados obtidos pelos estudos revistos ao longo deste artigo são dados de produção (e muitos se basearam em análises mais refinadas), pelo que uma comparação direta entre eles se

torna desigual. Assim, a partir deste estudo, não se pode obter conclusões sobre a razão de “V” ocorrer mais do que “T2”. No entanto, dados prévios sugerem que a pressão dos dialetos dominantes do PB (por meio dos materiais didáticos e dos meios de comunicação) pode ter o seu papel. A grande maioria das produções televisivas brasileiras concentram-se em São Paulo e no Rio de Janeiro, locais em que a forma “você” é a predominante, pelo menos dentre as classes média e alta (cf. LOPES, 2008; SCHERRE *et al.*, 2015) e é a forma “exportada” diariamente pelo meio televisivo a todas as partes do país.

Outra hipótese pode ter que ver justamente com a forma verbal. Ou seja, enquanto as outras três formas estudadas utilizam a mesma forma verbal, a flexão canônica consiste em morfologia flexional própria, pelo que é necessária a sua aquisição para a sua produção.<sup>19</sup> Além disto, o PB segue uma tendência de maior preenchimento do sujeito pronominal (cf. DUARTE, 1993; dentre outros) e Loregian-Penkal (2004) aponta como favorecimento da flexão canônica o sujeito nulo. Ora, se o PB vem apresentando mais sujeitos realizados, a necessidade da morfologia flexional reduz-se.<sup>20</sup> Havendo uma cisão entre o nível de formalidade atrelado à flexão canônica com o pronome “tu”, é mais natural que a forma não marcada do verbo seja priorizada, recaindo sobre o pronome (no caso, “você”) os valores semânticos de formalidade. Por fim, não se pode ignorar a associação de valores semânticos (ou até mesmo de prestígio/estigma) à forma “T2” mais além da formalidade. Assim, “T2” e “T3” seriam dois itens lexicalmente iguais, mas semanticamente distintos, com valores atribuídos a partir da flexão verbal.

Considerando ser uma tendência do PB “você” passar a abarcar os valores de informalidade outrora pertencentes ao pronome “tu”, tal processo parece estar menos avançado no PBRS (comparativamente a outros dialetos do PB), apesar de poder estar mais ou menos avançado nos indivíduos. Os dados de produção expostos durante a revisão da literatura neste artigo mostram pouca variação nos indivíduos – há mais falantes que utilizam uma forma categoricamente do que outros. Pode-

---

<sup>19</sup> No sentido em que, se não estiver presente no *input*, dificilmente será adquirida pelos falantes de modo que seja ativamente utilizada.

<sup>20</sup> De facto, vê-se que nesse empobrecimento do paradigma verbal é a forma de 3ª pessoa do singular que tende a inserir-se em pessoas e números que antes possuíam morfologia própria.

se interpretar esses dados da seguinte forma: os falantes que possuem “tu” e “você” como formas distintas estabelecem o nível de formalidade da tarefa (ou entrevista) ao início e mantêm a forma correspondente do início ao fim. Os falantes não categóricos utilizam as formas de acordo com outros fatores externos (i.e., género do discurso ou determinação do referente).

Quanto às variáveis sociais analisadas, postulava-se que o aumento da escolaridade teria alguma influência nos resultados dos falantes. Em médias de julgamentos, aqueles com mais anos de estudo, em comparação aos menos escolarizados, atribuem julgamentos mais formais a “T2”, o que pode ocorrer devido ao ambiente social pelo qual transitam ou por notarem que o seu uso acaba por ter situações específicas de ocorrência (i.e., mais formais). Quanto a “T3”, consideram-no mais informal do que os menos escolarizados o consideram, o que deve ocorrer por uma pressão da gramática normativa. No entanto, podia-se esperar que “V” fosse mais formal por estar presente em materiais didáticos e meios de comunicação impressos – a forma com que a maioria dos falantes (principalmente mais escolarizados) tem contacto com ele. As médias de julgamento, entretanto, indicam o oposto. No entanto, o teste estatístico não apontou a influência da escolaridade sobre os julgamentos, pelo que os valores das médias possam ser apenas um indício de alguma diferença, a qual, com a amostra utilizada, não mostrou ser relevante. É importante ressaltar que os falantes menos escolarizados tinham o Ensino Médio completo, ou seja, pelo menos 11 anos de escolarização, e não compunham nem 20% da amostra, pelo que, de certo modo, a amostra possa ter sido mais homogênea nesse quesito, explicando a falta de diferença entre os grupos, apesar das diferenças nas médias.

Outro fator com impacto semelhante ao da escolaridade é a renda. Por exemplo, é possível que falantes com maior renda tenham maior escolaridade, além de transitarem por ambientes em que a linguagem é mais cuidada. De facto, quanto maior renda, mais informal “T3” é considerado. Porém, são os participantes com renda entre 4 e 7 salários mínimos os que atribuem o maior nível de formalidade ao “T2”. Novamente, apesar dos valores nominais distintos, o teste estatístico não demonstrou diferença entre os resultados. Ressalta-se, também, que a análise deste contexto não pôde contemplar todos os participantes, visto que 15 deles não responderam em que faixa de renda se incluíam, o que pode ter influenciado, ainda que marginalmente, os resultados.

Quanto ao género, Loregian-Penkall (2004) notou que as mulheres em todas as localidades que estudou tendiam a preservar o pronome “tu” (i.e., mais mulheres usaram-no categoricamente) e, na cidade de Panambi, as que alternavam entre as duas formas tendiam a privilegiá-lo. Por outro lado, consoante a flexão verbal, dentre as cidades em que houve recolha de dados, a tendência a preservá-la só ocorreu dentre os homens de Porto Alegre. Neste estudo com dados de interpretação, todavia, não houve influência do género sobre os resultados. No entanto, salienta-se o facto de que a amostra continha mais de 75% de mulheres, o que pode ter contribuído para enfraquecer o poder do teste estatístico.

A única variável sociolinguística selecionada estatisticamente neste trabalho foi a idade dos participantes e só houve efeito sobre as formas de “tu”. Este resultado pode refletir os valores prévios deste pronome, até mesmo indicando uma certa diacronia da inserção de “T3”. Por exemplo, a análise estatística apontou que os falantes com idade entre 40 e 49 anos acompanham os falantes com mais de 50 anos de idade quanto à informalidade de “T2” quando comparados aos outros dois grupos de idade (18-29 anos e 30-39 anos). No entanto, já não há diferença entre os grupos dos mais jovens ao dos de idade entre 40 e 49 anos consoante “T3”. Ou seja, isso pode indicar que, para os falantes de 40 a 49 anos, a forma “T2” foi adquirida relacionada a situações informais, pelo que lhe atribuem esse valor. Já quanto ao “T3”, comportam-se como os outros participantes mais jovens, pois, provavelmente, essa forma já se teria normalizado (quem sabe com perda do estigma atrelado a ela). Por outro lado, os falantes com mais de 50 anos de idade continuam a considerar “T2” mais informal do que o resto dos falantes, já que lhe era mais comum (i.e., estava mais presente no *input* mesmo em situações informais).

O estudo mais antigo que analisou a questão da flexão verbal com o pronome “tu” foi o de Guimarães (1979), pesquisa na qual os participantes utilizaram a concordância canónica em 70% das vezes nos seus textos que envolviam interações entre amigos. Note-se que um participante no referido estudo que estivesse no início dos seus estudos universitários, teria por volta de 19 anos à época e por volta de 60 aquando da realização do presente estudo. Foram justamente os falantes de nível universitário os que mais realizaram a flexão canónica em Guimarães (1979). Esses dados, aliados aos de interpretação obtidos

neste estudo, parecem sugerir que em meados do século passado, “tu” com a flexão canônica ainda ocorria em proporções consideráveis e em situações informais no PBRs. Já com os dados orais, Loregian-Penkal (2004) também verificou que os mais velhos eram mais propícios a utilizá-la, além de preferirem a forma “você” como estratégia para maior formalidade. Os dados da tarefa de julgamento de formalidade parecem ir ao encontro dessa ideia.<sup>21</sup> Destarte, a *Hipótese 3* tampouco se mostrou válida, pois somente a variável “idade” foi selecionada como fator relevante para a interpretação do traço de formalidade.

## 9 Considerações Finais

O paradigma dos pronomes pessoais de 2ª pessoa no PB sofreu modificações ao passar dos anos. É inegável que a forma “você” o adentrou, transformando-se na forma de eleição da grande maioria dos falantes. Aliás, não é de se estranhar que se tenha a sensação de que o pronome “você” é predominante no PB: levando-se em consideração a região considerada por Scherre *et al.* (2015) com predominância daquele pronome e consultando os dados populacionais brasileiros (IBGE, 2020), 110 milhões de habitantes, ou seja, mais de 50% da população brasileira, são abrangidos pela zona em que “você” aparece com o pronome de eleição. A região com predominância do “tu” por outro lado, tem pouco mais de 30 milhões de falantes, menos de 15%. Portanto, aproximadamente 35% da população alterna entre “você” e “tu”, sendo que em grande parte das zonas de variação o pronome “você” ocorre em taxas superiores.

Os dados obtidos neste estudo mostram que os falantes interpretam claramente o valor de formalidade atrelado a cada pronome, evidenciando um sistema triádico, em que “você” compete pela forma intermédia (e não pela informal) com o pronome “tu” acompanhado da flexão canônica. Da mesma forma que nos dados de produção, a variação parece estar na comunidade e não no indivíduo, pois, como referido nos resultados, em 102 casos, falantes foram categóricos no seu julgamento da formalidade relativamente a um determinado pronome.

---

<sup>21</sup> Observe-se que as metodologias foram distintas, pelo que uma comparação *vis-à-vis* não é possível. No entanto, os dados parecem caminhar na mesma direção.

Não obstante, por se ter baseado em um teste que avaliava a interpretação dos falantes, uma das limitações deste estudo é o de não poder constatar o comportamento dos mesmos na produção dos pronomes envolvidos, por exemplo, em que situações os utilizariam e por que razão, o que só poderá ser analisado por meio de outros estudos com metodologias específicas para esse fim. Além disso, o teste era explícito quanto ao fator “formalidade” a ser julgado. No entanto, outros valores estão associados ao modo como os pronomes são utilizados (i.e., intimidade, distanciamento, respeito, etc.) que podem ou não serem interpretados como ligados à formalidade.

Por fim, sugere-se avaliar como falantes de outros dialetos (i.e., com predominância de “você”), além de falantes do PBRS cujo perfil social não tenha sido contemplado neste estudo (i.e., menores de 18 anos e adultos com baixa escolaridade) interpretam tais formas, sem deixar de lado a obtenção de dados de produção, enriquecendo o conhecimento sobre o funcionamento das diferentes variedades da língua portuguesa.

### Agradecimentos

Agradece-se à Profa. Dra. Ana Madeira e à Profa. Dra. Maria Lobo do Departamento de Linguística da Universidade Nova de Lisboa pelos muito úteis comentários dados tanto no planeamento do estudo, quanto na escrita deste artigo, e à Dra. Luísa Pilz do Laboratório de Cronobiologia e Sono do Hospital de Clínicas de Porto Alegre pela realização dos testes estatísticos. Agradece-se também a todos aqueles participaram ou contribuíram com a divulgação do *link* para a realização do teste.

### Referências

ALTENHOFEN, C. V. Áreas lingüísticas do português falado no Sul do Brasil: Um balanço das fotografias lingüísticas do ALERS. In: VANDRESEN, P. (org.). *Variação e mudança no português falado da Região Sul*. Pelotas: Educat, 2002. p. 115-145.

AMARAL, L. I. *A concordância verbal de segunda pessoa do singular em Pelotas e suas implicações lingüísticas e sociais*. 2003. 181f. Tese (Doutoramento em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

BARCIA, L. R. *As formas de tratamento em cartas de leitores oitocentistas: peculiaridades do gênero e reflexos da mudança pronominal*. 2006. 142f. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

BECHARA, E. *Moderna gramática portuguesa*. Rio de Janeiro: Lucerna, 1999.

BIDERMAN, M. T. Formas de tratamento e estruturas sociais. *Alfa*, São Paulo, v. 18/19, p. 339-382, 1972-1973.

BROWN, R.; GILMAN, A. The Pronouns of Power and Solidarity. In: SEBEEK, T. A. (org.). *Style in Language*. Cambridge: MIT, 1960. p. 253-276.

CARDOSO, S. A. *et al. Atlas Linguístico do Brasil*. Londrina: EDUEL, 2014. v. 2.

CAVALHEIRO, V. M. *As diferentes regras de uso das formas tu e você e suas influências na compreensão de narrativas literárias*. 2016. 338f. Tese (Doutoramento em Ciências da Linguagem) – Faculdade de Letras, Universidade do Porto, Porto, 2016.

COAN, M.; FREITAG, R. M. Sociolinguística variacionista: pressupostos teórico-metodológicos e propostas de ensino. *Domínios da Linguagem*, Uberlândia, v. 4, n. 2, p. 173-194, 2010.

CUNHA, C.; CINTRA, L. *Nova gramática do português contemporâneo*. Lisboa: Livraria Sá da Costa, 1984.

DUARTE, M. E. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: ROBERTS, I.; KATO, M. A. (org.). *Português Brasileiro: uma viagem diacrônica (Homenagem a Fernando Tarallo)*. Campinas: Editora da UNICAMP, 1993. p. 107-128.

FARACO, C. A. O tratamento você em português: uma abordagem histórica. *Fragmenta*, Curitiba, n. 13, p. 51-82, 1996.

FARACO, C. A. *Linguística Histórica: uma introdução ao estudo da história das línguas*. São Paulo: Parábola Editorial, 2005.

FLECK, L. S.; SIMIONI, T. Para uma pedagogia da variação no ensino médio: pesquisa sociolinguística sobre pronomes pessoais. *Diadorim*, Rio de Janeiro, v. 19, n. 2, p. 190-208, 2017.

GARCIA, C. F. *A variação linguística no uso da desinência verbal -sse na segunda pessoa do singular na escrita dos pelotenses*. 2018. 44f. Trabalho de Conclusão de Curso (Graduação em Letras) – Faculdade de Letras, Universidade Federal de Pelotas, Pelotas, 2018.

GOUVEIA, C. As dimensões da mudança no uso das formas de tratamento em Português Europeu. In: OLIVEIRA, F. D (org.). *O fascínio da linguagem*: actas do colóquio de homenagem a Fernanda Irene Fonseca. Porto: CLUP/FLUP, 2008. p. 91-100.

GUIMARÃES, A. M. *A ocorrência de 2ª pessoa*: estudo comparativo sobre o uso de tu e você na linguagem escrita. 1979. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, 1979.

HEAD, B. J. Social factors in the use of pronouns for the addresses in Brazilian Portuguese. In: SCHMIDT-RADEFELDT, J. (ed.). *Readings in Portuguese Linguistics*. Amesterdão: North-Holland Publishing Company, 1976. p. 289-343.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Estimativa Populacional, 2020*. Brasília: IBGE, 2020. Disponível em: [ftp://ftp.ibge.gov.br/Estimativas\\_de\\_Populacao/Estimativas\\_2020/estimativa\\_dou\\_2020.xls](ftp://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2020/estimativa_dou_2020.xls). Acesso em: 7 jan. 2021.

JENSEN, J. B. *Forms of Address in Brazilian Portuguese*: Standard European or Oriental Honorifics? Amersterdão: John Benjamins, 1981. DOI: <https://doi.org/10.1075/z.13.06jen>

KELLER, T.; FONTANA, P. O uso de tu e você na posição de sujeito em posts de fan page do Facebook do restaurante universitário da UFSM. *Cadernos do Instituto de Letras*, Porto Alegre, v. 59, p. 220-240, 2019. DOI: <https://doi.org/10.22456/2236-6385.92314>

LABOV, W. *Padrões Sociolinguísticos*. Trad. Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues de Oliveira. 2. ed. São Paulo: Parábola Editorial, [1972] 2008.

LOPES, C. R. Correlações histórico-sociais e lingüístico-discursivas das formas de tratamento em textos escritos no Brasil – séculos XVIII e XIX. In: CIAPUSCIO, G.; JUNGBLUTH, K.; KAISER, D.; LOPES, C. R. (org.). *Sincronia y diacronia de tradiciones discursivas en Latinoamérica*. Frankfurt: Vervuert; Bibliotheca Ibero-Americana, 2006. p. 187-214.

LOPES, C. R. Retratos da variação entre “você” e “tu” no português do Brasil: Sincronia e diacronia. In: RONCARATI, C.; ABRAÇADO, J. (org.). *Português Brasileiro II* – contato linguístico, heterogeneidade e história. Niterói: EDUFF, 2008. v. 2. p. 55-71.

LOPES, C. R.; CAVALCANTE, S. A cronologia do voceamento no português brasileiro: expansão de você-sujeito e retenção do clítico-te. *Revista Linguísticas*, Madrid, v. 25, p. 30-65, 2011.

LOPES, C. R.; MARCOTULIO, L. L.; SILVA, A. S.; SANTOS, V. M. Quem está do outro lado do túnel? Tu ou Você na cena urbana carioca. *Neue Romania*, Berlim, v. 39, p. 49-67, 2009.

LOREGIAN, L. *Concordância verbal com o pronome tu na fala do sul do Brasil*. 1996. 134f. Dissertação (Mestrado em Letras/Linguística) – Programa de Pós-Graduação em Letras/Linguística, Universidade Federal de Santa Catarina, Florianópolis, 1996.

LOREGIAN-PENKAL, L. *(Re)análise da referência de segunda pessoa na fala da Região Sul*. 2004. 261f. Tese (Doutoramento em Letras) – Programa de Pós-Graduação em Letras, Universidade Federal do Paraná, Curitiba, 2004.

MARTINS, R. L.; MAURER, C. F.; SEVERO, P. F. O tu e você no paradigma pronominal do português brasileiro em cartas pessoais. *Revista (Con)textos Linguísticos*, Vitória, v. 10, n. 16, p. 86-102, 2016.

MENON, O. P. O sistema pronominal do português do Brasil. *Letras*, Curitiba, n. 44, p. 91-106, 1995.

MENON, O. P. Pronome de segunda pessoa no sul do Brasil: tu/você/o senhor em Vinhas da Ira. *Letras de Hoje*, Porto Alegre, v. 35, n. 1, p. 121-164, 2000.

MENON, O. P.; LOREGIAN-PENKAL, L. Variação no indivíduo e na comunidade: Tu/você no Sul do Brasil. In: VANDRESEN, P. (org.). *Variação e mudança no português falado na região Sul*, Pelotas: EDUCAT, 2002. p. 147-192.

MONTEIRO, J. L. *Para compreender Labov*. 2. ed. Petrópolis: Vozes, 2000.

NASCIMENTO, M. F.; MENDES, A.; DUARTE, M. E. Sobre formas de tratamento no português europeu e brasileiro. *Diadorim*, Rio de Janeiro, v. 20, p. 245-262, 2018. DOI: <https://doi.org/10.35520/diadorim.2018.v20n0a23276>

PAREDES, V. L. A abordagem laboviana. In: ENCONTRO NACIONAL DA ANPOLL, 7., 1993, Goiânia. *Anais [...]* Goiânia: Anpoll, 1993. p. 882-886.

RAMOS, M. P. B. *Formas de tratamento no falar de Florianópolis*. 1989. 106f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Letras, Universidade Federal de Santa Catarina, Florianópolis, 1989.

ROCHA LIMA, L. E. *Gramática normativa da língua portuguesa*. 49. ed. Rio de Janeiro: José Olympio, 2011.

RUMEU, M. C. B. A variação “tu” e “você” no português brasileiro oitocentista e novecentista: Reflexões sobre a categoria social gênero. *Alfa*, São Paulo, v. 57, n. 2, p. 545-576, 2013. DOI: <https://doi.org/10.1590/S1981-57942013000200010>

SALOMÃO, A. C. B. Variação e mudança linguística: Panorama e perspectivas da Sociolinguística Variacionista no Brasil. *Revista Fórum Lingüístico*, Florianópolis, v. 8, n. 2, p. 187-207, 2011. DOI: <https://doi.org/10.5007/1984-8412.2011v8n2p187>

SCHERRE, M. M. P. Aspectos sincrônicos e diacrônicos do imperativo gramatical no português brasileiro. *Alfa*, São Paulo, v. 51, n. 1, p. 189-222, 2007.

SCHERRE, M. M. P.; DIAS, E. P.; ANDRADE, C.; MARTINS, G. F. Variação dos pronomes “tu” e “você”. In: MARTINS, M. A.; ABRAÇADO, J. (org.). *Mapeamento sociolinguístico do português brasileiro*. São Paulo: Contexto, 2015. p. 133-172.

SILVA, G. M. O. Estudo da regularidade na variação dos possessivos no Rio de Janeiro. 1982. Tese (Doutoramento em Letras) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 1982.

WEINER, J.; LABOV, W. Constraints on the agentless passive. *Journal of Linguistics*, Cambridge, n. 19, p. 29-58, 1977. DOI: <https://doi.org/10.1017/S0022226700007441>

WEINREICH, U.; LABOV, W.; HERZOG, M. *Fundamentos empíricos para uma teoria da mudança linguística*. Trad. Marcos Bagno. São Paulo: Parábola, 2006.

YACOVENCO, L. C.; SCARDUA, J. R. A variação pronominal de segunda pessoa: Contribuições da sociolinguística para o ensino de língua portuguesa. *Working Papers Linguística*, Florianópolis, v. 18, n. 2, p. 171-191, 2017. DOI: <https://doi.org/10.5007/1984-8420.2017v18n2p171>



## Sentidos do discurso *coaching* financeiro no enunciado vídeo publicitário “Meu nome é Bettina” e possibilidade de cotejo

### *Senses of the financial coaching discourse in the utterance advertising video “My name is Bettina” and possibilities of collation<sup>1</sup>*

Grenissa Stafuzza

Universidade Federal de Catalão (UFCAT), Catalão, Goiás / Brasil

grenissa@gmail.com

<https://orcid.org/0000-0002-9077-0652>

Maximiano Antonio Pereira

Universidade Federal de Catalão (UFCAT), Catalão, Goiás / Brasil

maximianoantonio@icloud.com

<https://orcid.org/0000-0003-3125-0638>

**Resumo:** Especialmente a partir dos fundamentos filosóficos sobre a linguagem pensados por Bakhtin e o Círculo russo (Volóchinov e Medviédev), propõe-se analisar os sentidos do discurso *coaching* financeiro no enunciado vídeo publicitário, “Meu nome é Bettina”, publicado em março de 2019 no canal da Empiricus, no Youtube, e viralizado nas redes sociais. A hipótese de pesquisa é que o discurso *coaching* financeiro presente no enunciado vídeo “Meu nome é Bettina” ancora-se no diálogo com os discursos de autoajuda e neoliberal, para a realização da comunicação discursiva. No enunciado vídeo “Meu nome é Bettina” observa-se que, por meio do discurso *coaching* financeiro, o enunciado vídeo faz circular o enriquecimento fácil *do* e autocentrado *no* indivíduo, apresentando ser constituído de relações dialógicas com discursos produzidos pelo capitalismo, com ênfase no individualismo, como os discursos de autoajuda e neoliberal. Destaca-se, ainda, o cotejo de outros enunciados que circularam no Twitter

---

<sup>1</sup> Utiliza-se a palavra *collation* (compilação) como tradução para “cotejo” por aproximação de sentido, pois, na língua inglesa, não há uma palavra para tradução direta.

que respondem ao enunciado vídeo “Meu nome é Bettina”, de modo a ilustrar a característica da responsividade do enunciado na esfera midiática.

**Palavras-chave:** enunciado; discurso *coaching* financeiro; discurso de autoajuda; discurso neoliberal; “Meu nome é Bettina”.

**Abstract:** Especially based on the philosophical foundations about language thought by Bakhtin and the Russian Circle (Volóchinov and Medviédev), it is proposed to analyze the senses of the financial *coaching* discourse in the utterance advertising video, “My name is Bettina”, published in March 2019 on the Empiricus channel, in Youtube, and viralized on social networks. The research hypothesis is that the financial *coaching* discourse present in the video utterance “My name is Bettina” is anchored in the dialogue with the self-help and neoliberal discourses, for the realization of discursive communication. In the video utterance “My name is Bettina” it is observed that through the financial *coaching* discourse, the video utterance circulates the easy enrichment of the self-centered in the individual, presenting being constituted of dialogical relations with discourses produced by capitalism, with an emphasis on individualism, such as self-help and neoliberal discourses. It is also worth mentioning the collation of other utterance that circulated on Twitter that respond to the video utterance “My name is Bettina”, in order to illustrate the characteristic of the responsivity of the utterance in the media sphere.

**Keywords:** utterance; financial coaching discourse; self-help discourse; neoliberal discourse; “Meu nome é Bettina”.

Recebido em 29 de janeiro de 2021

Aceito em 25 de março de 2021

## 1 Introdução

O filósofo francês Gilles Lipovetsky (2004) observa por meio do conceito de hipermodernidade que o termo pós-moderno já não é capaz de exprimir os acontecimentos da atualidade. O “pós”, de pós-moderno, emana o sentido de que o passado é um tempo já findado, não se considerando os processos históricos concretizados no atual neoliberalismo globalizado, na mercantilização da vida, dos corpos e do narcisismo individualista. A hipermodernidade, também denominada de supermodernidade, não é negacionista quanto ao passado; há integração das práticas históricas que resultam em lógicas modernas de consumo. Lipovetsky (2004, p. 58), ao definir o conceito de hipermodernidade, afirma que “as operações e os intercâmbios se aceleram; o tempo é

escasso e se torna um problema, o qual se impõe no centro de novos conflitos sociais”. Assim, observa: “o culto ao presente se manifesta com força aumentada, mas quais são seus contornos exatos e que vínculos ele mantém com os outros eixos temporais? (...) Convém reabrir a questão do tempo social, pois este merece mais do que nunca uma inquirição”. (LIPOVETSKY, 2004, p. 58)

Ao refletir sobre o conceito de sociedade hipermoderna, Lipovetsky aponta três elementos que a constitui: i) o aprofundamento da economia de mercado; ii) a revolução tecnológica que invade o cotidiano; iii) uma autonomia individual sem precedentes.<sup>2</sup> Assim, o período que se enfrenta é de processos neoliberalistas, em que o sujeito se inscreve “na dinâmica de fundo das economias modernas que se caracterizam pela otimização dos resultados e pelo cálculo sistemático dos custos e benefícios” de seus atos (LIPOVETSKY; SERROY, 2015, p. 11). Sob essa perspectiva, percebe-se como o sujeito, em todos seus excessos, tem se submetido, a partir desse contexto, às exigências neoliberais.<sup>3</sup> Assim, a hipermodernidade é marcada por imprecisões e grande dubiedade resultante de processos neoliberalistas: as relações humanas passaram a ser reguladas por meio do avanço tecnocientífico e da mercantilização generalizada.

A característica que melhor define a hipermodernidade é a cultura do excesso. Todo processo é exagerado, grandioso, espetacular, pesado. As mudanças acontecem de forma frenética determinando um tempo efêmero e adaptações imediatas. Não se trata de um processo fluido e natural: hiperconsumo; hipermercado, hipercorpo, hipertexto, tudo só é considerável se for levado ao excesso e marcado pelo abrupto. O sujeito que se apresenta na hipermodernidade é aquele que precisa ter, e quanto mais rápido for este processo, mais garantias de ter um espaço fora da

---

<sup>2</sup> Entrevista disponível em: <https://www.fronteras.com/entrevistas/gilles-lipovestky-qual-o-significado-do-consumo-em-nossas-vidas>. Acesso em: 27 jan. 2021.

<sup>3</sup> Um exemplo dessa questão e que faz eco com o tema deste trabalho, encontra-se na matéria divulgada pelo Fantástico, programa da Rede Globo, em 10 de janeiro de 2021, em que uma jovem bancária revela que investiu trinta e cinco mil reais em *coaching*, não teve retorno algum, mas afirmou que se “motiva por entrega”. Em suma: uma trabalhadora que paga para responder às demandas de seu trabalho de modo que ela acredita ser mais eficiente (“motiva por entrega”) e, com isso, se diz realizada. Disponível em: <https://www.hypeness.com.br/2021/01/ela-investiu-r-35-mil-em-coaching-nao-teve-retorno-e-diz-que-se-motiva-por-entrega/>. Acesso em: 27 jan. 2021.

margem. O resultado sócio histórico da política neoliberal no Brasil<sup>4</sup> possibilita o aparecimento de fenômenos variados que vão desde a cura física em poucos dias, um corpo perfeito em poucas semanas ou, como no caso do *coaching* financeiro, ter um poder aquisitivo invejável. É em relação a estes acontecimentos que enunciados de múltiplas semioses (verbais, vocais, visuais) veiculam e, por vezes, dialogam com campos de atuações acadêmico-profissionais de forma polêmica, buscando convencer seu interlocutor de que é ciência, no entanto, sem fundamentos que demonstram sua cientificidade.

Esses discursos que se apresentam como ciência, não sendo ciência, atingem o mercado informal que se orienta conforme o sistema capitalista neoliberal: no Brasil, institutos com vínculos internacionais apresentam propostas para *coachees*<sup>5</sup> se tornarem *coachs* em poucos dias. Nessa conjuntura, é de suma importância entender como discursos pseudocientíficos se cristalizam e emergem em diversas esferas da atividade humana, inclusive científicas, não sendo incomum observar o aparecimento de *coachs* acadêmicos, *coachs* pedagógicos ou *coachs* para concursos. Por isso, há, aqui, uma inquietação acadêmico-profissional para que o discurso *coaching* seja compreendido em seu caráter de fenômeno e não de verdade científica.

Assim, a concepção de *coaching* divulgada pelo Instituto Brasileiro de Coaching (IBC), a seguir, apresenta-se como “processo”, “metodologia”, abordando de modo generalizado precisamente para que o

---

<sup>4</sup> Conforme Almeida (2010), os ex-presidentes do Brasil, Fernando Collor (1990-1992) e Fernando Henrique Cardoso (1995-2003), são exemplos de governos que adotaram políticas neoliberais no Brasil. Michel Temer (2016-2019), por sua vez, marcou a volta do neoliberalismo, após o capitalismo no país ter “dependido de uma certa participação das classes populares para se desenvolver” nos Governos Lula e Dilma (2003-2016), conforme Souza e Hoff (2019, p. 1). Atualmente, com o governo Bolsonaro (2019-), um novo capítulo do neoliberalismo está sendo escrito no Brasil.

<sup>5</sup> De acordo com o dicionário Oxford (SIMPSON, 2017, p. 136) da língua inglesa, o termo *coach* significa “treinador, instrutor, tutor”. O *coachee* é também conhecido como cliente, aquele que busca o suposto profissional *coach* para maximizar resultados pessoais, de carreira e/ou profissionais. *Coach*, por sua vez, é a suposta pessoa especializada que, através de instruções e direcionamentos, busca ajudar uma ou várias pessoas a alcançar um objetivo ou a desenvolver suas habilidades em determinada área. Nesse sentido, mantém-se a ideia de treinamento, instrução e tutoria da etimologia da palavra *coach* em inglês.

público a ser alcançado possa ser o mais abrangente possível: o máximo de pessoas podem se identificar quando se lança formulações generalizadas.

*Coaching* é um processo, uma metodologia, um conjunto de competências e habilidades que podem ser aprendidas e desenvolvidas por absolutamente qualquer pessoa pra alcançar um objetivo na vida pessoal ou profissional, até 20 vezes mais rápido, comprovadamente.

E o Coaching também é uma excelente oportunidade de carreira para quem quer ajudar outras pessoas e ser muito bem remunerado por isso, atuando como Coach Profissionalmente.

O Coaching é um processo definido como um mix de recursos que utiliza técnicas, ferramentas e conhecimentos de diversas ciências como a administração, gestão de pessoas, psicologia, neurociência, linguagem ericksoniana, recursos humanos, planejamento estratégico, entre outras. A metodologia visa a conquista de grandes e efetivos resultados em qualquer contexto, seja pessoal, profissional, social, familiar, espiritual ou financeiro.<sup>6</sup>

O processo de treinamento apresenta-se como uma possibilidade para toda e qualquer pessoa, sem considerar suas diferenças, suas condições materiais de existência. Há, sobretudo, uma homogeneização: “*coaching* é um processo, uma metodologia, um conjunto de competências e habilidades que podem ser aprendidas e desenvolvidas por absolutamente qualquer pessoa”, “é uma excelente oportunidade de carreira para quem quer ajudar outras pessoas e ser muito bem remunerado”. A ideia de que o *coaching* pode ser útil para tudo e resolver todos os problemas, desde natureza pessoal até os de natureza profissionais, financeiros de uma determinada pessoa/empresa, equaliza as relações humanas e de trabalho, dando a falsa ideia de que o *coaching* pode ser a solução para qualquer problema de qualquer natureza.

Não há uma explicação sobre as metodologias adotadas e sua relação científica, apesar de afirmar que se trata de “um processo definido como um mix de recursos que utiliza técnicas, ferramentas e conhecimentos de diversas ciências” e que “a metodologia visa a conquista de grandes e efetivos resultados em qualquer contexto”. O discurso *coaching* se diz fundamentado em diversas ciências para sua aplicabilidade na vida

---

<sup>6</sup> Disponível em: <https://www.ibccoaching.com.br/portal/coaching/o-que-e-coaching/>. Acesso em: 26 mar. 2021.

de “absolutamente qualquer pessoa”. No entanto, torna-se esvaziado de sentido sobre o saber científico quando não elucida seus fundamentos científicos, podendo funcionar no gênero publicitário como um placebo para seu interlocutor: ao se autoafirmar científico, o discurso *coaching* se vende com esse rótulo, mesmo que seu efeito não seja científico.

Ainda, o IBC divulga em sua página que “o processo de *coaching* é realizado através de sessões”, nas quais o *coach* irá estimular e despertar a capacidade do *coachee* em “seu potencial infinito para que este conquiste tudo o que deseja”, podendo “ser aplicado em qualquer contexto e direcionado a pessoas, profissionais das mais diversas profissões e empresas de diferentes portes e segmentos.”<sup>7</sup> Ao adotar as generalizações e a homogeneização, o discurso *coaching* não se responsabiliza pelo seu próprio dizer, pois cabe ao seu público consumidor permitir ser estimulado e despertado pelo *coach*, de modo quase espiritual (“potencial infinito”) para que o “processo de coaching” funcione para transformar a vida do indivíduo/da empresa. Ainda, ao não fazer distinção entre pessoas e empresas, mostrando-se aplicável tanto para o indivíduo quanto para empreendimentos, o discurso *coaching* se coloca como um discurso capitalista, promovendo a ideia da educação neoliberal. Uma educação neoliberal possui característica tecnicista e tem como objetivo formar pessoas competitivas para o mercado de trabalho, ou seja, o indivíduo não se desvincula do capital, ele é engrenagem do empreendimento (EISENBACH NETO; CAMPOS, 2017). De modo contrário, uma educação libertadora compreende a autonomia, a coletividade e a intelectualidade, carecendo muito mais do que ensinar estratégias, “técnicas”, “habilidades”, “metodologia” para o sujeito em sua vida.

A partir das contribuições teórico-metodológicas de Bakhtin e do Círculo (Volóchinov e Medviédev),<sup>8</sup> considera-se, no item 2 deste trabalho, a análise do enunciado vídeo publicitário da Empiricus de um (1) minuto e dez (10) segundos, publicado em março de 2019 no canal

---

<sup>7</sup> Disponível em: <https://www.ibccoaching.com.br/portal/coaching/o-que-e-coaching/>  
Acesso em: 27 jan. 2021.

<sup>8</sup> Denomina-se Círculo neste artigo o conjunto de produções das décadas de 20 e 30 do século XX, especialmente de autoria de Bakhtin, Volóchinov e Medviédev, com a troca de pseudônimos no contexto da episteme soviética. Intenta-se valorizar tanto a questão da autoria como a questão do contexto de produção do conhecimento dos autores em estudo.

oficial da própria empresa no Youtube, intitulado “Meu nome é Bettina”,<sup>9</sup> que apresenta características do discurso *coaching* da IBC supracitado. A publicidade gerou polêmica em torno da veracidade do relato pessoal promovido pela empresa, sendo, já em abril, a Empiricus multada pelo PROCON em 40 mil reais por veiculação de publicidade enganosa.<sup>10</sup> Além disso, a empresa foi obrigada a excluir o vídeo do Youtube, sendo que em outubro, sete meses depois, a Empiricus publica novamente em seu canal no Youtube, um vídeo com a própria Bettina pedindo desculpas pelo vídeo anterior.<sup>11</sup>

Para Bakhtin (2017), a ideia de diálogo compreende a alternância dos sujeitos do discurso para o acabamento do enunciado, ou seja, para a sua possibilidade de resposta ou de compreensão responsiva, determinando “as posições responsivas dos outros nas complexas condições de comunicação discursiva de um dado campo da cultura” (BAKHTIN, 2017, p. 279). Assim, não se trata da linguagem sem considerar sua natureza dialógica e sua capacidade de resposta na comunicação discursiva.

No enunciado vídeo publicitário, a jovem Bettina Rudolph, de 22 anos, afirma ter conseguido 1 milhão e 42 mil reais de patrimônio acumulado partindo da quantia de R\$ 1.520 “sem nenhum segredo” (00min18s-00min19s). O aparecimento de Bettina ganhou enorme repercussão, tornando-se um dos assuntos mais comentados no *twitter* e gerando uma série de *memes* e vídeos respostas que a contrapunham. O aparecimento de diversos e numerosos enunciados que respondiam ao discurso *coaching* financeiro da campanha publicitária “Meu nome é Bettina” fez desse discurso alvo de críticas bastante oportunas como a questão do trabalhador assalariado que não consegue nem fechar o mês “no azul”, quanto menos “investir”: as respostas evidenciam que o enunciado vídeo publicitário enuncia para um determinado público, apesar de mostrar-se equivocadamente acessível para qualquer público.

<sup>9</sup> Como a empresa Empiricus excluiu o vídeo de sua conta oficial no Youtube, para recorrer a este enunciado foi necessário acessá-lo em contas não oficiais à propaganda. Link para acesso ao vídeo disponível em: <https://www.youtube.com/watch?v=knIHvor2gHs>. Acesso em: 27 jan. 2021.

<sup>10</sup> Disponível em: <https://www.cartacapital.com.br/economia/procon-multa-empiricus-por-propaganda-enganosa-da-milionaria-bettina/> Acesso em: 28 jan. 2021.

<sup>11</sup> Disponível em: [https://www.youtube.com/watch?reload=9&v=0wN6\\_3VjOfk](https://www.youtube.com/watch?reload=9&v=0wN6_3VjOfk) Acesso em: 27 jan. 2021.

Na sequência do estudo, especificamente no item 3, observa-se que o enunciado vídeo publicitário “Meu nome é Bettina” gerou inúmeros outros enunciados que o respondem e que com ele dialogam, promovendo diversos outros discursos (de classe social, do humor, da denúncia, da matemática financeira, entre outros) que movimentam os sentidos na comunicação discursiva, conforme apresentado por meio do cotejo de enunciados. Para isso, acionam-se três *tweets* a partir de publicações na rede social *Twitter*, de modo a ilustrar a característica da responsividade do enunciado. Sob essa perspectiva, a obra de Bakhtin e do seu Círculo (Volóchinov e Medviédev) apresenta potencialidades teóricas para fundamentar análises de diversos discursos que circulam na atualidade, especificamente, neste estudo, na esfera midiática.

A ideia de discurso inicialmente é bastante abstrata, sendo necessário materializar-se, para que o pesquisador possa analisá-lo. Para Bakhtin (2017, p. 274) “o discurso sempre está fundido em forma de enunciado pertencente a um determinado sujeito do discurso, e fora dessa forma não pode existir”. Isso significa dizer que um enunciado (material) comporta discursos possíveis de serem analisados a partir das relações dialógicas que estabelece com outros discursos que desse enunciado participam. O discurso *coaching* financeiro, por sua vez, é o discurso que se manifesta de modo mais contundente no enunciado vídeo publicitário “Meu nome é Bettina”, da empresa Empiricus, sendo possível observar suas características e modos de funcionamento neste enunciado, além dos diálogos possíveis com outros discursos, também gerados pelo capitalismo, que participam deste enunciado. É importante afirmar que a análise aqui empreendida não se encerra em si mesma, sendo, sobretudo, um catalizador para outras análises.

## 2 O enunciado vídeo “Meu nome é Bettina” em análise

A amplitude da noção de diálogo não o torna apenas um espaço de mediação dos conflitos sociais ou de interação face a face entre interlocutores, mas um espaço em que a comunicação se realiza de modo mais abrangente, considerando os processos históricos, ideológicos e culturais. Nesse sentido, de acordo com Volóchinov (2017, p. 118):

O diálogo, no sentido estrito do termo, não constitui, é claro, senão uma das formas, é verdade que das mais importantes, da interação discursiva. Mas pode-se compreender a palavra “diálogo” num

sentido amplo, isto é, não apenas como a comunicação em voz alta, de pessoas colocadas face a face, mas toda comunicação verbal, de qualquer tipo que seja.

Diálogo na concepção de Volóchinov (2017) apresenta duas faces, um conteúdo intrínseco e interno ao enunciado e sua valoração, que é exterior, para o outro, sendo ambos indissociáveis: o conteúdo interno e suas formas de valoração não se limitam por um centro de organização, mas dependem dos sujeitos sociais e dos processos de interação, inseridos na história e na cultura, com determinadas ideologias. Assim, o diálogo não deve ser compreendido simplesmente como reverberações, mas também como processos de interações com múltiplos sentidos dependendo do enunciado que participa. Para Bakhtin (2017, p. 275), “o diálogo é a forma clássica de comunicação discursiva”, ou seja, o diálogo pressupõe o outro no discurso, sendo que as réplicas do diálogo apresentam uma posição de quem enuncia.

Parte-se da premissa de que todo e qualquer enunciado advém de uma materialidade constituída por discursos que dialogam, fazendo emergir sentidos que respondem a outros enunciados (posição responsiva). Ao considerar o enunciado a unidade real da comunicação discursiva, entende-se que se deve considerar para análise o todo arquitetônico que significa, sendo que “a arquitetônica do mundo da visão artística não ordena só os elementos espaciais e temporais, mas também os de sentido” (BAKHTIN, 2017, p. 127). Medviédev (2012), por sua vez, compreende que, por meio do ato individual de realização do enunciado, seu sentido começa a participar da história e torna-se um fenômeno histórico. Logo, “o fato de que foi esse sentido que se tornou um objeto de discussão aqui e agora, que é dele que estão falando e que falam justamente assim e não de outra forma, que precisamente esse sentido entrou no horizonte concreto dos que falam” (MEDVIÉDEV, 2012, p. 184) é o que determina conjuntamente as condições sócio-históricas do enunciado.

Assim, por meio da análise do enunciado vídeo “Meu nome é Bettina”, um dos elos dessa cadeia infindável que se estabelece entre tempo-espaco-sentido na comunicação discursiva, podem ser também observadas as relações dialógicas que o constituem. No enunciado vídeo publicitário, Bettina se apresenta como uma jovem investidora bem-sucedida que oferece conselhos financeiros. Apresenta-se a seguir um *print* de tela do vídeo (FIGURA 1) para ilustração e, na sequência, a transcrição completa de seu áudio como recursos para operacionalizar a análise.

FIGURA 1 – Bettina Rudolph em propaganda para Empiricus



Fonte: Youtube<sup>12</sup>

Oi, meu nome é Bettina, eu tenho 22 anos e 1 milhão e 42 mil reais de patrimônio acumulado. Desculpa a indiscrição é que o tempo aqui embaixo não joga a meu favor e eu precisava chamar sua atenção. Ninguém acha normal eu ter juntado mais de 1 milhão de reais assim, tão nova e começando com muito pouco. Mas sabe o que chama a minha atenção? É que o que eu fiz não é nenhum segredo. Eu vivo falando por aí, pra todo mundo, eu comprei ações, na bolsa de valores. Não foi sorte, eu não herdei uma bolada, nem ganhei na loteria. Comecei com 19 anos e 1.520 reais. Três anos depois, tenho mais de 1 milhão, simples assim. Sabe qual é o problema? A maioria das pessoas vai pular esse anúncio e continuar se chocando com histórias fora da curva e a minoria vai clicar no botão azul que vai aparecer no fim desse vídeo. Se você for uma das pessoas que vai clicar no botão, você vai acessar o passo a passo que eu segui pra você chegar no seu primeiro ou próximo milhão. Os resultados, eu garanto, serão os mesmos. Não tem como ser diferente. Se você tiver as mesmas ações que eu tenho, vai lucrar proporcionalmente o mesmo que eu. Isso vale para as perdas também. É com você! Resultados diferentes, exigem atitudes diferentes! Botão azul! (YOUTUBE, “Meu nome é Bettina”, 1min10seg)

A Empiricus<sup>13</sup> foi fundada no Brasil no ano de 2009, sua sede se encontra na cidade de São Paulo, atuando como prestadora de serviço

<sup>12</sup> Disponível em: <https://www.youtube.com/watch?v=knIHvor2gHs>. Acesso em: 15 jan. 2021.

<sup>13</sup> Site: <https://www.empiricus.com.br/>

especializada na publicação de conteúdo financeiro, de modo que os assinantes tenham acesso a ideias de investimento. A empresa possui mais de 350 mil assinantes e 24 títulos abordando sobre temas como ações, fundos imobiliários e investimentos na bolsa, considerada aqui também como promotora do discurso *coaching* financeiro no Brasil

A empresa atua com o trabalho de 30 analistas provenientes de instituições como Credit Suisse, Itaú e Santander. A Empiricus promoveu eventos que trouxeram ao Brasil personalidades famosas, tais como o vencedor do Nobel de Economia Daniel Kahneman e o escritor Nassim Nicholas Taleb.<sup>14</sup> Outras personalidades que também marcaram presença foi o ex-presidente do Brasil Fernando Henrique Cardoso e o economista Eduardo Giannetti.<sup>15</sup> Os dois professores universitários que fundaram a Empiricus, Felipe Miranda e Rodolfo Amstalden decidiram empreender a Empiricus quando conheceram um antigo corretor da financiadora Link, Marcos Elias. Os três decidiram empreender e após alguns escândalos houve o afastamento de Elias em 2012 e a empresa seguiu em seu conceito de divulgar ideias de investimento às pessoas.

A Empiricus se posiciona como uma espécie de antagonista ao sistema financeiro tradicional. Normalmente, o aconselhamento está dentro dos grandes bancos e há um conflito de interesse já que os bancos recomendam aquilo que é melhor para eles. Há uma grande falta de transparência no mercado e a Empiricus se posiciona contra esse sistema. Como casa de pesquisa, ela recomenda um investimento e apenas isso. Não recebe nenhuma comissão por aquela recomendação. (*PASCOWITCH, Bernardo. Fundador do buscador de investimentos Yubb*)<sup>16</sup>

---

<sup>14</sup> Nassim Nicholas Taleb é um autor, ensaísta, estatístico, e analista de riscos comerciais e de investimento (coach), libano-americano, matemático de formação. Atualmente residente nos Estados Unidos, Reino Unido e Líbano, é conhecido por ser investidor no mercado financeiro, sendo professor do Instituto Politécnico da Universidade de Nova Iorque e presidente da empresa de investimentos *Empirica*, também atuando como conselheiro do grupo *Empiricus* no Brasil.

<sup>15</sup> Eduardo Giannetti da Fonseca (Belo Horizonte, 23 de fevereiro de 1957) é um economista e professor brasileiro, formado na Faculdade de Economia, Administração e Contabilidade (FEA) e em Ciências Sociais pela Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) ambas da Universidade de São Paulo.

<sup>16</sup> Disponível em: <https://blog.yubb.com.br/quem-e-a-empiricus/>. Acesso em 15 de jan. 2021.

No ano de 2017 houve um crescimento considerável e o número de assinantes da empresa se aproximava a 180.000, tendo um aumento de seis vezes desde 2014. Em relação a esse crescimento considerável, a empresa comprou metade do site O Antagonista, conhecido principalmente pelo posicionamento político de direita: a compra ocorreu em março de 2016 por 5 milhões de reais. Enquanto a empresa crescia e se tornava conhecida, cresciam também os escândalos e as queixas como, por exemplo, a orientação da empresa para que os investidores passassem a investir na empresa de Eike Batista,<sup>17</sup> empresário brasileiro condenado pelos crimes de uso de informação privilegiada e manipulação de mercado.

Em 2019, com a publicação do vídeo publicitário “Meu nome é Bettina”, surge Bettina Rudolph, profissional da equipe publicitária da Empiricus. A imagem de Bettina como a jovem investidora bem sucedida que aconselha as pessoas em como ganhar seu milhão virou meme, alvo de críticas em comentários e postagens que ridicularizaram as ideias de investimento propostas pela empresa. O excesso de publicidade com Bettina também foi alvo de críticas, sendo que a empresa passasse de “excelente” para “boa” na avaliação no site “Reclame Aqui”.

Não se trata de uma escolha aleatória a Empiricus ter contratado uma locutora mulher, jovem, branca, loira, padrão mídia para a sua publicidade de proposta de investimento, uma vez que isso aponta tanto para a falsa ideia de que as pessoas têm “livre escolha” para investir, apagando-se as condições materiais de existência da grande maioria dos usuários da rede que a publicidade alcança, quanto para o real público-alvo da publicidade. Nesse sentido, o enunciado vídeo publicitário evidencia o interesse da Empiricus em um público específico: “jovens investidores” que possuam capital para investir. Nesse caso, Bettina representa um grupo específico em que ela se insere, podendo ser observado inclusive nas escolhas lexicais pertencentes ao gênero publicitário, pois no enunciado, ao se apresentar para o público, Bettina reitera sobre sua idade e sobre seu capital:

00min1seg-00min-15seg: [Bettina] Oi, meu nome é Bettina, eu tenho 22 anos e 1 milhão e 42 mil reais de patrimônio acumulado. Desculpa a indiscrição é que o tempo aqui embaixo não joga a meu favor e eu precisava chamar sua atenção. Ninguém acha normal eu

<sup>17</sup> Disponível em: <https://www.infomoney.com.br/onde-investir/veja-os-maiores-erros-dos-analistas-com-a-ogx/> Acesso em: 15 jan. 2021.

ter juntado mais de 1 milhão de reais assim, tão nova e começando com muito pouco.

Ao anunciar sua pouca idade em relação ao seu patrimônio acumulado, o sujeito, que ao enunciar se denuncia, fala a partir de um grupo social, sendo a sua voz representante de um grupo de jovens pertencentes à elite empresarial, filhos de investidores, de proprietários de grandes empresas multinacionais e corporações, donos de latifúndios etc. Sendo sob essas condições e não outras vivenciadas por filhos de trabalhadores, também trabalhadores, com suas vidas precarizadas pelo sistema capitalista, financiador da desigualdade, que Bettina enuncia a partir do lugar de “jovem investidora”. A presença de Bettina, enquanto mulher, marca também o interesse da Empiricus em convidar o público feminino para investir, dado que, de acordo com uma pesquisa realizada pela revista *Exame*, as mulheres investem 29% menos que os homens, portanto, são minoria também no mercado financeiro,<sup>18</sup> no entanto, as mulheres são as que mais crescem no mercado de trabalho, conforme o blog SBCoaching.<sup>19</sup>

Medviédev (2012) aponta que para compreender um enunciado é preciso entendê-lo no contexto da sua contemporaneidade e do próprio pesquisador, caso esses contextos não coincidam. Isso significa dizer que, “é necessário compreender o sentido no enunciado, o conteúdo do ato e a realidade histórica do ato em sua união concreta e interna” (MEDVIÉDEV, 2012, p. 185), pois, sem tal abrangência, o sentido do enunciado se assemelhará ao sentido dicionarizado, sem refletir a interação dialógica da linguagem. Quando o sujeito Bettina marca seu dizer apresentando seu poder aquisitivo, há um posicionamento que determina seu lugar sócio-histórico sobre fazer parte de um pequeno grupo economicamente milionário, uma vez que o sujeito propõe uma

---

<sup>18</sup> Disponível em: <https://exame.com/seu-dinheiro/mulheres-investem-29-menos-que-homens-mostra-guiabolso/> Acesso em: 15 de jan. 2021.

<sup>19</sup> Disponível em: <https://www.sbcoaching.com.br/blog/mulher-mercado-trabalho/> Acesso em: 15 de jan. 2021. O discurso *coaching* se interessa pela mulher em vários de seus nichos (pessoal, profissional, nutricional, financeiro etc.) porque a mulher avançou no processo de emancipação financeira e os dados apresentam crescimento da mulher no mercado de trabalho. Logo, se a mulher se encontra economicamente emancipada, ela se torna um alvo do discurso *coaching* financeiro.

categoria de vida por meio do investimento: “(...) eu tenho 22 anos e 1 milhão e 42 mil reais de patrimônio acumulado”.

Ainda, ao apresentar o estranhamento do público sobre ser tão jovem e já ter uma quantia milionária (“Ninguém acha normal eu ter juntado mais de 1 milhão de reais assim, tão nova e começando com muito pouco”), Bettina aciona o discurso de autoajuda de modo a apresentar uma receita de como seu público pode fazer para conseguir os mesmos resultados que ela obteve (“começando com muito pouco”). Aqui, observa-se o discurso de autoajuda em diálogo com o discurso neoliberal, quando se coloca como possibilidade para qualquer público, indistintamente, que é possível qualquer pessoa, independentemente de sua posição econômica, juntar um milhão “começando com muito pouco”.

Assim, quando Bettina enuncia, parte de um ato individual de fala, ensaiado, claro, pois se trata de uma publicidade. No entanto, seu posicionamento reflete a posição de um determinado grupo social, como dito anteriormente, sendo que sua voz é povoada de outras vozes que enunciam, por exemplo, o discurso de autoajuda e o discurso neoliberal. Para além de se pensar a paternidade do capitalismo sobre os discursos de autoajuda e o neoliberal, o diálogo que se instaura no enunciado vídeo “Meu nome é Bettina” reflete uma determinada coletividade de vozes que se filiam especificamente a este momento vivenciado, a essa atualidade histórico-social de sentidos do enunciado. Por isso, o enunciado se torna único e irrepetível na comunicação discursiva. Ao mesmo tempo em que atende, o enunciado vídeo publicitário responde a uma demanda hipermoderna do desejo pelo enriquecimento individual, evidenciando esses sentidos: a promessa de que se vive em um tempo em que ser jovem e rico nunca foi tão fácil encontra respaldo nos diálogos com os discursos de autoajuda e neoliberal *para e na* construção do discurso *coaching* financeiro.

Ao apresentar a ideia de mercado relacionada à crença humana em uma divindade, Dufour (2008, p. 88-89) aponta que o “divino mercado” é uma invenção do liberalismo fundamentado no capitalismo que incorpora os próprios atributos da divindade: o mercado possui onipotência desde que o deixem de fato operar, apresentando-se como um lugar de verdade. Nesse sentido, o filósofo francês reconhece alguns aspectos do mercado, dentre eles, “a crença, constantemente mantida pelas inumeráveis vinhetas produzidas pelo catecismo publicitário, de que a salvação individual passa pelo consumo, motor do Mercado”. (DUFOUR, 2008, p. 89).

Chagas (1999), por sua vez, ao estudar a literatura de autoajuda, observa que seu nascimento data em meados do século XIX, em um momento caracterizado pelo culto à singularidade do indivíduo moderno, passando a ter um valor soberano e central na cultura ocidental, como nunca visto até então. Nesse sentido, a literatura de autoajuda origina-se e desenvolve-se como resultado do deslocamento dos referenciais coletivos para o individual, como um fenômeno cultural de massa e, ainda:

(...) pelo que caracterizou as estruturas modernas das sociedades industriais (ou, como dizem alguns, pós-industriais) capitalistas em seus novos modos de produção industrial: produção em massa (que pode ser aqui correlacionada como a indústria cultural), pelo funcionamento do regime capitalista, do mercado, do consumo e, sobretudo, do recalcamento da cultura tradicional, pela qual o sujeito já não mais pode orientar-se, visto que os referenciais coletivos não oferecem mais um mundo seguro, ordeiro e estável. Desse modo, o sujeito volta-se para si próprio, numa tentativa de sobreviver subjetivamente ao seu mal-estar, outrossim, para que possa enfrentar as adversidades do mundo contemporâneo, do progresso técnico e científico, da competição e do consumo exagerado. Enfim, para enfrentar esse mundo que reserva aos homens um futuro incerto (CHAGAS, 1999, p. 34)

Observado como produto do individualismo capitalista, Chagas (1999) caracteriza o discurso de autoajuda como individualista ao identificar as relações deste discurso com os discursos dominantes da sociedade. Por ancorar o estudo na perspectiva da psicanálise, Chagas compreende as determinações psíquicas que orientam o sujeito “mediante sua crença, na direção de um ideal impossível” (CHAGAS, 1999, p.18), apresentando, assim, processos inconscientes que permitem pensar sobre a filiação dos sujeitos no discurso de autoajuda.

No enunciado vídeo publicitário em estudo, ao mesmo tempo em que se promove uma surpreendente realização (conseguir juntar “mais de 1 milhão de reais assim, tão nova”), o enunciado antecipa uma possível resposta, no caso de o público reagir com descrédito sobre o feito de Bettina, colocando em dúvida a credibilidade da Empiricus. Entende-se que esta pode ser vista como uma estratégia de convencimento pela crença: ao antecipar uma resposta para uma reação negativa sobre a possibilidade de ganhar um milhão com baixo investimento em pouco tempo, o discurso *coaching* financeiro não explica matematicamente

como isso é possível, mas reconhece que é preciso fazer seu público “crer” no discurso *coaching* financeiro do enriquecimento fácil (“ninguém acha normal”), tal qual Dufour (2008) apresenta em seu “divino mercado”.

Na sequência do vídeo publicitário, o diálogo entre os discursos de autoajuda e o neoliberal aparece no enunciado, produzindo sentidos sobre a crença na simplicidade do ganho milionário idealizado, conforme transcrição a seguir:

00min16seg-00min35s: [Bettina] Mas sabe o que chama a minha atenção? É que o que eu fiz não é nenhum segredo. Eu vivo falando por aí, pra todo mundo, eu comprei ações, na bolsa de valores. Não foi sorte, eu não herdei uma bolada, nem ganhei na loteria. Comecei com 19 anos e 1.520 reais. Três anos depois, tenho mais de 1 milhão, simples assim.

A ideia de facilidade, de que não há segredos para obter bons resultados, que basta seguir o que a locutora afirma ter feito, são aspectos advindos do discurso de autoajuda, que age de modo a revelar fórmulas bastante simples, idealizadas, que dão certo para toda e qualquer pessoa (“Mas sabe o que chama a minha atenção? É que o que eu fiz não é nenhum segredo. Eu vivo falando por aí, pra todo mundo...”). Na sequência, aparecem indícios do discurso neoliberal sobre a crença no mercado (“Eu vivo falando por aí, pra todo mundo, eu comprei ações, na bolsa de valores. Não foi sorte, eu não herdei uma bolada, nem ganhei na loteria”), desconsiderando um método científico matemático que comprove a ideia exuberante de vir a ser milionário com bem pouco investimento. Consequentemente, a ideia de automatização da ciência revela a não necessidade de apresentação de cálculos sobre o tema de se conseguir em três anos mais de um milhão, investindo apenas 1.520 reais (“Comecei com 19 anos e 1.520 reais. Três anos depois, tenho mais de 1 milhão, simples assim”), como se os lucros fossem gerados automaticamente assim que o cliente inicia o investimento, de forma espontânea e certa.

De acordo com Bakhtin (2010, p. 142), “a vida conhece dois centros de valores, diferentes por princípio, mas correlatos entre si: o eu e o outro, e em torno destes centros se distribuem e se dispõem todos os momentos concretos do existir”. Assim sendo, todo sujeito enuncia atitudes (auto)avaliativas: ao mesmo tempo em que o sujeito avalia o outro, ele é também avaliado pelo outro. Isso significa dizer que o sujeito em sua singularidade sempre adentra em relações de valoração do mundo;

a valoração é estar no mundo e se posicionar nele. Medviédev (2012, p. 184), por sua vez, denomina de avaliação social “essa atualidade histórica que reúne a presença singular de um enunciado com abrangência e a plenitude do seu sentido”. Nesse sentido, compreende-se que “no enunciado, cada elemento da língua tomado como material obedece às exigências da avaliação social” (MEDVIÉDEV, 2012, p. 185).

A avaliação (ou valoração) participa das diversas situações de comunicação que se instauram entre os sujeitos participantes do diálogo na construção do enunciado para a comunicação discursiva. Assim, é por meio da (auto)avaliação que o enunciado revela posicionamentos sociais, históricos, econômicos, culturais, políticos etc., permitindo a compreensão dos sentidos tanto por meio dos diálogos que se instauram quanto nas próprias relações intrínsecas ao enunciado. Compreende-se que o sujeito Bettina, no enunciado vídeo publicitário, não assume sua voz singular, mas apresenta uma valoração no mundo por meio dos interesses comerciais da empresa Empiricus, aqui, considerada sujeito do discurso, pois possui um projeto de dizer. Logo, Bettina se torna uma participante do diálogo, como representante da voz da financiadora, por meio do discurso *coaching* financeiro.

Por meio do diálogo com os discursos neoliberal e de autoajuda, o discurso *coaching* financeiro funciona, no enunciado vídeo “Meu nome é Bettina”, vendendo para as pessoas a ilusão mercadológica de que é possível tornar-se milionário com baixo investimento. A garantia de ganho real fácil em pouco tempo (“simples assim”) através da compra de ações na bolsa de valores, conforme Bettina afirma, promove também o apagamento de que investir em ações corretamente deve demandar capital, estudo, conhecimento e, conseqüentemente, tempo. Eduardo Cubas Pereira, especialista financeiro, no artigo “É possível ficar rico investindo na Bolsa?”, publicado na InfoMoney,<sup>20</sup> explica que:

Seguindo esta lógica, não é normal alguém montar um negócio hoje para vendê-lo mais caro amanhã. É preciso entender o funcionamento do setor e da empresa onde se vai investir, além de ter conhecimentos macroeconômicos gerais. Isso tudo leva tempo. Além disso, é necessário esperar a maturação e o crescimento do negócio. Portanto, investir em ações corretamente deve demandar

---

<sup>20</sup> Disponível em: <https://www.infomoney.com.br/onde-investir/e-possivel-ficar-rico-investindo-na-bolsa/> Acesso em: 28 jan. 2021.

conhecimento, estudo e prazo. É preciso querer ser sócio de um determinado negócio e participar do seu crescimento ao longo do tempo.

Diante disso, observa-se que, ao garantir lucros grandiosos com investimento mínimo, o enunciado vídeo publicitário “Meu nome é Bettina” desconsidera, para além das condições materiais de existência do público que sua publicidade alcança, a importância de se conhecer amplamente a dinâmica econômica, de estudar e entender o funcionamento do setor e da empresa que se pretende investir. Logo, não é tão “simples assim” e nem tão rápido assim. Isso acontece porque o discurso *coaching* financeiro é um discurso que dialoga, na e para a sua constituição, com discursos gerados pelo capitalismo, tais sejam, o discurso da autoajuda e o discurso neoliberal. Observa-se que são discursos que se apoiam na certeza da vulnerabilidade (econômica, social, psicológica etc.) de seu interlocutor em potencial, e um interlocutor fragilizado tem menor possibilidade de questionar e maior possibilidade de se tornar alvo/cliente.

Ao final da publicidade, utilizando-se da estratégia comercial de Bettina ter pouco tempo no anúncio (“Desculpa a indiscrição é que o tempo aqui embaixo não joga a meu favor e eu precisava chamar sua atenção”) para contar para o público sobre suas escolhas de investimento que geraram mais de um milhão, a empresa aposta que o público seja direcionado pela curiosidade/vulnerabilidade de se tornar um milionário de modo facilitado e busque pelo curso de investimento ofertado pela empresa:

00min35seg-01min10seg: [Bettina] Sabe qual é o problema? A maioria das pessoas vai pular esse anúncio e continuar se chocando com histórias fora da curva e a minoria vai clicar no botão azul que vai aparecer no fim desse vídeo. Se você for uma das pessoas que vai clicar no botão, você vai acessar o passo a passo que eu segui pra você chegar no seu primeiro ou próximo milhão. Os resultados, eu garanto, serão os mesmos. Não tem como ser diferente. Se você tiver as mesmas ações que eu tenho, vai lucrar proporcionalmente o mesmo que eu. Isso vale para as perdas também. É com você! Resultados diferentes, exigem atitudes diferentes! Botão azul!

A pergunta retórica “Sabe qual é o problema?” indica um conhecimento prévio da Empiricus sobre a relação que os internautas

possuem com anúncios na internet. Não obstante, o discurso *coaching* financeiro apresenta a expressão “fora da curva”, fazendo emergir o sentido de que de fato esse anúncio, ao considerar o conjunto de suas propostas, é diferente do que é usual nas redes sociais. Nesse sentido, o enunciado vídeo publicitário “Meu nome é Bettina” antecipa uma resposta para o interlocutor que pode, de alguma forma, ainda estar se perguntando sobre a confiabilidade do relato (“A maioria das pessoas vai pular esse anúncio e continuar se chocando com histórias fora da curva e a minoria vai clicar no botão azul que vai aparecer no fim desse vídeo”) e promove a ideia neoliberal de crença no mercado, essa entidade onipresente que “escolhe” uma minoria para enriquecer, pois, de fato, somente um grupo bastante restrito da população brasileira possui condições materiais para investimentos.

Como já abordado anteriormente, o discurso *coaching* financeiro, ao fazer uso da ideia de automatização da ciência, promove sentidos de exatidão e acerto de modo que o cliente não precisa se preocupar, pois garante-se ganhos certos. Quando colocado em diálogo com o discurso da autoajuda, pode-se observar a promessa de acesso a um “passo a passo” que indica um caminho a ser traçado pelo interlocutor para se alcançar o resultado: o milhão. No entanto, ao enunciar que “não tem como ser diferente”, Bettina indica que seu interlocutor “vai lucrar proporcionalmente o mesmo” que ela, apontando que “isso vale para as perdas também”. A ideia gerada anteriormente de ganho fácil, certo e simples se desestabiliza com a nova informação de que há perdas no processo, logo, de alguma forma, o método “passo a passo” é contradito no próprio enunciado, uma vez que não funciona de forma equivalente e sim “proporcionalmente”.

O discurso de autoajuda, assim como o discurso neoliberal são discursos que homogeneízam os sujeitos e suas existências, pois adotam o individualismo, silenciando que o homem participa de estruturas coletivas na sociedade. Ao ancorar-se no diálogo com o discurso de autoajuda, o discurso *coaching* financeiro evidencia certas características deste discurso que circula socialmente, desde a revolução industrial capitalista do século XIX. Nesse sentido, é possível verificar algumas estratégias da autoajuda no discurso *coaching* financeiro em estudo como a receita, já mencionada (“você vai acessar o passo a passo que eu segui pra você chegar no seu primeiro ou próximo milhão. Os resultados, eu garanto, serão os mesmos. Não tem como ser diferente”), e a ideia do homem

seguro, autoconfiante, determinado e autocentrado, voltado “para os seus objetivos e interesses e que age em busca de seu próprio benefício, bem de acordo com o individualismo” (BRUNELLI, 2004, p. 62) da hipermodernidade (conforme conceito formulado por Lipovetsky que se adota neste estudo).

A estratégia do discurso de autoajuda de responsabilizar o interlocutor sobre suas próprias conquistas e fracassos também é acionada no discurso *coaching* financeiro, quando Bettina responsabiliza seu interlocutor por conseguir alcançar o primeiro milhão ou não (“É com você!”). Isso produz sentidos que, situações distintas a de Bettina, de não conseguir alcançar o primeiro milhão, origina-se pela falta de vontade ou de uma atitude não positiva em relação à proposta da Empiricus por parte do interlocutor. Essa proposição se evidencia quando Bettina afirma que “Resultados diferentes exigem atitudes diferentes”, fazendo emergir sentidos de que mudanças são mediadas por autoconvencimento ou crença nas propostas inovadoras e “empíricas” da Empiricus.

### **3 A responsividade do enunciado vídeo “Meu nome é Bettina” no Twitter: o cotejo como estratégia**

A análise do enunciado vídeo publicitário “Meu nome é Bettina”, considerando seus diálogos e sentidos na esfera midiática, realizada neste estudo, traz a ideia de que:

(...) todo e qualquer enunciado diz respeito a um “acontecimento”, a uma situação de linguagem, seja imediata ou pelo contexto mais amplo, pertencente ao conjunto das condições da vida social de uma determinada comunidade linguística. Isso significa dizer que, mesmo que o enunciado tenha um aspecto autônomo de acabamento, seus sentidos são orientados na e pela comunicação discursiva, pois operam em um movimento incessante entre a sociedade e a história. (STAFUZZA, 2018, p. 138)

Ao considerar o enunciado vídeo publicitário “Meu nome é Bettina” sob a perspectiva filosófica da linguagem, entende-se que, de acordo com Bakhtin (2016), o enunciado é a unidade real do discurso, pode ser composto por múltiplas semioses (verbal, sonoro, imagético), implica sempre um ato de comunicação social. Ainda, o enunciado encontra-se repleto de enunciados proferidos em outros momentos, outras situações de interações, as quais o sujeito apoia-se para elaborar seu discurso. O

enunciado caracteriza-se tanto pela alternância de atos de fala quanto pela sua conclusibilidade específica. Isso significa dizer que um falante, ao terminar o seu turno, dá lugar a fala do outro (não especificamente entre sujeitos, mas também entre sujeito e entre enunciados), permitindo a possibilidade de resposta. Daí o enunciado ser responsivo.

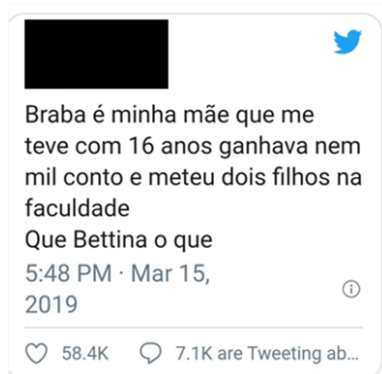
A responsividade possui dois aspectos: constituir e gerar resposta e constituir um ato responsável. Assim, o ato responsável baseia-se no reconhecimento da singularidade do próprio dever ser, como fundamento da vida, uma vez que ser na vida significa agir: “é ser não indiferente ao todo na sua singularidade” (BAKHTIN, 2010, p. 99). Isso significa dizer que, ao responder a um enunciado que o antecede, o enunciado *tweet*, constitui um ato responsável, o “*não-álibi no existir* que constitui a base da existência” (BAKHTIN, 2010, p. 99, grifos do autor) de seus autores criadores.

As redes sociais apresentam-se potencialmente como lugares de debates e funcionam como um termômetro de reações, sendo na atualidade as principais ferramentas de divulgação de anúncios, produtos ou conteúdos para se obter uma reação imediata do público. Assim, ao promover a publicidade da Empiricus em redes sociais, para além da rede de compartilhamento de vídeos Youtube, sejam elas quais forem, no caso do vídeo publicitário “Meu nome é Bettina”, os enunciados que respondem apontando as desvantagens de se investir em *coaching* financeiro apresentavam dúvidas frequentes sobre o teor das afirmativas propostas no vídeo. Elege-se, portanto, como estratégia de análise da responsividade do enunciado vídeo “Meu nome é Bettina”, o cotejo de três enunciados *tweets* publicados por usuários da rede social *Twitter* de modo a ilustrar a característica da responsividade do enunciado.

Diversos enunciados que circularam no *Twitter* suscitaram debates sobre privilégios sociais, econômicos e a classe social de Bettina, de modo que as interações em rede apontavam essas questões como o verdadeiro motivo que possibilitou a ela seu primeiro milhão: pertencer à classe alta, sendo, portanto, privilegiada economicamente. Um desses posicionamentos que dialogam com o tema da classe social e que respondem socialmente ao enunciado vídeo “Meu nome é Bettina” aparece na Figura 2.<sup>21</sup>

---

<sup>21</sup> Por razões de natureza ética, nomes e imagens de pessoas físicas estão omitidos das figuras presentes neste trabalho.

FIGURA 2 – *Tweet* publicado no *Twitter*

Fonte: Twitter.

O cotejo de enunciados pode ser considerado um recurso metodológico bakhtiniano para analisar como o diálogo entre enunciados se materializa *na e pela* linguagem, especialmente para mostrar como funciona a característica da responsividade do enunciado e seus sentidos, ou seja, da possibilidade de resposta que todo e qualquer enunciado possui, de ser, sobretudo, “um elo na corrente complexamente organizada de outros enunciados” (BAKHTIN, 2016, p. 26). Na tradução realizada do francês para o português da obra *Estética da criação verbal*, de Bakhtin (1997), Maria Ermantina Galvão G. Pereira traduz como “cotejo” a palavra-chave que orienta a formulação metodológica do capítulo “Observações sobre a epistemologia das ciências humanas”. Já na tradução realizada por Paulo Bezerra diretamente do russo para o português, em sua edição de 2017, o referido capítulo intitula-se “Metodologia das ciências humanas” e a palavra “cotejo” apresenta-se agora como “correlacionamento”. “Cotejar”, “correlacionar” textos, discursos, enunciados é um princípio metodológico bakhtiniano sobre a linguagem, podendo orientar o pesquisador que: i) a linguagem considera o outro sujeito; ii) por considerar o outro sujeito, a linguagem é dialógica; iii) não há textos, discursos, enunciados “sem possibilidade de diálogo, isto é, sem possibilidade de *resposta*.” (AMORIM, 2004, p. 95, grifo da autora)

Nesse sentido, o *tweet* presente na figura 2, considerado aqui como enunciado, apresenta uma resposta ao enunciado vídeo “Meu nome é Bettina” fundamentada em uma existência sócio-histórica-econômica

de luta, privações e de enfrentamentos em sua vida cotidiana (“ganhava nem mil conto”). Tal resposta é construída a partir de um diálogo com o discurso social, especialmente, sobre o tema dos privilégios de classe, e constitui um ato responsável que emerge sentidos sobre a impossibilidade de a classe trabalhadora brasileira conseguir vir a ser “investidora”: uma mãe adolescente com filhos para criar/educar não teria a oportunidade de classe para prover investimentos.

Bakhtin (2017, p. 275) observa que “essa alternância dos sujeitos do discurso (...) no diálogo (...), em que se alternam as enunciações dos interlocutores (parceiros do diálogo)” denomina-se “réplicas” (BAKHTIN, 2017, p. 275). Assim, cada réplica “possui uma conclusibilidade específica ao exprimir certa posição do falante que suscita resposta, em relação à qual se pode assumir uma posição responsiva” (BAKHTIN, 2017, p. 275). Nesse sentido, a resposta ao privilégio de classe econômica a qual Bettina pertence é mensurada pela contestação ao enunciado vídeo “Meu nome é Bettina” (“Que Bettina o que”) ao mesmo tempo em que se destaca o papel social desta mãe na educação por ter conseguido colocar “dois filhos na faculdade”. Nota-se que, ao dialogar com o discurso social de classe, o enunciado *tweet* em análise toma uma posição responsiva contrária ao discurso *coaching* financeiro promovido pelo vídeo publicitário. A resposta gerada e a constituição do ato responsável do enunciado *tweet* evidenciam as fragilidades econômico-sociais da maior parte da população brasileira, apontando para qual grupo social destina-se o discurso *coaching* financeiro veiculado pelo enunciado vídeo “Meu nome é Bettina”: uma minoria pertencente a classe abastada.

No *Twitter*, ainda, muitos usuários postavam montagens, fazendo uso do vídeo publicitário da Empiricus com piadas e ironias sobre seu conteúdo, gerando denúncia e humor sobre a publicidade. Além disso, a sentença que inicia o vídeo, “Oi, meu nome é Bettina, eu tenho 22 anos e 1 milhão e 42 mil reais de patrimônio acumulado”, acabou ganhando incontáveis variações realizadas por usuários que a adaptavam de acordo com suas realidades como, por exemplo, o segundo enunciado *tweet* a seguir, a ser analisado, publicado pelo perfil da revista *Globo Rural* que se utiliza do diálogo com o discurso *coaching* financeiro de Bettina para promover o discurso agropecuário de produtores de laticínios:

FIGURA 3 – *Tweet* do perfil da revista *Globo Rural* @Globo\_Rural, publicado no Twitter



Fonte: Veja São Paulo<sup>22</sup>

O perfil da revista *Globo Rural* no *Twitter* é amplamente conhecido por seu humor como estratégia de alcance aos leitores da rede social, que nem sempre são leitores do conteúdo da revista: notícias e reportagens sobre agronegócios, agricultura, pecuária, meio ambiente, entre vários outros referentes ao meio rural. Para anunciar a matéria sobre o aumento na compra do leite no ano de 2018 em comparação ao ano de 2017, o perfil da revista *Globo Rural* produz um enunciado com efeito de humor ao fazer dialogar o discurso agropecuário<sup>23</sup> com o

<sup>22</sup> Os *tweets* mais compartilhados e comentados que respondiam ao vídeo publicitário da Empiricus, “Meu nome é Bettina”, foram publicados pela Veja São Paulo em matéria publicada em 15 de março de 2019, sendo que este *tweet* do perfil da revista Globo Rural, assim como o *tweet* abordado anteriormente (figura 2), estão entre os mais comentados/compartilhados. Link para acesso: <https://vejasp.abril.com.br/cidades/bettina-propaganda-meme/> Acesso em: 20 jan. 2021.

<sup>23</sup> Para ler a matéria completa na revista *Globo Rural* acesse o link: <https://revistagloborural.globo.com/Noticias/Criacao/Leite/noticia/2019/03/globo-rural-aquisicao-de-leite-sobe-05-em-2018-ante-2017-diz-ibge.html#:~:text=A%20aquisi%C3%A7%C3%A3o%20de%20leite%20por,14%2C%20pelo%20Instituto%20Brasileiro%20de> Acesso em 19 de jan. 2021.

discurso *coaching* financeiro do enunciado vídeo “Meu nome é Bettina”. Para isso, o perfil da revista utiliza-se de formulação paródica (“Oi. Meu nome é *Globo Rural*, tenho 33 anos e o leite tem 116,60 milhões de litros de patrimônio acumulado”) como estratégia publicitária de divulgação da notícia de que no Brasil, a aquisição de leite por estabelecimentos produtores de laticínios obteve um crescimento importante e de destaque no cenário agropecuário brasileiro. Ao se promover a partir de um assunto que estava em foco nas mídias e redes sociais, tal seja o vídeo publicitário da Empiricus, “Meu nome é Bettina”, o enunciado *tweet* do perfil da revista *Globo Rural* tenta garantir maior alcance dos usuários do Twitter por meio do discurso do humor, ao mesmo tempo em que divulga a notícia do volume da produção de leite no Brasil do ano de 2018.

O enunciado *tweet* do perfil da revista *Globo Rural*, ainda, faz humor às avessas com o discurso *coaching* financeiro, uma vez que os dados, matematicamente falaciosos apresentados de modo automático, sem explicação, no enunciado vídeo “Meu nome é Bettina”, são substituídos no *tweet* por dados do IBGE (Instituto Brasileiro de Geografia e Estatística) sobre o aumento da compra e volume de leite produzido. Sendo o IBGE o principal fornecedor de dados e informações do Brasil, isso significa dizer que, ao se utilizar de uma fonte segura de dados, o enunciado *tweet* se posiciona contrário ao posicionamento descomprometido com a ciência do discurso *coaching* financeiro.

De acordo com Volóchinov (2017, p. 232) “cada elemento semântico isolável do enunciado, assim como o enunciado em sua totalidade, é traduzido por nós para outro contexto ativo e responsivo”, assim, “*toda compreensão é dialógica*” (VOLÓCHINOV, 2017, p. 232; grifos do autor). Ao mesmo tempo em que o enunciado *tweet* em cotejo provoca humor pela paródia do enunciado vídeo “Meu nome é Bettina”, dele se distingue e se distancia por operar com dados matemáticos concretos para divulgar a matéria sobre o aumento da compra e volume de leite produzido em 2018 em relação ao ano de 2017.

Ainda, diversos *tweets* apresentavam cálculos realizados por especialistas em matemática financeira e economia na tentativa de demonstrar a impossibilidade de se enriquecer em tão pouco tempo. Na figura 4, apresentamos o *tweet* de um dos economistas que apresentaram argumentos contra a veiculação e a ideia de investimento apresentada na publicidade da Empiricus.

FIGURA 4 – *Tweet* de economista contra a publicidade “Meu nome é Bettina”

Daqui a 15 anos, nossa heroína terá 37 anos de idade e um patrimônio de 157 quintilhões de reais: 2 milhões de vezes o PIB americano de 2018 e 316 milhões de vezes a fortuna Jeff Bezzos - homem mais rico do mundo segundo a Forbes.

Idade	Valor (R\$)
19	1.520
22	1.042.000
25	714.318.421
28	489.684.075.485
31	335.691.320.167.845
34	230.125.233.957.167.000
37	157.756.903.804.848.000.000

11:22 PM - 16 de mar de 2019 ⓘ

Fonte: Twitter.

O enunciado *tweet* em cotejo apresenta a voz de um especialista, no caso, trata-se de um economista, que apresenta um argumento com base em cálculos que rompe com a ideia de facilidade em investir e lucrar. Ao trazer dados de cálculo matemático, o enunciado *tweet* do economista posiciona-se responsivamente contrário ao vídeo publicitário “Meu nome é Bettina” e constitui um ato responsável quando demonstra a impossibilidade dos lucros em relação ao tempo de investimento garantido por Bettina. Para isso, utiliza-se de um tom irônico, sugerindo uma progressão absurda de valores que causa humor. Desse modo, o enunciado *tweet* oportuniza um debate sobre a ideia de ganhos fáceis sem antes averiguar a veracidade, as possibilidades e as exceções em relação ao que Bettina afirma ter ocorrido com ela: “...o que eu fiz não é nenhum segredo. Eu vivo falando por aí, pra todo mundo, eu comprei ações, na bolsa de valores. Não foi sorte, eu não herdei uma bolada, nem ganhei na loteria.” (00min17seg-00min28seg). Ou seja, ao tratar da Empiricus enquanto uma empresa que instiga o investimento por meio do convencimento, sem constatações e métodos claros, percebe-se que ela assume o discurso *coaching* financeiro do Instituto Brasileiro de Coaching (IBC) apresentado na introdução deste estudo.

O historiador dinamarquês Niklas Olsen, em uma entrevista de Daniel Zamora, publicada na revista argentina de ciências sociais *Nueva Sociedad*, sobre o seu mais recente trabalho, *Sovereign Consumer: A New Intellectual History of Neoliberalism*,<sup>24</sup> aponta que a retórica da escolha é geralmente enganosa no discurso neoliberal. Ao mesmo tempo

<sup>24</sup> Consumidor soberano: uma nova história intelectual do neoliberalismo (tradução nossa). (Basingstoke: Palgrave Macmillan, 2019)

em que é virtualmente impossível se opor à ideia de livre escolha para todos, na realidade, a maioria das pessoas tem muito pouco dinheiro para gastar e poucos itens para escolher em uma economia dominada pela desigualdade generalizada e pelas grandes empresas monopólicas. E uma vez que essa retórica nos convence, destrói a nossa capacidade de fazer demandas coletivas por direitos sociais.

Sob essa perspectiva, a empresa Empiricus, juntamente com seus analistas, apresenta propostas que são facilmente desacreditadas quando são postas frente à ciência ou se deparam com acadêmicos e estudiosos que difundem conhecimento com base em evidências e não em relatos pessoais. Mediante este acontecimento, abre-se espaço para a compreensão de que o discurso *coaching* financeiro, promovido pelo vídeo publicitário “Meu nome é Bettina”, dialoga com o tema da automatização da ciência, como se não fossem mais necessários estudos analíticos de cálculo sobre o sistema econômico e a ciência fosse automática, ou seja, realizada de forma mecânica, sem nenhuma intervenção humana, ignorando os processos científicos da matemática que fundamentam os cálculos econômicos que dependem *de* e oscilam *por* diversas condições. Por exemplo, em: “Comecei com 19 anos e 1.520 reais. Três anos depois, tenho mais de 1 milhão, simples assim.” (00min28seg-00min34seg) há, sobretudo, uma simplificação do funcionamento do sistema econômico que produz sentidos de apagamento das condições materiais de existência do interlocutor que a campanha publicitária alcançou, usuários da rede de compartilhamento de vídeos Youtube, que possuem os mais diversos perfis sócio-econômicos, gerando sentidos de banalização sobre poder aquisitivo. Ao narrar sua trajetória financeira e seus ganhos, Bettina a faz de modo reducionista, como se ela não pertencesse ao sistema econômico financeiro que orienta as práticas de mercado, sendo possível para qualquer um, independentemente de sua condição e existência material, o mesmo feito.

#### 4 Considerações finais

O sistema capitalista tem como característica o individualismo, que marca historicamente o processo de revolução industrial do século XIX, sendo central tanto no discurso de autoajuda quanto no discurso neoliberal, ambos discursos produzidos pelo capitalismo e colocados em diálogo na construção do discurso *coaching* financeiro. O culto ao individualismo enfraquece a coletividade e, conseqüentemente, as

lutas por direitos sociais, colocando o sujeito como único responsável por si mesmo, por suas condições materiais de existência, silenciando os processos sócio-históricos e o sistema político que atuam sobre os indivíduos. Sob essa perspectiva, o discurso *coaching* financeiro, ao dialogar com os discursos de autoajuda e neoliberal, encontra condições sociais oportunas para fazer circular sentidos que promovam a ideia do enriquecimento individual, que é contrária à ideia progressista de distribuição igualitária da riqueza para a população, de modo a equalizar as oportunidades e acesso aos bens de consumo e culturais. Lipovetsky e Serroya (2015, p. 09), ao descreverem o capitalismo, afirmam que:

Abraçando unicamente a rentabilidade e o reinado do dinheiro, o capitalismo aparece como um rolo compressor que não respeita nenhuma tradição, não venera nenhum princípio superior, seja ele ético, cultural ou ecológico. Sistema comandado por um imperativo de lucro que não tem outra finalidade senão ele próprio, a economia liberal apresenta um aspecto niilista cujas consequências não são apenas o desemprego e a precarização do trabalho, as desigualdades sociais e os dramas humanos, mas também o desaparecimento das formas harmoniosas de vida, o desvanecimento do encanto e da graça da vida em sociedade (...).

O discurso neoliberal tem origem no discurso capitalista em que se coloca o lucro, o consumo e a geração de renda, capital e bens acima da equidade social, como se essas práticas fossem a salvação do indivíduo e da sociedade, sendo que é bem o contrário. O mesmo fenômeno ocorre com o discurso de autoajuda, que responsabiliza unicamente o indivíduo, tanto por suas conquistas quanto por seus fracassos. Se o capitalismo se realiza por meio de práticas predatórias, apagando as crises econômicas e sociais que o próprio sistema gera, “provocando catástrofes ecológicas de grandes proporções, reduzindo a proteção social, aniquilando as capacidades intelectuais e morais, afetivas e estéticas dos indivíduos” (LIPOVETSKY; SERROYA, 2015, p. 09), têm-se no diálogo com os discursos neoliberal e de autoajuda suas reverberações na hipermodernidade, no projeto de dizer do discurso *coaching* financeiro veiculado pelo enunciado vídeo publicitário “Meu nome é Bettina”.

Sob essa perspectiva, uma das características do discurso *coaching* financeiro, ao se basear no diálogo com os discursos neoliberal e de autoajuda, é o individualismo, que na comunicação discursiva publicitária do enunciado vídeo “Meu nome é Bettina” provoca a homogeneização

do seu interlocutor em potencial: o cliente que investe ou que poderia investir. Assim, o discurso *coaching* financeiro silencia a heterogeneidade das condições sociais e econômicas, ao anunciar produtos e ideias de investimento em redes sociais com a proposta de alcance e acesso aos mais diversos públicos de diferentes classes econômicas, como se todos tivessem poder aquisitivo de investidor em potencial.

Nesse contexto, o discurso *coaching* financeiro caracteriza-se pela promoção publicitária de investimento mínimo com lucro certo para alcançar maior público possível, enunciando a falsa ilusão de enriquecimento fácil em pouco tempo. Por ser um discurso que dialoga com o discurso de autoajuda e o discurso neoliberal, o discurso *coaching* financeiro encontra na hipermodernidade condições de realização, de modo que diversos enunciados advindos de múltiplos gêneros publicitários (anúncios, vídeos, posts, postagens etc.) da internet e fora dela, fazem circular a (hiper) facilidade do indivíduo vir a ser (hiper) rico: um milionário.

O aparecimento do enunciado vídeo “Meu nome é Bettina” pode ser compreendido como uma forma sistemática de publicidade predatória que o contexto sócio-histórico permite, assim como revela uma política de captura do indivíduo como empreendedor de seu próprio capital, podendo ser observada em diversas publicidades de setores financeiros no cenário nacional. O sujeito Bettina deve ser compreendido enquanto investidor pertencente a uma classe elitista que possui capital para investir em ações na bolsa de valores. No entanto, o enunciado revela dizeres enganosos que garantem facilidade de enriquecimento para qualquer sujeito que acessar o material da Empiricus, disponível após o anúncio publicitário promovido por Bettina, sendo que a empresa não se responsabiliza, no caso de o sujeito não lucrar.

Os diálogos possíveis no enunciado vídeo publicitário “Meu nome é Bettina” revelam redes de memória na esfera política econômica neoliberal que apagam o saber científico, o conhecimento, produzindo sentidos que operam na urgência da transformação do *status* social do indivíduo. De acordo com Volóchinov (2017, p. 232), “em cada palavra de um enunciado compreendido, acrescentamos como que uma camada de nossas palavras responsivas”. Nesse sentido, “comprei”, “ganhei”, “lucrar” apresentam-se como palavras do discurso neoliberal que também participam do discurso *coaching* financeiro; assim como “passo a passo”, “É com você!”, “atitudes diferentes!”, são palavras recorrentes do discurso de autoajuda que dialoga e faz funcionar o discurso *coaching* financeiro no enunciado vídeo em estudo.

É em relação ao aparecimento do enunciado vídeo “Meu nome é Bettina” que múltiplos enunciados advindos do discurso *coaching* financeiro mostram-se constituídos por discursos produzidos pelo capitalismo, como o discurso neoliberal e o discurso de autoajuda arrolados na hipótese deste estudo. A hipótese de que o discurso *coaching* financeiro ancora-se no diálogo com os discursos de autoajuda e neoliberal no enunciado vídeo “Meu nome é Bettina”, para a realização da comunicação discursiva publicitária da Empiricus, apresenta-se de modo afirmativo, confirmando-se na análise aqui realizada. Diante disso, a partir dos fundamentos filosóficos sobre a linguagem advindos dos escritos de Bakhtin e do Círculo russo (Volóchinov e Medviédev), entende-se que o discurso *coaching* financeiro apresenta-se constituído por diálogos, especialmente, com os discursos de autoajuda e neoliberal, sendo por meio desses diálogos que o discurso *coaching* financeiro funciona na esfera da comunicação midiática publicitária. Ainda, enquanto enunciado, o enunciado vídeo “Meu nome é Bettina” responde ao mesmo tempo que é respondido por tantos outros enunciados em diálogos com diversos outros discursos como, por exemplo, o científico, o da economia, o do humor, o de classe social – para mencionar alguns que aparecem no cotejo de enunciados realizado neste estudo, como forma de ilustrar a responsividade do enunciado na comunicação discursiva.

Na esfera *coaching*, de modo mais amplo e para além do que se propôs neste estudo, observa-se um momento político fértil para a proliferação de discursos *homo economicus*: um sujeito livre para investir, comprar e lucrar, promovendo um modo de vida individualista, baseado no ganho e bem-estar exclusivamente pessoal. Ao assumir o discurso neoliberal, gradativamente o Estado deixa de assumir suas funções de seguridade social, de incentivo e promoção de ações de desenvolvimento coletivo nas áreas da ciência, da saúde, da educação, da habitação etc., abandonando a concepção democrática de políticas públicas efetivas.

### **Contribuição dos Autores**

Grenissa Stafuzza contribuiu com a análise do discurso *coaching* financeiro, escreveu a fundamentação teórica de perspectiva bakhtiniana e realizou o trabalho de revisão e correção teórica e textual do artigo. Maximiano Antonio Pereira contribuiu com o tema do discurso *coaching* financeiro no enunciado vídeo publicitário “Meu nome é Bettina”, coletou o *corpus* e escreveu a fundamentação teórica que embasa as discussões.

## Referências

- ALMEIDA, M. P. *Reformas neoliberais no Brasil: a privatização nos governos Fernando Collor e Fernando Henrique Cardoso*. 2010. 427f. Tese (Doutorado em História Social) – Programa de Pós-Graduação em História, Universidade Federal Fluminense, Niterói, 2010. Disponível em: <http://nupehic.net.br/wp-content/uploads/2018/08/a-privatiza%C3%A7%C3%A3o.pdf> Acesso em: 29 jan. 2021.
- AMORIM, M. *O pesquisador e o seu outro: Bakhtin nas ciências humanas*. Rio de Janeiro: MUSA, 2004.
- BAKHTIN, M. M. *Estética da criação verbal*. Trad. Maria Ermantina Galvão G. Pereira a partir do francês. São Paulo: Martins Fontes, 1997.
- BAKHTIN, M. M. *Para uma filosofia do ato responsável*. Trad. Vladimir Miotello e Carlos Alberto Faraco. 2. ed. São Carlos: Pedro e João Editores, 2010.
- BAKHTIN, M. M. *Os gêneros do discurso*. Trad. Paulo Bezerra. São Paulo: Editora 34, 2016.
- BAKHTIN, M. M. *Estética da criação verbal*. 6. ed. Trad. Paulo Bezerra a partir do russo. São Paulo: Editora Martins Fontes, 2017.
- BRUNELLI, A. F. *O sucesso está em suas mãos: análise do discurso de auto-ajuda*. 2004. 149f. Tese (Doutorado em Linguística) – Programa de Pós-Graduação em Linguística, Universidade Estadual de Campinas, Campinas, 2004. Disponível em: [http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/268906/1/Brunelli\\_AnaFlora\\_D.pdf](http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/268906/1/Brunelli_AnaFlora_D.pdf) Acesso em: 29 jan. 2021.
- CHAGAS, A. T. S. *A ilusão no discurso da auto-ajuda e o sintoma social*. Ijuí: Ed. UNIJUÍ, 1999.
- DUFOUR, D. R. *O divino mercado: a revolução cultural liberal*. Trad. Procópio Abreu. Rio de Janeiro: Companhia de Freud, 2008.
- EISENBACH NETO, F. J.; CAMPOS, G. R. de. O impacto do neoliberalismo na educação brasileira. In: CONGRESSO NACIONAL DE EDUCAÇÃO: FORMAÇÃO DE PROFESSORES: CONTEXTOS, SENTIDOS E PRÁTICAS, XIII., 2017, Curitiba. *Anais [...]*. Curitiba: EDUCERE, 2017. p. 10985-10999. Disponível em: [https://educere.bruc.com.br/arquivo/pdf2017/24420\\_12521.pdf](https://educere.bruc.com.br/arquivo/pdf2017/24420_12521.pdf) Acesso em: 26 mar. de 2021.

LIPOVETSKY, G. Tempo contra tempo, ou a sociedade hipermoderna. In: LIPOVETSKY, G.; CHARLES, S. (org.). *Os tempos hipermodernos*. Tradução de Mario Vilela. São Paulo: Barcarolla, 2004. p. 49-104.

LIPOVETSKY, G; SERROYA, J. *Estetização do mundo: viver na era do capitalismo artista*. Trad. Eduardo Brandão. São Paulo: Editora Companhia das Letras, 2015.

MEDVIÉDEV, P. N. *O método formal nos estudos literários: introdução crítica a uma poética sociológica*. Trad. Sheila Camargo Grillo e Ekaterina Vólkova Américo. São Paulo: Contexto, 2012.

SIMPSON, J. (ed.). *Oxford English Dictionary*. 3. ed. Oxford: Oxford University Press, 2017.

SOUZA, M. B.; HOFF, T. S. R. O governo Temer e a volta do neoliberalismo no Brasil: possíveis consequências na habitação popular. *urbe. Revista Brasileira de Gestão Urbana*, Curitiba, v. 11, p. 1-14, 2019. DOI: <https://doi.org/10.1590/2175-3369.011.e20180023>. Disponível em: <https://www.scielo.br/pdf/urbe/v11/2175-3369-urbe-11-e20180023.pdf>. Acesso em: 8 out. 2020.

STAFUZZA, G. B. Sentidos do enunciado verbovisual em pôsteres publicitários de Bastardos inglórios. *SCRIPTA*, Belo Horizonte, v. 22, n. 45, p. 137-150, 2018. DOI: <https://doi.org/10.5752/P.2358-3428.2018v22n45p137-150>. Disponível em: <http://periodicos.pucminas.br/index.php/scripta/article/view/17133/13828> Acesso em: 29 jan. 2021.

VOLOCHÍNOV, V. *Marxismo e filosofia da linguagem*. Trad. Sheila Vieira de Camargo Grillo e Ekaterina Vólkova Américo. São Paulo: Editora 34, 2017.

ZAMORA, D. Cómo el neoliberalismo reinventó la democracia. Entrevista a Niklas Olsen. *Nueva Sociedad*, Buenos Aires, n. 282, jul./ago. 2019. Disponível em: <https://nuso.org/articulo/como-el-neoliberalismo-reinvento-la-democracia/>. Acesso em: 15 jan. 2021.



## Uma gramática computacional de um fragmento do nheengatu

### *A computational grammar for a fragment of Nheengatu*

Leonel Figueiredo de Alencar

Universidade Federal do Ceará (UFC), Fortaleza, Ceará / Brasil

leonel.de.alencar@ufc.br

<https://orcid.org/0000-0001-8148-6994>

**Resumo:** A disponibilidade de recursos para o processamento computacional constitui um dos fatores de sobrevivência de uma língua. O objetivo deste trabalho foi implementar um fragmento do nheengatu no formalismo *Grammatical Framework*, especialmente projetado para o desenvolvimento de aplicações multilíngues. Outrora mais falado que o português na Amazônia, o nheengatu está ameaçado de extinção, embora ainda conte com estimados 14000 falantes. O fragmento restringe-se a orações que expressam estados contingentes e não-contingentes, mas inclui fenômenos gramaticais estruturalmente complexos típicos da família tupi-guarani, os quais contrastam fortemente com as construções equivalentes em português e inglês. Constitui um dos módulos da GrammYEP, uma gramática computacional multilíngue que integra módulos análogos do inglês e do português. A implementação tomou como ponto de partida as descrições gramaticais não formalizadas de Navarro (2011) e Cruz (2011). A formalização revelou lacunas e inconsistências nessas abordagens, em parte sanados por meio de uma reanálise dos dados. A GrammYEP alcançou resultados bastantes satisfatórios na tradução do e para o nheengatu. Traduziu para o português e o inglês a totalidade de um conjunto-teste de 142 sentenças dessa língua. Inversamente, verteu para o nheengatu 98,18% e 84,11% dos conjuntos-teste correspondentes em português e inglês. Por outro lado, analisou apenas dois exemplos de um conjunto-teste negativo com 171 construções agramaticais em nheengatu. Desta avaliação resultou um *treebank* com 243 sentenças do nheengatu, emparelhadas com as sentenças equivalentes em português e inglês.

**Palavras-chave:** língua geral amazônica (LGA); tupi moderno; predicação qualificativa; construção possessiva; tradução automática; linguística computacional; processamento de linguagem natural.

**Abstract:** The availability of resources for computational processing is one of the survival factors of a language. The goal of this work was to implement a fragment of Nheengatu in the Grammatical Framework formalism, specially designed for the development of multilingual applications. Once more widely spoken than Portuguese in the Amazon region, Nheengatu is threatened with extinction, although it still has an estimated number of 14,000 speakers. The fragment is restricted to sentences that express contingent and non-contingent states, but includes structurally complex grammatical phenomena typical of the Tupí-Guaraní family, which strongly contrast with the equivalent constructions in Portuguese and English. It constitutes one of the modules of GrammYEP, a multilingual computational grammar comprising equivalent English and Portuguese modules. The starting point of the implementation was the non-formalized grammatical descriptions of Navarro (2011) and Cruz (2011). The formalization revealed gaps and inconsistencies in these approaches, which were partly remedied through a reanalysis of the data. GrammYEP achieved quite satisfactory results in the translation from and to Nheengatu. It translated into Portuguese and English all examples from a test set with 142 Nheengatu sentences. Conversely, 98.18% and 84.11% of the corresponding Portuguese and English test sets were rendered into Nheengatu. On the other hand, it parsed only two examples from a negative test set with 171 ungrammatical constructions in Nheengatu. This evaluation resulted in a treebank with 243 Nheengatu sentences, paired with the equivalent sentences in Portuguese and English.

**Keywords:** Amazonian Lingua Franca; Modern Tupí; qualifying predication; possessive construction; machine translation; computational linguistics; natural language processing.

Recebido em 27 de agosto de 2020

Aceito em 04 de novembro de 2020

## 1 Introdução

Uma gramática computacional é uma descrição formal das estruturas sintáticas e lexicais de uma língua capaz de ser utilizada pelo computador para análise ou geração de sentenças (DUCHIER; PARMENTIER, 2015). Esse tipo de recurso integra a arquitetura de diversas tecnologias de linguagem natural, como tradutores automáticos, sistemas de diálogo etc. (JURAFSKY; MARTIN, 2009). Por outro lado, constitui valioso instrumento para a linguística teórica, ao permitir a verificação automática da consistência interna e da validade empírica de uma determinada modelagem de fenômenos gramaticais (BENDER, 2008, 2010; MÜLLER, 2015).

Um dos fatores de sobrevivência de uma língua nos dias de hoje é possuir recursos para o processamento computacional (PIRINEN *et al.*, 2017). Entre os mais urgentes figuram gramáticas computacionais, por permitirem a construção de vários outros recursos, como *treebanks* (HAJIČOVÁ, 2010).

Neste artigo, apresentamos um fragmento de gramática do nheengatu no formalismo computacional *Grammatical Framework* (RANTA, 2011), doravante GF, especialmente projetado para facilitar o desenvolvimento de aplicações multilíngues, incluindo tradutores automáticos, sistemas de diálogo e ferramentas de aprendizagem de línguas mediada por computador. Esse fragmento integra a GrammYEP, uma gramática computacional multilíngue da qual fazem parte fragmentos equivalentes do inglês e do português. Nessa sigla, as três línguas do sistema estão identificadas pela primeira letra dos respectivos códigos no padrão ISO 639-3, respectivamente, *yrl*, *eng* e *por* (EBERHARD; SIMONS; FENNIG, 2020). O primeiro deriva de *yeral*, designativo em espanhol do nheengatu, também conhecido como geral, língua geral amazônica ou tupi moderno, entre outros termos (RODRIGUES, 1996; EBERHARD; SIMONS; FENNIG, 2020).

Descendente do tupinambá, língua tupi-guarani que era falada no norte do Brasil até o século XVIII (FREIRE, 2011; RODRIGUES, 1996), o nheengatu é a L1 ou L2 de estimadas 6000 pessoas no município de São Gabriel da Cachoeira, na região do alto rio Negro, e 8000 em território colombiano, pertencentes a etnias originalmente falantes de línguas não tupis, quase todas extintas ou moribundas (CRUZ, 2011; EBERHARD; SIMONS; FENNIG, 2020). Praticamente desaparecido na Venezuela e ameaçado de extinção tanto na Colômbia quanto no Brasil, ainda é usado neste último pela população em idade reprodutiva, porém, a transmissão às crianças está sendo interrompida (EBERHARD; SIMONS; FENNIG, 2020). Felizmente, nos últimos anos, tem passado por um processo de revitalização, impulsionado por traduções literárias (NAVARRO; ÁVILA; TREVISAN, 2017) e cursos em universidades brasileiras, despertando o interesse também do público não indígena.

Vários outros fatores concorreram para a escolha do nheengatu como protótipo inicial de um projeto maior de implementação de gramáticas computacionais de línguas indígenas brasileiras. Em primeiro lugar, pesou a sua importância histórica, por ter sido, durante dois séculos e meio, “a principal língua da Amazônia”, posição que perderia para o

português apenas na segunda metade do século XIX, conforme Freire (2011, p. 16-17), certamente o mais abrangente levantamento da história social da língua geral amazônica da sua origem no século XVII ao século XXI. Afirmar Aryon Dall'Igna Rodrigues no prefácio desse livro:

A história da Língua Geral Amazônica é, sem dúvida, uma das mais interessantes nas Américas e deve ser conhecida não só pelos estudiosos da História do Brasil, mas também por todos os que estudam as Ciências Sociais, as Letras e a Linguística neste país. (RODRIGUES, 2011, p. 13)

O segundo fator decisivo foi a existência de uma descrição normativa didática de acesso fácil e gratuito em formato de PDF editável (NAVARRO, 2011). Finalmente, o GF não havia sido aplicado antes a qualquer língua ameríndia. A implementação de fenômenos gramaticais de uma língua do tronco tupi oferece a oportunidade de testar a abrangência do formalismo, possibilitando detectar eventual aspecto de difícil ou impossível implementação.

A GrammYEP é um projeto em andamento. Cobre, na versão atual, cerca de um quinto do conteúdo gramatical de Navarro (2011), incluindo diversos fenômenos de considerável complexidade gramatical típicos da família tupi-guarani, exemplificados em (1)-(17).<sup>1</sup>

- (1) kambi s-aku                      u-iku  
leite 3s.INACT-quente 3s.ACT-estar  
'o leite está quente'

- (2) kuá                      meíú-itá ta                      r-aku u-iku  
DEM.PROX beiju-PL 3p.INACT N3s-quente<sup>2</sup> 3p.ACT-estar  
'estes beijus estão quentes'

<sup>1</sup> Na glosas interlineares, seguimos as *Leipzig Glossing Rules* (LGR), complementado-as com símbolos propostos por Gynan (2017) para determinados morfemas do guarani, a saber: ACT=ativo, INACT=inativo, IMPR=impessoal (*impersonal* em inglês) e RLL=relacionalizador.

<sup>2</sup> A abreviatura N3s segue convenção das LGR, onde N- corresponde ao prefixo *non-* 'não'. Marca concordância com todas as pessoas, exceto a 3ª do singular.

- (3)   nhaã           t-imbiú           sepiasu   retana  
DEM.DIST IMPR-comida caro       muito  
‘aquela comida é muito cara’
- (4)   kuá-itá           pirá       piranga   sepiasuíma u-iku  
DEM.PROX-PL peixe   vermelho barato       3p.ACT-estar  
‘estes peixes vermelhos estão baratos’
- (5)   s-endaua                   puranga  
3s.INACT-comunidade bonito  
‘a comunidade dele é bonita’
- (6)   s-uka s-uri  
3s.INACT-casa 3s.INACT-alegre  
‘a casa dele é alegre’
- (7)   i           igara   i           pusé   u-iku  
3s.INACT canoa 3s.INACT   pesado 3s.ACT-estar  
‘a canoa dele está pesada’
- (8)   (ixé)<sup>3</sup> se           pusé  
(eu) 1s.INACT   pesado  
‘(eu) sou pesado’
- (9)   (ixé) se r-uri  
(eu) 1s N3s-alegre  
‘(eu) estou alegre’
- (10) (aé) s-uri<sup>4</sup>  
(ele) 3s.INACT-alegre  
‘(ele) é alegre’

<sup>3</sup> Nos exemplos deste artigo, constituintes entre parênteses são opcionalmente realizáveis.

<sup>4</sup> O *nheengatu* não marca gênero nos pronomes e prefixos flexionais pessoais. Nas traduções para o português, traduzimos esses elementos indistintamente pelo masculino ou feminino.

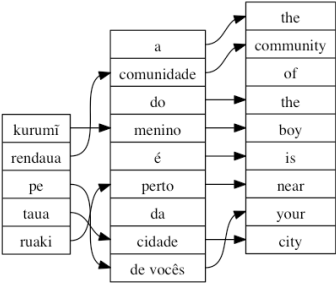
- (11) (ixé) a-iku iké  
(eu) 1s.ACT-estar aqui  
'(eu) estou aqui'
- (12) (iandé) ti ia-iku t-endaua upé  
(nós) NEG 1p.ACT-estar IMPR-comunidade em  
'(nós) não estamos na comunidade'
- (13) (indé) ti ne r-uri re-iku  
(tu) NEG 2s.INACT N3s-alegre 2s.ACT-estar  
'(tu) não estás alegre'
- (14) Maria r-uka ti s-uaki  
Maria RLL-casa NEG 3s.INACT-perto.de  
'a casa da Maria não é perto dele'
- (15) kurumĩ r-endaua pe taua r-uaki  
menino RLL-comunidade 2p.INACT cidade RLL-perto.de  
'a comunidade do menino é perto da cidade de vocês'
- (16) nhaã x-imiriku kisé ti u-iku uka pupé  
DEM.DIST 3s.INACT-esposa faca NEG 3s.ACT-estar casa dentro.de  
'aquela faca da esposa dele não está dentro da casa'
- (17) nhaã Pedro pindá u-iku igara kuara upé  
DEM.DIST Pedro anzol 3s.ACT-estar canoa buraco em  
'aquele anzol do Pedro está dentro da canoa'

Destaquemos três desses fenômenos. O primeiro é a distinção entre verbos ativos e inativos, flexionados por duas séries diferentes de prefixos de concordância. O segundo é a realização do argumento interno de substantivos e posposições pelos prefixos da série inativa. Finalmente, temos a divisão de verbos, substantivos e posposições em uniformes e multiformes. Enquanto em todas as traduções a mesma forma básica dos substantivos *comunidade* e *casa* é utilizada, nos exemplos correspondentes do nheengatu em (5)-(16) ocorrem três formas distintas

de cada lexema. Analogamente, os verbos inativos traduzidos pelos adjetivos *quente* e *alegre* e as posposições manifestam duas formas diferentes, que compartilham os mesmos prefixos *s-* e *r-* desses dois substantivos.

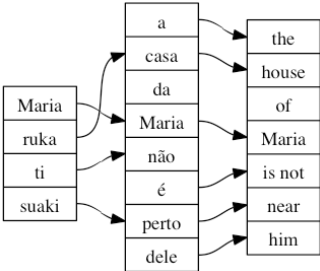
Características tipológicas constituem uma das principais dificuldades para a tradução automática (JURAFSKY; MARTIN, 2009). Um sistema como o que propomos precisa lidar não apenas com diferenças significativas entre o português e o inglês, de ramos distintos do tronco indo-europeu, mas também com discrepâncias estruturais ainda maiores em relação ao tronco tupi, como se pode constatar em (1)-(17). Não obstante isso, a GrammYEP possibilita a tradução automática entre as três línguas de forma satisfatória, como exemplificam as Figuras 1 e 2.

FIGURA 1 – Tradução automática de (15) com alinhamento das palavras



Fonte: Elaboração própria.

FIGURA 2 – Tradução automática de (14) com alinhamento das palavras<sup>5</sup>



Fonte: Elaboração própria.

<sup>5</sup> Na versão em inglês, a cópula e a negação ocupam uma mesma célula porque foram amalgamadas num único *token* para simplificar a implementação da contração *isn't*.

Na próxima seção, especificamos o fragmento do *nheengatu* abrangido pela *GrammYEP*. A seção seguinte trata do formalismo GF, utilizado na modelação computacional desses fenômenos. Na seção 4, descrevemos os aspectos mais importante da implementação. A seção 5 apresenta os resultados da aplicação do sistema na análise e tradução automáticas de sentenças. A última seção traz as conclusões, problemas remanescentes e perspectivas para a expansão da gramática.

## 2 Cobertura do fragmento gramatical

A implementação de uma gramática computacional de qualquer língua natural é uma tarefa extremamente complexa, que impõe a necessidade de segmentar o desenvolvimento em sucessivas fases de crescente nível de complexidade e abrangência (FRANCEZ; WINTNER, 2012). De fato, não se limita a converter, para um dado formalismo, as descrições linguísticas disponíveis, mas implica também dirimir inconsistências e preencher lacunas destas, decorrentes da formalização insuficiente. Nesta seção, apresentamos o recorte do *nheengatu* do qual a *GrammYEP*, atualmente, implementa a maior parte. A quantificação do DP e construções possessivas com múltiplos núcleos nominais encaixados ficaram para uma próxima versão.

Na especificação desse fragmento, utilizamos a gramática independente de contexto (doravante CFG, do inglês *context-free grammar*), formalismo computacional que, pela sua simplicidade conceitual, é mais acessível a um público não especializado do que os formalismos mais complexos como a Gramática Léxico-Funcional (BRESNAN, 2001) e o GF. A limitação da CFG é dificultar a implementação de fenômenos que envolvem propriedades de subclasses de categorias sintáticas, como a concordância e a valência, tratados de forma elegante nesses outros modelos por meio de mecanismos mais expressivos (SAG; WASOW; BENDER, 2003). Desse modo, o fragmento apresentado nesta seção gera também muitos exemplos agramaticais, problema que será enfrentado com a modelação em GF descrita na seção 4.

Utilizamos como fontes de conhecimento sobre a língua, principalmente, duas das exposições gramaticais mais recentes e livremente disponíveis na Internet em formato de PDF editável, a saber, Cruz (2011) e Navarro (2011). Para dirimir dúvidas, consultamos Casasnovas (2006), exposição gramatical sucinta com muitos exemplos

traduzidos seguida de uma coletânea de lendas e glossários, que tem sido utilizada em cursos para falantes.<sup>6</sup> Constitui importante documentação da língua falada no rio Negro, mas as explicações gramaticais não são suficientemente detalhadas para não falantes. Por exemplo, não há qualquer menção à flexão de verbos inativos, da qual se apresentam apenas exemplos, como na p. 26, para ilustrar o tópico “Verbo SER, ESTAR (IKÚ)”. Apresenta-se apenas a conjugação dos verbos ativos.

Cruz (2011) é uma tese de doutorado sobre a língua falada no alto rio Negro. A transcrição dos inúmeros exemplos, extraídos sobretudo de um corpus oral compilado a partir de textos autênticos, procura geralmente refletir a pronúncia da língua falada.

Navarro (2011, p. 7) segue uma orientação oposta. Propõe uma gramática normativa da língua com base “nos seus vários autores, mas respeitando os fatos linguísticos da língua geral falada hoje em dia, principalmente nos centros urbanos do médio e alto rio Negro.” Consiste de treze lições que introduzem o vocabulário e as estruturas gramaticais de forma progressiva por meio de diálogos, narrativas, canções etc., acompanhados de exercícios. Esse manual é utilizado em cursos ministrados pelo autor na Universidade de São Paulo.<sup>7</sup>

Todos esses trabalhos são de inestimável valor e, ao nosso ver, complementam-se. Por um lado, Cruz (2011) permite dirimir lacunas de Navarro (2011), embora também suscite outras dúvidas, como veremos mais adiante. No caso de Navarro (2011), essas questões resultam, provavelmente, de uma preocupação pedagógica em evitar o jargão linguístico, atendo-se à nomenclatura tradicional da gramática do português. Casasnovas (2006), por sua vez, constitui um rico repositório de dados para cotejo com esses dois trabalhos.

---

<sup>6</sup> Conforme um dos avaliadores anônimos, “a gramática mais utilizada pelos falantes de Nheengatu”. Infelizmente, quando da implementação da GrammYEP, não encontramos na Internet o arquivo completo desse manual, que é citado tanto por Navarro (2011) quanto por Cruz (2011). Agradecemos o envio de link para a imagem de um exemplar físico da segunda edição (CASASNOVAS, 2006). Esse arquivo, porém, apresenta alguns problemas. Faltam as páginas 94 e 95. Anotações manuscritas e trechos apagados dificultam a conversão para o formato de PDF editável e, conseqüentemente, a compilação de um corpus para buscas automáticas de exemplos. Uma edição de 2014, indicada pelo parecerista, não foi encontrada no site da editora.

<sup>7</sup> Disponível em: <http://tupi.fflch.usp.br/curso-de-lingua-nheengatu-e-de-cultura-amazonica>. Acesso em: 27 ago. 2020.

Por outro lado, acreditamos que o estabelecimento de uma norma padrão, sem inibir a riqueza da variação intralinguística, constitui fator de vitalidade de uma língua. Em primeiro lugar, permite o uso da língua em contextos não coloquiais, por exemplo, em textos legais, científicos etc. Em segundo lugar, promove o intercâmbio entre falantes de diferentes variedades, incluindo o diálogo intergeracional propiciado pela literatura de épocas passadas, da qual o *nheengatu* é particularmente rico.<sup>8</sup> Em terceiro lugar, favorece o desenvolvimento de aplicações de processamento de linguagem natural, para o qual a variação linguística constitui um entrave, ainda mais no caso de uma língua minoritária como o *nheengatu*. De fato, a relação custo-benefício de aplicações customizadas para determinadas variedades seria bastante desfavorável, dado o pequeno número de falantes de cada uma. Finalmente, facilita o aprendizado do idioma como língua estrangeira (LE) ou como segunda língua (L2), sobretudo por pessoas sem acesso a falantes nativos. Esse último aspecto é particularmente relevante no caso do *nheengatu*, que parece ocupar uma posição singular no quadro das línguas indígenas brasileiras pela expressiva quantidade de textos produzidos por usuários da língua como LE. Nesse repertório, sobressaem as traduções para o *nheengatu* de clássicos da literatura (NAVARRO; ÁVILA; TREVISAN, 2017).

Há muita divergência na representação ortográfica do *nheengatu*. Cruz (2011), Navarro (2011) e Ávila (2016), por exemplo, discrepam no uso do acento agudo. Diferentemente da primeira, os dois últimos, continuando tradição que remonta pelo menos a Sympson (1877), separam das respectivas bases os prefixos do Quadro 1, ver (7) e (18).<sup>9</sup> Casanovas (2006, p. 26), por sua vez, grafá *peyumasí* ('vocês são famintos' na nossa tradução), amalgamando o prefixo de 2ª pessoa do plural *pe* ao verbo inativo *yumasí* 'ser faminto'. Em *pe igara* 'canoa de vocês', porém, o mesmo prefixo é grafado separado (CASANOVAS, 2006, p. 34).

(18)	a-iku	i	irũmu
	1s.ACT-estar	3s.INACT	com
	'estou com ela'		

<sup>8</sup> Ver levantamentos de Freire (2011) e Ávila (2016).

<sup>9</sup> A segunda edição de Navarro (2011), datada de 2016, atualmente indisponível, adota convenções ortográficas parecidas com as de Ávila (2016).

QUADRO 1 – Prefixos inativos silábicos conforme Navarro (2011)

1s.INACT	2s.INACT	3s.INACT	1p.INACT	2p.INACT	3p.INACT
se	ne	i	iané	pe	aintá (ta)

Fonte: Elaboração própria.

Os prefixos do Quadro 1 integram a série estativa ou série II, distinguindo verbos estativos ou inativos de verbos dinâmicos ou ativos (CRUZ, 2011; PRAÇA; MAGALHÃES; CRUZ, 2017). Desempenham três funções em nheengatu, conforme o tipo de base a que se adjungem. Quando a base é um verbo inativo flexionável, como na segunda ocorrência de *i* em (7), marcam a concordância com o sujeito da sentença. No caso de posposições e substantivos, realizam o argumento interno, como em (18) e na primeira ocorrência de *i* em (7).

A GrammYEP objetiva inicialmente auxiliar o autoestudo da língua a partir de Navarro (2011) por meio da implementação de um sistema de tradução automática e da compilação de *treebank*. Desse modo, preferimos seguir tanto suas convenções ortográficas quanto sua orientação normativa na definição do fragmento a ser modelado computacionalmente. Certamente, o estabelecimento de uma ortografia unificada e a adoção de uma norma padrão para uma língua indígena devem resultar de decisão por parte das comunidades de falantes.<sup>10</sup> A GrammYEP, porém, constitui software livre e de código aberto, podendo vir a ser adaptada por qualquer um com conhecimentos de programação em GF para refletir outras convenções gramaticais e ortográficas.

Dois critérios orientaram a delimitação do fragmento. Por um lado, tomamos como ponto de partida a minigramática bilíngue do inglês e do italiano com que Ranta (2011) introduz as noções fundamentais do GF, formalismo utilizado na especificação computacional objeto da seção 3. Essa minigramática (doravante ITALENG) é capaz de analisar e traduzir comentários sobre alimentos nessas duas línguas análogos aos exemplos (1)-(4). Por outro lado, seguimos, em linhas gerais, a progressão gramatical de Navarro (2011), de que destacamos, no Quadro 2, apenas os pontos que implementamos. Os títulos dos tópicos foram adaptados à terminologia adotada neste artigo. Entre parênteses, indicamos exemplos de cada fenômeno. A terceira coluna indica o percentual aproximado implementado de cada lição.

<sup>10</sup> Agradecemos a um dos pareceristas por essa observação.

QUADRO 2 – Tópicos gramaticais das lições de Navarro (2011) implementados

Lição	Conteúdo gramatical	Cobertura
1	I: flexão de verbos ativos (11)-(13) II: pronomes pessoais livres e prefixos inativos (7)-(13) III: modificadores nominais adjetivais (4) e predicação com verbos inativos (1) (5) (8) IV: substantivo como genitivo adnominal (14)-(16)	100%
2	I: posposições (12) (14)-(18) II: negação sentencial (12)-(14) IV: marcação de plural no domínio nominal (4)	60%
3	I: prefixos inativos na função de genitivo adnominal (5)-(7) II: prefixos inativos como complementos de posposições (14) (18) VI: determinantes demonstrativos (2)-(4)	40%
4	II – substantivos multiformes (5) (12) (15)	20%
5	V – verbos inativos multiformes (1) (6) (9)	10%
6	IV – posposições multiformes (14) (15)	20%
9	III – posposição <i>pupé</i> e locução posposicional <i>kuara upé</i> (17) (16)	10%

Fonte: Elaboração própria.

Coincidentemente, abstraindo das diferenças sintáticas superficiais entre inglês, italiano e nheengatu, as duas introduções, Ranta (2011) e Navarro (2011), esta à língua objeto, aquela à metalinguagem, focam inicialmente o mesmo tipo fundamental de construção linguística: a predicação qualificativa (*qualifying predication*) do tipo copulativo, que, conforme Mathesius (1975, p. 114), consiste na atribuição de uma qualidade (*qualificans*) a uma entidade (*qualificandum*), expressas, respectivamente, pelo predicado e pelo sujeito, ligados por um verbo do tipo de *ser* (inexistente em nheengatu) ou *estar*.

Preferindo, como Casasnovas (2006), uma nomenclatura mais próxima da portuguesa, Navarro (2011) diverge de Cruz (2011) no tratamento dos pronomes pessoais, verbos inativos e prefixos do Quadro 1, exigidos por parte desses verbos. Aos verbos inativos não flexionáveis denomina adjetivos de primeira classe e classifica os inativos flexionáveis em dois grupos, a saber, adjetivos e verbos de segunda classe, segundo correspondam em português a um adjetivo, como em (1), (2) e (6)-(10), ou a um verbo, como no caso de *kérpi* ‘sonhar’. Trata os prefixos flexionais

desses dois grupos, bem como na função de complemento de posposições, como pronomes pessoais de segunda classe, mas os denomina pronomes adjetivos possessivos quando adjungidos a substantivos. Os pronomes pessoais livres, na acepção de Cruz (2011), como *ixé* ‘eu’ em (8), são para Navarro (2011) pronomes de primeira classe.

Analogamente a *ser* e *estar* em português, mas diferentemente de *be* em inglês, o *nheengatu* lexicaliza, em orações predicativas do tipo de (1)-(18), a distinção entre predicado de nível de indivíduo (*individual level predicate*) e predicado de nível de fase (*stage level predicate*),<sup>11</sup> ou, conforme Cruz (2011, p. 475), entre estados não-contingentes e contingentes. Estes são marcados pelo auxiliar *iku* ‘estar’, enquanto aqueles não são marcados, compare-se (7) com (8) e (4) com (3).

Utilizando a nomenclatura de Navarro (2011), definimos em (19) um fragmento de gramática do *nheengatu* capaz de gerar sentenças como (1)-(10), que transcendem o recorte da ITALENG, uma vez que incluem pronome de primeira classe (Pron1) como *qualificandum* na função de sujeito da sentença e pronome possessivo adnominal (Poss). Pron2 refere-se aos pronomes de segunda classe.

- (19) S→(NP) VP  
 VP→AP (V)  
 NP→(Det) (Poss) (AP) N (AP)  
 NP→Pron1  
 AP→(Pron2) A (Adv)  
 V→“aiku”|“uiku”  
 N→“igara”|“kambi”|“timbiú”|“sendaua”|“suka”|“meiú-itá”|“pirá”|“pirá-itá”  
 A→“pusé”|“saku”|“raku”|“ruri”|“suri”|“sepiasu”|“sepiasuíma”|“piranga”  
 Det→“kuá”|“nhaã”|“kuá-itá”  
 Poss→“i”|“se”|“ta”  
 Pron1→“ixé”|“aé”  
 Pron2→“i”|“se”|“ta”  
 Adv→“retana”

<sup>11</sup> Segundo Bentley (2017, p. 343), essa é apenas uma entre outras explicações para a alternância entre as cópulas ESSE e STARE em construções predicativas copulativas das línguas românicas, questão cujo aprofundamento ultrapassaria o escopo deste artigo.

Nos formalismos gramaticais derivados da CFG, como a LFG, a estrutura gramatical de uma língua natural é modelada em duas dimensões distintas (SAG; WASOW; BENDER, 2003). O nível sintagmático modela a combinação de categorias lexicais e funcionais para formar sintagmas, enquanto estruturas de traços implementam restrições relacionadas à concordância e à valência, bloqueando combinações que as violam. Como veremos na próxima seção, o GF faz distinção análoga.

O fragmento (19) limita-se à primeira dimensão. Desse modo, hipergera, produzindo muitas sentenças agramaticais como (20)-(23). Em (20), a flexão do verbo auxiliar conflita com a do principal, ver correções em (25) e (26). O exemplo (21) também admite duas correções alternativas: ou adjungimos um Pron2 a *pusé* para obter (26) ou tratamos esse constituinte como modificador do NP nucleado por *igara*, satisfazendo as exigências valenciais de *iku* por meio de um outro núcleo A, ver (27). O exemplo (22), corrigido em (28), é agramatical porque o substantivo *timbiú* e o verbo inativo *raku* são multiformes,<sup>12</sup> incompatíveis com o prefixo *i*, exigindo, em vez disso, o alomorfe assilábico *s-*, ver (1), realizado como *x* antes de *i*.<sup>13</sup> Finalmente, (23) e (24) não constituem sentenças por carecerem de um sujeito (CRUZ, 2011, p. 142), comparem-se (12) e (29).

- (20) \*se            pusé    u-iku  
      1s.INACT   pesado   3s.ACT-estar
- (21) \*igara       pusé    u-iku  
      canoa       pesado   3s.ACT-estar
- (22) \*i            timbiú    i            raku  
      3s.INACT   comida   3s.INACT   quente
- (23) \*t-tendaua                    upé  
      IMPR-comunidade       em

<sup>12</sup> A esse respeito, seguimos Cruz (2011, p. 137-138). Navarro (2011) não considera *timbiú* multiforme.

<sup>13</sup> Para Cruz (2011, p. 137), os alomorfes *s-* e *x-* estão em distribuição complementar. Ao contrário, Navarro (2011) apresenta *ximiriku* ‘esposa dele’ como variante de *simiriku*, forma que não ocorre em nenhum exemplo de Cruz (2011).

- (24) \*puranga  
bonito
- (25) se            pusé      a-iku  
1s.INACT pesado 1s.ACT-estar  
'estou pesado'
- (26) igara i            pusé      u-iku  
canoa 1s.INACT pesado 3s.ACT-estar  
'a canoa está pesada'
- (27) igara      pusé      sepiasuíma    u-iku  
canoa      pesado barato      3s.ACT-estar  
'a canoa está barata'
- (28) x-imbiú<sup>14</sup>            s-aku  
3s.INACT-comida 3s.INACT-quente  
'a comida dele é quente'
- (29) aé puranga  
ela bonito  
'ela é bonita'

Seguindo a progressão do Quadro 2, expandimos o fragmento de (19) nas duas dimensões referidas. Na primeira, modificamos as regras sintagmáticas de modo a também gerar exemplos do tipo de (12)-(18), que envolvem os tópicos 1.IV, 2.I, 2.II e 9.III, e exemplos análogos do tipo de (11), em que o complemento locativo é expresso por um advérbio.

Desses quatro tópicos, o primeiro envolve um grau considerável de complexidade, mas os demais são de implementação bastante simples, abstraindo da questão das posposições multiformes, que abordaremos no final desta seção. A negação sentencial é feita em nheengatu por meio do clítico *ti*, que se adjunge à primeira posição do rema (CRUZ, 2011,

<sup>14</sup> Ou *s-imbiú*, se considerarmos livre a variação entre <s> e <x>, como sugere Navarro (2011), ver (175).

p. 406), ocupada por um verbo ou por um sintagma posposicional em (13)-(16).<sup>15</sup>

Tanto prefixos inativos quanto NPs plenos podem realizar o argumento interno de posposições, ver (12)-(18). Para gerar esse tipo de sintagma, incluímos (30) e (31) na gramática de (19), onde P designa tanto uma posposição quanto uma locução posposicional, compare (16) e (17). Essas regras produzem também estruturas agramaticais, uma vez que não modelam a checagem de caso e a alomorfia entre *i* e *s-* na realização de 3s.INACT, que abordaremos mais adiante.

(30) PP → NP P

(31) NP → Pron2

A construção possessiva expressa uma relação de posse *lato sensu* entre dois participantes, um possuidor (PSOR, do inglês *Possessor*) e uma entidade possuída (PSUM, do inglês *Possessum*). Essa relação abstrata envolve uma série de relações mais específicas como posse (*ownership*), parte-todo, parentesco etc. (KARVOVSKAYA, 2018).

Analogamente a outras línguas ameríndias, o nheengatu distingue entre nomes relativos e nomes autônomos (CRUZ, 2011). Enquanto estes são intrinsecamente monovalentes, licenciando, em caráter facultativo, um complemento genitivo, aqueles são intrinsecamente divalentes, sendo obrigatória a realização do argumento interno, mesmo quando recuperável contextualmente, como em (32), extraído de Cruz (2011, p. 155).<sup>16</sup>

(32) u-riku        i                paia  
       3s.ACT-ter 3s.INACT pai  
       ‘Ele tem pai.’

<sup>15</sup> Navarro (2011, p. 18) consigna também a forma *niti*, que, numa contagem por meio da ferramenta *grep*, ocorre 98 vezes no texto, contra 28 da forma clítica *ti*. Um dos pareceristas observa que *niti* ocorre em documentos do século XIX, mas não nos dados de Cruz (2011). No nheengatu do rio Negro, conforme Casasnovas (2006, p. 49), a negação do verbo é feita apenas por meio de *ti* (*te* no imperativo). No entanto, implementamos também a forma *niti* para possibilitar a análise de textos mais antigos.

<sup>16</sup> Em todos os exemplos extraídos da literatura, adaptamos a ortografia e as glosas às utilizadas neste artigo. As traduções são dos autores citados, salvo indicação contrária.

A valência dos nomes em *nheengatu* constitui um ponto que mereceria um aprofundamento, diante de contradições internas das abordagens de Cruz (2011) e Navarro (2011). Embora aquela classifique todos os nomes multiformes como relativos, apresenta diversos exemplos de *uka* ‘casa’ e *timbiú* ‘comida’ na forma absoluta, sem complemento genitivo, ver exemplos análogos em (16) e (3). O segundo autor, por sua vez, trata os nomes de parentesco e as partes do corpo como necessariamente possuíveis, portanto, incompatíveis com a forma absoluta. No entanto, apresenta, de diversos desses nomes que são multiformes, uma forma absoluta, como *tamunha* ‘avô’.

Uma comparação superficial entre os NPs na posição de sujeito em (14) e (15) com os exemplos equivalentes em (33) e (34) sugere que a construção possessiva do *nheengatu* organiza-se da mesma forma que a construção com genitivo *saxão* do inglês, ressalvada a inexistência, naquela língua, de marca morfológica correspondente:

(33) the boy's community

(34) Mary's house

Há, porém, diferenças fundamentais na realização dos participantes nos dois tipos de construção. Para mostrar isso, recorreremos à hipótese DP, proposta em sintaxe gerativa no quadro da teoria X-barra. Na construção do inglês, o PSOR é um DP completo, gerado na posição de especificador do DP nucleado pelo determinante *'s*, enquanto o PSUM realiza-se como NP complemento desse núcleo (CARNIE, 2002, p. 145-146), conforme (35). Desse modo, o PSUM não admite determinação por meio de um demonstrativo como no exemplo do *nheengatu* em (17), pelo que os exemplos análogos (36) e (37) são agramaticais. Na construção genitiva *normanda*, porém, não há essa restrição, ver (38).<sup>17</sup>

(35) [DP [DP the boy] [D' [D 's] [NP community]]]

(36) \*Peter's that hook

(37) \*that the boy's hook

(38) that hook of the boy

<sup>17</sup> Sobre os dois genitivos do inglês, ver Comrie (1983, p. 85).

Inversamente ao que se observa no genitivo saxão, na construção genitiva do nheengatu o PSUM é um DP completo, ao passo que o PSOR corresponde a um NP, pelo que (40) não constitui tradução de (39), mas de (38).

(39) that boy's hook

(40) nhaã            kurumĩ pindá  
DEM.DIST menino anzol  
'aquele anzol do menino'

Antes de detalhar essa análise, formalizamos, por meio do fragmento de gramática (41)-(48), o esquema que Cruz (2011, p. 282), no quadro da abordagem tradicional do sintagma nominal, propõe para esse constituinte em nheengatu. Três tipos de constituintes podem anteceder um núcleo nominal NOME na ordem especificada: (i) quantificadores (Quant), (ii) determinantes (Det), incluindo demonstrativos (Dem), indefinidos (Indf) e numerais (Num), e (iii) nomes (N) ou prefixos inativos (Pron2) na função de complemento nominal (CN), ver (7) e (14). NOME, por sua vez, segundo Cruz (2011, p. 282), é um “nome dêítico”, i.e., um Pron1, ou um “nome substantivo”, ou seja, um N.

(41) NP → (Quant) (Det) (CN) NOME

(42) Quant → “panhẽ”...

(43) Det → Dem | Indf | Num

(44) Dem → “nhaã”...

(45) Num → ...

(46) CN → N | Pron2

(47) N → “igara”|“kambi”...

(48) NOME → N | Pron1

Complementada com regras de inserção de Pron1 e Pron2, essa minigramática gera qualquer dos sintagmas nominais dos exemplos anteriores, assim como estruturas mais complexas como (49), em que se realizam todas as categorias facultativas de (41). No entanto, gera também estruturas agramaticais como (50) e (51), em que um NOME realizado como pronome pessoal rege um CN.<sup>18</sup>

- (49) [<sub>Quant</sub> panhẽ] [<sub>Det</sub> nhaã] [<sub>CN</sub> kurumĩ] [<sub>NOME</sub> igara-itá]  
 todo DEM.DIST menino canoa-PL  
 ‘todas aquelas canoas do menino’

- (50) \* [<sub>CN</sub> se] [<sub>NOME</sub> indé]  
 1s.INACT tu

- (51) \* [<sub>CN</sub> kurumĩ] [<sub>NOME</sub> penhẽ]  
 menino vós

Por outro lado, (41)-(48) não geram os exemplos (52)-(54) de Cruz (2011, p. 258, 184, 176), que contrariam a sua própria abordagem, como evidenciamos por meio da análise estrutural dos dois primeiros. Em ambos os exemplos, o CN de *nheenga* constitui-se de um nome que, por sua vez, possui um prefixo inativo como CN, ao passo que, no segundo, o próprio núcleo *nheenga* recebe um desses prefixos, configurações não contempladas pelo esquema estrutural formulado por Cruz (2011). Retomaremos (54) mais adiante.

- (52) [<sub>CN</sub> [<sub>CN</sub> ne] [<sub>NOME</sub> mena]] [<sub>NOME</sub> dheenga]]  
 2s.INACT marido língua  
 ‘língua do teu marido’

- (53) [<sub>CN</sub> [<sub>CN</sub> ne] [<sub>NOME</sub> kiuíra]] [<sub>CN</sub> ta [<sub>NOME</sub> dheenga]]  
 2s.INACT irmão 3p.INACT fala  
 ‘o conselho dos teus irmãos’<sup>19</sup>

<sup>18</sup> Karvovskaya (2018, p. 3) chama atenção para o fato de que o PSUM “geralmente não é um elemento pronominal: \**John’s she*.” Esse tipo de construção somente é possível com substantivos homônimos de pronomes pessoais: *O eu do João*.

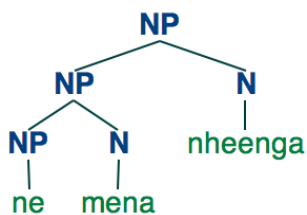
<sup>19</sup> Uma tradução mais próxima da estrutura original seria talvez ‘o conselho deles de irmão teu’.

- (54) aitenhaã paa sukuriu pedasu-itá kuera  
 DEM.DIST QUOT sucuri pedaço-PL resquício  
 ‘Diz que aquilo eram pedaços de sucuri.’

No quadro do debate sobre recursividade nas línguas ameríndias, Leandro e Amaral (2014) defendem a existência de genitivos recursivos em wapichana, da família arawak. O exemplo (54) com dois genitivos nominais, extraído de Cruz (2011), sugere que esse fenômeno também ocorre em nheengatu. Desse modo, reformulamos (41)-(48), com base na hipótese DP, inicialmente como (55)-(59). Conforme (56), o núcleo D pode não se realizar foneticamente, pelo que o DP assume, via de regra, uma interpretação definida.<sup>20</sup>

- (55)  $DP \rightarrow (QP) D'$   
 (56)  $D' \rightarrow (D) NP$   
 (57)  $NP \rightarrow (NP) N$   
 (58)  $N \rightarrow \text{“igara”} | \text{“kambi”} | \text{“nheenga”} | \text{“mena”} | \dots$   
 (59)  $NP \rightarrow \text{“se”} | \text{“ne”} | \text{“i”} | \dots$

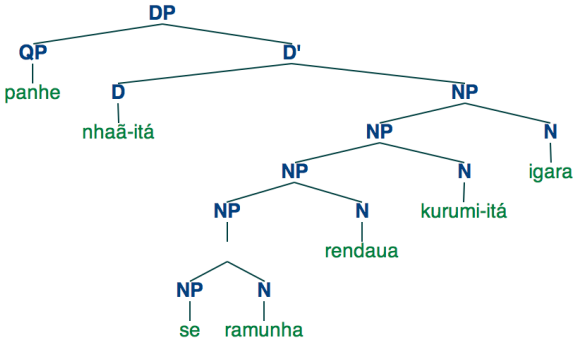
FIGURA 3 – Representação arbórea do NP de (52) conforme (55)-(59)



Fonte: Elaboração própria.

<sup>20</sup> Cruz (2011, p. 265) mostra que, em narrativas, DPs com D zero assumem interpretação indefinida quando não se referem ao tópico discursivo, marcado, apenas na primeira menção, pelo D indefinido *yepé* ‘um’.

FIGURA 4 – Representação arbórea do NP de (60) conforme (55)-(59)<sup>21</sup>



Fonte: Elaboração própria.

Além de (52), essa gramática gera, entre infinitos outros, exemplos como (60), produzindo representações recursivas como as da Figura 3 e da Figura 4, onde NPs funcionam como complementos nominais de outros NPs.

- (60) panhẽ nhaã-itá se r-amunha r-endaaua kurumi-itá igara  
todo DEM.DIST-PL 1s.INACT RLL-avô RLL-comunidade menino-PL canoa  
‘todas aquelas canoas dos meninos da comunidade do meu avô’

A gramática de (55)-(59), contudo, não gera (53), pois analisa o PSUM *ta nheenga* como NP, não admitindo, portanto, o complemento *ne kiuíra*. Substituímos, então, (58) e (59) por (61)-(63). Nessa reformulação, nomes com prefixos inativos são formados no léxico, em consonância com o Princípio da Integridade Lexical (BRESNAN, 2001). Conforme (61), um N é formado pela concatenação opcional de um prefixo inativo a um radical nominal. Esse prefixo satura o lugar vazio de complemento dos nomes relativos. Nomes em que o prefixo de 3ª pessoa não é *i*, mas *s-*, compare (5)-(7), parecem corroborar essa abordagem.

- (61)  $N \rightarrow (\text{Prf}) N_{\text{Stem}}$   
(62)  $\text{Prf} \rightarrow \text{“se”} | \text{“ne”} | \text{“i”} | \text{“s”} \dots$   
(63)  $N_{\text{Stem}} \rightarrow \text{“nheenga”} | \text{“kiuíra”} | \text{“mena”} \dots$

<sup>21</sup> O til sobre *e* e *i* não é suportado pelo programa que utilizamos para gerar esse gráfico.

Uma série de questões, contudo, resta investigar por meio de *corpora* mais extensos ou de experimento nos moldes do proposto por Leandro e Amaral (2014). Primeiro, se a construção de (54) é produtiva em *nheengatu*, dada a gramaticalização de *kuera* (CRUZ, 2011). Segundo, se um número maior de encaixes é licenciado. Finalmente, se um outro tipo de construção é utilizado para expressar relações de posse mais complexas, por exemplo, orações relativas ou sintagmas posposicionais.<sup>22</sup>

Independentemente dessas questões, a Figura 4 permite visualizar as características fundamentais da construção possessiva genitiva do *nheengatu*: (i) quantificadores e determinantes têm escopo sobre o NP mais alto, que representa o PSUM; (ii) o PSOR, realizado pelo NP complemento (genitivo) do N núcleo do PSUM, não é passível de quantificação nem determinação; (iii) admite, porém, adjunção da partícula de plural *itá*, como evidencia o exemplo (64) de Cruz (2011, p. 209).

- (64) aitenhaã      ia-seruka      kariua-itá      nheenga rupi      Martim Pescador  
DEM.DIST 1p.ACT-chamar não.indígena-PL palavra      PERL Martim Pescador  
'Aquele lá, chamamos pela língua dos brancos de Martim Pescador.'

As propriedades (i) e (ii) são reconstruções da análise de Cruz (2011), dirimindo uma aparente incoerência dessa abordagem. De fato, por um lado, analisa o complemento nominal genitivo em exemplos do tipo de (49) como N. Por outro, afirma que a partícula de plural incide sobre o NP com núcleo contável (CRUZ, 2011, p. 164). Logo, se um N funcionando como complemento nominal pode estar sob o escopo dessa partícula, deve constituir um NP, exatamente como propomos. A análise do PSOR como NP prevê, também, que possa ser modificado por um AP, o que de fato se verifica em (65).

- (65) piranga      tuí      kue(ra)      nhaã      i      iuru=pe (CRUZ, 2011, p. 139)  
vermelho sangue resquício DEM.DIST 3s.INACT boca=LOC  
'O resquício de sangue vermelho ficou na boca dele.'

<sup>22</sup> Na variedade descrita por Sympson (1877, p. 66), ocorre, ao lado do genitivo pré-nominal do tipo de (14), construção com possuidor pós-nominal introduzido por posposição: *kisaua Mandu ressé* 'rede de Manuel'.

Seguindo Navarro (2011, p. 12), para quem “adjetivos qualificativos”, i.e., verbos inativos na classificação de Cruz (2011), podem anteceder ou subseguir o núcleo nominal, formalizamos a modificação adjetival do NP por meio de (66), reformulação da terceira regra de (19).

(66) NP→(AP) N (AP)

Ao contrário, Cruz (2011, p. 194) nega não só a existência de adjetivos em *nheengatu*,<sup>23</sup> como também a possibilidade de verbos inativos, sem a intermediação do relativizador *waa*, funcionarem como modificadores nominais, contrariamente ao que (65) evidencia e Navarro (2011) preconiza, no que é corroborado por exemplos como (67) de Casasnovas (2006, p. 53) (adaptamos a ortografia e incluímos as glosas, mas mantivemos a tradução original).

(67) A-putari            apukuitaua miri      mirá tauá      suiudara.  
1s.ACT-querer remo            pequeno pau    amarelo de  
‘Quero um remo pequeno feito de pau amarelo.’

Voltando à comparação entre a construção possessiva com *ʔ* do inglês e a construção genitiva do *nheengatu*. Na primeira, conforme a análise no âmbito da hipótese DP, tanto o PSOR quanto o PSUM são DPs, sendo que esse último tem *ʔ* como núcleo D. Na segunda língua, o PSUM é um DP que admite como núcleo qualquer membro da categoria funcional D (respeitadas restrições semânticas), ao passo que o PSOR é um NP. Há uma diferença fundamental entre DPs e NPs: apenas os primeiros são referenciais (BERNSTEIN, 2003). Isso se reflete no contraste entre os modificadores genitivos de (68) e (69). No primeiro exemplo, a categoria PP domina um DP que introduz um participante da situação referida pelo nome modificado, ao passo que, no segundo, domina um NP, não introduzindo participante, mas funcionando apenas como classificador do nome modificado, à semelhança de um adjetivo denotativo de qualidade (FÁBREGAS, 2017).

<sup>23</sup> Como salienta um dos pareceristas anônimos, Cruz (2011, p. 130) classifica palavras como *puku* ‘comprido’ como verbos estativos, e não adjetivos, porque, analogamente aos verbos dinâmicos, recebem um sufixo nominalizador para funcionarem como argumentos.

(68) este retrato da criança

(69) este retrato de criança

Dada a interpretação do complemento genitivo em (49) como uma expressão referencial, não haveria um núcleo D foneticamente vazio dominando esse NP? Ao nosso ver, essa interpretação não é determinada na sintaxe, mas no nível discursivo. De fato, como evidenciam exemplos como (70) de Cruz (2011, p. 260), em consonância com exemplos análogos de Navarro (2011), noutros contextos o genitivo não introduz participante, funcionando apenas como classificador. Dada a sua natureza de NP, o genitivo não é especificado quanto à determinação, pelo que (71) traduz-se como (68) ou (69), dependendo da situação comunicativa concreta.

(70) mirá rakanga  
 árvore galho  
 ‘galho de árvore’

(71) kuá taína r-angaua  
 DEM.PROX criança RLL-retrato

Na segunda dimensão da formalização gramatical, modelamos as relações de compatibilidade sintagmática entre os diferentes itens lexicais, de modo a excluir exemplos agramaticais tais como exemplificados em (20)-(22). Para tanto, implementamos restrições em diferentes domínios: (i) concordância verbo-nominal, (ii) marcação de plural no DP, (ii) checagem de caso e (iii) seleção dos alomorfes de substantivos, verbos inativos e posposições multiformes.

Vejam os detalhes de cada um desses fenômenos no âmbito do recorte definido. Em nheengatu, os verbos subdividem-se em dois grupos principais: flexionáveis e não-flexionáveis. Os primeiros classificam-se em ativos e inativos, recebendo prefixos das séries ativa ou inativa, respectivamente. Esses prefixos expressam a concordância número-pessoal com o sujeito, tanto sob a forma de um DP pleno ou pronome livre quanto de um pronome nulo. Verbos não-flexionáveis do mesmo modo que construções sem verbo não admitem sujeito nulo, licenciado pela flexão verbal, ver (24) e (23).<sup>24</sup>

<sup>24</sup> Utilizamos a noção de pronome nulo, simbolizado como *pro*, e sujeito nulo da gramática gerativa (CARNIE, 2002, p. 273).

Na construção de verbo inativo flexionável com o auxiliar *iku*, ambos se flexionam, marcando a concordância com o sujeito. Na 3ª pessoa da série ativa, porém, não há, propriamente, concordância de número na variedade descrita por Navarro (2011): o prefixo é *u-* tanto no singular quanto no plural, como em Casasnovas (2006, p. 27-28). Exemplos de Casasnovas (2006, p. 34):

- (72) kuá                      apigá    u-puraki                      puranga

DEM.PROX homem 3s.ACT-trabalhar bom

‘este homem trabalha bem’

- (73) musapíri kumurĩ-itá    u-musarai                      u-iku

três                      menino-PL 3p.ACT-brincar 3p.ACT-estar

‘três meninos estão brincando’

Essa era a situação no século XIX, conforme Cruz (2015), que mostra que a concordância de número na 3ª pessoa da série ativa constitui um fenômeno mais recente. Para Cruz (2011, p. 133), ao prefixo *u-* no singular opõe-se *ta(u)-* ~ *tu-* no plural. Segundo Cruz (2015, p. 430-31), no nheengatu do século XXI, a combinação do clítico *ta=* (pronomes de 3ª pessoal do plural) e do prefixo *u-* (3ª pessoa da série ativa) é característica da fala dos mais velhos, cedendo lugar, na fala dos mais jovens, a *tu-* no dialeto do Xié e a *ta-* no dialeto do Negro e do Içana. Tanto Navarro (2011, p. 89) quanto Casasnovas (2006, p. 28) contemplam a marcação da 3ª pessoa do plural por meio de *ta=u-*.

O plural do DP é marcado formalmente por meio da partícula *itá*, sufixada ao núcleo nominal ou ao núcleo D, compare (2) e (4). Embora Navarro (2011, p. 25-26), na exposição dos demonstrativos (tópico L3.VI do Quadro 2), se limite ao segundo caso, ele abona o primeiro com o exemplo (74) (NAVARRO, 2011, p. 48).

- (74) Re-rasu                      kuá                      maniaka-itá    memũitendaua kiti.

2s.ACT-levar DEM.PROX mandioca-PL cozinha                      para

‘Leve estas mandiocas para a cozinha.’

Para Navarro (2011, p. 19), somente se usa a partícula *-itá* quando absolutamente indispensável, como em (75), onde, devido à ambiguidade da flexão verbal, a sua omissão resultaria na leitura de (76). Segundo ele, a partícula é omitida quando outro elemento indica pluralidade, como em

*mukûi apigaua* ‘dois homens’. Na língua falada no alto rio Negro, ocorre, com quantificadores discretos, tanto a marcação redundante quanto a não redundante, como nos exemplos (77) e (78) extraídos de Cruz (2011, p. 269-270). Casasnovas (2006, p. 34-35) apresenta exemplos análogos.

(75) *kunhã-itá puranga u-iku*  
mulher-PL bonito 3p.ACT-estar  
‘as mulheres estão bonitas’

(76) *kunhã puranga u-iku*  
mulher-SG bonito 3s.ACT-estar  
‘a mulher está bonita’

(77) *ia-uasému mukûi pessoa-itá*  
1p.ACT-encontrar dois pessoa-PL  
‘Encontramos duas pessoas.’

(78) *aikué musapíri pessoa u-iku uaá ape*  
*há dois pessoa 3s.ACT-estar<sup>25</sup> REL lá*  
‘Havia três pessoas, que estavam lá.’

Não há, portanto, concordância de número entre os constituintes do DP em *nheengatu*: um quantificador plural admite tanto um NP com *-itá* quanto um NP sem essa partícula, como evidenciam, respectivamente, os exemplos (77) e (78) de Cruz (2011). A GrammaYEP ainda não contempla DPs com quantificadores nem numerais como nesses dois exemplos, apenas com núcleo D demonstrativo. Nesse caso, estipulamos que um DP é plural se o núcleo D (demonstrativo) ou o núcleo N é marcado com *itá*. Admitindo que a marcação do plural é um fenômeno lexical, núcleos D demonstrativos e substantivos constituem paradigmas que variam conforme o número, a forma de plural diferindo da de singular por meio do sufixo *itá*. Na sintaxe, um demonstrativo no plural seleciona um substantivo no singular, ao passo que um demonstrativo no singular não impõe restrição sobre o número do NP complemento.

<sup>25</sup> Cruz (2011) analisa essa forma como 3ª pessoa do singular, pois, para ela, como vimos acima, a flexão de 3ª do plural é *ta(u)-*.

Essa análise da partícula *itá* é uma reconstrução das abordagens de Cruz (2011) e Navarro (2011), falseável por dados de corpora ou julgamentos de aceitabilidade corroborando exemplos do tipo de (79), que, por ora, consideramos agramaticais. De fato, não encontramos, nesses dois trabalhos, DP com iteração da partícula *-itá* como em (79). Cruz (2011, p. 377) parece negar a possibilidade desse tipo de construção ao caracterizar *-itá* como “partícula independente” com escopo sobre o sintagma nominal (DP na nossa abordagem) que pode adjungir-se tanto ao núcleo D ou ao núcleo N.

(79) \*kuá-itá kunhã-itá

Aplicando a análise do DP e do NP formalizada em (55)-(57) e (61) a exemplos como (5)-(12) e (80)-(84), as seguintes generalizações emergem sobre a distribuição de pronomes pessoais livres (Pron1) e prefixos inativos (Pron2): os primeiros ocupam a posição de sujeito de verbos ativos e inativos e complemento (objeto direto) de verbos ativos, enquanto os segundos funcionam como marcas de concordância de verbos inativos flexionáveis e complementos (i.e., argumentos internos) de nomes e posposições.

(80) aé i pusé  
ele 1s.INACT pesado  
‘ela é pesada’

(81) (\*k unhã) s-uka taua upé  
mulher 3s.INACT-casa cidade em

(82) (\*k unhã) i igara paranã upé  
mulher 3s.INACT canoa rio em

(83) ixé a-iku (\*k unhã) s-uaki

(84) \*a-iku aé irũmu  
1s.ACT-estar ele.ACC com

(85) aé s-aku  
ele 1s.INACT quente  
‘ela é quente’

- (86) ixé            a-maã            aé  
 eu.NOM 1s.ACT-ver ela.ACC  
 ‘eu a vejo’

O Quadro 3 evidencia que, em posições argumentais de V ou P, os itens das colunas 2 e 3 estão em distribuição complementar. Os elementos da coluna 4, por sua vez, como vimos em (61), saturam, no nível sublexical, a posição de complemento de  $N_{Stem}$ . Estendendo essa análise aos elementos da coluna 3, reformulamos (30) e (31) como (87). Tanto no caso de  $N_{Stem}$  quanto de  $P_{Stem}$ , a saturação da posição de complemento no nível sublexical impede que um NP ou DP pleno venha a ocupar essa mesma posição na sintaxe, ver (81) e (82).<sup>26</sup> As marcas de concordância de sujeito, ao contrário, coocorrem com DP pleno ou Pron1, ver, por exemplo, (8), (80) e (85). Desse modo, em relação aos itens das colunas 3 e 4, as flexões verbais inativas constituem um paradigma à parte de formas homônimas.

- (87)  $P \rightarrow (Prf) P_{Stem}$   
 $PP \rightarrow (DP) P$

QUADRO 3 – Distribuição de Pron1 e Pron2 no singular em posições argumentais

Pessoa	DP argumento externo ou interno de V ou externo de P (Pron1)	Argumento interno de $P_{Stem}$ (Pron2)	Argumento interno de $N_{Stem}$ (Pron2)
1s	ixé	se	se
2s	indé	ne	ne
3s	aé	i s-	i s-

Fonte: Elaboração própria.

Para modelar a distribuição dos elementos do Quadro 3, recorreremos à noção de caso: Pron1 satura posições argumentais associadas aos casos nominativo e acusativo, ao passo que Pron2 realiza o complemento genitivo de nomes e posposições, compare (7), (18) e

<sup>26</sup> Em (53), o primeiro NP não ocupa a mesma posição que o prefixo do segundo, dada a discrepância de número entre esses constituintes.

(86). Pron1 manifesta, portanto, sincretismo sistemático entre nominativo e acusativo, fenômeno bastante comum nas línguas do mundo (ZOMPI, 2017). DPs e NPs plenos não exibem caso morfológico.

Tanto os nomes relativos quanto os autônomos admitem um complemento genitivo, assim como as posposições de modo geral. No nosso fragmento, limitamo-nos às posposições locativas *upé* ‘em’, *pupé* ‘dentro de’ e *ruaki* (*suaki*) ‘perto de’, à comitativa *irūmu* e à locução pospositiva *kuara upé* ‘dentro de’. Segundo Cruz (2011, p. 199), a primeira é incompatível com prefixos inativos, analogamente à posposição *arama* ou (*a*)*rã* do dativo intralocutivo, que exige como complemento um pronome livre de 1ª ou 2ª pessoa, ver (88). A posposição *supé* do dativo extralocutivo, porém, formalmente relacionada a *upé*, admite prefixo inativo de 3ª pessoa (CRUZ, 2011, p. 201).<sup>27</sup> Ao contrário, para Navarro (2011), *arama* é a única posposição que não admite pronome pessoal de segunda classe. Apesar de Cruz e Navarro não apresentarem qualquer exemplo de *upé* regendo pronome (seja de primeira, seja de segunda classe), seguimos a exposição desse último, não tratando essa posposição como excepcional no que tange à regência de caso. Desse modo, consideramos (89) como gramatical e (90) como agramatical.

(88) re-rúri                      t-imbiú                      ixé                      arama (NAVARRO, 2011, p. 23)  
2s.ACT.IMP-trazer IMPR-comida eu.ACC para  
‘traz-me comida’

(89) tukandira-itá u-iku                      i                      upé  
tocandira-PL 3s.ACT-estar 3s.INACT em  
‘as tocandiras estavam nele’

(90) \*tukandira-itá uiku                      aé upé  
tocandira-PL 3s.ACT-estar ele em

<sup>27</sup> A abordagem de Cruz (2011) parece contraditória nesse aspecto. Por um lado, apresenta exemplos de *supé* regendo prefixo inativo de 3ª pessoa, ver (304) na p. 201 e (363) na p. 218. Por outro, em mais de uma passagem, explicitamente nega essa possibilidade, por exemplo quando afirma na p. 197 que as posposições locativas do grupo *pé*, entre as quais classifica *supé*, são incompatíveis com prefixos inativos, impossibilidade que confirma na p. 200 e no Quadro 28 na p. 223.

Do ponto de vista da implementação computacional, um dos aspectos mais complexos do nheengatu são as classes de palavras multiformes, que constituem subgrupos de substantivos, verbos inativos e posposições. Distinguem-se pela incompatibilidade com o prefixo silábico *i* do Quadro 1. Em vez disso, selecionam o alomorfe assilábico *s-* (*t-* no caso de uns poucos substantivos).

Os substantivos multiformes apresentam, no caso canônico, representado por *tendaaua* em (5), (12) e (15), três formas, que se distinguem, conforme Navarro (2011), pelos prefixos de relação *t-*, *r-* e *s-*, representados nas glosas por IMPR, RLL e INACT, respectivamente.<sup>28</sup> A terceira forma resulta da concatenação com *s-*. A segunda forma ocorre com todos os outros tipos de PSOR, isto é, NP pleno ou prefixo inativo de 3ª pessoa do plural e de 1ª e 2ª pessoas, ver (91). A primeira forma, chamada absoluta, que constitui a forma de citação do substantivo, assinala que o substantivo não realiza PSUM de construção possessiva.

- (91) se            r-uka      nhaã      kiá s-apé      upé  
 1s.INACT RLL-casa DEM.DIST sujo 1s.INACT-rua em  
 ‘a minha casa é naquela suja rua dele’

O Quadro 4 sistematiza os diferentes tipos de substantivos multiformes descritos por Navarro (2011). Os microgrupos (ii)-(v) resultam de desvios do padrão canônico representado pelo microrupo (i), decorrentes de alomorfias do primeiro ou terceiro prefixos.

QUADRO 4 – Microgrupos de substantivos multiformes

Grupo	Forma absoluta	PSOR ≠ 3s.INACT	PSOR = 3s.INACT
	<i>tendaaua</i> ‘comunidade’	<i>rendaaua</i>	<i>sendaaua</i>
	<i>sangaua</i> ‘retrato’	<i>rangaua</i>	<i>sangaua</i>
	<i>uka</i> ‘casa’	<i>ruka</i>	<i>suka</i>
	<i>pé</i> ‘caminho’	<i>rapé</i>	<i>sapé</i>
	<i>tui</i> ‘sangue’	<i>ruí</i>	<i>tui</i>

Fonte: Elaboração própria.

<sup>28</sup> Cf. nota 1.

As posposições e os verbos inativos multiformes exibem comportamento análogo ao dos substantivos multiformes prototípicos, com a diferença de que não possuem a forma em *t-*, uma vez que são sempre empregados relacionalmente, ver (9), (10), (14) e (15).

Os verbos inativos multiformes possuem tema iniciado por vogal (CRUZ, 2011, p. 135). O alomorfe *s* do prefixo de concordância de 3ª pessoa do singular concatena-se diretamente ao tema, ver (1), (6), (10), (85). Todas as demais flexões pessoais se adjungem à forma prefixada por *r-*, ver (2), (9) e (13).

### 3 O formalismo GF

O termo GF possui três acepções distintas, mas intimamente relacionadas (RANTA, 2010). Em primeiro lugar, designa um formalismo para especificação de gramáticas que constitui, em si mesmo, uma linguagem de programação. Isso significa que uma gramática de uma língua como nheengatu formalizada no GF pode ser não só interpretada por um especialista humano com conhecimento do formalismo, mas também compilada num código binário passível de ser processado por um computador na execução de tarefas como análise (*parsing*) e geração de sentenças, tradução automática etc. Por extensão metonímica, o termo GF designa também o software que permite realizar essas tarefas. Finalmente, na terceira acepção, GF designa uma teoria gramatical, i.e., um modelo da organização e funcionamento das gramáticas de línguas naturais.<sup>29</sup>

De um ponto de vista formal, a sintaxe de uma língua natural consiste no conjunto de regras para formação de unidades complexas (sintagmas) a partir de unidades elementares (palavras) (JURAFSKY; MARTIN, 2009). Essa noção estende-se à morfologia, com a diferença de que as unidades elementares são, tipicamente, morfemas e as maiores, palavras.

---

<sup>29</sup> Diferentemente de outras teorias gramaticais computacionalmente implementadas como a LFG (BRESNAN, 2001), Ranta (2011) não reivindica qualquer validade tipológica ou plausibilidade psicológica para a GF. A justificação da teoria advém da engenharia de software, que procura otimizar a elaboração e a eficiência de programas de computador. Sobre a relação entre teorização linguística e engenharia da gramática, ver Müller (2015).

O formalismo GF baseia-se na teoria matemática dos tipos, que trata da combinação de termos elementares de determinados tipos para formar termos complexos de outros tipos. Enquanto linguagem de programação, o GF enquadra-se no paradigma funcional. Desse modo, a construção dos tipos dá-se por meio da aplicação de funções sobre zero, um ou mais argumentos.

A característica distintiva do GF é a fatoração da combinatória de tipos em dois componentes, a sintaxe abstrata e a sintaxe concreta. No primeiro componente, abstrai-se da realização morfossintática dos elementos envolvidos, que é especificada apenas no segundo.

Na sintaxe abstrata da ITALENG, Ranta (2011) propõe os tipos Comentário, Item, Classe (*Kind*) e Qualidade, que se combinam por meio das funções em (92)-(97). Comentário é o tipo de sentenças declarativas do tipo de (1), (2) etc., em que se atribui uma qualidade a uma entidade (ou conjunto de entidades) de uma determinada classe.

- (92) *Pred*: Item  $\rightarrow$  Quality  $\rightarrow$  Comment;
- (93) *This*, *That*, *These*, *Those*: Kind  $\rightarrow$  Item;
- (94) *Mod*: Quality  $\rightarrow$  Kind  $\rightarrow$  Kind;
- (95) *Very*: Quality  $\rightarrow$  Quality;
- (96) *Food*, *Milk*, *Fish*: Kind;
- (97) *Warm*, *Cheap*, *Expensive*, *Red*: Quality;

Nessa notação, os termos à esquerda dos dois pontos são os nomes das funções e os tipos à direita separados por setas as caracterizam em termos dos tipos de argumentos e do valor que produzem, de tal modo que o último (ou o único) desses elementos é o tipo do valor da função. Por exemplo, a função *Pred* produz, a partir de um Item e de uma Qualidade, um Comentário, que é o tipo mais complexo, não entrando na composição de nenhum outro tipo.

A função *Mod* produz uma Classe a partir de uma Classe e uma Qualidade. Por exemplo, em (4), a classe designada por *peixes vermelhos* resulta da aplicação de *Mod* sobre a Qualidade e a Classe designadas pelo adjetivo e substantivo, respectivamente. O tipo Item resulta da aplicação, a uma Classe, de uma das funções nomeadas pelos demonstrativos do

inglês *This* ‘este’, *That* ‘aquele’, *These* ‘estes’ e *Those* ‘aqueles’. Em outras palavras, essas funções extraem indivíduos ou grupos de indivíduos das Classes designadas pelos seus argumentos. Por exemplo, em (4), a aplicação da função *These* à Classe *peixes vermelhos* resulta no Item *estes peixes vermelhos*. O Comentário expresso por (3) exemplifica a função *Very* ‘muito’, que, aplicada à Qualidade *caro*, gera a Qualidade *muito caro*.

Em (96) e (97), temos funções sem argumentos que produzem as Classes e as Qualidades correspondentes aos conceitos básicos designados, respectivamente, pelos substantivos e adjetivos do inglês que nomeiam essas funções. Por exemplo, a função *Fish* produz a Classe correspondente ao conceito de peixe, enquanto a função *Red* gera a Qualidade correspondente ao conceito de vermelho.

Com base nessa sintaxe abstrata, Ranta (2011) define, na ITALENG, duas sintaxes concretas: uma para o inglês e outra para o italiano. Cada um desses módulos especifica como os diferentes tipos e funções da sintaxe abstrata são *linearizados*, ou seja, expressos nessas duas línguas.

Para especificar as propriedades formais das categorias que realizam os diferentes tipos da sintaxe abstrata, utilizam-se *registros*, uma estrutura de dados que permite organizar em campos as diferentes propriedades dessas categorias. Exemplifiquemos. Em inglês, a função *Red* lineariza-se como (98), ou seja, um registro cujo único campo *s* tem como valor “red”, que constitui uma cadeia de caracteres (tipo *Str*, do inglês *string*, doravante Forma). Em português, a função *Fish* lineariza-se por meio do registro (99), que possui dois campos, *s* e *g*, para representar o paradigma flexional e o gênero do substantivo *peixe*, respectivamente. O valor do primeiro campo é uma tabela bidimensional do tipo Número=>Forma (QUADRO 5), que associa, a cada número gramatical, a cadeia de caracteres correspondente. O valor do segundo campo indica o gênero masculino. Os traços gramaticais gênero, número, caso etc. precisam ser definidos por meio de parâmetros, exemplificados em (100) e (101).

(98) Red={s=“red”};

(99) Fish={s=table{Sg=>“peixe”; Pl=>“peixes”}; g=Masc};

(100) Gender=Masc|Fem ;

(101) Number=Sg|Pl ;

QUADRO 5 – Paradigma do lexema *peixe*

Número	Forma
Sg	“peixe”
Pl	“peixes”

Fonte: Elaboração própria.

Diferentemente de línguas desprovidas de flexão adjetival como o inglês, *Red* lineariza-se em português como (102), cujo único campo tem como valor uma tabela tridimensional do tipo Gênero=>Número=>Forma. Para extrair um membro de um paradigma  $q$  desse tipo, GF utiliza a sintaxe  $q.s!g!n$ , onde  $s$  é o valor do campo  $s$  de  $q$ , o qual constitui uma tabela da qual se extrai, por meio da aplicação sucessiva do operador “!”, a forma de gênero  $g$  e número  $n$ .

(102) Red={s=table{Masc=>table{Sg=>“vermelho”; Pl=>“vermelhos”};  
table{Fem=>table{Sg=>“vermelha”; Pl=>“vermelhas”}}};

A exemplo das linguagens de programação comuns, como Java, C ou Python, GF permite definir operações parametrizadas (funções) para execução de tarefas repetidas. Esse recurso torna desnecessário especificar, no léxico, todas as formas de cada adjetivo do português, uma vez que se flexionam conforme padrões predizíveis a partir do lema. Desse modo, definimos a operação mkAdj, que gera os paradigmas dos diferentes tipos de adjetivo em português, permitindo simplificar (102) como (103). Essa operação utiliza o operador de concatenação de cadeias “+”, chamado “operador de colagem” (*glue operator*), que permite formar palavras a partir de unidades menores, no caso em tela, radical adjetival e flexão de gênero e número. GF permite modelar não só processos morfológicos concatenativos, mas também não concatenativos, como a metáfora. Operações análogas a (103) podem ser definidas para gerar todos os paradigmas de todas as classes de palavras flexionáveis.

(103) Red=mkAdj “vermelho”;

A linearização dos tipos complexos da sintaxe abstrata é especificada de forma análoga, utilizando registros para representar as informações morfossintáticas dos constituintes envolvidos e operações

para manipular esses registros, de modo a dar conta de fenômenos como concordância, regência, variações na ordem dos constituintes etc. Vejamos alguns exemplos. Em português, o tipo Item possui o tipo de linearização de (104), onde o valor do campo *s* é uma tabela do tipo Caso=>Forma. A especificação de caso é necessária porque elementos do tipo Item são realizados não só por sintagmas nominais com núcleo substantival, mas também por pronomes pessoais, que variam em caso conforme a função sintática exercida. Os outros três campos especificam as propriedades inerentes do sintagma nominal. Em (105), temos o tipo de lexicalização de Itens em inglês, que difere de (104) apenas pela ausência de especificação de gênero, que não participa de processos de concordância nessa língua.

(104) {s: Case=>Str; n: Number; p: Person; g: Gender};

(105) {s: Case=>Str; n: Number; p: Person};

As definições de linearização de funções que combinam dois ou mais tipos especificam não somente a ordem em que se concatenam as suas linearizações, mas também, para tipos expressos por tabelas, as células compatíveis. Por exemplo, no caso de *peixes vermelhos*, é exigida compatibilidade de gênero e número. Para concatenar cadeias que representam formas de paradigmas lexicais, GF utiliza o operador ++, exemplificado na função de linearização de *Very* em inglês definida em (106). Como vimos em (92), a função *Very* aplica-se a uma Qualidade para produzir uma Qualidade. Por exemplo, em nheengatu, de *sepiasu* ‘caro’ obtém-se *sepiasu retana* ‘muito caro’, ver (3). Em inglês, o intensificador precede o adjetivo. Desse modo, a função (106) especifica que, em inglês, a cadeia “very” é concatenada ao valor do campo *s* do argumento da função, i.e., à forma do adjetivo, representado pela variável *a*.

(106) *Very* a={s=”very”++a.s};

No caso da função *Mod* de (92), é preciso especificar, para o inglês, que Qualidade precede Classe. Em português, de um modo geral, assim como em nheengatu, Qualidade pode tanto preceder quanto subseguir Classe. Ao contrário do inglês e do nheengatu, o português exige concordância de gênero e número entre os dois constituintes.

A função de linearização (107) da gramática concreta do português aplica-se aos dois registros que constituem as linearizações de Qualidade e Classe, referidos pelas variáveis  $q$  e  $k$ , respectivamente. Na primeira linha, é atribuído, por meio do operador *let*, o valor  $k.g$  à variável  $g$  do tipo *Number*. Na segunda linha, o valor de  $g$  é atribuído ao campo  $g$  do registro que resulta da combinação dos dois argumentos da função. O valor do campo  $s$  desse registro é uma tabela do tipo Número=>Forma, análoga à de (102), com a diferença de que se constitui de uma única linha, visto que uma combinação particular de Qualidade e Classe (por ex., *peixes vermelhos*) possui um número específico, representado pela variável  $n$ . O valor dessa tabela é o resultado da operação *shuffle* aplicada sobre as cadeias que constituem as linearizações de  $q$  e  $k$ . Essa operação, que definimos no módulo Oper.gf, gera todas as permutações possíveis entre os seus argumentos. Desse modo, a gramática gera adjetivos tanto pós-nominais quanto pré-nominais (ordem marcada no caso de adjetivos de cores).

(107) Mod  $q\ k=let\ g: Gender=k.g\ in$   
 $\{s=\backslash n=>shuffle\ (q.s!n!g)\ (k.s!n); g=g\};$

A linguagem de programação GF determina uma estrita separação entre a concatenação de *tokens* (palavras) para formar listas de *tokens* (sintagmas) por meio do operador “++”, exemplificado em (106), e a colagem de cadeias para formar *tokens* (palavras individuais ou locuções) por meio do operador “+”, utilizado pela operação (103).<sup>30</sup> Este último processo ocorre em tempo de compilação do código-fonte, ao passo que o primeiro ocorre em tempo de execução do código binário. A exigência de todos os *tokens* estarem formados em tempo de execução confere maior eficiência à análise e geração de sentenças e tem uma consequência importante para a implementação linguística: regras de geração de formas representadas por *tokens* individuais, como as formas com prefixo inativo *s-* do *nheengatu*, precisam ser modeladas como regras lexicais. Não é possível, portanto, na definição da função de linearização de *Pred* numa sintaxe concreta do *nheengatu*, concatenar *s-* com uma base verbal para formar, por exemplo, *suri*, como em (10). Por outro lado, é possível gerar

<sup>30</sup> Limitamo-nos aqui a uma descrição simplificada, informal dos dois operadores, definidos formalmente em Ranta (2011).

formas do tipo de *se ruri* de (9) ou na sintaxe por meio da concatenação dos *tokens* “se” e “ruri”, solução que adotamos, ou no léxico como um único *token* “se ruri” (análogo a uma locução).

#### 4 Descrição da implementação

Uma descrição detalhada das sintaxes concretas do nheengatu, português e inglês e da sintaxe abstrata subjacente às três línguas extrapolaria os limites de um artigo. O leitor familiarizado com o GF pode consultar os códigos-fonte de todos os componentes, disponíveis *on-line* sob licença de uso de software livre e suficientemente autoexplicativos.<sup>31</sup> Desse modo, limitamo-nos aqui a destacar as particularidades mais importantes da implementação.

Um princípio fundamental da programação é a modularização, a divisão de um programa complexo em componentes organicamente articulados. Desse modo, dividimos a sintaxe abstrata nos seguintes componentes: (i) Gra.gf, (ii) Func.gf e (iii) Cont.gf. As sintaxes concretas do nheengatu, português e inglês constituem os módulos GraYrl.gf, GraPor.gf e GraEng.gf, respectivamente, cada um dos quais recorre a operações definidas num módulo Oper.gf comum e nos módulos específicos OperYrl.gf, OperPor.gf e OperEng.gf.

O módulo Gra.gf constitui-se de regras de formação de tipos complexos, como as exemplificadas em (108)-(112). Algumas dessas funções utilizam tipos definidos nos demais componentes, como Polaridade, Locação e Qualidade.

(108) Pred: Polarity → Item → State → Comment;

(109) StageLevelState: Property → State;

(110) IndLevelState: Property → State;

(111) mkPropLoc: Location → Property;

(112) mkPropQual: Quality → Property;

---

<sup>31</sup> Disponível em: <http://github.com/leoalenc/nheengatu>. Acesso em: 27 ago. 2020.

O módulo Func.gf contém funções que definem tipos linearizados por palavras funcionais ou sinsemânticas (BUSSMANN, 2002), ou seja, em GraYrl.gf, o advérbio de intensidade *retana* ‘muito’, clíticos de negação, posposições, pronomes e determinantes. Por exemplo, as funções de (113), análogas a (93), aplicam-se ao tipo Classe para formar o tipo Não-Dêitico, de que trataremos mais adiante, linearizado sob a forma de um DP pleno. A presente versão da gramática não contempla DPs com quantificadores ou numerais, exemplificados em (49), (77) e (78). *TheSG* e *ThePL* permitem construir DPs definidos no singular e no plural, respectivamente. Em GraYrl.gf, essas duas funções linearizam-se como DPs com artigo zero. Desse modo, exemplos do nheengatu com esse tipo de constituinte serão sempre traduzidos em português e inglês como DPs plenos nucleados por artigo definido. Essa solução, porém, tem um caráter apenas prático, uma vez que, em narrativas em nheengatu, uma interpretação indefinida é, por vezes, preferível.<sup>32</sup> No entanto, a solução teoricamente mais adequada, que seria implementar as duas interpretações possíveis, provocaria um aumento exponencial da ambiguidade de sentenças com esse tipo de DP.<sup>33</sup>

(113) TheSG, ThePL, This, That, These, Those: Kind → NonDeictic;

O módulo Cont.gf define tipos expressos por palavras lexicais (*content words*) ou autosssemânticas (BUSSMANN, 2002), no caso, apenas substantivos e adjetivos, pois verbos plenos não foram ainda incluídos. As cópulas são tratadas como palavras sincategoremáticas, definidas por Ranta (2011, p. 100) como palavras sem representação própria na sintaxe abstrata. Desse modo, são introduzidas na regra de linearização da função *Pred* de (108).

Para a elaboração da sintaxe abstrata, tomamos como ponto de partida os tipos e as funções de (92), extraídos da minigramática do inglês e do italiano de Ranta (2011). Essa gramática, porém, restringe-se a uma fração dos fenômenos gramaticais do Quadro 2. Desse modo,

---

<sup>32</sup> Cf. nota 20.

<sup>33</sup> Abstraindo de outras fontes de ambiguidade, o número de leituras de uma sentença com *n* DPs com núcleo zero, pressupondo a ambiguidade desse núcleo entre uma interpretação definida e outra indefinida, seria 2<sup>n</sup>. Desse modo, uma sentença com quatro DPs desse tipo teria, pelo menos, 16 leituras.

foi necessário criar diversos novos tipos, reformular várias das regras propostas por Ranta (2011) e elaborar novas regras.

Para dar conta da negação, da predicação locativa e da distinção entre nível de fase e nível de indivíduo, substituímos a regra (92) pelas regras de (108)-(112). Conforme (108), um Comentário é constituído de uma Polaridade, um Item e um Estado. O tipo Polaridade, por sua vez, é formado em Func.gf a partir dos tipos *Yes* ou *No*, que codificam polaridade positiva e negativa, respectivamente. Apenas esta última é expressa na linearização.

O tipo Estado constitui-se de uma Propriedade, que, por sua vez, pode ser uma Locação ou uma Qualidade, conforme (111) e (112). As funções (109) e (110) geram estados contingentes e não-contingentes, respectivamente. Qualidades são geradas nos módulos Cont.gf e Func.gf por funções análogas a (95) e (97), respectivamente, Locações no módulo Func.gf pelas funções de (114) e (115). Por exemplo, em (15), a aplicação da função *Near* ao Item designado por *pe taua* ‘a cidade de vocês’ gera a Locação referida por *pe taua ruaki* ‘perto da cidade de vocês’. Em (11), o advérbio *iké* ‘aqui’ constitui em si mesmo uma Locação, dado que lineariza a função *Here* de (115).

(114) On, With, In, Inside, Near: Item -> Location;

(115) Here, There: Location;

Do ponto de vista da semântica referencial, é indiferente se um Item é linearizado como pronome pessoal ou DP pleno, compare (8) e (7). Ambos os exemplos consistem numa predicação qualificativa que atribui uma qualidade a uma entidade. Pronomes pessoais e DPs plenos, porém, não compartilham exatamente a mesma distribuição, uma vez que os primeiros não realizam PSUM, ver (50) e (51). Desse modo, como meio de evitar a hipergeração, a implementação, em sintaxe abstrata, das construções possessivas e dos sujeitos pronominais, inclusive nulos, distingue, no âmbito dos Itens, por meio das regras (116) e (117), entre os subtipos Dêítico e Não-Dêítico, linearizados como pronomes pessoais e DPs plenos, respectivamente.<sup>34</sup>

---

<sup>34</sup> Seguimos aqui Lyons (1995, p. 307), que considera dêíticos tanto os pronomes pessoais de 1ª e 2ª pessoas quanto os anafóricos de 3ª.

(116) mkItemDeictic: Deictic -> Item;

(117) mkItemNonDeictic: NonDeictic -> Item;

O tipo Dêítico, por sua vez, é construído por meio das funções de (118), cujos nomes se baseiam nas formas dos pronomes pessoais do inglês. Dada a ambiguidade da forma pronominal *you*, a 3ª pessoa do singular e a do plural são designadas por meio de *YouSG* e *YouPL*, respectivamente, analogamente a *TheSG* e *ThePL* de (113).

(118) He, She, It, They, I, YouSG, YouPL, We: Deictic;

Completam o módulo Gra.gf as regras de (119)-(124), responsáveis pela geração de construções possessivas do tipo de (5), (7), (14), (15), (52) e (53). Nesta versão da gramática, limitamo-nos a um subconjunto das construções possessivas exemplificadas por Cruz (2011), deixando, para uma versão futura, construções recursivas com mais de um núcleo nominal encaixado, como em (54).

(119) Poss: Psor -> NonDeictic -> Item;

(120) Poss\_: PossPro -> SimpleKind -> PossKind;

(121) mkPsor: Num -> SimpleKind -> Psor;

(122) mkPsor\_: Num -> PossKind -> Psor;

(123) mkKind: SimpleKind -> Kind;

(124) mkKind\_: PossKind -> Kind;

Conforme (123) e (124), há dois tipos de Classes: Classe Simples (*SimpleKind*), como *tendaua* ‘comunidade’ em (12), e Classe Possuída (*PossKind*), como *sendaua* ‘comunidade dele’ e *pe taua* ‘cidade de vocês’ em (5) e (15). O primeiro subtipo é gerado no módulo Cont.gf por meio de regras do tipo de (125). O segundo é construído por meio de (120) pela combinação dos tipos Classe Simples e PossPro, este último construído por meio das funções de (126), análogas às de (118).

(125) Food: SimpleKind;

(126) His, Her, Its, Their, My, YourSG, YourPL, Our: PossPro;

Por que pronomes possessivos (prefixos inativos em *nheengatu*) não foram implementados como linearizações de Itens? A razão é que, diferentemente do inglês e do português, o *nheengatu* não licencia um DP pleno como PSOR, mas apenas um NP ou um prefixo inativo, conforme (55) e (61). De modo a possibilitar a tradução, entre as três línguas, de construções possessivas do subconjunto referido, evitando, ao mesmo tempo, a hipergeração em *nheengatu*, criamos o tipo *Psor*, construído pelas regras (121) e (122) a partir do tipo *Num*, por um lado, e do tipo Classe Simples ou Classe Possuída, por outro. O tipo *Num* é construído no módulo *Func.gf* por meio de (127), linearizando-se como morfema de número, o qual determina a forma de plural ou singular da expressão que realiza o segundo argumento das funções (121) e (122).

(127) PL, SG: Num;

Em *nheengatu*, conforme (57), o PSOR é um NP, podendo traduzir-se em português tanto por um NP quanto por um DP nucleado por artigo definido, ver (71). Como essa última opção tradutória é a mais comum em Navarro (2011), implementamo-la nas funções de linearização de (119) em português e inglês, seguindo o tratamento conferido ao DP pleno de núcleo zero.

A função (128) define o tipo de linearização de Classe como registro cujo único campo constitui uma tabela do tipo Número=>*NForm*=>Forma. O parâmetro *NForm*, definido em (129), modela a distribuição das formas de substantivos multiformes, tendo dois construtores, Rel e Abs. O primeiro indica que a forma é relativa, ou seja, constitui complemento genitivo, o segundo que se trata da forma absoluta. Como há duas formas no primeiro caso, conforme o tipo de argumento interno, uma para prefixo inativo de 3ª pessoa do singular, a outra para todos os demais casos, o construtor Rel possui como argumento o parâmetro *ArgForm*, definido em (130), cujos valores correspondem, respectivamente, aos dois tipos de argumento interno.

(128) KIND: Type={s: Number=>*NForm*=>Str};

(129) *NForm*=NRel *ArgForm*|NAbs;

(130) *ArgForm*=SG3|NSG3;

A função (131) define o tipo de linearização de Classe Simples. Constitui uma extensão de (128), diferindo pela inclusão do campo *nc*, que codifica a classe nominal. O parâmetro *NClass*, definido em (132), classifica os substantivos em uniformes e multiformes. Assume um dos dois valores *NCI* e *NCS*, segundo o alomorfe do prefixo inativo de 3ª pessoa do singular exigido, *i* e *s-*, respectivamente, permitindo a linearização correta de (120).

(131) SIMPLEKIND: Type=KIND\*\*{nc: NClass};

(132) NClass=NCI|NCS;

Como evidenciam os exemplos (1)-(17), o tipo Item lineariza-se em nheengatu tanto como DP pleno quanto como pronome livre, pronome nulo ou prefixo inativo, dependendo de uma série de fatores. O primeiro é a função sintática, que determina caso nominativo ou genitivo. O segundo é a classe do predicado regente: apenas verbos inativos flexionáveis licenciam incondicionalmente um sujeito nulo, os outros tipos de predicado o fazem condicionalmente, dependendo do terceiro fator, o nível da predicação. Por exemplo, em (18), o sujeito nulo é licenciado porque, não obstante o predicado regente consistir num PP, a predicação é de nível de fase, marcada, portanto, pelo auxiliar *iku*, compare com (24) e (23). O quarto fator é a classe flexional (*NClass*), que regula a alomorfia entre *i* e *s-*.

O tipo de linearização de Item em nheengatu, definido em (133), que incorpora (134), leva em conta todos esses fatores. Embora desprovida, como em inglês, de especificação de gênero gramatical, essa definição, que se aplica também ao tipo Dêitico, é bem mais complexa do que as correspondentes nas outras duas línguas, compare-se com (104) e (105).

(133) ITEM: Type={s: Class=>Level=>FORM; n: Number; p: Person}\*\*{pos: POS};

(134) FORM: Type=Case=>NClass=>Str;

Tipicamente, para linearizar uma função que combina dois ou mais tipos, como *Pred* e *Mod*, basta concatenar sequencialmente as cadeias que representam os diferentes argumentos da função, como em (106) e (107). Em nheengatu, esse também é geralmente o caso. No

entanto, para linearizar corretamente (119), a representação ortográfica do PSOR deve ser enxertada dentro da do PSUM. Por exemplo, em (17), o DP sujeito *Pedro* (PSOR) deve ser inserido em *nhaã pindá* ‘aquele anzol’ (PSUM). A simples concatenação desses dois constituintes resultaria numa construção agramatical, ver (135).

- (135) \*Pedro nhaã            pindá  
           Pedro DEM.DIST   anzol

- (136) {s: {d: Str; h: NForm=>Str}; n: Number; p: Person} \*\* {pos: POS};

A definição de linearização do tipo Não-Dêitico em (136) permite gerar exemplos como (17), evitando, ao mesmo tempo, a geração de exemplos como (135). Nessa definição, o campo *s* consiste num registro com dois campos. O primeiro abriga a forma do determinante, o segundo, uma tabela do tipo *NForm=>Forma*, que determina a forma do substantivo, conforme (129). Na linearização de (119), a forma do PSOR é inserida entre as dos campos *d* e *h*.

Vejam agora como o léxico é codificado na sintaxe concreta. Em GF, entradas lexicais possuem o formato CONCEITO=LINEARIZAÇÃO, como vimos em (98)-(102). Consideremos primeiro estes exemplos:

- (137) River=regNoun (“paranã”|”paraná”);  
 (138) Son\_Of\_Woman=regNoun (“mimbira”|”mbira”);  
 (139) Brother\_Of\_Woman=regNoun “kiuíra”;  
 (140) Language=regNoun “nheenga”;  
 (141) Word=regNoun “nheenga”;

O léxico modela uma relação de muitos para muitos, porque um dado conceito, numa dada língua, pode ter mais de uma linearização, ver (137) e (138),<sup>35</sup> ao mesmo tempo que diferentes conceitos podem compartilhar a mesma linearização, ver (140) e (141).

<sup>35</sup> GF permite implementar esse tipo de variação como resultado da aplicação de regras que alteram representações ortográficas de itens lexicais, análogas a regras fonológicas. Deixamos isso para uma versão futura da gramática.

A notação especial dos conceitos de (138) e (139) decorre de que não são lexicalizados em inglês (nem em português), apenas em nheengatu. O termo *mimbira* (ou *mbira*) designa tanto o filho quanto a filha de uma mulher. No nheengatu, tanto os designativos de parentes colaterais de 2º grau (i.e., irmãos) quanto os dos demais descendentes imediatos (i.e., filho e filha de homem) lexicalizam não somente o gênero da própria pessoa, mas também do parente.

Para facilitar a codificação das entradas lexicais, seguindo o exemplo de (103), implementamos no módulo OperYrl diversas operações que geram os paradigmas flexionais de substantivos e adjetivos. Nas entradas acima, a operação regNoun gera as formas de singular e plural de substantivos regulares, ou seja, uniformes.

Conforme o Quadro 4, os substantivos multiformes distribuem-se em cinco microgrupos. A operação RelPrefNoun, exemplificada em (142)-(147), permite reduzir esses microgrupos a apenas dois macrogrupos, facilitando a codificação de novas entradas lexicais. O primeiro engloba os microgrupos (i)-(iv), reunindo, portanto, a maioria dos substantivos multiformes, com entradas lexicais no formato de (142)-(146). Nesse caso, a operação RelPrefNoun tem como argumento apenas a forma de citação do substantivo, a partir da qual as demais formas são computadas. O segundo macrogrupo é constituído pelos membros do microgrupo (v), para os quais é necessário especificar, além da primeira, a terceira forma do Quadro 4, ver (147).

(142) Community=RelPrefNoun “tendaua”;

(143) Picture=RelPrefNoun “sangaua”;

(144) Street=RelPrefNoun “pé”;

(145) Path=RelPrefNoun “pé”;

(146) House=RelPrefNoun “uka”;

(147) Blood=RelPrefNoun “túi” “túi”;

As entradas das demais classes de palavras possuem uma estrutura análoga. Vejamos alguns exemplos. Para codificação das posposições locativas, implementamos a operação mkLoc, exemplificada em (148)-(151). Essa operação possui duas variantes, uma monoargumental para as

posposições uniformes e outra biargumental para as multiformes, compare (148)-(150) com (151). As duas primeiras entradas correspondem às preposições *on* e *in* do inglês e *em* do português. Enquanto *upé* lineariza os dois conceitos, o segundo é passível de linearização também pela posposição *pupé* e pela locução pospositiva *kuara upé*, que expressam localização no interior de algo. Na última entrada, o segundo argumento da operação indica que se trata de posposição multiforme.

(148) On=mkLoc “upé”;

(149) In=mkLoc (“upé”|”pupé”|”kuara upé”);

(150) With=mkLoc “irūmu”;

(151) Near=mkLoc “ruaki” NCS;

Verbos inativos são codificados por meio de variantes da operação mkQual: a monoargumental se aplica aos não flexionáveis, a biargumental, aos flexionáveis, compare (152) e (153) com (154) e (155). Nesse último caso, o parâmetro *NClass* permite distinguir verbos uniformes de multiformes, por meio de NCI e NCS, respectivamente.

(152) New=mkQual “pisasu”;

(153) Red=mkQual “piranga”;

(154) Heavy=mkQual “pusé” NCI;

(155) Hot=mkQual “raku” NCS;

Pronomes pessoais são codificados por meio da operação *PersPron*, que recebe três argumentos, a saber, número, pessoa e forma ortográfica, como nos exemplos (156)-(160). Observe que, na 3ª pessoa, uma única forma do nheengatu lineariza três pronomes diferentes do inglês, dado que a primeira língua não marca o gênero gramatical.

(156) YouSG=PersPron Sg P2 “indé”;

(157) YouPL=PersPron Pl P2 “penhê”;

(158) He=PersPron Sg P3 “aé”;

(159) She=PersPron Sg P3 “aé”;

(160) It=PersPron Sg P3 “aé”;

De forma análoga, os possessivos de (126) são linearizados em *nheengatu* por meio da operação *PossPron*, que possui apenas dois argumentos, número e pessoa, uma vez que a forma é gerada por uma outra operação, responsável pela geração dos prefixos inativos.

(161) YourSG=PossPron Sg P2;

(162) YourPL=PossPron Pl P2;

(163) His=PossPron Sg P3;

(164) Her=PossPron Sg P3;

(165) Its=PossPron Sg P3;

Concluímos esta seção com dados quantitativos da cobertura lexical da gramática do *nheengatu*. Conforme a Tabela 1, foram implementados apenas cerca de um sexto dos substantivos e posposições e um quinto dos adjetivos do glossário de Navarro (2011). A Tabela 2 apresenta um levantamento comparativo das quantidades de conceitos implementados na GrammYEP e na ITALENG. Pode-se constatar que a primeira é bem mais abrangente do que a segunda.

Há ainda um longo caminho a percorrer na codificação do léxico do *nheengatu*, de modo a cobrir todo o vocabulário das 13 lições de Navarro. Essa tarefa, contudo, será bastante facilitada pelas operações lexicais implementadas, as quais geram as representações de todas as subclasses de substantivos e adjetivos a partir unicamente da forma de citação na maioria dos casos, necessitando de apenas um argumento adicional para os demais.

TABELA 1 – Percentual implementado das principais categorias do glossário de Navarro (2011)

<b>Categoria</b>	<b>Número de lexemas</b>	<b>Percentual implementado</b>
Substantivo	305	15,1%
Adjetivo	72	20,8%
Verbo	186	0,5%
Advérbio	62	3,2%
Posposição	26	15,4%
Numeral	23	0%

Fonte: Elaboração própria.

TABELA 2 – Quantidades dos principais conceitos na GrammYEP e na ITALENG

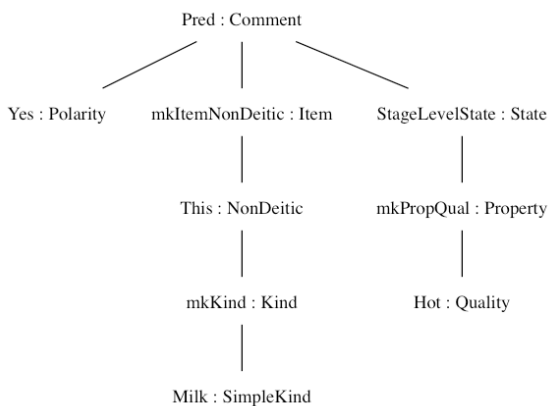
<b>Tipo</b>	<b>GrammYEP</b>	<b>ITALENG</b>
Classe	50	4
Qualidade	15	6
Locação	5	0

Fonte: Elaboração própria.

## 5 Avaliação da gramática

Nesta seção, apresentamos dados que permitem avaliar a qualidade da gramática computacional do nheengatu na análise e geração de sentenças. No primeiro caso, cada sentença definida como gramatical recebe uma ou mais representações semânticas, correspondentes às diferentes leituras da sentença conforme a sintaxe abstrata. No segundo, são geradas todas as linearizações possíveis de uma dada representação semântica, conforme as sintaxes concretas dadas.

FIGURA 5 – Análise da sentença (1) em termos de funções e tipos semânticos

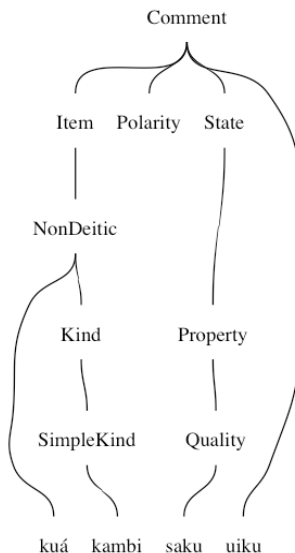


Fonte: Elaboração própria.

São três os tipos de formatos de representação semântica que o software GF pode produzir para cada leitura de sentença passível de geração por uma gramática dada: (i) árvore em formato textual, (ii) gráfico arbóreo de funções e tipos e (iii) gráfico arbóreo ligando palavras e tipos semânticos. O primeiro formato é a base para geração dos dois últimos, os quais constituem variações notacionais mais amigáveis. Exemplifiquemos com base em (1). Para essa sentença não ambígua, o analisador gera unicamente a árvore (166). Essa árvore indica as sucessivas aplicações funcionais para construir os tipos de que se constitui a representação semântica da sentença. A partir dessa representação, podem ser gerados os gráficos da Figura 5 e na Figura 6. No primeiro, cada nó constitui um par ordenado  $f:t$ , onde  $f$  é a função correspondente de (166), e  $t$  é o tipo produzido por essa função, cujos argumentos, no caso de funções com aridade maior que zero, são os respectivos nós filhos. Por exemplo, *Pred*, a função hierarquicamente mais alta, gera um Comentário a partir dos tipos Polaridade, Item e Estado, gerados, por sua vez, respectivamente, pelas funções indicadas à esquerda dos dois pontos em cada caso.

(166) Pred Yes (mkItemNonVar (This (mkKind Milk))) (StageLevelState (mkPropQual Hot))

FIGURA 6 – Correspondência entre palavras e tipos semânticos do exemplo (1)



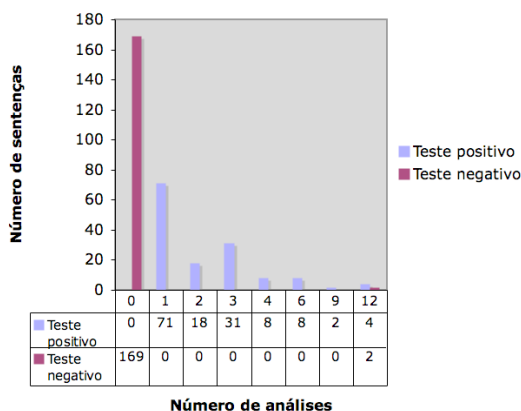
Fonte: Elaboração própria.

No gráfico da Figura 6, os nós não terminais são os tipos semânticos, enquanto os terminais, ou seja, as “folhas” da árvore, são as palavras da sentença, que constituem linearizações desses tipos. Conforme esse gráfico, a sentença é do tipo Comentário, o nó raiz da árvore, que se constitui dos tipos dos nós filhos Item, Polaridade e Estado, os quais, por sua vez, se constituem dos tipos dos respectivos filhos. O gráfico distingue formalmente as palavras sincategoremáticas *kuá* ‘este’ e *uiku* ‘está’ das categoremáticas *kambu* ‘leite’ e *saku* ‘quente’. Enquanto estas são os únicos filhos dos respectivos pais, aquelas têm tipos como irmãos.

Um fragmento de gramática consiste num modelo de um dado recorte de fenômenos gramaticais. Desse modo, deve satisfazer duas exigências simultâneas: (i) analisar todas as sentenças gramaticais passíveis de ser construídas utilizando os recursos desse recorte, (ii) não gerar nenhuma sentença agramatical. Para tanto, foram compilados dois conjuntos-teste iniciais, um negativo e outro positivo, denominados GraYrl-Neg e GraYrl-Pos, constituídos, respectivamente, de 171 sentenças agramaticais, como (20)-(23), (81)-(84) e (90), e de 142 sentenças

gramaticais, incluindo todos os exemplos deste trabalho, com exceção de dois grupos. O primeiro contém DPs quantificados ou múltiplos possuidores sob a forma de núcleo nominal, ver (49), (54) e (60). O segundo consiste de construções que extrapolam o domínio das predicções qualificativas, ver (32), (64), (65), (67), (72)-(74), (77), (78), (86) e (88).

GRÁFICO 1 – Número de análises por número de sentenças nos conjuntos-teste positivo e negativo



Fonte: Elaboração própria.

O Gráfico 1 apresenta os resultados da aplicação da gramática do nheengatu na análise desses dois conjuntos-teste. Todas as sentenças gramaticais foram analisadas, com uma média de 2.44 análises por sentença. Apenas duas sentenças do conjunto negativo foram analisadas, a saber, os exemplos (81) e (82), cada um dos quais recebeu doze análises. Essas análises, porém, referem-se a leituras em que o prefixo inativo na função de complemento genitivo de *uka* ‘casa’ e *igara* ‘canoa’ não é correferencial do DP asteriscado entre parênteses, como nas linearizações em português de (81) em (167) e (168). Esses exemplos não parecem semanticamente muito felizes, mas não violam nenhuma regra sintática ou morfológica. De fato, as sentenças estruturalmente análogas (169) e (170) são aceitáveis.

(167) a casa dele da mulher é na cidade

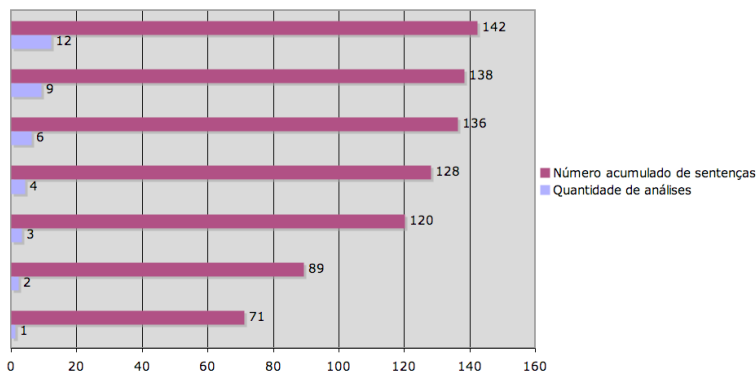
(168) a mulher é na cidade da casa dele

(169) a fotografia dele da canoa está ali

(170) a casa é na cidade da irmã dele

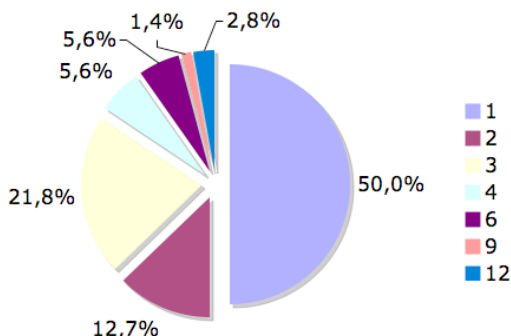
Uma das maiores dificuldades da análise sintática automática é o elevado número médio de análises geradas para as sentenças de entrada, uma vez que, via de regra, apenas uma interpretação está em jogo num determinado contexto (LJUNGLÖF; WIRÉN, 2010). O Gráfico 2 e o Gráfico 3 mostram que o grau de ambiguidade produzido pela gramática nesse conjunto-teste é baixo, uma vez que a grande maioria das sentenças (120 do total ou 84,5%) recebeu entre uma e três análises, sendo que a metade recebeu apenas uma análise.

GRÁFICO 2 – Quantidade de análises por número acumulado de sentenças do conjunto-teste positivo



Fonte: Elaboração própria.

GRÁFICO 3 – Percentual de sentenças do conjunto-teste positivo por número de análises



Fonte: Elaboração própria.

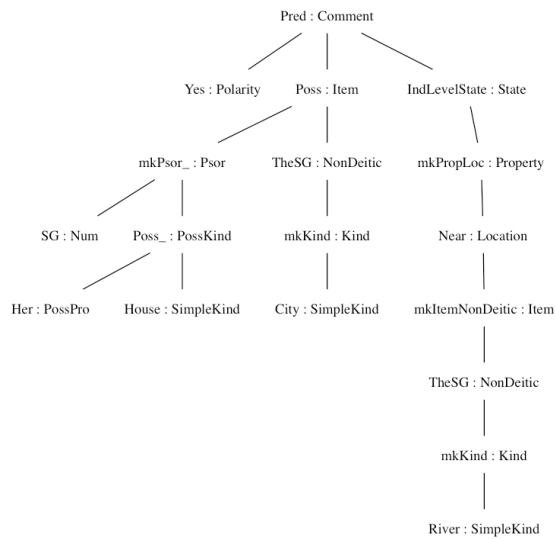
A ambiguidade pode ser de dois tipos: lexical e gramatical (LYONS, 1995). No primeiro caso, uma palavra lineariza mais de um conceito, como em (144) e (145) ou (158)-(160). No segundo, uma mesma sequência de palavras corresponde a duas ou mais configurações arbóreas, como nas leituras gramaticais de (81) e (82) acima referidas. Com frequência, os dois tipos de ambiguidade interagem, o número de leituras de um multiplicando o do outro. Esse é o caso de (171). Por um lado, *taua* ‘casa’ liga-se ou à palavra precedente, funcionando como seu núcleo regente, ou à subsequente como seu complemento nominal, conforme as representações da Figura 7 e da Figura 8, respectivamente, para as quais o sistema gera as linearizações em português (172) e (173). Por outro lado, o prefixo inativo *s-*, na função de possessivo nesse exemplo, possui três leituras, conforme (163)-(165). A interação entre todas essas ambigüidades produz  $2 \times 3 = 6$  leituras.

(171) s-uka                    taua    paranã        r-uaki  
 3s.INACT-casa cidade rio    RLL-perto.de

(172) a cidade da casa dela é perto do rio

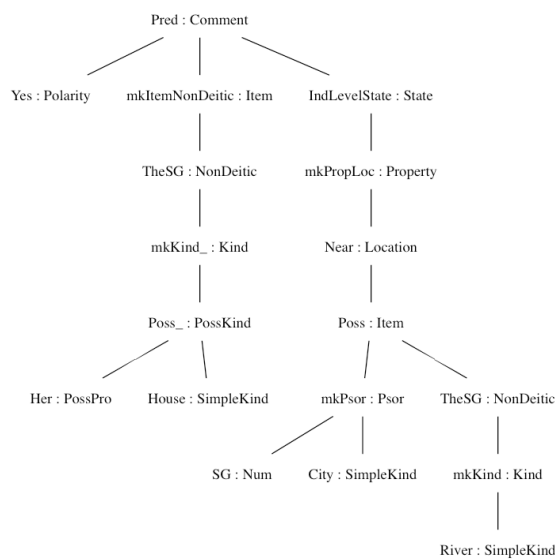
(173) a casa dela é perto do rio da cidade

FIGURA 7 – Leitura de (171) correspondente a (172)



Fonte: Elaboração própria.

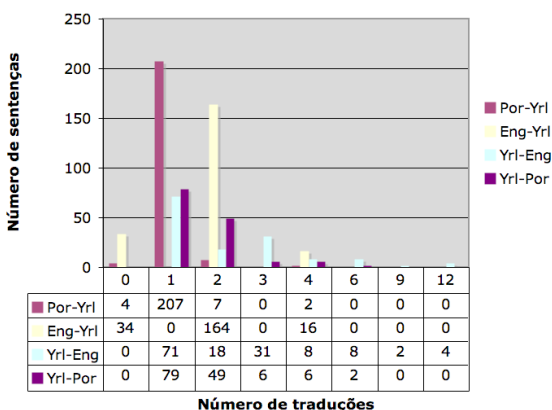
FIGURA 8 – Leitura de (171) correspondente a (173)



Fonte: Elaboração própria.

A gramática foi avaliada também como recurso de tradução automática entre o nheengatu, o inglês e o português, com o primeiro idioma tanto como língua-fonte quanto língua-alvo. No primeiro caso, o sistema produz análises do tipo de (166) para as sentenças do nheengatu, a partir das quais gera as linearizações correspondentes do inglês e do português. No segundo caso, o processo é inverso: as representações semânticas geradas para as sentenças do inglês e do português são linearizadas em nheengatu. O Gráfico 4 apresenta os resultados. Os 4 pares de língua-fonte e língua-alvo (nesta ordem) estão identificados pelos respectivos códigos no padrão ISO 639-3 (EBERHARD; SIMONS; FENNIG, 2020), aqui com inicial maiúscula.

GRÁFICO 4 – Número de traduções por número de sentenças nos conjuntos-teste Por-Yrl, Eng-Yrl, Yrl-Eng e Yrl-Por



Fonte: Elaboração própria.

Todas as 142 sentenças do conjunto-teste positivo foram traduzidas para o português e o inglês, com uma média de 1,63% e 2,44 traduções por sentença, respectivamente, eliminando as repetições.<sup>36</sup> Comentemos primeiro os resultados obtidos com o par Yrl-Por. Mais da metade das sentenças-fonte (79 ou 55.63%) recebeu uma única tradução,

<sup>36</sup> As repetições decorrem de distinções semânticas que não são expressas em todas as três línguas, por exemplo, a distinção entre gênero masculino, feminino e neutro do inglês ou entre predicado de nível de indivíduo e predicado de nível de fase do nheengatu e do português.

mais de um terço (49 ou 34,51%) exatamente duas e a pequena parcela restante (14 ou 9,86%) entre 3 e 6. Apenas duas sentenças não foram traduzidas corretamente, a saber (174), baseada em (53), e variante com o alomorfe *aintá* em vez de *ta*. Na tradução dessas sentenças, a gramática do português produziu a forma agramatical *delea* em vez de *deles*.

- (174) ne                      kiuíra                      ta                      nheenga puranga  
       2s.INACT irmão.de.mulher 3p.INACT palavra bonito  
       ‘a palavra deles do teu irmão é bonita’

Como mostra o Gráfico 4, na tradução do nheengatu para o inglês (identificada por Yrl-Eng), 120 sentenças, portanto, a grande maioria (84,51% do total), receberam entre uma e três traduções, ao passo que as 22 restantes (15,49%), entre quatro e doze. Todas as traduções permitem compreender as diferentes leituras das sentenças-fonte. No entanto, 59 traduções (de 7 sentenças-fonte) apresentam desvios em relação ao inglês padrão contemporâneo, no que tange à ordem de demonstrativos e possessivos. Por exemplo, em vez de (177), (175) é traduzida como (176), ao lado de duas outras possibilidades, dada a não especificação de gênero do possessivo na sentença-fonte. Enquanto essa construção ocorria normalmente em estágios anteriores do inglês, hoje em dia é considerada obsoleta (COMBINING, 2019). Por outro lado, o sistema traduz (91) como (178), quando (179) seria o correto. Esses dados apontam para a necessidade de corrigir a linearização dos possessivos na sintaxe concreta do inglês.

- (175) nhaã                      x-imbiú                      puranga  
       DEM.DIST 3s.INACT-comida                      bonito  
       ‘aquela comida dele é bonita’

(176) ?that his food is beautiful

(177) that food of his is beautiful

(178) \*my house is in that dirty his street

(179) my house is in that dirty street of his

A partir das traduções em português e inglês corrigidas, foram compilados os conjuntos-teste Por-Pos e Eng-Pos, com 220 e 214 sentenças cada, a fim de avaliar a tradução automática para o nheengatu. Conforme o Gráfico 4, 216 e 180 sentenças, representando 98,18% e 84,11% de cada conjunto, obtiveram traduções em nheengatu, com uma média de 1,04 e 1,83 tradução por sentença, respectivamente.

Todas as traduções do inglês estão corretas. No caso do português como língua-fonte, o único erro cometido pelo sistema foi usar o substantivo necessariamente possível *simiriku* ‘esposa’ em (182) e (183) como tradução de (180) e (181), respectivamente, ao lado das traduções que realizou corretamente com o substantivo *kunhã* ‘mulher’, capaz de ser usado autonomamente. A fonte desse problema é a ambiguidade da palavra *mulher* em português, que lineariza tanto o conceito *Wife* quanto *Woman*.

(180) a mulher está bonita

(181) aqueles filhos da mulher estão com eles

(182) simiriku            puranga   uiku  
3.INACT-esposa bonito   3p.ACT-estar  
‘a esposa dele está bonita’

(183) nhaã-itá            s-imiriku            mimbira            uiku            aintá            irũmu  
DEM.DIST-PL 3.INACT-esposa filho.de.mulher 3p.ACT-estar 3p.INACT com  
‘aqueles filhos da esposa dele estão com eles’

O pior desempenho da tradução a partir do inglês, em que 15,89% das sentenças-fonte não foram traduzidas, contra apenas 1,82% com o português como língua-fonte, explica-se pelo problema apontado na sintaxe concreta dos possessivos, ver (176) e (178). Essa questão representa uma dificuldade também na tradução a partir do português, porém, em menor grau, uma vez que só afeta os possessivos de 3ª pessoa, no caso de sentenças-fonte como (174).

Desta avaliação da gramática resultou um *treebank* com 243 sentenças do nheengatu, reunindo o conjunto-teste positivo e as traduções dos conjuntos Por-Yrl e Eng-Yrl, emparelhadas com as sentenças equivalentes em português e inglês.

## 6 Considerações finais

Neste artigo, apresentamos a implementação computacional de um fragmento do nheengatu abrangendo cerca de um quinto do conteúdo gramatical de Navarro (2011). Esse fragmento integra a GrammYEP, uma gramática computacional multilíngue no formalismo GF, da qual fazem parte fragmentos análogos do português e do inglês. Com isso, o sistema traduz do nheengatu para essas duas línguas e vice-versa. No momento, limita-se a orações que atribuem qualidades e localizações contingentes e não-contingentes a pessoas e coisas. O vocabulário é reduzido, mas pode ser facilmente expandido por meio das operações lexicais implementadas.

Pré-requisito para a implementação computacional de uma língua é uma rigorosa formalização das estruturas gramaticais e lexicais. Dado o caráter não formalizado de Navarro (2011) e Cruz (2011), as duas descrições do nheengatu utilizadas, formalizamos inicialmente as estruturas de constituintes na CFG. Em seguida, modelamos formalmente as restrições de concordância e valência na combinatória de constituintes. Finalmente, integramos ambas as dimensões na implementação em GF.

A formalização revelou lacunas e inconsistências daquelas duas abordagens no tratamento das expressões nominais, que em parte sanamos com base nos dados desses mesmos autores. No entanto, algumas questões levantadas ficaram por esclarecer à luz de novos dados, entre as quais destacamos as seguintes:

- i. A complementação genitiva é recursiva?
- ii. Qual o estatuto do possuidor nessa construção? DP ou NP?
- iii. A dupla marcação de plural com *-itá* é licenciada?
- iv. Os substantivos necessariamente possuíveis possuem uma forma absoluta?

A GrammYEP obteve resultados bastante satisfatórios tanto na análise quanto na tradução de sentenças das três línguas. Traduziu para o português e o inglês todas as 142 sentenças do conjunto-teste positivo do nheengatu. Inversamente, verteu para o nheengatu 98,18% e 84,11% dos conjuntos-teste correspondentes do português e do inglês, com 220 e 214 sentenças, respectivamente. A partir do conjunto-teste do nheengatu e das traduções do português e do inglês constituiu-se um *treebank* do

nheengatu com 243 sentenças, emparelhadas com as equivalentes nas duas outras línguas.

O sistema ainda padece de hipergeração, problema a ser enfrentado nas próximas versões. Nesse quesito, o *nheengatu* representa o maior desafio, por conta de particularidades lexicais e sintáticas sem equivalência nas outras duas línguas. A primeira dessas dificuldades são os substantivos necessariamente possuíveis, cujo complemento genitivo precisa sempre realizar-se, ver (32). A segunda são os termos designativos de descendentes imediatos e parentes colaterais de 2º grau, que lexicalizam o gênero do parente.

Dado seu caráter de software livre, a GrammYEP oferece diversas oportunidades de colaboração por parte da comunidade de linguistas e programadores em GF visando à solução dos problemas apontados e à expansão do sistema.

### Agradecimentos

Agradecemos aos dois pareceristas anônimos pelos comentários e sugestões.

### Referências

ÁVILA, M. T. *Estudo e prática da tradução da obra infantil A terra dos meninos pelados, de Graciliano Ramos, do português para o dheengatu*. 2016. 199f. Dissertação (Mestrado em Estudos da Tradução) – FFLCH, Universidade de São Paulo, São Paulo, 2016.

BENDER, E. M. Grammar Engineering for Linguistic Hypothesis Testing. In: GAYLORD, N. et al. (org.). *The Proceedings of the Texas Linguistics Society 10: Computational Linguistics for Less-Studied Languages*. Stanford: CSLI, 2008. p. 16-36.

BENDER, E. M. Reweaving a Grammar for Wambaya. *Linguistic Issues in Language Technology*, Stanford, v. 3, n. 3, p. 1-36, 2010.

BENTLEY, D. Copular and Existential Constructions. In: DUFTER, A.; STARK, E. (org.). *Manual of Romance Morphosyntax and Syntax*. Berlin: De Gruyter, 2017. p. 332-366. DOI: <https://doi.org/10.1515/9783110377088-009>

BERNSTEIN, J. B. The DP Hypothesis: Identifying Clausal Properties in the Nominal Domain. In: BALTIN, M.; COLLINS, C. (org.). *The Handbook of Contemporary Syntactic Theory*. Malden: Blackwell, 2003. p. 536-561. DOI: <https://doi.org/10.1111/b.9781405102537.2003.00019.x>

BRESNAN, J. *Lexical-Functional Syntax*. Malden: Blackwell, 2001.

BUSSMANN, H. (org.). *Lexikon der Sprachwissenschaft*. 3. ed. Stuttgart: Kröner, 2002.

CARNIE, A. *Syntax: A Generative Introduction*. 3. ed. Malden: Blackwell, 2012.

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

COMBINING Demonstrative and Possessive Pronoun. [S.l.]: [S.n.], 2019. Disponível em: <https://english.stackexchange.com/questions/476384/combining-demonstrative-and-possessive-pronoun>. Acesso em: 19 jun. 2020.

COMRIE, B. *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell, 1983.

CRUZ, A. *Fonologia e gramática do nheengatú: a língua falada pelos povos Baré, Warekena e Baniwa*. Utrecht: LOT, 2011.

CRUZ, A. The Rise of Number Agreement in Nheengatu. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas*, Belém, v. 10, n. 2, p. 419-439, 2015. DOI: <https://doi.org/10.1590/1981-81222015000200011>

DUCHIER, D.; PARMENTIER, Y. High-level Methodologies for Grammar Engineering, Introduction to the Special Issue. *Journal of Language Modelling*, Warszawa, Poland, v. 3, n. 1, p. 5-19, 2015. DOI: <https://doi.org/10.15398/jlm.v3i1.117>

EBERHARD, D. M.; SIMONS, G. F.; FENNIG, C. D. (org.). *Ethnologue: Languages of the World*. 23. ed. Dallas: SIL International, 2020. Disponível em: <http://www.ethnologue.com>. Acesso em: 12 jun. 2020.

FÁBREGAS, A. Adjectival and Genitival Modification. In: DUFTER, A.; STARK, E. (org.). *Manual of Romance Morphosyntax and Syntax*. Berlin: De Gruyter, 2017. p. 771-803. DOI: <https://doi.org/10.1515/9783110377088-021>

FRANCEZ, N.; WINTNER, S. *Unification Grammars*. Cambridge: CUP, 2012.

FREIRE, J. R. B. *Rio Babel: a história das línguas na Amazônia*. 2. ed. Rio de Janeiro: EdUERJ, 2011.

GYNAN, S. Morphological Glossing Conventions for the Representation of Paraguayan Guaraní. In: ESTIGARRIBIA, B.; PINTA, J. (org.). *Guarani Linguistics in the 21st Century*. Leiden: Brill, 2017. p. 86-130.

HAJIČOVÁ, E. et al. Treebank Annotation. In: INDURKHYA, N.; DAMERAU, F. J. (org.). *Handbook of Natural Language Processing*. 2. ed. Boca Raton: Chapman & Hall/CRC, 2010. p. 167-188.

JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2. ed. Upper Saddle River: Prentice Hall, 2009.

KARVOVSKAYA, L. *The Typology and Formal Semantics of Adnominal Possession*. Utrecht: LOT, 2018.

LEANDRO, W. M.; AMARAL, L. A. The Interpretation of Multiple Embedded Genitive Constructions by Wapichana and English Speakers. *Revista Linguística*, Rio de Janeiro, v. 10, n. 2, p. 149-162, 2014.

LJUNGLÖF, P.; WIRÉN, M. Syntactic Parsing. In: INDURKHYA, N.; DAMERAU, F. J. (org.). *Handbook of Natural Language Processing*. 2. ed. Boca Raton: Chapman & Hall/CRC, 2010. p. 59-91.

LYONS, J. *Linguistic Semantics: An Introduction*. Cambridge: CUP, 1995. DOI: <https://doi.org/10.1017/CBO9780511810213>

MATHESIOUS, V. *A Functional Analysis of Present Day English on a General Linguistic Basis*. Haia: Mouton, 1975. DOI: <https://doi.org/10.1515/9783110813296>

MÜLLER, S. The CoreGram Project: Theoretical Linguistics, Theory Development and Verification. *Journal of Language Modelling*, Warszawa, Poland, v. 3, n. 1, p. 21-86, 2015. DOI: <https://doi.org/10.15398/jlm.v3i1.91>

NAVARRO, E. A. *Curso de Língua Geral (nheengatu ou tupi moderno): a língua das origens da civilização amazônica*. São Bernardo do Campo: Paym, 2011.

NAVARRO, E. 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://doi.org/10.35572/rlr.v6i2.768>

PIRINEN, T. *et al.* Introduction. In: INTERNATIONAL WORKSHOP FOR COMPUTATIONAL LINGUISTICS OF URALIC LANGUAGES, 3., 2017, St. Petersburg. *Proceedings* [...]. Stroudsburg, USA: Association for Computational Linguistics, 2017. p. iii.

PRAÇA, W. N.; MAGALHÃES, M. M. S.; CRUZ, A. Indicativo II da família Tupi-Guaraní: uma questão de modo? *Liames*, Campinas, v. 17, n. 1, p. 39-58, 2017. DOI: <https://doi.org/10.20396/liames.v17i1.8646480>

RANTA, A. *Grammatical Framework Tutorial*. [S.l.]: [S.n.], 2010. Disponível em: <https://www.grammaticalframework.org/doc/tutorial/gf-tutorial.html>. Acesso em: 15 jun. 2020.

RANTA, A. *Grammatical Framework: Programming with Multilingual Grammars*. Stanford: CSLI, 2011.

RODRIGUES, A. D. As línguas gerais sul-americanas. *Papia*, São Paulo, v. 4, n. 2, p. 6-18, 1996.

RODRIGUES, A. D. Prefácio. In: FREIRE, J. R. B. *Rio Babel: a história das línguas na Amazônia*. 2. ed. Rio de Janeiro: EdUERJ, 2011. p. 13-14.

SAG, I. A.; WASOW, T.; BENDER, E. *Syntactic Theory: A Formal Introduction*. 2. ed. Stanford: CSLI, 2003.

SYMPSON, P. L. *Grammatica da lingua brazílica geral, fallada pelos aborigines das provincias do Pará e Amazonas*. Manaus: Typographia do Commercio do Amazonas, 1877.

ZOMPI, S. *Case Decomposition Meets Dependent-Case Theories*. 2017. 108 f. Dissertação (Mestrado em Linguística) – Corso di Laurea Magistrale in Linguistica, Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Pisa, 2017. Disponível em: <https://ling.auf.net/lingbuzz/003421>. Acesso em: 14 jun. 2020.





## **Garimpendo palavras: produção e circulação de conhecimento no século XIX a partir do gesto de autoria de Alencar**

### ***Digging for Words: the production and circulation of knowledge in the 19th Century from the gestures of Alencar***

Vanise Medeiros

Universidade Federal Fluminense (UFF), Niterói, Rio de Janeiro / Brasil

vanisegm@yahoo.com.br

<https://orcid.org/0000-0002-6998-9377>

Thais Costa

Universidade Federal Fluminense (UFF), Niterói, Rio de Janeiro / Brasil

Universidade Estadual de Campinas (Unicamp), Campinas, São Paulo / Brasil

araujo\_thais@yahoo.com.br

<https://orcid.org/0000-0002-8599-3528>

Raphael Mendes

Universidade Federal Fluminense (UFF), Niterói, Rio de Janeiro / Brasil

raphael\_mendes@live.com

<http://orcid.org/0000-0002-7256-0707>

**Resumo:** Com este artigo, que se ancora na Análise de Discurso materialista na sua relação com a História das Ideias Linguísticas, partimos de um conjunto de notas produzidas por Alencar para os romances *Iracema* e *O Guarani* as quais tinham como característica comum a referenciação explicitada de suas fontes. Nossa proposta foi a de ir em busca de um horizonte de retrospecto projetado nas notas, a fim de refletir sobre a (re)produção e a circulação do conhecimento a partir das notas. Para a análise das relações entre as notas e suas fontes, lançamos mão ainda do aparato teórico-analítico de Authier-Revuz, o que nos permitiu observar o jogo que aí se tece entre o que se atribui a si e ao outro em suas diferentes formas e funcionamentos. O percurso de análise empreendido nos permitiu observar os gestos de autoria do escritor sobre a língua.

**Palavras-chave:** notas de José de Alencar; horizonte de retrospectção; autoria; História das Ideias Linguísticas; Análise de Discurso materialista.

**Abstract:** With this article, which is anchored in the Analysis of Materialist Discourse and in its relation with the History of Linguistic Ideas, we started from a set of notes produced by Alencar for the novels *Iracema* and *O Guarani* which had as their common characteristic the explicit referencing of their sources. Our proposal was to go in search of a retrospection horizon projected in the notes in order to reflect on the (re) production and circulation of knowledge by the notes. For the analysis of the relationships between the notes and their sources, we also used Authier-Revuz's theoretical-analytical apparatus, which allowed us to observe the dynamic that is woven between what is attributed to you and the other in its different forms and functions. The analysis path undertaken allowed us to observe the author's gestures about the language.

**Keywords:** notes by José de Alencar; horizon of retrospection; authorship; History of Linguistic Ideas; Materialist Discourse Analysis.

Recebido em 15 de outubro de 2020

Aceito em 28 de dezembro de 2020

Observar a história dos discursos, os percursos que eles realizam, leva a compreender a produção do conhecimento.

(NUNES, 2008, p. 87)

## 1 Primeiras palavras

É longo e complexo o processo de dicionarização brasileiro. Somente nas décadas de 30 e 40 do século XX teremos no Brasil o aparecimento dos primeiros dicionários apresentados como brasileiros (NUNES, 2002, p. 110). Tal processo não se fez sem, por exemplo, dicionários parciais produzidos no século XIX por brasileiros e publicados no Brasil,<sup>1</sup> tampouco sem vocabulários e glossários que lhe foram contemporâneos ou anteriores. As inúmeras e extensas notas de José de Alencar (1829-1877), comportando saberes sobre as línguas

---

<sup>1</sup> Por dicionários parciais, Nunes (2006, p. 205) indica pequenos dicionários monolíngues, como os de regionalismos, caso de Coruja (1853). São, ainda, dicionários ou vocabulários que “complementam os dicionários portugueses” (NUNES, 2010, p. 10), caso de Costa Rubim (1853).

em solo brasileiro ao lado de saberes outros vários sobre as gentes, os costumes, as terras, as culturas, a fauna e a flora, se inscrevem, de forma saliente e distinta, neste processo de produção dicionarística. Elas resgatavam, produziam e faziam circular saberes postos às margens de romances, transitando, assim, por espaços nos quais muitas vezes instrumentos linguísticos<sup>2</sup> não compareciam.<sup>3</sup> No caso das notas aqui em foco, a saber, aquelas dos romances *O Guarani* (1857) e *Iracema* (1865)<sup>4</sup> em que comparecem representações de dizeres (de) outros, se faz ainda notar sua inscrição em uma das primeiras práticas lexicográficas produzidas no Brasil, qual seja, a das

listas de palavras português-Tupi e Tupi-português, como listas de nomes de plantas e animais, de partes do corpo humano, de objetos da cultura indígena, dentre outras. Essas listas de palavras deram origem aos primeiros dicionários brasileiros, que foram dicionários bilíngues português-Tupi elaborados por missionários jesuítas dos séculos XVI ao XVIII. (NUNES, 2010, p. 9.)

Em poucas palavras, uma prática que se inscreve na produção de dicionários bilíngues – institucionalizada no Brasil Imperial (NUNES, 2006) – e que, alinhando-a ao projeto literário romântico, Alencar faz significar em um momento de descolonização linguística<sup>5</sup> ao dizer sobre

---

<sup>2</sup> Conforme Auroux (1992), “instrumentos linguísticos são tecnologias que descrevem e instrumentalizam uma língua e prolongam a ‘competência’ do locutor, dando acesso respectivamente a um corpo de regras e a um conjunto de palavras que um mesmo locutor não domina completamente” (AUROUX, 1992, p. 69, aspas do autor).

<sup>3</sup> Ademais, não se pode deixar de considerar o valor performativo de tais saberes inscritos em um romance, isto é, de fazer saber tais saberes, funcionando de forma a torná-los evidentes.

<sup>4</sup> Nessas obras, as notas comparecem como uma espécie de apêndice, são notas de fim encontradas, em *Iracema*, após o romance e antes do posfácio e, em *O Guarani*, após cada uma das três partes do romance, fazendo, em ambos, referência à página em que o termo de entrada apareceu pela primeira vez.

<sup>5</sup> Orlandi (2009, p. 172) define o processo de descolonização linguística como o imaginário no qual se dá um acontecimento linguístico “sustentado no fato de que a língua faz sentido em relação a sujeitos não mais submetidos a um poder que impõe uma língua sobre sujeitos de uma outra sociedade, de um outro Estado, de uma outra Nação”. Trata-se, em outras palavras, do processo por meio do qual a língua passa a ser significada como “signo de nacionalidade, ou seja, em sua relação com a nação” (ORLANDI; GUIMARÃES, 2001, p. 24) e, portanto, com os sujeitos a ela identificados.

a chamada cor local. Nas notas de tais romances, línguas indígenas estão, sobremaneira, em foco, diferentemente do que ocorre em outros romances do autor, como *Diva* ou *O Gaúcho*.

O objetivo geral deste artigo é, então, à luz da Análise de Discurso materialista na sua relação com a História das Ideias Linguísticas, compreender o funcionamento do gesto de autoria de Alencar no que concerne à mobilização de dizeres (de) outros, especialmente, sobre aquilo que os românticos chamaram de cor local, tendo como *corpus* as notas de Alencar nos romances aludidos. Com isso, a partir da descrição do horizonte de retrospecto<sup>6</sup> projetado pelas notas inscritas nesse nome de autor, buscaremos também, em última análise, refletir sobre a (re)produção, circulação e formulação do conhecimento (científico) no/do Brasil no século XIX. Para tanto, com vistas a mapear as redes de filiações significativas em que se inscreve ao se significar/ser significado como autor de romances românticos indianistas e como intelectual brasileiro, buscaremos responder às seguintes perguntas: (i) que tipos de alteridade se fazem presentes nas notas de rodapé dos romances de Alencar?; (ii) a quem o autor romântico dá a palavra ou que autores e, por conseguinte, que saberes (não) podem e (não) devem em suas obras comparecer?; (iii) como se dá em seu dizer a representação do dizer (do) outro?; (iv) o que esse comparecimento nos diz sobre as condições de (re)produção, circulação e formulação do conhecimento (científico) no/do Brasil no século XIX?; e (v) como funciona nas notas o gesto de autoria da posição sujeito lexicógrafo que nelas se inscreve? Com efeito, é necessário dizer que tal empreendimento não se deu sem um trabalho de garimpo sobre o gesto de garimpo de um grande pesquisador de gabinete, como era conhecido Alencar.<sup>7</sup>

## 2 Do arquivo e da garimpagem

Ler as notas de Alencar e ir em busca dos livros indicados por Alencar, perseguir as fontes, confrontar as partes referidas com aquelas

---

<sup>6</sup> Este conceito, tomando como referência Aurox (1992), diz respeito ao fato de que o conhecimento guarda necessariamente uma relação com a temporalidade. Ou seja, há a construção de uma certa memória na produção do conhecimento, uma vez que “todo conhecimento é uma realidade histórica” e que “sem memória e sem projeto, simplesmente não há saber.” (AUROUX, 1992, p. 11-12).

<sup>7</sup> As notas deixam entrever uma longa e árdua pesquisa de Alencar em documentos e fontes antigos. Daí a metáfora do garimpo no gabinete.

indicadas nas notas. Essa foi a garimpagem empreendida com esta pesquisa a partir das pistas nas notas de Alencar. Há ainda um outro gesto de garimpagem sobre o qual o nosso se deteve: o gesto de garimpar palavras – eis uma imagem-síntese do trabalho que, da posição sujeito lexicógrafo pesquisador, empreendeu Alencar, ao se debruçar sobre línguas e sobre línguas indígenas brasileiras. As notas de rodapé que estão dispostas em *O Guarani* e *Iracema* são tomadas como verbetes (MEDEIROS, 2019b), uma vez que funcionam como glossários e se constituem, desta maneira, como instrumento linguístico. Neles se assume que a produção de saberes sobre a língua pelo escritor afeta o imaginário de língua. Dizem sobre a língua, etimologia, fauna, flora e os mais diversos saberes sobre a configuração brasileira no século XIX.

O processo de seleção, leitura e recorte, denominado por nós como “garimpo” dos documentos que constituem o arquivo sobre o qual nos debruçamos neste artigo se deu em três etapas, a saber: a escavação do solo, a purificação do minério e a transformação em barra. Em analogia com a garimpagem, a primeira etapa consiste no uso do maquinário para extração do minério ainda em estado bruto do solo, fazendo escavações e adentrando as suas mais diversas camadas. Em um segundo momento, é realizado um trabalho de separação dos cascalhos e impurezas, adicionando mercúrio para formar uma liga com o material extraído, a fim de levá-lo ao fogo de modo que reste somente o metal precioso. Por fim, esse material é transformado em barras para que possa circular e ser vendido.

No que tange à “escavação do solo”, vemos, por exemplo, que, em suas vastas notas, o autor realiza gestos de busca, de mergulho em fontes, que traduz, cita e faz referências para validar o seu posicionamento. Assim, fez-se preciso ir às fontes indicadas.<sup>8</sup> Muitas, diversas, em várias línguas (francês e latim são duas delas). Foi necessário buscar a produção bibliográfica de cada autor e ir mapeando livro por livro indicado nas notas de Alencar a fim de encontrar a referida publicação. Um trabalho de escavação<sup>9</sup> para chegar aos títulos originais referidos por Alencar. Além disso, algumas entradas das notas, que estamos tomando como verbetes,

---

<sup>8</sup> E vale lembrar que o acesso a livros antigos não é tão simples, o que demandou uma busca em bibliotecas físicas e virtuais de obras em vários idiomas.

<sup>9</sup> Por muitas vezes, Alencar traduz o nome do livro, mesmo que ele não tenha sido publicado no Brasil, o que faz com que alguns dados sejam perdidos por conta da tradução ou da ocultação de algum termo importante, por exemplo.

também foram traduzidas, como é o caso de *Carbeto*, ou possuíam ainda grafias diferentes, caso de *Ibiapaba*. É preciso tempo para escavar, tempo para apurar, tempo para conferir, tempo para saber. Tempo da pesquisa.

No segundo momento, a purificação do minério – teste de fogo – significa procurar as primeiras edições dos materiais ou a versão citada pelo autor. O nosso objetivo era encontrar o fragmento da mesma edição utilizada pelo autor. Em um caso específico, porém, só foi possível achar uma reedição. Isso é importante porque, em versões outras ou mais recentes, muitas palavras, expressões e notas de rodapé são modificadas substancialmente e fazem com que se percam informações, dados e nuances que outrora compareciam em edições anteriores, afetando a (re) produção dos sentidos e de seus efeitos. Uma vez que investigamos o horizonte de retrospectão, debruçarmo-nos sobre os exteriores referidos pelo autor e usarmos o mesmo verbete em que se baseou para fazer as análises nos permite mapear e confrontar dizeres, considerando as condições de produção que determinam o dizer de Alencar.

Para terminar o garimpo, é o momento de transformar o metal refinado em barras. Ou seja, é momento de colocar o metal em circulação. No nosso caso, significa realizar um segundo e adensado gesto analítico sobre esses verbetes, ao observar rupturas e regularidades, ao enveredar, no nosso caso, pelo funcionamento no jogo de representação do dizer (do) outro, ao perseguir o gesto de autoria e, não menos importante, ao refletir sobre a produção e circulação dos conhecimentos a partir de Alencar.

### 3 Sobre cor local e alteridade

“É visível (...) o desejo do autor de lutar pela formação de uma consciência nacional, pela busca da chamada ‘cor local’”, afirma Citelli (1990, p. 52) sobre Alencar. Mas o que estaria em jogo quando se diz “cor local”? É a Alencar que recorreremos para responder a essa pergunta. Em “Benção paterna”, prefácio de *Sonhos d’Ouro*, diz o autor:

Onde não se propaga com rapidez a luz da civilização, que de repente cambia a cor local, encontra-se ainda em sua **pureza original**, sem mescla, esse viver singelo de nossos pais, tradições, costumes e linguagem, com um sainete todo brasileiro. Há, não somente no país, como nas grandes cidades, até mesmo na corte, desses recantos, que guardam intacto, ou quase, o passado. (ALENCAR, [1872] s/d, p. 21, negrito nosso.)

Em *Iracema*, é possível observar, conforme podemos ler no prefácio supracitado, o que seria a tal “pureza original”, “primitiva”, isto é, anterior ao contato com outros povos, notadamente europeus. Em *O Guarani*, temos um momento posterior em que se dá “o consórcio do povo invasor com a terra americana, que dele recebia a cultura, e lhe retribuía nos eflúvios de sua natureza virgem e nas reverberações de um solo esplêndido” (ALENCAR, 1857, p. 20). É desse contato com outras nacionalidades, dessa “luta entre o espírito conterrâneo e a invasão estrangeira” (ALENCAR, 1857, p. 23), que se forma, nas palavras de Alencar, “a nova e grande nacionalidade brasileira” (ALENCAR, 1857, p. 22). Nesse processo, a independência política, associada à “cor local”, é concebida, como podemos ler no posfácio à 2ª edição de *Iracema*, como um fator determinante:

Quando povos de uma raça habitam a mesma região, a independência política só por si forma sua individualidade. Mas se esses povos vivem em continentes distintos, sob climas diferentes, não se rompem unicamente os vínculos políticos, opera-se, também, a separação nas ideias, nos sentimentos, nos costumes, e, portanto, na língua, que é a expressão desses fatos morais e sociais. (ALENCAR, [1870] s/d, p. 114-115.)

Especificamente sobre a questão linguística – que se coloca nesse imaginário como, diremos, uma questão de descolonização linguística –, considera ainda Alencar residir no “espírito popular” o germe da tendência incontestavelmente existente no Brasil, “não para a formação de uma nova língua, mas para a transformação profunda do idioma de Portugal” (ALENCAR, [1870] s/d., p. 114). Daí concluir o prefácio de *Sonhos d’Ouro* com a seguinte questão: “O povo que chupa o caju, a manga, o cambucá e a jabuticaba, pode falar uma língua com igual pronúncia e o mesmo espírito do povo que sorve o figo, a pera, o damasco e a nêspera?” (ALENCAR, [1872] s/d, p. 28).

Posto isso, ousamos afirmar que “cor local” diz daquilo que, do lugar do escritor romântico, considera-se constituir a nacionalidade brasileira. Tal nacionalidade, no entanto, ainda em processo de formação, é delineada na relação e no confronto com outras nacionalidades, colocando, assim, em jogo a diferença, diferença esta estabelecida nas fronteiras – sempre porosas – entre o interior e o seu exterior constitutivo, no batimento e no atravessamento entre o nacional e o estrangeiro, entre o um e o outro.

Que todo dizer é atravessado por outros dizeres, por algo que fala “antes, em outro lugar e independentemente”, nos ensina Pêcheux (2009, p. 149), lembrando-nos ainda de que as evidências a partir das quais se produz o efeito de que “todo mundo sabe” e, por conseguinte, a ilusão de transparência e de completude da linguagem são produtos do funcionamento ideológico. Mas as evidências, no que concerne à identidade nacional, ainda não estão postas quando se trata de uma “sociedade nascente”, de “povos não feitos”, para jogarmos com expressões de Alencar no prefácio supracitado, fazendo-se, pois, necessário dizer sobre o que se toma por seus traços distintivos para significá-la.

Como anunciamos anteriormente, neste artigo, nos propomos a refletir sobre as notas<sup>10</sup> presentes em dois romances de Alencar, *Iracema* e *O Guarani*, ambos publicados no século XIX. Trazemos para esta reflexão, todavia, não quaisquer notas, mas, na esteira de Authier-Revuz (1990) no tocante à relação entre heterogeneidade constitutiva e heterogeneidade mostrada, especificamente aquelas que, funcionando como glossários, representam de diferentes formas dizeres (de) outros. Sobre o funcionamento das notas no período em questão, convém recordarmos o que pontua Medeiros (2020) especificamente sobre as notas de rodapé, e que é também pertinente em relação às notas de fim aqui analisadas. Diz a autora:

Os rodapés falam muito no século XIX. Falam nos jornais, com seus folhetins que se dirigiam a um público vasto, falam nos livros de história, nos romances, nas poesias. Falam, falam, falam... tagarelam, discutem, debatem, argumentam, contestam, bravejam, citam, questionam, divulgam, alardeiam, instauram, acolhem e repelem ideias, propostas, opiniões. As notas atordoam, revelam, omitem, trapaceiam. Espaço de agitado de vozes e silêncios ruidosos. Espaço outro de escrita e de leitura. (MEDEIROS, 2020, p. 213.)

As notas-glossário sobre as quais aqui nos debruçamos não são diferentes. Filiam-se ao imaginário que se institui com o projeto literário romântico e contribuem para a constituição disso que se toma

---

<sup>10</sup> As notas, presença quase constante nos romances do século XIX, podem comparecer ao pé de página ou ao final.

por identidade nacional. Assim, ao dizerem sobre a “cor local”, como pontuamos anteriormente, dizem sobre línguas indígenas e sobre outras línguas com as quais se relaciona a língua no/do Brasil, dizem sobre a fauna e a flora brasileira, sobre os rituais e os costumes indígenas e sobre o que dizem outros enunciadores, brasileiros e estrangeiros, a partir do lugar daquilo que à época se tinha por ciência. É, portanto, tendo em vista essa tagarelice que consideramos que as notas, ao mesmo tempo que funcionam como sintoma da não transparência e completude da linguagem e, portanto, como lugares de dispersão dos sujeitos e dos sentidos, buscam contornar esse funcionamento, como explica Medeiros (2016, p. 261), a partir da leitura de Authier-Revuz, promovendo “a demarcação da alteridade (...) dentro e às margens do texto”.

Alteridade é uma noção importante que assume diferentes concepções, conforme o terreno teórico onde se situa. Em Authier-Revuz,<sup>11</sup> ela é pensada e articulada a partir de uma heterogeneidade fundante que constitui o sujeito, efeito da linguagem. Uma heterogeneidade que lhe é constitutiva, a qual não se tem acesso e com a qual se lida no complexo jogo da denegação e da representação de uma heterogeneidade outra: a heterogeneidade mostrada, que pode ser ou não marcada no dizer. Esta, mostrada, se instaura no jogo imaginário – sujeito a falhas e tropeços – que se estabelece entre um dizer que se considera do outro e um que se considera de si. Um jogo que trabalha a relação com o outro, que tece uma imagem do outro e que denuncia uma imagem de si, o que pode se dar pelas várias formas de representação do discurso outro (RDA, conforme AUTHIER-REVUZ (2020), aqui traduzida por RDO) e/ou ainda pelas outras várias formas de modalização autonímica, denunciando as não coincidências do dizer.<sup>12</sup> É no que concerne à heterogeneidade mostrada, no que ela traz do encontro e do trabalho com o que vai sendo posto como do outro (em minúsculo, diferentemente do Outro<sup>13</sup> da heterogeneidade constitutiva) que podemos tecer uma relação com a alteridade de que fala Aurox (1992).

---

<sup>11</sup> São várias os textos de Authier-Revuz nos quais nos apoiamos, conforme se lê na bibliografia.

<sup>12</sup> São quatro as não coincidências propostas por Authier-Revuz (1998, 2000): interlocutiva, interdiscursiva, entre as palavras e as coisas e das palavras consigo mesmas.

<sup>13</sup> Outro, não localizável e não representável, “onde estão em jogo o interdiscurso e a ideologia” (AUTHIER-REVUZ, 1990, p. 32).

Conforme este autor e considerando aqui nosso objeto, os glossários nascem de uma alteridade que lhes antecede: texto outro ou língua outra, por exemplo, que se deseja desvendar, conhecer, apreender. A alteridade diz respeito, então, em Auroux (1992), ao outro que lhe antecede,<sup>14</sup> ou seja, ao outro, tecido no imaginário da atividade linguageira. Com efeito, lançar mão do aparato teórico de Authier-Revuz colabora para compreensão de como vai sendo costurada na língua uma relação que delimita os espaços do outro. Neste sentido, não podemos nos esquecer de que ir em busca do horizonte de retrospectção significa percorrer as pegadas que antecedem e constituem a tal ou qual saber linguístico. E resulta, com isso, em mergulhar nas relações entre os dizeres e nas formas como comparecem no dizer. As reflexões de Authier-Revuz, em seus diversos textos, nos permitem ainda ir além, uma vez que possibilitam observar as imagens de si neste tenso jogo entre o que se atribui a si e ao outro, ou seja, ajudam, na análise do nosso *corpus*, a refletir sobre o lugar de autoria em se fazendo na relação com os outros dizeres.

#### 4 O jogo da representação do dizer (do) outro

“Toda enunciação, então, se inscreve deliberadamente ou inocentemente em um horizonte de completude”<sup>15</sup> – trazemos de Flauhaut (1978, p. 101) para lembrar o intrincado jogo que se tece na enunciação na ilusão de uma separação entre, lançando mão de Authier-Revuz (1990), um dizer que seria de si de um dizer que seria do outro e que, assim, sustenta a ilusão de um domínio de seu dizer. Marcar, indicar, discriminar o que advém de outra enunciação, então, parte deste mecanismo que caminha de braços dados com o horizonte de uma (in)completude, ainda que se indique o não-tudo dizer. Se faz parte do labor do historiador ou do pesquisador, por exemplo, não se pode não observar que fez parte do trabalho de Alencar também em seus romances e, portanto, do seu gesto de autoria. Ao ler as notas de Alencar que, com suas diversas instâncias de referenciação, nos revelam, como dissemos, uma densa atividade de leitura e de garimpagem, deparamo-nos com a história das leituras (re)produzidas pela posição-

---

<sup>14</sup> E sobre a distinção da noção de alteridade em Auroux e em Authier-Revuz, cf. Medeiros (2019a).

<sup>15</sup> Do original: “Tout énonciation, donc, s’inscrit délibérément ou naïvement dans un horizon de complétude”.

sujeito leitor inscrita sob o nome Alencar em se fazendo significar na história do dizer (re)produzido a partir da posição-sujeito autor.<sup>16</sup> Aqui privilegiamos o conjunto dessas notas que se debruça sobre a língua, compondo um glossário, embora não nos tenhamos limitado somente a ele. Para nossa análise, de uma seleção de vinte e nove verbetes em que comparece uma representação explícita do dizer (do) outro, isto é, em que se encontram formas mostradas e marcadas no fio discursivo, e referidas, ou seja, a cuja exterioridade se faz remissão,<sup>17</sup> procedemos à análise sobre um *corpus* de exemplificação, conforme Mazière,<sup>18</sup> constituído por cinco verbetes que portam especificidades na relação com a alteridade. São, pois, as seguintes voltas sobre palavras, expressões, sintagmas em notas, situadas ao final das partes do livro ou mesmo ao final do livro, que constituem as entradas/verbetes do que sinalizamos como glossário de Alencar: *carbeto*, *ibypaba*, *inubia*, *cabuiba* e *saio de algodão*. As três primeiras de *Iracema*; as demais de *O Guarani*.

Citações, discurso indireto, aspas sobre palavras ou sintagmas, discurso segundo, modalização autonímica, denunciam trajetos de leitura, trajetos de formulação e de circulação de saberes. São marcas enunciativas que dão a saber das fontes e do modo como com elas se relaciona a posição-sujeito lexicógrafo que se inscreve nas notas. É preciso lembrar que as notas nestes romances funcionam também como um importante espaço de ligação entre o que se indica como fonte e o que e como nelas se discursiviza. Atravessadas por uma heterogeneidade que se apresenta e se representa em suas formulações, elas iluminam um diálogo, diríamos, de pares, na medida em que são trabalhadas seja no tocante a uma aquiescência dos sentidos – que não se dá sem reflexão e sem inscrição de autoria –, seja no que tange a uma discordância, seja ainda nos cortes e deslocamentos que promove.

---

<sup>16</sup> Costa (2019) nos fala de uma relação indissociável entre gesto de leitura e gesto de autoria. Segundo a autora, “o gesto de autoria, a significação do sujeito em autor, é determinado não só pela memória do dizer, mas também por esta na sua relação com a memória social de leitura (NUNES, 1998), colocada em funcionamento quando da significação do sujeito em leitor, e, por conseguinte, pela história das leituras (ORLANDI, 2007) por este (re)produzidas” (COSTA, 2019, p. 86).

<sup>17</sup> Estamos considerando aquelas em que são indicadas as fontes no corpo das notas.

<sup>18</sup> Trata-se de uma noção feita por Francine Mazière em conversa com Vanise Medeiros que diz respeito à montagem de um *corpus* para pensar tanto questões teóricas quanto para iluminar e refletir sobre determinados aspectos concernentes ao objeto em estudo.

Nas notas fala-se *com*, fala-se *de* e fala-se *sobre* as palavras do outro, para lançarmos mão de algumas das especificações que Authier nos possibilita ao mapear, exaustivamente, as formas de heterogeneidade marcada. No caso das notas em Alencar aqui recortadas em função de referirem a um horizonte de retrospecção, o que se destaca sobremaneira é um trabalho que se situa no campo da representação do discurso outro (RDO, cf. AUTHIER-REVUZ, 1999, 2001, 2020), e, por vezes, de modalização autonímica. Em geral, de uma modalização que faz fronteira com a RDO (cf. AUTHIER-REVUZ, 1999), qual seja, aquela que diz respeito à não coincidência do discurso consigo mesmo,<sup>19</sup> isto é, que assinala para a presença de outros discursos. Mas não apenas. São notas que vão em busca de fontes de outros que aqui estiveram e se lançaram à tarefa de compreensão do que para eles, europeus, se apresentava como um mundo a descrever, a compreender. Neste sentido, nelas comparecem, com efeito, a não coincidência entre as palavras e as coisas; uma não coincidência que decorre do gesto de nomeação.

Nosso caminho de análise será, como anunciamos, um estudo de cinco casos. Passemos ao primeiro

#### 4.1 Carбето

Começemos lendo a nota de Alencar, em *Iracema*, e, em seguida, o fragmento extraído (no trabalho de garimpagem que resgata do texto em francês e pela palavra *carbet* aquela outra, *carbeto*, que se acha em Alencar) da fonte à qual a nota se refere, a saber, o livro *Viagem ao norte do Brasil* de Ives D'Evreux, um naturalista e religioso que percorreu o norte do Brasil no século XVII:

II.—*Carбето*.—Especie de serão que fazião os indios á noite em uma cabana maior, onde todos se reunião para conversar. Leia-se Ives D'Evreux: *Viagem ao norte do Brasil*.

Fonte: Alencar (1865, p. 186).

---

<sup>19</sup> Cf Authier-Revuz (1999), a modalização autonímica não pertence ao quadro da RDO, mas há uma relação de interseção (e não de inclusão) entre tais formas. É neste espaço de fronteira que entram as formas de não coincidência do discurso consigo mesmo, ou ainda, como ela denomina, da não coincidência interdiscursiva.

Chacun l'environnoit pour l'escouter quand il alloit au Carbet. p. 55.

C'est un Tabajara qui parle, mais nous ferons observer que le mot *Carbet* n'appartient pas à la *lingoa geral*. Le P. Ruiz de Montoya ne l'a pas inséré dans son précieux *Tesoro de la lingua Guarani*. Il est plus particulièrement en usage parmi les Galibis et d'autres peuples de la Guyane. Le voisinage de notre colonie se fait sentir dans le récit du P. Yves, rien que par cette expression. Il faut faire une certaine différence entre les Carbets et les *Ocas* ou *Tabas*, qui constituaient l'architecture rudimentaire des autres peuples du Brésil. Écoutons à ce sujet le P. du Tertre: „Au milieu de toutes ces cases, ils en font une grande commune qu'ils appellent *Carbet*, laquelle a toujours 60 ou 80 pieds de longueur et est composée de grandes fourches hautes de 18 ou 20 pieds, plantés en terre. Ils posent sur ces fourches un latanier ou un autre arbre fort droit qui sert de faît, sur lequel ils ajustent des chevrons qui viennent toucher la terre, et les courent de roseaux ou de fuësilles de latanier, de sorte qu'il fait fort obscur dans ces Carbets, car il n'y entre aucune clarté que par la porte, qui est si basse qu'on ne sauroit y entrer sans se courber.“

Les détails que nous venons de donner ici sont empruntés à un ouvrage qui date de l'année 1643, et ils se rapportent plus spécialement à l'architecture rustique des Caraïbes insulaires. Nous avons choisi cet exemple à peu près contemporain du livre publié par notre auteur, parce que il n'y avait pas en réalité de notables différences entre les Carbets des îles et ceux du continent. Si l'on faisait une histoire de ces cases de feuillage si promptement élevées, on pourrait en constater néanmoins certaines variétés, selon les usages auxquels on les destinait. (Voy. à ce sujet, *le voyage pittoresque au Brésil de Debret*, puis les gravures du livre d'André Thevet, publ. en 1558.) Il y avait les petits et les grands Carbets, ceux où les Piayes faisaient leurs jongleries, et ceux où se tenaient les grands conseils. Ces derniers affectaient la forme d'un de nos vastes hangars, et pouvaient contenir jusqu'à 150 ou 200 guerriers. Au XVII<sup>e</sup> siècle, dans le langage de nos colonies, parmi les îles ou sur le continent, tenir un conseil quelconque, c'était *Carbeter*; le terme était consacré et se trouve dans tous les voyageurs. (Voy. entre autres Biet, *Voyage de la France équinoxiale*. Paris, 1654, in-4.)

Fonte: D'Evreux (1864, p. 406-407).

As fontes em Alencar percorrem o universo científico da época, no qual os naturalistas têm destaque ao lado de historiadores, e denunciam a posição-sujeito pesquisador. Não são obtidas por um trabalho de contato com indígenas, como ocorre, por exemplo, com Taunay, mas nos livros vários em que mergulha e referencia.<sup>20</sup>

<sup>20</sup> Daí uma de suas alcunhas: pesquisador de gabinete.

Longo trecho no livro em nota de Ferdinand Denis,<sup>21</sup> um importante historiador do século XIX, no livro de D'Évreux, no qual nos deparamos com camadas arqueológicas de circulação de dizeres e de atribuição de sentidos a um léxico posto como advindo de língua indígena. Fala atribuída a um tabajara, recusa do pertencimento de tal léxico a uma língua geral, a explicação é dada da posição-sujeito de saber do europeu. Se o indígena fala (*C'est un Tabajara qui parle*), é ao europeu que devemos escutar: *Écoutons à ce sujet le P. de Tertre* introduz uma longa citação acerca do que se chama *carbet*; léxico, portanto, capturado, grafado e significado em francês, língua europeia. Gestos significativos de silenciamento do outro nas formas de atribuição de dizer (do) outro (no caso, do indígena); afinal a voz a ser escutada é aquela posta no lugar do saber europeu.

Neste fragmento se assinala para uma construção diferenciada das ocas e das tabas, se descreve como seriam os *carbets*, se indica a circulação dos termos, no século XVII, nas colônias francesas e nos relatos dos viajantes e se define o verbo *carbeter*: *tenir un conseil quelconque, c'était Carbeter*. Se confrontarmos com o que lemos na nota de *Iracema*, de imediato, observamos o gesto de autoria sobre a fonte. Uma volta ao substantivo, agora para o léxico posto em língua portuguesa (*carbeto*), um enunciado definidor que acena não mais para uma construção mas para um *serão* (substantivo) e uma *reunião para conversar*. É interessante observar na introdução do verbete a expressão *Espécie de*. Por um lado, *espécie de* nos joga num universo de outras possibilidades, ou melhor, nos remete para um conjunto de outros termos possíveis agasalhados por um hiperônimo, inscrevendo assim um pré-construído.<sup>22</sup> Por outro lado, *espécie de* não deixa de trabalhar uma imprecisão, uma falta de justeza. É como se tateasse no dizer, no léxico trazido de um universo que não aquele indígena, o que se pode notar no movimento parafrástico em busca de precisão – *serão à noite, reunião para conversar*. Em outros termos, posto que uma definição funciona como glosa à palavra sobre a qual se volta, sobre esta podemos dizer que se inscreve uma não coincidência

<sup>21</sup> A capa do livro de Yves D'Évreux já indica que, ao final, constam notas elaboradas por Ferdinand Denis (1798-1890), historiador francês responsável por elaborar diversos estudos sobre a América do Sul e, em especial, o Brasil.

<sup>22</sup> A partir do qual o que se toma por “cor local” (aqui a língua e a cultura indígena) é delineado pela relação com o outro (a língua e a cultura europeia).

entre as palavras e as coisas. Falha e falta que dizem respeito também ao descompasso entre línguas no processo de nomeação.

Por fim, duas breves observações acerca do enunciado introdutor da fonte na nota: *Leia-se Ives D'Evreux: Viagem ao norte do Brasil*. Estamos diante de uma formulação que, seguindo Authier-Revuz (2001, p. 199), evoca um outro ato enunciativo sem informação sobre aquilo que neste outro lugar é enunciado. Tal remissão ocorre ainda após a definição do verbete. Estas são duas regularidades nas notas dos dois romances em foco que trabalham o lugar do outro como caucionando seu dizer. Em outras palavras, a referenciação a outros autores não encabeça a predicação relativa à palavra em foco, como vemos em algumas poucas notas;<sup>23</sup> ao contrário, o que encontramos com certa regularidade é uma fórmula (nome de entrada: definição + fonte) no qual uma enunciação alhures funciona como abonação ao que se diz e que vai de passo com o gesto de autoria que se apropria de outras enunciações como forma de atestação de seu dizer. Trata-se de um funcionamento comum ao discurso científico. Indo adiante, diríamos que as notas se inscrevem em uma formação discursiva científica que vai significar sujeitos e língua, para ficarmos com dois pontos importantes nas notas destes romances. Não podemos, neste sentido, não observar que glossários são instrumentos linguísticos que, como tal, servem à (re)produção e à circulação de conhecimento.

## 4.2 Ibyapaba

Uma nota encabeçada por palavra indígena, como tantas em ambos os romances, nas quais encontramos o gesto de decomposição da palavra e de busca de sua etimologia (cf. MEDEIROS, 2019b), um gesto que inscreve numa mesma prática científica lexicógrafos e naturalistas, com suas descrições e decomposições.

---

<sup>23</sup> Caso de *Quixeramobim*, que se inicia com “Segundo Dr. Martius, traduz-se...” (*Iracema*), ou ainda, para dar outro dos poucos exemplos que se tem, caso de *Curaré* (*O Guarani*), que começa com uma longa citação em francês.

Pap. 9.—*Ibyapaba*.—Grande serra que se prolonga ao norte da provincia e a extrema com Piauhy. Significa terra aparada. O Dr. Martius em seu *glossario* lhe attribue outra athmologia. *Iby*-terra — e *pabe* — tudo. A primeira porém tem a authoridade de Vieira.

Fonte: Alencar (1865, p. 166).

Definição que segue um padrão clássico (NUNES, 2003), em que se tem nome de entrada (*Ibyapaba*), nome cabeça (*Grande serra*) seguido de adjetiva (*que se prolonga ao norte da provincia e a extrema com Piauhy*). Após tal definição segue-se outra (*Significa terra aparada*) e, em seguida, marcam-se discursos (de) outros. Tal funcionamento, isto é, definição inicial seguida de discursos (de) outros, que se faz notar não apenas nesta nota-verbete,<sup>24</sup> se inscreve em um fazer lexicográfico que separa definição de comentários, citações ou mesmo de exemplos. Estes, quando comparecem (e cabe lembrar que nosso *corpus* reside sobre tal comparecimento), funcionam, grosso modo, como acréscimos posteriores, o que sinaliza para uma posição-sujeito lexicógrafo na qual a definição é produzida sob o efeito de um dizer monológico que trabalha o efeito da evidência: X é isto. Nesta, contudo, o que se segue são definição outra e conflito de interpretações posto entre Dr. Martius e Vieira, com a opção por Vieira. Vejamos Dr. Martius:

*Hibiapaba, Ipiapaba* (Ceará, Cordilheira) — *iby* terra, *pabe* tudo.  
Terreno descoberto. Omne terra.

Fonte: Martius (1867, p. 501)<sup>25</sup>

De Dr. Martius, traz-se parte de sua definição (*iby* terra, *pabe* tudo) para refutá-la em prol de outra indicada como de Vieira,<sup>26</sup> que não se mostra mas se faz supor como tendo sido *terra aparada*. É interessante notar que “significa terra aparada” também comparece de

<sup>24</sup> Visto também em carbetto, por exemplo.

<sup>25</sup> Dr. Martius (1794-1868) foi um importante médico, botânico e explorador alemão que realizou uma expedição no Brasil com intuito de documentar a fauna e flora desta região, sendo a *Flora Brasiliensis* sua obra mais famosa.

<sup>26</sup> Não foi possível obter esta fonte, a despeito de inúmeras buscas.

forma definidora monologicamente. É ao seguir na leitura da nota que nos voltamos e a equivocamos: de onde provém tal definição? Da fonte em Vieira ou da posição lexicográfica das notas? Mais uma vez o gesto de autoria se faz notar no corte a partir da escrita outra, na equivocação das vozes na segunda definição e no deslocamento de *terra aparada*, em que se diz um trabalho sobre a terra, para a nomeação de um topônimo. Eis-nos diante de outro equívoco: quem nomeia tal serra?

### 4.3 Cabuiba

Recortar, da cadeia falada de línguas desconhecidas, e desenhar, a partir de tais recortes, palavras; grafar em uma língua estranha àquela que se escuta; impor a letra escrita a uma língua falada, eis o movimento que desde o século XVI viajantes, jesuítas, naturalistas, entre outros, efetuaram sobre línguas indígenas. Alteridade na escuta; alteridade na língua; alteridade no processo de objetivação da língua; alteridade que se inscreve na carne da palavra ao grafá-la. Alteridade que se impôs, no que se refere aos habitantes destas terras antes dos europeus, às suas línguas, às suas culturas. Alteridade que se impõe também em direção outra, isto é, em direção ao que seriam língua e sujeitos nacionais, incidindo sobre outras porções de língua falada em solo brasileiro. “A existência da escrita transforma profundamente o estatuto da fala humana”, nos lembra Auroux a partir de Platão (AUROUX, 1998, p. 69), isto para agudizar o gesto que incide sobre o que vai se promovendo e indicando como língua nacional; gesto do qual as notas são testemunhas, escritas. Observemos *cabuiba*.

**Cabuiba.**— A cabuiba ou cabureiba—*Balsamum Peruvianum* de Pison, *cabuiba iba* Marcgrave e *Miroxilem Cabriuva* de outros naturalistas — é uma arvore das nossas mattas de mais de cem palmos, e a que vulgarmente se chama arvore do balsamo.

Destilla um licor louro de um cheiro agradável, que dizem milagroso para cura de feridas frescas. (Gabriel Soares, B. Lisboa e Ayres de Casal).

Fonte: Alencar (1857, s.p.)

“O que define uma língua é a soma de seus equívocos, sobretudo quando eles não são frutos do acaso, mas sim fundados (...) na longa história desta língua”,<sup>27</sup> como nos fala Cassin (2012, p. 28). Em *cabuiba*, o equívoco da apreensão se lê na alternativa que inicia a predicação do verbete – *cabuiba* ou *cabureiba*. Um *ou* que não interroga, mas que abre para as duas possibilidades na relação de forças com aquilo que se nomeia – no caso, indicado como um já nomeado alhures – *cabuiba* ou *cabureiba* – e que vai ser renomeado, agora pela posição cientista naturalista na língua de sua ciência,<sup>28</sup> isto é, em latim – *balsamum peruvianum* e *miroxilem cabriuva*. Língua de valor na nomeação da natureza, esta vem em destaque, *itálico*; não corre no fio discursivo sem marcas que a iluminam em sua diferença, como se dá com *cabuiba*, *cabureiba* ou ainda como árvore do bálsamo. O *itálico* denuncia seu lugar outro.

Se na nomeação incide uma não coincidência entre as palavras e as coisas, como sabemos com Authier-Revuz (1997), aqui, neste verbete, se encontra, além desta, uma outra não coincidência, aquela do discurso consigo mesmo, “afetado pelo jogo em si mesmo de outros discursos” (AUTHIER-REVUZ, 1997, p. 259), que se faz ver na incisa que se dobra, como volta metaenunciativa, sobre a nomeação a partir de línguas indígenas – “*cabuiba* ou *cabureiba*” – trazendo a nomeação produzida do lugar de um discurso outro, aquele da ciência em latim. Neste verbete, a nomeação não se esgota na oposição entre duas posições discursivas postas a partir de língua indígena – *cabuiba*, *cabureiba* – e de língua latina – *balsamum peruvianum* e *miroxilem cabriuva*. Ela resgata ainda o que seria uma outra ordem de nomeação: “a que vulgarmente se chama árvore do bálsamo”, uma posição discursiva que remete para circulação do termo no território brasileiro. É esta nomeação na relação sinonímica com *cabureiba* que se lê nas referências feitas na nota de Gabriel Soares:

---

<sup>27</sup> Do original: “Ce qui définit une langue, c’est la somme de ses équivoques, surtout quand elles ne sont pas le fruit d’un hasard, mais qu’elles sont fondées, (...) dans la longue histoire de cette langue”.

<sup>28</sup> Não podemos esquecer que as ciências também instauram suas línguas; e elas mudam, como sabemos.

Não se podiam arrumar em outra parte que melhor estivessem as arvores de virtude que apoz das que dão fruto; e seja a primeira arvore do balsamo que se chama *cabureiba*; que são arvores mui grandes de que se fazem eixos para engenhos, cuja madeira é pardaça e incorruptivel. Quando lavram esta madeira cheira a rua toda a balsamo, e todas as vezes que se queima cheira muito bem. D'esta arvore se tira o balsamo suavissimo, dando-lhe piques até um certo lugar, donde começa de chorar este suavissimo licor na mesma hora, o qual se recolhe em algodões, que lhe mettem nos golpes; e como estão bem molhados do balsamo, os espremem em uma prensa, onde lhe tiram este licor, que é grosso e da côr do arroze; o qual é milagroso para curar feridas frescas, e para tirar os sinacs d'ellas no rosto. O carunchio d'este pão, que se cria no lugar donde sahiu o balsamo, é preciosissimo no cheiro; e amassa-se com o mesmo balsamo, e fazem d'esta massa contas, que depois de seccas ficam de maravilhoso cheiro.

Fonte: Souza (1851, p. 195-196.)

Uma nomeação que trabalha a transparência do nome pelas suas propriedades – fazer cheirar a rua toda a bálsamo e ser milagrosa para curar feridas frescas – e que, em Aires do Casal ([1817] s/d),<sup>29</sup> agasalhada na categoria de “árvores resinentas”, recebe outros nomes, postos em itálico: *Cabureigba* – com o qual mais uma vez nos expõe a variação na língua indígena ou a equivocação na sua escuta –, e ainda *bálsamo-do-Espírito Santo*, no qual o determinativo (Espírito Santo) abre para outras espécies de árvores com tais qualidades.

Entre as árvores resinentas, notam-se o *Ángico*, a que produz a goma copai, a da almêcega, a do beijoim, a do estoraque; e entre as que destilam bálsamo, nomeiam-se em primeiro lugar a do *Cabureigba*, mais conhecido pelo nome de *bálsamo-do-Espírito Santo*, a do *Cupaiba*, a do *Cumaru*.

Fonte: Casal ([1817] s.d., p. 58.)<sup>30</sup>

<sup>29</sup> Manuel Aires de Casal (1754-1821) foi um padre português, dedicou-se a estudar a corografia (geografia e história) brasileira e foi o primeiro a transcrever, impressa, a carta feita por Pero Vaz de Caminha.

<sup>30</sup> Não foi possível encontrar essa obra em sua 1ª edição. Na internet, encontramos apenas a edição disponibilizada pela biblioteca digital da PUC-Campinas, disponível no endereço: <http://bibliotecadigital.puc-campinas.edu.br/services/e-books/Aires%20de%20Casal-1.pdf>. Acesso em: 10 set. 2020.

Uma flutuação na nomeação que também se evidencia em Bento Lisboa (*Balsamum ex-Peré, Ancabureiba sive Balsamum Peruvim, Pison, Cabuiba Iba, Mascgrave*) e nos assinala um tatear sobre o mundo que se abria aos europeus e que urgia nomear, compreender.

58° *Balsamum ex-Perú Ancabureiba sive Balsamum Peruvianum, Pison = Cabuiba Iba, Mascgrave* ; he huma arvore de alta grandeza de oitenta a cem palmos, e mais de comprimento, de dous a seis e mais de grossura , cuja casca he cinzenta , grossa , manchada como de pontos ferrugineos , que contém hum licor louro; ferida na Lua cheia de Fevereiro e Março distila esse oleo conhecido por *Balsamo do Perú* : he macio no lavar ; serve para obras que se confundem com o molgamo : a resina dá o cheiro de pastilhas.

Fonte: Lisboa (1834, Tomo I, p. 214)<sup>31</sup>

Nunes nos lembra que:

Ao se deparar com as coisas do Novo Mundo, o viajante não tem as palavras adequadas para descrevê-las. A realidade exige que o discurso se acomode a ela. Em tais condições, os referentes encontram-se em um outro lugar, a que poucos têm acesso. Por conseguinte, há um trabalho de nomeação em que se negocia a adequação das palavras às coisas. Ao admitir que não há essa adequação, essa coincidência, e distante da diversidade referencial que se apresenta, o viajante constrói um espaço de nomeação por meio de um trabalho que vai em direção à coincidência.

Segundo Authier-Revuz, na adequação entre as palavras e as coisas jogam duas forças: a força da língua e a força das coisas. (NUNES, 1994, p. 129)

As fontes em Gabriel Soares de Souza (1851), Aires Casal ([1817] s.d.), e Bento Lisboa (1834), nos jogam diante desta necessidade de nomeação e apreensão do que se descortinava.

<sup>31</sup> Bento da Silva Lisboa (1793-1864), nascido no estado da Bahia, foi um diplomata brasileiro e um dos membros fundadores do Instituto Histórico e Geográfico Brasileiro – a mais antiga instituição de preservação da história do Brasil, fundada em 1838.

Voltando à nota em Alencar, nela lemos sobre sua característica (altura) e suas propriedades, já citadas, as quais se seguem parênteses que nos remetem às fontes. Tal como em *carbeto*, evocam-se, pelo nome, outros atos enunciativos que simulam um “eu falo pelo outro” na ilusão da reprodução de um dizer outro. Um funcionamento que vai comparecer no discurso da ciência, assim como no discurso dicionarístico para ficarmos com dois campos por onde transitam as notas produzidas pelo autor Alencar nestes romances.

#### 4.4 Saio de algodão

Lembramos, a partir de Authier-Revuz (2020, p. 253), com *cabuiba*, que, além da não coincidência entre a palavra e a coisa, que denuncia a dimensão de uma “perda”, de uma “falha” inerente ao ato de nomear, na nomeação pode estar em jogo ainda uma outra não coincidência, qual seja, a interdiscursiva. Esta nos impele a reformular aqui a questão que levantamos com *Ibyapaba*: afinal, quem tem o direito de nesse imaginário nomear?

Há singularidades que distinguem este verbete recortado de *O Guarani* dos demais e que devemos de imediato anunciar. *Saio de algodão* não é uma palavra, mas um sintagma. *Saio de algodão* não traz à baila uma outra língua marcada pela nota como corpo estranho à escrita em língua portuguesa. Posto isso, vamos ao verbete:

PAG. 19.

**Saio de algodão.** — Referem os chronistas que muitas tribus indias fiavão o algodão para vestir-se, fazer redes e outros objectos. No « Diccionario da Lingua Brasilica » encontramos a palavra « guarina » significando « camizas, gibão. » Isto nos autorizou a apresentar um selvagem assim trajado, sem faltar em nada á verdade; devendo-se notar que os Goytacazes erão uma das nações mais industriosas.

Fonte: Alencar (1857, notas, p. I.)

Entrada – comentário – definição/tradução – comentário. É esse o percurso com o qual nos deparamos no verbete em análise, percurso a partir do qual o sintagma de entrada é primeiramente decomposto ([saio] /de/ [algodão]) para em seguida ser recomposto ([saio de algodão]), reiterando a ilusão de transparência da linguagem e das línguas.

O primeiro comentário volta-se sobre o significante *algodão*. Nele, tem-se, por meio do discurso indireto, uma representação de um discurso (de) outro (RDO) (AUTHIER-REVUZ, 2020), a saber, o dos *chronistas*, viajantes que se aventuravam nas terras desconhecidas e cujos relatos tinham no século XIX, como dito anteriormente, *status* científico.<sup>32</sup> A despeito de a alteridade ser aqui referida de forma genérica, o que nos impede de ir à fonte, interessa-nos o que com ela está em jogo no plano constitutivo: o discurso da história sobre a cultura indígena – discurso esse representado, não só como evidente e unísono (*como é sabido, todos os cronistas se referem...*), mas também como generalizante, uma vez que apaga a alteridade das chamadas *tribus índias* (*muitas tribos fiavam algodão para...*)

A definição, por sua vez, volta-se sobre *saio*. Nela, mobilizando-se um exterior identificado, qual seja, o *Diccionario da língua brasilica*,<sup>33</sup> estabelece-se uma equivalência entre esse significante e *guarina*, palavra que, segundo o dicionário citado, significaria *camizas, gibão*. Tem-se aqui, diferentemente do que ocorreu no primeiro comentário, uma modalização autonímica sob a forma de ilha textual (AUTHIER-REVUZ, 2000, 2001). Em Alencar, ao se dizer sobre a língua indígena, não se diz apenas *sobre* as palavras do outro, o lexicógrafo-autor do *Diccionario da língua brasilica*, mas, filiando-se ao discurso dicionarístico, diz-se *com* palavras outras.

O discurso indireto, ao simular a tradução de um dizer (de) outro, tem como efeito a produção de um imaginário de fidelidade não às palavras, mas ao conteúdo relatado. As ilhas textuais – pontos de alteridade delimitados pelas aspas no fio do dizer que remetem a uma exterioridade, o discurso dicionarístico – produzem um efeito de objetividade e de literalidade dos fragmentos do dizer convocado. Em ambos os casos, o dizer mobilizado diz do lugar da ciência. Em ambos os acasos, a mobilização desses dizeres tem como efeito a atribuição de legitimidade ao dizer de Alencar sobre o índio – legitimidade esta

---

<sup>32</sup> Segundo Orlandi (2008), no século XIX vê-se um grande número de reedições de relatos de missionários e viajantes europeus sobre o Brasil. Esses relatos, como demonstra a autora em suas análises, uma vez que se apresentam, ao mesmo tempo, como lugar de conhecimento e como manifestação literária, constituem-se como uma forma de produção carregada de ambiguidade na qual verdade e mentira se tensionam.

<sup>33</sup> Língua brasilica ou língua geral era o nome por meio do qual era genericamente chamado no início da colonização, conforme Mariani (2004), o Tupinambá ou Tupi antigo, língua falada na costa brasileira.

reiterada no comentário final no qual o sintagma *saio de algodão* é, por fim, recomposto com a conclusão de um raciocínio silogístico constituído não por duas, mas por três premissas,<sup>34</sup> como demonstramos a seguir:

Premissa A → *Referem os chronistas que muitas tribus indias fiavão o algodão para vestir-se...*

Premissa B → *No “Diccionario da Lingua Brazilica”, encontramos a palavra “guarina” significando “camizas, gibão”.*

Premissa C → *... os Goytacazes erão uma das nações mais industriosas.*

Conclusão → *Logo... isto nos autorisou a apresentar um selvagem [goitacá]<sup>35</sup> assim trajado [isto é, de saio de algodão], sem faltar em nada á verdade.*

Algumas observações aqui devem ser feitas. Em primeiro lugar, não foi possível encontrar as fontes citadas por Alencar nesse verbete. No caso dos relatos dos viajantes, porque, como dito, a referência se dá de forma genérica. Quanto ao *Diccionario da Lingua Brazilica*, encontramos com esse nome um manuscrito anônimo tupi-português do século XVIII,<sup>36</sup> mas nele não há o verbete *guarina*. Imaginando que talvez tivesse havido um equívoco em relação ao nome da obra citada, ampliamos nossa investigação, buscando o verbete *guarina* também em outros vocabulários e dicionários em circulação nas condições de produção em questão. No *Vocabulário da Língua Brasileira*, manuscrito anônimo português-tupi

<sup>34</sup> Nessas premissas, o interdiscurso se faz significar de diferentes formas. Na A e na B, como vimos, com a identificação de uma alteridade, genérica ou específica, está em jogo a mobilização do dizer da ciência. Na premissa C, diferentemente, uma vez que não há identificação de uma alteridade que responderia como fonte do dizer, tem-se a irrupção de um pré-construído que se coloca como uma evidência na qual opera a ilusão de transparência da linguagem: *todo mundo sabe que os Goytacazes eram uma das nações mais industriosas*.

<sup>35</sup> Trata-se de Peri. É a primeira vez em que ele aparece na Segunda Parte do romance para, heroicamente, salvar Cecília da morte: “De pé, fortemente apoiado sobre o respaldo estreito que formava a rocha, um selvagem coberto com um ligeiro saio de algodão, mettia o horabro á uma lasca de pedra que se desencravára do seu alveolo, e ia rolar pela encosta” (ALENCAR, 1857, p. 19).

<sup>36</sup> Disponível em: [https://bdigital.sib.uc.pt/bg3/UCBG-Ms-94/UCBG-Ms-94\\_item1/index.html](https://bdigital.sib.uc.pt/bg3/UCBG-Ms-94/UCBG-Ms-94_item1/index.html). Acesso em: 23 set. 2020.

produzido pelos jesuítas do século XVI-XVII,<sup>37</sup> não há também registro dessa palavra. Ela comparece, todavia, no *Dicionário da língua geral no Brasil*, de 1771, como sinônimo de *casaca* e, ao lado de *goarina* e *sayo* ou *saio*, em outros dicionários do português também produzidos por europeus no século XVIII e início do século XIX (BARROS; LESSA, [1771] 2015).

Em Bluteau (1712/1720, p. 436), *sayo* aparece como sinônimo de *saia*. Não há o registro da palavra *guarina*, apenas o de *goarina*, definida nos seguintes termos: “Era huma roupetona que dava pelos juelhos, fechada por todas as partes, & aberta por diante, a moda das que trazem em algumas partes os Carneiros” (BLUTEAU, 1712/1720, p. 85). Na reedição ampliada desse dicionário publicada por Silva, em 1789, passa a comparecer, para *saio*, antecedendo ao significado de *saia*, o de “vestidura antiga, espécie de roupa larga, ou casacão usado na guerra; e depois na paz dos cavalheiros” (BLUTEAU; SILVA, 1789b, p. 367). A definição de *goarina* comparece de forma mais enxuta – “roupeta aberta por diante, que dava pelo juelho” (BLUTEAU; SILVA, 1789a, p. 661) – e é incluída a entrada *guarina* – “túnica militar curta” (BLUTEAU; SILVA, 1789a, p. 674). Na reedição de 1813 desse dicionário, assinada somente por Silva, todos os verbetes comparecem sem alteração, salvo o acréscimo, no final da definição de *goarina*, da seguinte indicação: “melhor *guarina*” (SILVA, 1813, p. 90).

Três breves comentários sobre esses comparecimentos se fazem necessários. Em primeiro lugar, eles atestam que *saio*, *sayo*, *goarina*, *guarina* eram significantes em circulação quando da formulação de *O Guarani*. Em segundo lugar, é interessante observar como esses comparecimentos mostram, por meio do funcionamento parafrástico, o deslocamento e a progressiva aproximação, a partir da filiação a uma formação discursiva militar, entre os sentidos de *saio*, casacão de guerra usado pelos cavalheiros europeus, e *guarina*, túnica militar curta. Em terceiro lugar, importa observar que essa circulação não se dá sem disputa, disputa entre o significante europeu e o que é posto como indígena, mas também entre as duas formas de se grafar este último, com preferência por uma delas, aquela cujo significado historicamente aproximou-se do associado ao significante europeu.

Lembremos que é o significante europeu, e não o indígena, que comparece no corpo do romance de Alencar. Lembremos ainda que o

<sup>37</sup> Disponível em: [http://purl.pt/28813/4/1-12470-v\\_PDF/1-12470-v\\_PDF\\_24-C-R0150/1-12470-v\\_0000\\_capa-cap\\_a\\_t24-C-R0150.pdf](http://purl.pt/28813/4/1-12470-v_PDF/1-12470-v_PDF_24-C-R0150/1-12470-v_0000_capa-cap_a_t24-C-R0150.pdf). Acesso em: 23 set. 2020.

significante dito indígena comparece à margem, após a segunda parte do romance, na nota de fim, fazendo funcionar, por meio da ilusão de possibilidade de tradução, em que os dois significantes são tomados como equivalentes, o imaginário de transparência das línguas e de correspondência entre os dois mundos. Na esteira de Mariani (2004, p. 30), consideramos que, com o estabelecimento de denominações, tomadas aqui enquanto produto do processo de construção discursiva de referentes, para o que se toma por cor local, “organiza-se uma taxionomia semântica a partir da representação linguística para os termos indígenas, misturados a termos provenientes do colonizador”, através da qual “sítios de significação são codificados em termos do domínio de pensamento” deste, e não do índio, que, em *O Guarani*, como vimos, se traça então à semelhança de um cavaleiro medieval. Em Alencar, essa taxionomia semântica, como tem demonstrado nossas análises, se dá por meio da mobilização do dizer da ciência. Esta, de acordo com Orlandi (2008, p. 67), ao lado da política e da religião, consiste num dos modos de “domesticar a diferença”. Por meio do conhecimento dito científico, apaga-se a identidade do índio enquanto cultura diferente e constitutiva da identidade nacional. Diz a autora:

A ciência torna o índio observável, compreensível, e sua cultura, legível; o indigenismo se torna administrável (...). Diremos, pois, que a compreensão amansa o conceito índio (...). Essa domesticação representa o processo pelo qual ele deixa de funcionar, com sua identidade, na constituição da consciência nacional. (ORLANDI, 2008, p. 57)

Sigamos com a análise de um último verbete.

#### 4.5 Inubia

Assim como vimos nos verbetes anteriores, em *inubia*, de *Iracema*, a nota-glossário nos diz da cultura indígena, notadamente de um artefato bélico. Leiamos a nota:

Pag. 42.—*Inubia*.—Trombeta de guerra. Os indígenas, segundo Lery, as tinham tão grandes que medião um deametro na abertura.

Fonte: Alencar (1865, p. 174.)

Também assim como vimos nos demais verbetes, um certo conhecimento sobre a “cor local” é dado a saber por meio da mobilização da voz de um estrangeiro, um viajante europeu, o pastor calvinista e missionário francês Jean de Léry, que esteve no Brasil no século XVI. Segundo Orlandi (2008), em Léry ([1578]1880), a função colonizadora e a catequese estão associadas a uma certa prática científica através da qual o índio, sua língua e sua cultura são observados e descritos. É, pois, mais uma vez o discurso da ciência que se faz significar aqui.

Nesse verbete, observamos pelo menos três voltas metaenunciativas. A nota é composta de duas partes: uma primeira em que há uma tradução da palavra dita indígena, corpo estranho que comparece alusivamente no romance e que funciona como entrada do verbete na nota,<sup>38</sup> e uma segunda em que o discurso outro é retomado, nos termos de Authier-Revuz (2000, 2001), por meio de uma modalização. A nota volta-se sobre a palavra de entrada recortada do corpo do romance. O comentário volta-se sobre a tradução – gesto de interpretação de Alencar sobre o dizer de Léry – legitimando-a. Ainda no comentário, o sintagma “segundo Lery” volta-se sobre o enunciado modalizando-o. O efeito desses enlaces estabelecidos no fio discursivo é a ilusão de desopacificação da palavra posta como indígena, ilusão esta produzida à medida que ocorre a construção discursiva do seu referente (MARIANI, 2004).

Mobilizado por meio da forma de modalização que Authier-Revuz (2000, 2001) nomeia como discurso segundo sobre o conteúdo, o dizer de Léry é tomado como empréstimo, como fonte do enunciado com o qual e a partir do qual Alencar, inscrevendo-se na posição escritor-lexicógrafo, diz sobre “as coisas do Brasil”. Em nosso gesto retrospectivo, com vistas a compreender o funcionamento da função-autor que organiza as notas-glossários aqui analisadas, buscamos a passagem de Léry ([1578]1880) em que o dizer se volta sobre o significante *inubia*:

---

<sup>38</sup> Sobre o funcionamento das notas, cabe ressaltar aqui, com Authier-Revuz (2011, p. 17), que, na relação estabelecida entre o corpo do romance e as notas, no caso de fim, faz-se significar, como está sendo possível observar a partir das nossas análises, “o jogo solidariamente interlocutivo e interdiscursivo da alusão”. De acordo com essa autora, as notas, repatriam a exterioridade do já-dito representado no dizer que havia comparecido no corpo do texto alusivamente “à convivência de um compartilhamento de memórias”.

...  
 nent : à fin d'avertir & tenir les autres en cervelle,  
 il y en a toujours quelques-uns, qui avec des cornets,  
qu'ils nomment Inubia, de la grosseur & longueur  
 d'une demie pique, mais par le bout d'embas large  
 d'environ demi pied comme un haubois, sonnent au  
 milieu des troupes. Mesmes aucuns ont des fifres

Fonte: Lery ([1578] 1880, p. 35, grifo nosso.)

Trata-se de uma passagem em que se diz sobre os armamentos terrestres e navais dos tupinambás. Nesse dizer, é pela ótica do estrangeiro que o então mundo desconhecido é significado. “Escrever sobre a nova terra – diz Mariani (2004, p. 66) sobre os relatos dos viajantes produzidos no século XVI – era uma forma de fazer o pensamento europeu apropriar-se dela através de suas próprias categorias”. Nesse movimento, o significante *inubia*, também corpo estranho na escrita francesa, é marcado, delimitado, com itálico: “avec des cornets, qu'ils nomment *Inubia*”.

Escrito por um francês para leitores, a princípio, franceses, no relato de Léry ([1578] 1880), o outro é o indígena. Assim, por meio de um trajeto centrífugo, do que se coloca como interior para o exterior, promove-se, por acréscimo da palavra indígena, conforme Authier-Revuz (2000, p. 344), “uma excursão em território estrangeiro”, que só é possível porque a nomeação está de antemão assegurada no interior, ainda que ilusoriamente. Manifesta-se, assim, no plano da representação, a não coincidência interlocutiva de que nos fala a autora – nós (europeus) x eles (indígenas) – e, no plano da constituição, por meio da mobilização da língua indígena como discurso associado, a não coincidência interdiscursiva – nós dizemos *cornet*, eles dizem *inubia*. As duas palavras, porém, também não coincidem: *inubia* não é exatamente *cornet*. Faz-se, pois, necessário dar continuidade à explicação, lançando mão de descrições – *de la grosseur & longueur d'une demie pique...* – e comparações – *comme un haubois*. Assim, partindo-se de uma semelhança – *cornets* –, busca-se tornar o diferente, o desconhecido – *inubia* – conhecido, familiar, e contornar, por meio da gestão dos sentidos, “a opacidade da nova terra” (MARIANI, 2004) e da(s) língua(s). A diferença é, pois, como pontua Mariani (2004, p. 71-72), colocada ao lado da semelhança “como uma espécie de complemento, de complementação de um sentido já configurado, ainda que vagamente”.

Há ainda um último ponto que gostaríamos de observar sobre o dizer de Léry ([1578] 1880). Assim como vimos em *carbeto*, também

aqui o significante da língua indígena, ouvido pelo viajante estrangeiro, é transcrito (ORLANDI, 2008), apreendido, numa língua escrita europeia, o francês. Conforme Mariani (2004, p. 78), essa apreensão, que tem como efeito a construção de “um simulacro” desse encontro com a oralidade da língua indígena, produz a transformação, quer fonética, quer semântica, das palavras e expressões dessa língua. Escrevê-las como soam é “trabalhar uma sua imagem fora da sua história, de seu modo de existência” – diz Orlandi (2008, p. 101). É apagar a alteridade, promovendo “a redução do outro ao um” (ORLANDI, 2008, p. 101). E é – faz preciso lembrar – por meio da língua francesa escrita, e não diretamente da língua indígena, que a palavra *inubia* e os sentidos a ela filiados a partir da posição do viajante europeu chegam a Alencar. Eis, pois, aí a contradição: “o brasileiro se cria pelo fato de fazer falarem os outros” (ORLANDI, 2008, p. 25).

Nessa formação imaginária em que circulam sentidos de nação e sujeitos nacionais, embora se diga do índio para dizer do brasileiro, o índio não tem voz. Se a ele se atribui uma fala, dele ela não é escutada. O índio não é incluído. O índio não é brasileiro (cf. LEAL, 2012).

## 5 Palavras finais

Lendo as notas de Alencar, garimpando palavras e perseguindo os seus trajetos de leitura, ao analisar o horizonte de retrospecto projetado nas notas de *Iracema* e *O Guarani*, pudemos observar o encontro entre o que estamos indicando como Alencar-leitor, pesquisador de gabinete, e como Alencar-autor, escritor romântico. Fez-se possível observar como a história das leituras produzidas pelo Alencar-leitor se fizeram significar em seu gesto de autoria, como sentidos filiados a diferentes posições discursivas são mobilizados e alinhavados no dizer sobre a língua, que é também um dizer sobre a chamada “cor local”.

Posição-sujeito lexicógrafo, posição-sujeito tradutor, posição-sujeito cientista foram algumas das posições que pudemos depreender com a análise das notas nos romances em cena. Posto em outros termos, podemos agora dizer que a posição lexicógrafo não é sem as duas outras e, indo adiante, diremos ainda que decorrem do gesto de autoria sobre a língua. Colombat, Fournier e Puech nos lembram que “a língua não está diante daquele que se propõe a descrevê-la ou a ‘reduzi-la em regras’ como um objeto bem delimitado que lhe seria suficiente delimitar e explorar” (2017, p. 166, aspas dos autores), e, continuando, as gramáticas e os dicionários

funcionam de modo a fabricar uma língua (COLOMBAT; FOURNIER; PUECH, 2017, p. 166). Mas não são apenas estes instrumentos linguísticos a funcionarem de tal modo. Glossários também o fazem e, considerando as condições de produção das notas no cenário brasileiro do século XIX (de ausência de dicionários, como indicado no início deste artigo; do lugar do escritor e de seu gesto sobre a língua; e ainda de circulação do romance – afinal, produção de conhecimento não se faz sem sua circulação), podemos arriscar nossa hipótese de que as notas de Alencar em seus romances também funcionam de modo senão a construir ao menos a colaborar na construção de uma língua que se quer nacional.

É já deveras conhecido um fragmento de Auroux e, ainda assim, precisamos retomá-lo aqui:

Todo conhecimento é uma realidade histórica, sendo que seu modo de existência real não é a atemporalidade ideal da ordem lógica do desfraldamento do verdadeiro, mas a temporalidade ramificada da constituição cotidiana do saber. Porque é limitado, o ato de saber possui, por definição, uma espessura temporal, um horizonte de retrospecção (Auroux, 1987b), assim como um horizonte de projeção. O saber (as instâncias que o fazem trabalhar) não destrói seu passado como se crê erroneamente com frequência; ele o organiza, o escolhe, o esquece, o imagina ou o idealiza, do mesmo modo que antecipa seu futuro sonhando-o enquanto o constrói. Sem memória e sem projeto, simplesmente não há saber. (AUROUX, 1992, p. 12)

Fomos em busca deste passado que se inscreve nas notas dos romances; fomos em busca do horizonte de retrospecção que se fazia registrar pelas fontes indicadas. E nos deparamos com o horizonte de projeção; afinal, o horizonte de retrospecção não se descola do horizonte de projeção. No caso, nos defrontamos com um projeto de nação que passa pela transmissão de saberes, com sua reelaboração, e pela forja da língua; daí o que assinalamos como gesto de autoria. Um gesto que não se dá sem outros dizeres e saberes; que não é sem o outro.

Nas notas, saltam aos olhos as inúmeras vozes que se entrelaçam no fio do discursivo. Para persegui-las e compreendê-las, lançamos mão ainda do aparato teórico de Authier-Revuz (1997, 1998, entre outros textos indicados na bibliografia). São, pois, heterogeneidades marcadas por meio da qual a heterogeneidade constitutiva de todo e qualquer dizer se permite representar, na ilusão fundante do sujeito de poder

separar o que é de si e o que é do outro. Trouxemos (e nos debruçamos analiticamente sobre) as diversas alteridades convocadas para construir (seja de forma a atestar ou a refutar) os saberes dispostos nas notas. É hora de falar do outro a partir do qual se fala mas que não se faz falar: o elemento indígena, que serve à “cor local”, que serve à construção de nação e de língua. Mas de cuja voz no dizer de Alencar só há pistas...

E isto porque, como foi possível depreender a partir de nossas análises, embora se diga do índio para dizer da chamada “cor local”, por meio da mobilização do discurso científico – sobretudo o produzido a partir de uma posição naturalista –, o elemento indígena é silenciado. Na “luta entre o espírito conterrâneo e a invasão estrangeira” (ALENCAR, 1857, p. 23) a partir da qual se diz formar “a nova e grande nacionalidade brasileira” (ALENCAR, 1857, p. 22), vence o elemento europeu. É, portanto, contraditoriamente, a partir do seu olhar – que é o olhar do colonizador, não podemos esquecer – que se diz sobre as coisas do Brasil, significando-as no seu domínio de pensamento e, com isso, apaziguando as diferenças à medida que o outro – o elemento indígena – é reduzido ao um.

### **Declaração de contribuição de cada autor**

A realização do artigo compreendeu busca de fontes, recorte, análise e escrita. Raphael Mendes procedeu ao primeiro levantamento e catalogação de fontes. Thaís Costa procedeu a um segundo levantamento de fontes em função do que já tínhamos lido e do que ainda nos interessava investigar. O cotejo e confirmação dos dados foram feitos por Vanise Medeiros e por Thaís Costa. A partir daí, nós três fizemos várias reuniões de leitura para seleção dos verbetes e recortes. Mapeamos pontos que nos interessavam atacar e fizemos roteiros de análise. Dividimos a escrita do artigo em algumas partes, e cada um começou a escrever uma parte, intervindo na escrita do outro quando necessário. Grosso modo, a Raphael Mendes coube a descrição da montagem do arquivo; a Thaís Costa e Vanise Medeiros, a análise dos verbetes. Este foi um trabalho tecido de fato a seis mãos.

### **Agradecimentos**

Vanise Medeiros agradece ao CNPq (PQ, Proc. 305428/2018-7) e à Faperj CNE (Proc. n.º E-26/211.459/2019) a concessão de bolsas de pesquisa. Raphael Mendes agradece ao CNPq (projeto 305428/2018-7) a concessão da bolsa de iniciação científica.

## Referências

- ALENCAR, J. de. Benção paterna (Prefácio). In: ALENCAR, J. de. *Sonhos d'Ouro*. São Paulo: E-book Brasil, [1872] s/d. Disponível em: <http://www.ebooksbrasil.org/adobeebook/sonhosdoro.pdf>. Acesso em: 11 set. 2020.
- ALENCAR, J. de. *Iracema*. Rio de Janeiro: Typ. de Viana & Filhos, 1865.
- ALENCAR, J. de. *O Guarani*. Rio de Janeiro: Empreza Nacional do Diario, 1857.
- ALENCAR, J. de. Posfácio. In: ALENCAR, J. de. *Iracema*: lenda do Ceará. 2. ed. [1870]. [s.l.]: Editora Klick, [s.d.].
- AUROUX, S. *A escrita*: Filosofia da Linguagem. Campinas: Unicamp, 1998.
- AUROUX, S. *A revolução tecnológica da gramatização*. Campinas: Unicamp, 1992.
- AUTHIER-REVUZ, *La Représentation du Discours Autre*. Principes pour une description. Berlim; Boston: De Gruyter, 2020. DOI: <https://doi.org/10.1515/9783110641226>
- AUTHIER-REVUZ, J. Dizer ao outro no já-dito: interferências de alteridades – interlocutivas e interdiscursiva – no coração do dizer. *Letras Hoje*, Porto Alegre, v. 46, n. 1, p. 6-20, 2011.
- AUTHIER-REVUZ, J. Modalisation par discours autre et bivocalité. In: TOMASSONE, R. (org.). *Une langue*: le français. Paris: Hachette, 2001. p. 199-201.
- AUTHIER-REVUZ, J. Duas palavras para uma coisa: trajetos de não-coincidência. *Universa*, Brasília, v. 8, n. 2, p. 333-359, 2000.
- AUTHIER-REVUZ, J. Algumas considerações sobre modalização autonímica e discurso outro. *Letras de hoje*, Porto Alegre, v. 34, n. 2, p. 7-30, 1999.
- AUTHIER-REVUZ, J. *Palavras incertas*: as não-coincidências do dizer. Campinas: Editora da Unicamp, 1998.

AUTHIER-REVUZ, J. Falta do dizer, dizer da falta: as palavras do silêncio. In: ORLANDI, E. P. (org.). *Gestos de leitura*. Campinas: Unicamp, 1997. p. 257-280.

AUTHIER-REVUZ, J. Heterogeneidade(s) Enunciativa(s). *Caderno de Estudos Linguísticos*, Campinas, v. 19, p. 24-42, 1990.

BARROS, C.; LESSA, A. L. (org.). *Dicionário da língua geral no Brasil (1771)*. Belém: MPEG, 2015.

BLUTEAU, R. *Vocabulario portuguez e latino*. Coimbra: Real Collegio das Artes da Companhia de Jesus, 1712/1720.

BLUTEAU, R.; SILVA, A. M. *Diccionario Língua Pôrtugueza*. Composto pelo padre d. Rafael Bluteau, reformado e accrescentado por Antônio de Moraes Silva natural do Rio de Janeiro. Tomo primeiro (A-K). Lisboa: Officina de Simão Thaddeo Ferreira, 1789a.

BLUTEAU, R.; SILVA, A. M. *Diccionario Língua Pôrtugueza*. Composto pelo padre d. Rafael Bluteau, reformado e accrescentado por Antônio de Moraes Silva natural do Rio de Janeiro. Tomo segundo (L-Z). Lisboa: Officina de Simão Thaddeo Ferreira, 1789b.

CASAL, M. A. de. *Corografia Brasília ou Relação Histórica e Geográfica do Reino do Brasil*. Rio de Janeiro: Imprensa Régia, [1817] [s.d.].

CASSIN, B. *Plus d'une langue*. Paris: Bayard Éditions, 2012.

CITELLI, A. *Romantismo*. 2. ed. São Paulo: Editora Ática, 1990.

COLOMBAT, B.; FOURNIER, J.M.; PUECH, C. *Uma história das Ideias Linguísticas*. São Paulo: Contexto, 2017.

COSTA, T. de A. *Discurso gramatical brasileiro: permanências e rupturas*. Campinas: Pontes, 2019.

D'EVREUX, Y. *Voyage dans le nord du Brésil*. Leipzig; Paris: Librairie A. Franck, 1864.

FLAHAUT, F. *La parole intermediaire*. Paris: Ed. Du Seuil, 1978. DOI: <https://doi.org/10.3917/ls.flaha.1978.01>

LEAL, M. S. P. *Raposa Serra do Sol no discurso político roraimense*. Boa Vista: Editora UFRR, 2012.

LÉRY, J. de. *Histoire d'un voyage faict en la terre du Brésil*. Paris: Alphonse Merre, [1578] 1880.

LISBOA, B. S. *Annaes do Rio de Janeiro*. Rio de Janeiro: Typ. Imp. e Const. de Seignot-Plancher e Ca, 1834.

MARIANI, B. *Colonização Linguística*. Campinas: Pontes, 2004.

MARTIUS, C. F. P. *Wörterammlung Brasilianischer Sprachen*. Glossaria linguarum Brasiliensium: Glossarios de diversas lingoas e dialeotos, que fallao os Indios no imperio do Brazil. Leipzig: Friedrich Fleischer, 1867.

MEDEIROS, V. Notas de rodapé. In: MEDEIROS, V. et al. (org.). *Almanaque e fragmentos: ecos do século XIX*. Campinas: Pontes Editores, 2020. p. 213-217

MEDEIROS, V. Os homens fazem... mas...: língua e sujeito: uma reflexão em três tempos. In: FERRARI, A.; SCHERER, A.; MARIANI, B.; CAMPOS, L. (org.). *Discurso, interlocuções e...* Caxias do Sul; Educ, 2019a. p 2017-222.

MEDEIROS, V. A retórica da mediação: dois momentos. *Linguagem em (Dis)curso*, Tubarão, v. 19, n. 2, p. 355-371, 2019b. DOI: <https://doi.org/10.1590/1982-4017-190208-8118>

MEDEIROS, V. Língua e sujeito na captura da palavra. In: NUNES, S. R.; SILVA, A. R. da; KARIM, J. M.; MOTTA, A. L. A. (org.). *Sujeito e memória: lugares constitutivos*. Campinas: Pontes, 2016. p. 255-270.

NUNES, J. H. Dicionários: história, leitura e produção. *Revista de Letras*, Brasília, v. 3, n. 1/2, p. 6-21, 2010.

NUNES, J. H. O discurso documental na História das Ideias Linguísticas e o caso dos dicionários. *ALFA*, São Paulo, v. 52, n. 1, p. 81-100, 2008.

NUNES, J. H. *Dicionários no Brasil*. Campinas: Pontes Editores; São Paulo, São José do Rio Preto: FAPESP, 2006.

NUNES, J. H. Definição lexicográfica e discurso. *Revista Língua e Instrumentos Linguísticos*, Campinas, n. 11, p. 9-30, 2003.

NUNES, J. H. Dicionarização no Brasil: condições e processos. In: NUNES, J. H.; PETTER, M. (org.). *História do saber lexical e constituição de um léxico brasileiro*. São Paulo: USP/FFLCH; Humanitas; Campinas: Pontes Editores, 2002.

NUNES, J. H. Aspectos da Forma Histórica do Leitor Brasileiro Na Atualidade. In: ORLANDI, Eni Puccinelli (org.). *A leitura e os leitores*. Campinas: Pontes, 1998.

NUNES, J. H. *Formação do Leitor Brasileiro*: imaginário da leitura no Brasil Colonial. Campinas: Unicamp, 1994.

ORLANDI, E. *Interpretação*: autoria, leitura e efeitos do trabalho simbólico. 5. ed. Campinas, SP: Pontes, 2007.

ORLANDI, E. *Língua brasileira e outras histórias*. Campinas: RG Editores, 2009.

ORLANDI, E. *Terra à vista – discurso do confronto*: velho e novo mundo. 2. ed. Campinas: Editora da Unicamp, 2008.

ORLANDI, E.; GUIMARÃES, E. Formação de um espaço de produção linguística: a gramática no Brasil. In: ORLANDI, E. (org.). *História das ideias linguísticas*: construção do saber metalinguístico e constituição da língua nacional. Campinas: Pontes; Cáceres: Unemat Editora, 2001. p. 21-38.

PÊCHEUX, M. *Semântica e discurso*: uma crítica à afirmação do óbvio [1975]. 4. ed. Trad. Eni Orlandi *et al.* Campinas: Editora da Unicamp, 2009.

SILVA, A. M. *Diccionario da Língua Portuguesa*. Recopilado dos Vocabulários impressos ate agora, e nesta segunda edição novamente emendado, e muito acrescentado. Lisboa: Typographia Laceroina, 1813.

SOUZA, G. S. de. *Tratado descriptivo do Brazil em 1587*. Rio de Janeiro: Typographia Universal de Laemmert, 1851.



## A família lexical de *usura*: um estudo etimológico e morfossemântico

### *The lexical family of usura: an etymological and morphosemantic study*

Matheus Pinto

Universidade Federal da Bahia (UFBA), Salvador, Bahia / Brasil

matheus.machadopinto@gmail.com

<https://orcid.org/0000-0002-4097-9570>

Mailson Lopes

Universidade Federal da Bahia (UFBA), Salvador, Bahia / Brasil

mailson.lopes@ufba.br

<https://orcid.org/0000-0003-3926-0494>

**Resumo:** O estudo ora apresentado presta-se a descrever a trajetória da família lexical de *usura* (constituída pelas formas *usura*, *usurar*, *usurário*, *usureiro*, *usurável*, *usurento*, *usuroso*, *usurador*, *zura*, *zuraco* e *usurariamente*) no percurso histórico da língua portuguesa. Para a constituição de seu *corpus* de análise e como embasamento para a explicação dos processos de empréstimo e neologia em abordagem histórico-diacrônica e histórico-comparativa, fincou-se na integração de dados dialetais e de publicações da *web* com uma grande variedade de bancos de dados em diversas línguas românicas e em inglês. Faz o exame etimológico desse grupo vocabular pela análise gradual das transformações semânticas e morfológicas de cada unidade que o constitui, fornecendo evidência textual dos fenômenos discutidos. Definindo-se em um eixo interfacial entre Morfologia, Semântica e Léxico, o trabalho incorpora ao quadro descritivo resultante considerações sobre formação de palavras, sinonímia corradical, sinmorfismo e variação e mudança morfolexicais.

**Palavras-chave:** família lexical; etimologia; morfologia histórica; semântica histórica.

**Abstract:** This study is dedicated to describing the historical development of the word family of *usura* (which is constituted by *usura*, *usurar*, *usurário*, *usureiro*, *usurável*, *usurento*, *usuroso*, *usurador*, *zura*, *zuraco* e *usurariamente*) in the Portuguese language. It takes as a basis for its diachronic and comparative approach the integration of dialectal data and online posts with a variety of other data sources in several Romance languages and in the English language. These sources constitute its *corpus* and the foundation which underpins the elucidation of borrowing and neology processes. It examines the etymology of this lexical family via the gradual analysis of semantic and morphological changes in each word, which is supported by textual evidence for such phenomena. Being defined from an interfacial standpoint to Morphology, Semantics, and the Lexicon, it incorporates into the resultant etymological description the discussion of themes such as word formation, the synonymy of co-radical words, affix synonymy, and morpho-lexical variation and change.

**Keywords:** lexical family; etymology; historical morphology; historical semantics.

Recebido em 07 de setembro de 2020

Aceito em 14 de janeiro de 2021

## 1 Introdução

Volta-se este artigo à constituição da família lexical de *usura* no percurso histórico da língua portuguesa, norteado pelo intento de sumarizar algumas considerações e informações sobre a sua composição e as suas origens, bem como as transformações formais e semânticas que incidiram sobre seus elementos constituintes. Embora centrado no desenvolvimento de tal paradigma léxico no vernáculo, vale-se também de dados de outras línguas com as quais o português estabeleceu ou estabelece contatos, de caráter diverso; daí ser possível sustentar que este estudo se apoia também no cotejo entre línguas em seu fluxo histórico: em outras palavras, no método histórico-comparativo. Definindo-se nesse eixo, da Linguística Histórica e da Etimologia, o trabalho ora ensaiado se funda na integração de dados dialetológicos com o estudo etimológico e morfossemântico e se justifica pela consideração do que está ausente nos dicionários ditos “representativos” e na necessidade do estudo histórico-diacrônico para o desvelamento da língua em seu estado atual.

Trata-se de um estudo descritivo, uma primeira aproximação à constituição paradigmática da família léxica de *usura* no fluxo histórico da língua portuguesa (o que parece gozar de certo ineditismo), chegando-

se à sua feição e arquitetura na hodiernidade. Para tanto, considerou-se um rol de aportes do âmbito da morfologia lexical, da etimologia e da lexicologia e semântica históricas para a fundamentação teórica do estudo. Como hipótese inicial (que se viu confirmada), tinha-se o entendimento de que tal família de palavras seria quantitativamente modesta (com poucas lexias), embora possuidora de uma rica morfologia derivacional e de um relevante espectro polissêmico.

Uma etimologia é aqui compreendida nos termos de Viaro (2014), isto é, a descrição da trajetória de uma forma (palavra, morfema, expressão etc.) em dada língua, abrangendo a discriminação de aspectos formais e semânticos. Distingue-se **étimo** de **origem**, em conformidade com Viaro (2014), pois a descrição etimológica não precisa necessariamente remontar, por exemplo, às raízes indo-europeias. **Étimo** é a forma estudada tal qual se apresenta em uma sincronia pretérita da língua de que foi herdada ou de que foi tomada de empréstimo, sem aumento ou supressão de seus elementos formativos. Assim, por exemplo, uma forma de raiz românica em português pode ter étimo francês ou espanhol, por ter sido tomada de empréstimo dessas línguas, e não seria preciso remontar a uma raiz do proto-romance ou mesmo do protoindo-europeu para o estabelecimento de seu étimo. **Origem**, em sentido amplo, pode ser entendida como a explicação da proveniência imediata de uma forma herdada ou emprestada (confundindo-se, nesse sentido, com **étimo**) ou da **formação** (no caso de palavras, **derivação**) do item investigado, e, em sentido estrito, é entendida como o recuo à origem da raiz que gera determinada lexia. Tratar-se-á aqui de **étimo** e **formação**. Quanto ao significado, com base em Viaro (2009a, 2011a, 2014), considera-se **sentido prototípico** (sentido esperado de uma forma recém-criada, condizente com a base e o processo derivacional) e **transformação semântica** (metáfora, metonímia, especialização, generalização etc.) ocorrida após a formação da lexia (por vezes, imprevisível na base ou nos afixos que carrega). Estando, no entanto, em uma interface entre a Morfologia e a Semântica Históricas, o estudo considerará esses aspectos conjuntamente e nas relações que estabelecem entre si e na evolução das formas estudadas.

Em adição ao referencial teórico do campo da Etimologia, o artigo fundamenta-se também em estudos sobre formação de palavras (ARONOFF, 1976; DÍAZ HORMIGO, 2004-2005; ROCHA, 1998; RODRIGUES, 2016), léxico real e virtual (BASILIO, 2004), morfossemântica histórico-diacrônica (SOLEDADE, 2004; VIARO, 2011b) e variação morfológica e

sinonímia afixal (CAMBRAIA, 2010; PONCE DE LEÓN, 2010; SIMÕES NETO, 2018; SOLEDADE, 2012). Conhecendo a natureza do trabalho linguístico histórico-diacrônico, que não pode se valer de amostras da língua falada quando trata de sincronias pretéritas em que a aparelhagem de gravação inexistia, como assevera Mattos e Silva (2008), faz-se uso abundante de abonações como evidência textual para comprovação das transformações semânticas e datação dos fenômenos.

As buscas a que a família lexical de *usura* foi submetida, discutidas em detalhe na seção seguinte, permitiram a identificação de onze formas que constituem o conjunto e partilham um morfema lexical básico (MLB) que se apresenta sob os alomorfes *usur-* e *zur-* (este último em lexias como *zura* e *zuraco*). Em consulta a dicionários de referência da língua portuguesa, quais sejam, o *Novo dicionário Aurélio da língua portuguesa* (FERREIRA *et al.*, 2004), o *Dicionário Houaiss da língua portuguesa* (HOUAISS; VILLAR, 2009), o *Dicionário online Caldas Aulete*, o *Dicionário Priberam da Língua Portuguesa* e a *Infopédia – Dicionários Porto Editora* (incluídos no rastreamento das formas), observou-se que apenas seis dessas lexias são registradas em todas as cinco obras, uma só é registrada por três dicionários e quatro estão ausentes em todos eles. Vide o Quadro 1:

QUADRO 1 – Dicionarização da família lexical de *usura* em português

Formas	Ferreira <i>et al.</i> (2004)	Houaiss e Villar (2009)	Caldas Aulete	Priberam	Infopédia
usura	X	X	X	X	X
usurar	X	X	X	X	X
usurário	X	X	X	X	X
usureiro	X	X	X	X	X
usurável					
usurento					
usuroso					
usurador					
zura	X	X	X	X	X
zuraco	X	X	X	X	X
usurariamente			X	X	X

Fonte: Levantamento feito pelos autores.

*Usura*, *usurar*, *usurário* e *usureiro* pertencem ao léxico geral e têm ampla cobertura lexicográfica. *Usurariamente*, embora não encontrado em Ferreira *et al.* (2004) e Houaiss e Villar (2009), ocorre em outros repertórios lexicográficos (Caldas Aulete, Priberam e *Infopédia*) e possui constituição mórfica transparente (como advérbio derivado em *-mente*). *Zura* e *zuraco*, apesar de registradas nas obras lexicográficas, são indicadas como formas populares e regionalismos do Brasil e se distinguem das outras pela aférese da vogal inicial, fenômeno histórico da língua (lat *acumen* > port *gume*) que continua atuando no português popular brasileiro, a exemplo de port *agarrar* > [ga'ya]. *Usurável*, *usurento* e *usuroso*, ausentes em todos os dicionários examinados, são formas que ocorrem no primeiro atlas regional brasileiro, o *Atlas Prévio dos Falares Baianos – APFB* (ROSSI *et al.*, 1963), dividindo com *usurário* o espaço dialetal baiano na condição de sinônimos corradicais para ‘avarento, sovina’. *Usurável* também ocorre em Sergipe (CARDOSO; FERREIRA, 2000) junto com *usurário* (ARAGÃO, 2014), indicando que possivelmente os corradicais sinonímicos em *usur-/zur-* têm mais variedade e vitalidade, no português brasileiro, na área dos *falares baianos* (cf. divisão dialetal do Brasil em NASCENTES, 1955). *Usurador* só foi identificado em publicações recentes na *web*.

Essa família lexical apresenta oito vozes nominais que são sinônimos corradicais, i.e., formas de mesmo sentido, com o mesmo MLB e em que se manifesta a variação morfológica, aqui entendida, em consonância com Bagno (2007) e Simões Neto (2018), como a flutuação no uso de morfemas derivativos (prefixos ou sufixos) semanticamente congêneres, da qual resultaria dada coocorrência/concorrência de formas lexicais sinonímicas, como *tonteira* e *tontura* (SIMÕES NETO; SOLEDADE, 2013). Duas lexias são derivados deverbais (*usurador* e *usurável*) e seis são derivados denominais (*usurário*, *usureiro*, *usurento*, *usuroso*, *zuraco* e *zura*),<sup>1</sup> mas, apesar das bases diferentes, todas são utilizadas de maneira intercambiável na acepção de ‘avarento, sovina’ e em algumas acepções dadas a *usurário* e *usureiro* nos dicionários (ali apontados como sinônimos totais), ‘que(m) faz empréstimos a juros’, ‘ganancioso’. E não parece se estabelecer diferenciação diatópica em

---

<sup>1</sup> O que é possível apontar ao se considerar as lições de Rio-Torto *et al.* (2016) e Soledade (2004) quanto às seleções de base por tais sufixos, assim como a etimologia dos derivados em tela (o seu percurso formativo delineado no eixo temporal).

todos os casos: no próprio APFB (Cartas 104 e 105), p.ex., *usurável* e *usurento* ocorrem na resposta do mesmo informante em dois pontos (44 e 48), *usurável* e *usuroso* ocorrem no mesmo ponto, na fala de informantes diferentes (17) e *usurário* e *usurável* ocorrem juntos em dois pontos, na resposta de informantes diferentes (3 e 8).

Que razões poderiam ser elencadas para a explicação desse fenômeno de variação morfológica? Que implicações os dados empíricos aqui apresentados trazem na consideração de temas como o *bloqueio* para uma nova criação lexical, apontado por Aronoff (1976) e Rocha (1998)? A discussão dessas perguntas acrescenta-se ao objetivo geral deste trabalho, que é traçar o desenvolvimento histórico-diacrônico da família léxica de *usura* na língua portuguesa e descrever a etimologia de cada lexia que a compõe, considerando suas transformações no âmbito da forma e do significado.

## 2 Metodologia e corpora

A delimitação do objeto de análise deste artigo, a família lexical de *usura*, é, sobretudo, consequência do trabalho de inserção do APFB (ROSSI *et al.*, 1963) na plataforma do *Tesouro do léxico patrimonial galego e português* – TLPGP (ÁLVAREZ, 2014-2020),<sup>2</sup> objetivo central do projeto de iniciação científica (CNPq) no qual os coautores foram, respectivamente, bolsista e orientador, e do seminário interno de um componente curricular da graduação em Letras na instituição a que são filiados, *Introdução aos Estudos Dialetais*, no qual o bolsista apresentou um trabalho, em 2019.2, analisando as cartas fonéticas 103, 104, 105 e a carta-resumo 103/104/105 do APFB, todas referentes a denominações para *avarento*. A carta fonética 104 apresenta *usurário* e *usurável*, com várias realizações fonéticas diferentes, e *usura* em suas notas; já *usurento* e *usuroso* estão na carta fonética 105. A ocorrência dessas cinco formas e, especialmente, a coexistência de quatro corradicais sinonímicos inspirou a pesquisa, que agora engloba toda a família léxica.

É comum que alguns estudos em morfologia e formação lexical apontem que a existência de certa palavra pode bloquear a criação

---

<sup>2</sup> Projeto internacional e interinstitucional sediado no Instituto da Lingua Galega da Universidade de Santiago de Compostela, sob a coordenação geral da Profa. Dra. Rosario Álvarez Blanco. O portal do *Tesouro* pode ser acessado em: <http://ilg.usc.es/Tesouro>.

de novas formas sinônimas, sobretudo corradicais, pelos diversos processos derivacionais disponíveis na língua (cf. ARONOFF, 1976; ROCHA, 1998), mas os dados do APFB, trazendo 27 lexias diferentes para ‘avarento, sovina’ no mesmo espaço dialetal, entre elas quatro sinônimos corradicais, desafiam tal noção. O que geraria essa profusão de formas? O estudo do léxico não pode ignorar fato incontornável, aqui parafraseado de Viaro (2011a): as palavras que se busca descrever não foram criadas hoje. O *momento neológico* (VIARO, 2011a) reside em sincronias pretéritas. Tentar intuir donde vem esta ou aquela palavra apenas com o conhecimento da sincronia atual não somente seria um exercício acientífico, mas potencialmente falseador do funcionamento da língua (cf. VIARO, 2014).

Há razão, portanto, para defender que é o estudo etimológico e histórico-comparativo o mais acertado para explicar a diversidade e variação em qualquer família de palavras, e naquela que porta o MLB **usur-** (alomorfes *usur-* e *zur-*) em particular. A Etimologia (cf. VIARO, 2014, 2013), enquanto descrição do trajeto histórico de dada forma ou palavra, desde sua formação ou de seu *étimo* até o momento atual, terá de se ocupar em estabelecer os seguintes elementos-chave:

- i) o **étimo** de uma forma X que, como apontado anteriormente, corresponde a essa *mesma forma* tal como se apresenta em uma sincronia pretérita de dada língua ou de outra língua donde foi herdada ou tomada de empréstimo, sem ter experimentado aumento ou supressão de seus elementos formativos (VIARO, 2014). Assim, por exemplo, o galego-português *sodes* sXIII é um étimo válido para o português *sois* (2ª pessoa do plural do verbo *ser* no presente do indicativo), e seria possível ainda recuar à forma latina reconstruída *\*sutis*. Normalmente atribuir-se-á como étimo a atestação de maior antiguidade (lat *causam* sII a.C. > port *coisa* sXIII) ou uma reconstrução plausível, consideradas as histórias interna e externa e as relações genéticas da língua em questão (lat *\*sutis* > gal-port *sodes* sXIII > port *sois*). Há que se dizer ainda que esses recuos, no caso do étimo, são válidos apenas para o que se poderia considerar diferentes estágios da mesma língua. Por exemplo: o étimo do port *deletar* não é o lat *deletus*, e sim o ing (to) *delete*. Embora a forma participial latina seja a origem da forma inglesa, o verbo *deletar* entra na língua portuguesa como

empréstimo inglês, logo seu étimo é inglês. O ponto crucial é a forma de entrada na língua: o étimo de formas herdadas é encontrado na língua-mãe, e o étimo de formas emprestadas está na língua de contato. Para o presente estudo, formas atestadas terão absoluta primazia, e recorrer-se-á a reconstruções apenas nos casos em que evidências de outras línguas românicas suportem um étimo reconstruído. Nenhum étimo será proposto caso ambas as situações não se apliquem;

- ii) o ***terminus a quo***, isto é, a atestação mais antiga da forma estudada. Embora essa informação geralmente não permita o estabelecimento do *momento neológico*, que é quase sempre anterior, com raras exceções, ela testemunha a existência da forma naquela sincronia, dado fundamental para qualquer explicação das transformações e dos processos de transmissão da lexia investigada. Buscar-se-á fornecer uma datação para todas as formas, com suas respectivas abonações;
- iii) os **elementos formativos** das palavras estudadas, tais quais eram na língua e sincronia em que foram formadas, a saber: a) a **raiz**, elemento estritamente diacrônico, que pode reunir formas interpretadas como independentes na sincronia atual (lat *planum* > port *porão* e *chão* vs. port *plano* < lat *planus*), formas de relação bastante opaca (port *falar* e *fama*, cf. PENA; CAMPOS SOUTO, 2009) e ainda formas de línguas com alguma relação genética, ainda que distante (protoindo-europeu *\*pisk-* latim *piscis* ‘peixe’ = irlandês antigo *īasc* = gótico *fisks*). A raiz é definidora daquilo que Pena e Campos Souto (2009) tratarão como “família etimológica”; b) o **radical**, termo que agrega complexa polissemia cujos pormenores ultrapassam o escopo do artigo. Para evitar confusões, utilizar-se-á o sinônimo **morfema lexical básico** (MLB), que é mais preciso por definir um paradigma léxico mais ou menos homogêneo e estabelecer um significado básico que ainda pode ser recuperado em todas as formas derivadas. Um radical, por exemplo, pode ser tanto o MLB quanto qualquer base que servirá à adjunção de afixos derivacionais (***amig-o***, ***amicabil-idade***). Resta acrescentar que o MLB pode se sobrepor à raiz (cf. lat *pisc-is*), mas ambos se distinguem porque **raiz** diz respeito a uma noção diacrônica, enquanto **MLB** define um grupo de palavras em dada sincronia; e c) os **afixos** (prefixos, sufixos,

circunfixos etc.), estejam completamente fundidos ao MLB (*com-* em port *com-er*, sequência que correspondia ao prefixo em lat *com-edēre*) ou não (port *just-ez-a*, MLB *just-*). A identificação dos elementos formativos históricos da palavra estudada encerra uma *etimologia*: port *amigo* < lat *amīcum* ~ *amīcus* ← lat *am-* + *īcus,a,um* (cf. VAAN, 2008).

Entende-se neste artigo por *família lexical* (ou *família léxica*) o conjunto de vocábulos que se relacionam entre si pelo compartilhamento de um MLB, revelando um vínculo/parentesco etimológico e a manutenção de um significado central comum entre as formas ou, quando menos, alguma aproximação semântica. É esse segmento, o MLB (com suas possíveis variações alomórficas), que reverbera na palavra-base ou lexema matriz que encabeça a família lexical (por exemplo, família lexical de *morte*, família lexical de *lento*, família lexical de *andar* etc.), item responsável pela ativação de tal paradigma léxico, sendo o principal fator para sua caracterização e identificação.

Além da família de *usura*, objeto deste estudo, poder-se-ia mencionar, a modo de exemplificação, a família lexical de *outono*, que se estrutura sob o MLB *outon-*, tem como palavra-base ou lexema matriz *outono*, e é constituída pelas lexias *outono*, *outonada*, *outonal*, *outonar*, *outonear* e *outoniço*. Destarte, o termo *família lexical* corresponderia ao que Rodrigues (2016) designa por *família de palavras*, Haspelmath (2002) e Haspelmath e Sims (2010) por *word family* e *lexeme family*, Sánchez Martín (2008) por *família léxica* e a NGLE (RAE; AALE, 2009) por *familia de palabras*. Pode-se perceber, assim, com Pena e Campos Souto (2009) e Hernández Arocha (2020), que a noção de *família lexical* não se confunde com o que se só denominar nos estudos léxico-semânticos como *campo lexical*, *campo nocional*, *campo semântico* ou mesmo *família etimológica*. Trata-se, em suma, de um agrupamento de vocábulos analisado sob um ponto de vista morfológico, ainda que implique considerações de ordem semântica e etimológica.

Cabe ressaltar, ainda sobre a família lexical, que se trata de um paradigma léxico-morfológico caracterizável como um agrupamento de alcance potencialmente elástico, ou seja, aberto (MORERA, 2002), dinâmico, passível de ampliação, com a geração/inclusão de novos vocábulos. Sendo assim, se se concebe a família lexical de *tecla* como constituída por esse elemento e pelas lexias *teclado*, *teclar*, *teclear*,

*tecleador* e *tecladista*, pode-se afirmar que se mantém *potencial* ou *virtualmente* aberta, dada a *possibilidade* de criação neológica e difusão de formas como *tecleamento*, *tecleação*, *tecleadista*, *tecleiro*, *multitecla*, *megateclado*, *supertecleador*, *retecl(e)ar*, *porta-teclado/guarda-teclado* etc.

Embora o árduo trabalho de ir aos textos antigos para desvendar os caminhos de uma única palavra não raro baste como escopo de um artigo, há certa vantagem no estudo de uma *família lexical* (cf. PENA; CAMPOS SOUTO, 2009, e o verbete *família* em DUBOIS *et al.*, 2014), pois o labor etimológico exige que se enfoque “o todo que circunda a palavra pesquisada” (VIARO, 2014, p. 235), considerando que os vários membros dessa *família* tecem múltiplas relações entre si, desde sua criação até o presente, tanto no tocante à forma quanto ao sentido. Mesmo nos casos em que uma das formas dista substancialmente de seus corradicais pela atuação de transformações metafóricas, a própria história dessas relações é o que esclarecerá tal distância. O português, por sua vez, não é língua isolada, mas manteve e mantém contato, em suas diversas manifestações, com muitas e diferentes línguas de todo o mundo. Além disso, é uma língua românica, da família indo-europeia, e partilha origem comum com o castelhano, o galego, o francês, o italiano, o romeno, entre outras, sendo descendente do latim vulgar (BASSETTO, 2005; ILARI, 2018). Cabe mencionar também que o *latim literário* é fonte ininterrupta de empréstimos eruditos e técnicos ao longo de toda a história do grupo românico, o que Bassetto (2005) e Ilari (2018) chamam de *adstrato permanente*. Assim justifica-se a aplicação de método histórico-comparativo, para uma compreensão mais sólida da constituição do conjunto de palavras estudadas.

Ao discutir os graus de certeza de uma etimologia propostos por Jespersen (1954), dirá Viaro (2009a, p. 455): “Quanto maior o número de línguas envolvidas (...), mais próximo do seguro esse étimo estará”. O primeiro passo realizado foi o rastreamento de todas as formas com o MLB **usur-** em dicionários de língua portuguesa e em uma variedade de *corpora* linguísticos *online*, o que permitiu a construção gradual da lista de palavras apresentadas no Quadro 1. Em seguida, fez-se ao mesmo tempo o levantamento de novos *corpora* (para o registro da ocorrência de cada palavra em outras bases de dados e a coleta de abonações e datações) e a seleção das línguas que seriam incluídas na comparação. Essa escolha baseou-se sobretudo na disponibilidade de materiais fidedignos, na possibilidade de manejo de textos em língua

estrangeira e no potencial contributo de dada língua à comparação, isto é, o que o registro e a datação em uma e outra língua poderiam indicar para a história dessa família léxica, consideradas a história dos contatos de cada língua e as divisões dialetais propostas para a România (cf. ILARI, 2018). Por exemplo: um item de raiz românica identificado em inglês provavelmente terá existido em francês, e certamente não ocorreria isoladamente no português e no inglês; outro, identificado em galego e castelhano, poderá ser específico aos romances ibéricos; e ainda outro, sendo identificado em romeno ao lado das outras línguas românicas, sem apresentar características de um empréstimo ou decalque do francês ou do italiano, se caracterizará em um caso mais robusto para a proposição de um étimo latino, ainda que reconstruído. Fizeram parte da seleção: o latim (clássico, vulgar, medieval e renascentista), o castelhano, o catalão, o francês, o galego, o italiano, o provençal/occitano, o romeno, o sardo e o inglês (que, mesmo não sendo língua românica, recebeu expressivo aporte de léxico latino-românico).

Quanto aos *corpora*, o quadro de referências (que em número ultrapassa uma centena) pode ser dividido em sete tipos: 1) dicionários de língua monolíngues, bilíngues ou trilingues (*online* e impressos); 2) dicionários etimológicos (em sua maioria impressos, e alguns *online*); 3) *corpora* textuais *online*; 4) bancos de dados léxicos (como o TLPGP); 5) mecanismos de busca na *web* (a pesquisa avançada do Google<sup>3</sup> e o Google Livros);<sup>4</sup> 6) atlas linguísticos (entre eles o próprio APFB); e 7) *varia*, como livros ou estudos de menor extensão que contêm as lexias investigadas ou portais em diferentes línguas que reúnem referências, como o *Centre National de Ressources Textuelles et Lexicales*<sup>5</sup> ou o *Institut d'Estudis Catalans*.<sup>6</sup> Mesmo não tendo sido possível em todos os casos usar um *corpus* de cada tipo para cada língua, houve variedade significativa nos materiais de cada uma delas. Construído o referencial, a família lexical foi submetida a rastreio exaustivo de correspondências em todas as línguas escolhidas, em todos os materiais levantados.

Também não foram ignoradas formas que apresentam constituição morfológica diferente daquelas em língua portuguesa. O achado de

<sup>3</sup> Disponível em: [https://www.google.com/advanced\\_search](https://www.google.com/advanced_search). Acesso em: 21 ago. 2020.

<sup>4</sup> Disponível em: <https://books.google.com.br/>. Acesso em: 31 jul. 2020.

<sup>5</sup> Disponível em: <https://www.cnrtl.fr/>. Acesso em: 18 ago. 2020.

<sup>6</sup> Disponível em: <https://www.iec.cat/activitats/entrada.asp>. Acesso em: 21 ago. 2020.

*usurador* em galego, por exemplo, levou à inclusão de formas hipotéticas correspondentes em outras línguas na lista de palavras que seriam buscadas e a um novo rastreamento em todas as línguas selecionadas: em analogia ao resultado de outras palavras em *X-tôr*, previa-se *\*usurador* em português, castelhano, catalão e occitano, *\*usurateur* ou *\*usureur* em francês, *\*usuratore* em italiano, *\*usuradòri* em sardo, *\*usurător* em romeno, *\*usurator* ou *\*usuror* em inglês e ainda uma forma latina como *\*usurator*. Essas lexias foram então rastreadas nos materiais escolhidos, e com este procedimento um filtro bastante eficaz foi acrescentado à pesquisa, pois ele permitiu, neste caso, a atestação de *usurador* em português, castelhano e catalão, *usureur* em francês, *usuror* em inglês e *usurator* em latim. Para os dicionários de língua e etimológicos, a disposição dos verbetes foi um facilitador do trabalho: por estarem em ordem alfabética, bastou seguir a sequência *verbo ad verbum* até que todas as formas com o MLB pesquisado ali registradas fossem recuperadas. Em alguns casos, especialmente em dicionários *online*, quando não há índices alfabéticos, a solução foi digitar a sequência *usur-* (ou o equivalente na língua específica) e abrir cada verbete sugerido com essa mesma sequência até o registro de todas as formas de MLB **usur-** descobertas. No caso de *corpora* textuais e bancos de dados lexicais, diversas foram as estratégias. Alguns, como o *Corpus do Português*<sup>7</sup> (DAVIES, 2006-2018), permitem a busca lematizada, então foi suficiente digitar *usurário* em caixa alta e o *corpus* forneceu todas as ocorrências desta palavra, em todas as formas flexionadas, identificadas no banco de textos. Outros forneciam índices, e procedeu-se aí com a mesma técnica usada nos dicionários. Sem índices, buscou-se por sequência de letras. Quando nenhuma dessas opções estava disponível, somente formas conhecidas ou cuja existência se supunha por equivalentes em outras línguas foram buscadas.

A pesquisa avançada do Google permite refinar a busca por língua e período, assim é possível pesquisar, p.ex., *usura* apenas em páginas em português publicadas entre 2005 e 2007. Esse mecanismo foi pouco usado para datações, já que os outros materiais e o Google Livros serviam melhor a tal atribuição, mas permitiu o registro da ocorrência de palavras ausentes nos dicionários e textos. Somente páginas que poderiam ser abertas para verificação do contexto da palavra foram consideradas, pois

---

<sup>7</sup> Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 21 ago. 2020.

os resultados gerados nem sempre correspondiam à sequência digitada, ou ainda a sequência era a mesma, mas estava em duas palavras separadas. Também foi necessário, no caso do verbo *usurar* e correspondentes nas outras línguas, verificar o contexto sempre que se incluísse na pesquisa uma forma que fosse idêntica a alguma lexia nominal (*a usura* x *ele usura*, *as usuras* x *tu usuras*), para garantir que a ocorrência fosse de fato da forma verbal. O latim, o galego, o occitano/provençal e o sardo não foram pesquisados no Google por não haver a opção de restringir a busca a elas, e também o romeno por haver uma maior dificuldade dos investigadores de interpretar textos nessa língua e pela coocorrência de formas similares não facilmente distinguíveis da família investigada sem treinamento prévio sobre leitura e ortografia em romeno.

O Google Livros também admite o refinamento por língua e período e, pelas mesmas razões acima expostas, formas latinas, galegas, occitanas, sardas e romenas não foram buscadas com essa ferramenta. Somente obras datadas e em que era possível a visualização da página ou do texto integral foram consideradas, pois muitas vezes o mecanismo de busca gerava um resultado que tinha a sequência digitada, mas estava em duas linhas diferentes como duas palavras soltas (ex. *usura dor*) ou como partes de duas palavras com hífen identificadas como uma sequência só (ex. *usur-pador miser-ável*) ou ainda uma palavra com letras ilegíveis ou rasuradas (ex. *usurpador*). A ocorrência de cada forma e a datação, quando fornecida, foram anotadas em documento à parte, com a devida referência do material donde se extraíram essas informações. Findada essa fase, foi possível elaborar uma tabela *Excel* com os lemas de cada língua e seus respectivos *termini a quo*. A extração de abonações só foi feita para o português e, nos casos em que se identificou empréstimo ou léxico herdado, para a língua do étimo.

Para melhor exposição dos resultados do rastreamento, dois quadros foram elaborados: o primeiro deles (QUADRO 2) é organizado com formas que possuem correspondência direta às formas encontradas em português, i.e., que possuem base léxica equivalente ao port *usura* e afixos que correspondem àqueles identificados no vernáculo; o segundo (QUADRO 3) reúne todas as outras formas que não possuem equivalência direta em termos de suas bases ou de seus afixos. Em ambos os quadros, indica-se o *terminus a quo* de cada forma, quando este pôde ser certificado. Além disso, o Quadro 3 traz a tradução e a categoria gramatical das lexias encontradas.

QUADRO 2 – Equivalentes latinos, românicos e ingleses às formas de língua portuguesa identificados no rastreamento em sincronias atuais ou pretéritas<sup>8</sup>

português	latim	castelhano	catalão	francês	galego
usura 1264-84	ūsūra, ae 191 a.C.	usura 1218-50	usura sXIII	usure 1140	usura 1264-84
usurar sXIV	usurare 1188	usurar 1541	usurar sXV	usurer sXIII	-
usureiro 1264-84 usurário 1589	ūsūrārīus, a, um sII a.C.	usurero 1247 usurario 1325	usurer 1249 usurari 1385-86	usurier 1213 usuraire 1311	usureiro 1264-84 usurario
usurável 1963	-	usable 1895	usable 1913	usable 1336	-
usurento 1963	-	usuriento 2012	-	-	-
usuroso 1963	-	usuroso 2012	-	-	-
usurador 2013	usurator sXV	usurador 1913	usurador 1913	usureur sXII	usurador
usurariamente 1758	-	usurariamente 1744	usuràriament 1839	usurairement 1448	-
português	italiano	occitano	romeno	sardo	inglês
usura 1264-84	usura sXIII	usura sXII	uzură 1826	usura 1316	usure † <sup>1</sup> sXIV
usurar sXIV	usurare 1309	usurar	-	usurai 1832	to usure † 1757
usureiro 1264-84 usurário 1589	usuraio sXIII usurario sXIII usuriere sXIII	usurièr sXII usurari sXIII	uzurar 1826	usureri 1316 usuraju 1832 usuràriu 1832	usurer sXIV usurary † 1681
usurável 1963	usurabile 1850	-	-	-	usable 1879
usurento 1963	-	-	-	-	-
usuroso 1963	usuroso 2016	-	-	-	usurous † 1716
usurador 2013	-	-	-	-	usuror † 1686
usurariamente 1758	usurariamente 1717	-	-	-	-

Fonte: Rastreamento feito pelos autores.

<sup>8</sup> Nenhuma forma equivalente a *zura* ou *zuraco* foi encontrada fora do português no rastreamento, e por essa razão ambas não são trazidas no quadro, uma vez que não serviriam ao propósito da comparação.

<sup>9</sup> O símbolo «†» indica que a palavra é desusada, seguindo a simbologia adotada por Viaro (2014).

QUADRO 3 – Outras formas sem equivalentes diretos em português<sup>10</sup>

latim <sup>11</sup>	castelhano	catalão
<ul style="list-style-type: none"> <li>● <i>interusurium</i> (M) sVI subs. ‘juro que se acumula por um período’</li> <li>● <i>usuraticus</i> (R) 1513 adj. ‘relativo à usura, referente a juros’</li> <li>● <i>usuratio</i> (M) sXIII subs. ‘prática de usura, ato de emprestar a juros’</li> <li>● <i>usurella</i> (R) 1697 subs. ‘pequena usura, logro’</li> <li>● <i>usura</i> (M) 1710 subs. ‘direito de usar, uso, usagem’</li> <li>● <i>usurizare</i> (M) sXIII v. ‘fazer empréstimo a juros, usurar’</li> <li>● <i>usurolum</i> (M) 1710 adj. ‘usado, costumado’ (?)</li> <li>● <i>usurula</i> (R) 1697 subs. ‘pequena usura, logro’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>interusurio</i> † 1648 subs. ‘juro que se acumula por um período’</li> <li>● <i>usuravelmente</i> 1913 adv. ‘de maneira usurável’</li> <li>● <i>usuradamente</i> 1913 adv. ‘de maneira usurada’</li> <li>● <i>usurear</i> 1739 v. ‘dar ou tomar a usura’</li> <li>● <i>usureria</i> † 1845 subs. ‘prática da usura’ ou ‘lugar onde se usura’</li> <li>● <i>usúria</i> † 1378-1406 subs. ‘usura’</li> <li>● <i>usurioso</i> adj. ‘que tem usura, de que se cobra ou paga usura’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>usurablement</i> 1913 adv. ‘de maneira usurável’</li> <li>● <i>usuradament</i> 1913 adv. ‘de maneira usurada’</li> <li>● <i>usural</i> 1913 adj. ‘usurário’</li> <li>● <i>usurant</i> 1913 adj. ‘usurário’</li> <li>● <i>usurejar</i> 1913 v. ‘usurar’</li> <li>● <i>usurerament</i> 1913 adv. ‘de maneira usurária’</li> <li>● <i>usureria</i> 1913 subs. ‘prática da usura’</li> <li>● <i>usureta</i> 1840 subs. ‘usura pequena’</li> <li>● <i>usuriós</i> adj. ‘que tem usura, de que se cobra ou paga usura’</li> </ul>
francês	galego	italiano
<ul style="list-style-type: none"> <li>● <i>interusure</i> 1767 subs. ‘juro que se acumula por um período’</li> <li>● <i>usuratif</i> 1587 adj. ‘usurário’</li> <li>● <i>usuratoire</i> 1426 adj. ‘que diz respeito à usura’</li> <li>● <i>usurie</i> † subs. ‘usura’</li> <li>● <i>usurieux</i> 1497 adj. ‘usurário, usuroso’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>usúria</i> 1969 subs. ‘cobiça’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>usuraiaccio</i> subs. ‘usureirinho’</li> <li>● <i>usuraietto</i> subs. ‘usureirinho’</li> <li>● <i>usurale</i> adj. ‘usurário’</li> <li>● <i>usurarietà</i> 1876 subs. ‘qualidade de ser usurário’</li> <li>● <i>usuratico</i> adj. ‘que se baseia em usura’</li> <li>● <i>usureggiamento</i> sXIV subs. ‘ação de usurar, o resultado de usurar’</li> <li>● <i>usureggiante</i> adj. ‘que pratica a usura’</li> <li>● <i>usureggiare</i> 1334 v. ‘usurar’</li> <li>● <i>usuriare</i> 1313 ‘emprestar a juros de usura’</li> <li>● <i>usurioso</i> 1749 adj. ‘usurário, usuroso’</li> </ul>

<sup>10</sup> Não foram encontradas outras formas no occitano, e por esse motivo essa língua não está no Quadro 3.

<sup>11</sup> (M) e (R) indicam que a forma é atestada, respectivamente, no latim medieval ou no latim renascentista.

romeno	sardo	inglês
<ul style="list-style-type: none"> <li>● <i>usurare</i> 1823 subs. ‘prática da usura’</li> <li>● <i>usuras</i> 1823 subs. ‘usurário’ (ocupação)</li> <li>● <i>usurat</i> 1823 adj. ‘emprestado a juros’</li> <li>● <i>usuratoãre</i> 1823 subs./adj. ‘usurária’</li> <li>● <i>usuretiese</i> 1823 adj. ‘relativo à usura, usurário’</li> <li>● <i>usurez</i> 1823 v. ‘liquidar a usura’</li> <li>● <i>uzurãrie</i> subs. ‘usura, ocupação do usureiro’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>usuramentu</i> subs. ‘usura’</li> <li>● <i>usuresu</i> subs. ‘usurário’</li> <li>● <i>usuria</i> subs. ‘usura’</li> <li>● <i>usurittu</i> subs. ‘usurário’</li> </ul>	<ul style="list-style-type: none"> <li>● <i>usuress</i> 1648 subs. ‘usureira, mulher usureira’</li> <li>● <i>usurious</i> 1640 adj. ‘que empresta dinheiro a juros exagerados’</li> <li>● <i>usuriously</i> 1667 adv. ‘de maneira usurária’</li> <li>● <i>usuriousness</i> 1731 subs. ‘qualidade ou disposição usurária ou extorsiva’</li> <li>● <i>usury</i> sXIV subs. ‘usura’</li> </ul>

Fonte: Rastreamento feito pelos autores.

Não é supérfluo salientar, mais uma vez, que o estudo ora delineado tem como escopo o rastreamento das formas que constituem a família lexical de *usura* na língua portuguesa, sendo esse seu objeto, tendo-se como recorte empírico os dados extraídos das fontes várias mencionadas, atinentes ao vernáculo. Destarte, tomam-se informações/dados de outras línguas românicas e do inglês apenas como apoio (que se mostrou imprescindível) para algumas das análises morfológicas, semânticas e etimológicas delineadas nas próximas seções do artigo.

Dada a profusão de dados encontrados, a discussão dessa família de palavras em âmbito românico (com a adição do inglês) ficará para outra publicação. Isso de modo algum anula o valor do rastreamento panromânico para este trabalho. Seus resultados foram profícuos no esclarecimento do processo de transmissão das palavras para o português, permitiram a discriminação entre o que é herdado e o que surge e se manifesta em língua portuguesa e, sobretudo, deram respaldo empírico às etimologias que serão propostas a seguir.

### 3 A base lexical *usura*

Muito embora o agrupamento das onze lexias encontradas em uma mesma família se dê por um critério eminentemente formal (todas apresentam um *morfema* comum, o MLB **usur-**), a relação entre elas é de caráter dúplice, pois se refere à forma, mas também ao significado. É com essa duplicidade em mente que *usura* será definida como *base*

*lexical*, pois é baseando-se nessa palavra e em sua teia semântica que seus derivados e os derivados de seus derivados se definem formalmente e quanto ao leque de significados que possuem.

Faz-se imperioso ressaltar que *usura* não é tomada por base derivacional primeira da família de palavras por ser a forma “mais simples”, mas sim, por sua maior antiguidade em relação às outras ser respaldada pelos textos. Não somente é possível que formas complexas derivem formas simples, como também a existência de um *léxico virtual* internalizado permite que o falante aceda e utilize palavras de paradigmas lexicais que jamais se manifestam integralmente no *léxico real* (BASILIO, 2004). Notável é o exemplo mencionado por Viaro (2014): *prostrado* é a base donde deriva o verbo *prostrar*. Esse particípio é mais antigo que o infinitivo correspondente em português, e já existia em latim sob a forma *prostratus*, do verbo *prosternere*. Sobre isso, ele afirma: “As intuições do falante ou as regras práticas da Gramática valem muito pouco para a Etimologia.” (VIARO, 2014, p. 105). Interpretações logicizantes dos processos derivacionais, sem sustentação em evidências documentais, não se coadunam com o devido exame etimológico e serão rejeitadas aqui.

*Usura* é primeiro atestada na peça *Pseudolus* (191 a.C., cf. HARRISON, 2005), de Plauto, autor do período arcaico (para uma periodização do latim, cf. CLACKSON; HORROCKS, 2007). Seu significado original, conforme definido por Lewis e Short (1879), é ‘um uso, ou proveito de algo’,<sup>12</sup> como se vê no trecho abaixo, extraído do banco de dados do *Corpus corporum*,<sup>13</sup> da Universidade de Zurique:

*non potest usura usurpari*<sup>14</sup>  
(Plauto, *Pseudolus*, 1, 2; l. 191 a.C.)

Sua origem estaria na conversão do particípio futuro do verbo depoente *utor*, *ūtī* ‘usar, servir-se de’ 200 a.C.: *ūsūrus*, a, um ‘que (se) usará’. Esse processo – a conversão de particípios em nomes departicipiais – revela a atividade já no latim arcaico da tendência que mais tarde se consolidaria nas línguas românicas (p.ex., lat *futūrus* ‘que será’ part fut de *sum*, *esse* ‘ser, estar’ > port *futuro*; lat *nūtrītūra*

<sup>12</sup> Trad. nossa do original “a using, use, or enjoyment of a thing”.

<sup>13</sup> Disponível em: <http://mlat.uzh.ch/MLS/>. Acesso em: 15 jul. 2020.

<sup>14</sup> Em tradução livre: “o uso não pode ser usurpado”.

fem do part fut *nūtrītūrus* ‘que (se) nutrirá’, de *nūtriō*, *nūtrīre* ‘nutrir, amamentar’ > fr *nourriture* ‘comida’) e é análogo ao que aconteceu com o participio passado do mesmo verbo: lat *ūsus* ‘usado’ > port *uso*. Por sua raiz, *usura* está relacionada a *uso*, *usual*, *usuário*, *útil*, *usufruto* etc., mas bem cedo seu sentido geral foi restringido e as transformações semânticas a distanciaram de seus congêneres.

Por uma especialização semântica, o lat *ūsūra* não seria o uso de qualquer coisa, mas ‘um uso ou utilidade de dinheiro emprestado’<sup>15</sup> (LEWIS; SHORT, 1879), definição encontrada em Cícero, em meados do século I a.C., abonada na passagem seguinte, extraída do *corpus* de textos latinos clássicos do *Packard Humanities Institute*:<sup>16</sup>

*magistratus a publicanis pecuniam pro usurā auderet auferre?*<sup>17</sup>  
(Cícero, *Orationes* 3, *In Verrem*, 2, 3, 72; 4. sI a.C.)

Uma das coisas que se relaciona ao uso de dinheiro emprestado são os juros pagos por esse uso, assim, por metonímia, ‘um uso ou utilidade de dinheiro emprestado’ torna-se ‘juro pago pelo uso de dinheiro, usura’<sup>18</sup> (LEWIS; SHORT, 1879). Esse sentido é encontrado já no século I a.C., como se vê a seguir (exemplo retirado do *Corpus corporum*):

*ut sexenni die sine usuris creditae pecuniae solvantur*<sup>19</sup>  
(Júlio César, *De Bello Civili*, 3, 20, 5; 6. sI a.C.)

Por generalização, de ‘juro específico para empréstimos monetários’, o termo passa a referir-se a ‘juro’<sup>20</sup> (LEWIS; SHORT, 1879). Vide este exemplo, de Cícero (do *Corpus corporum*):

*nec unquam sine usurā reddit quod accepit*<sup>21</sup>  
(Cícero, *Cato Maior: de Senectute*, 15, 50; 2. sI a.C.)

<sup>15</sup> Trad. nossa do original “a use of money lent”.

<sup>16</sup> Disponível em: <https://latin.packhum.org/index>. Acesso em: 31 jul. 2020.

<sup>17</sup> Em tradução livre: “o magistrado ousaria tirar dinheiro dos publicanos por causa de seu uso?”

<sup>18</sup> Trad. nossa do original “interest paid for the use of money, usury”.

<sup>19</sup> Em tradução livre: “que em seis anos o dinheiro emprestado seja liquidado sem usuras”.

<sup>20</sup> Trad. nossa do original “interest”.

<sup>21</sup> Em tradução livre: “nenhuma vez dá ou recebe sem usuras”.

Os significados mais antigos, relacionados a *uso*, são perdidos ao longo do tempo e não passam às línguas românicas, mas os sentidos relacionados a *juros* permanecem, são transmitidos no latim vulgar e estão no cerne de neologismos românicos posteriores. As primeiras atestações de *usura* em textos galego-portugueses se referem a *juros*. É o que se vê nas *Cantigas de Santa Maria* (1264-84), de Afonso X (exemplo extraído do *Corpus Informatizado do Português Medieval – CIPM*):<sup>22</sup>

*Como Santa Maria deu o fillo a hũa bõa [dona] que o deitara  
em pennor, e creçera a **usura** que o non podia quitar*  
(Afonso X, CSM062, 1264-84)

O último sentido identificado na pesquisa como tendo se desenvolvido ainda em latim é *lucro* (definição também dada pelo *Dicionário de Latim-Português Português-Latim* da Porto Editora, 2014). Gerado por nova atuação de metonímia, ele está em textos do século IV d.C., na Antiguidade Tardia. O exemplo mais antigo é de Ambrósio (340-397 d.C.), Bispo de Milão, no período final do Império Romano (extrato do *Corpus corporum*):

*Itaque quanto uberior fenoris summa, tanto gravior sortis **usura***<sup>23</sup>  
(Ambrósio de Milão, *De excessu fratris sui Satyrus*, 16, 1291C. sIV d.C.)

Este significado foi herdado pelo português, e o exemplo mais antigo encontrado no rastreamento é de João de Barros, em 1540 (acedido por consulta ao *corpus* de DAVIES, 2006-2018):

*[...] ali está escrito de mi e de todo fiêl sérvô que quér dar  
a **usura** o talento do Senhor.*  
(João de Barros, *Gramática da língua portuguesa*, 1540)

A metonímia é o processo que mais atua nas transformações semânticas de *usura*. Cedo ganhou (junto com a profissão de *usureiro*) conotações morais e a palavra adquiriu espaço na denominação de vícios de caráter. Do juro ou lucro da prática *usurária*, a *usura* seria agora, por metonimização, uma característica dos que a cobram, a *cobiça* ou

<sup>22</sup> Disponível em: <https://cipm.fcsh.unl.pt/>. Acesso em: 21 ago. 2020.

<sup>23</sup> Em tradução livre: “e assim quanto mais abundante a soma de juros, mais agradável o capital de lucro”.

a *mesquinhez*. O exemplo mais antigo encontrado é de 1399, fornecido por Machado Filho (2019, p. 693):

*Esto se os que leuan os algos dos judeus que elles leuan dos cristãos por vsura*  
(*Tratado dos sacramentos da ley antiga e NOVA*, 1399. Fólio 34, reto, coluna 1)

Outro exemplo do mesmo processo é o significado de ‘ganância’, que *usura* adquiriu, cujo exemplo mais antigo é o texto *Sacramental*, de 1488 (CIPM). Caracterizando os sete pecados capitais, o *scriptor* afirma que um deles, a avareza (*avaricia*), tem sete filhas e a terceira delas é a *usura*. Veja-se o trecho abaixo:

*A terceira [filha da avaricia] he ussura e em outra maneira se chama*  
*apetito de ganho. E o sseu titolo he este: [...] Todo meu cuydado*  
*he enriqueçer com ganhos de usura.*  
(*Sacramental*, 1488)

Essa abonação revela ainda duas outras coisas de interesse. Primeiro, põe *usura* numa relação de hiponímia com *avareza*, seu hiperônimo, incluindo-a na “categoria” de vícios típicos de quem é *avarento*, o que ajuda a explicar como *usura* e *avareza* vieram a se tornar sinônimos (cf. verbete *usura*<sub>1</sub> em Ferreira *et al.* (2004) e Houaiss e Villar (2009)) e também como vários de seus derivados constam como denominações para ‘sovina’ no APFB. A segunda é a coexistência de significados diferentes para *usura*, uma vez que suas correspondências a *ganância* e a *juro* estão registradas no mesmo texto.

Amplamente dicionarizada e presente em todas as línguas e variedades de língua pesquisadas, até mesmo no *latim vulgar* (VÄÄNÄNEN, 1968), *usura* possui complexa polissemia. Sua última transformação de sentido que será relevante para as seções/discussões a seguir é a que se atesta no APFB (ROSSI *et al.*, 1963):

*[usurável] invejoso, tem usura pelo que é dos outros.*  
(APFB, 1963, Carta 104, P-29, inf. A)

Esse caso é outro exemplo de metonimização. Por contiguidade semântica, passa-se da ‘ganância ou cobiça ao dinheiro dos outros’ à ‘inveja pelo que os outros têm (em termos concretos ou abstratos) e/ou ao que são’.

Cabe destacar, por fim, a existência de um homônimo, *usura*<sub>2</sub>, definido como brasileirismo por Houaiss e Villar (2009), que significa ‘corrosão que sofrem os materiais em função do tempo de uso, de fricção ou atrito’, ‘desgaste’ (HOUAISS; VILLAR, 2009). A formação que sugerem para a forma dista dos resultados da pesquisa. Segundo esses lexicógrafos, seria formada “irregularmente” com *uso* + *-ura* em lugar de *usado* + *-ura*, porém lexias derivadas em *-ura* originam-se de 3 fontes principais: 1) da nominalização de participios futuros latinos (lat *apertūra* > port *abertura*, lat *armātūra* > port *armadura*); 2) da sufixação de participios passados (port *fechado* + *-ura* → *fechadura*, *atado* + *-ura* → *atadura*); e 3) da sufixação de adjetivos (port *formoso* + *-ura* → *formosura*, *feio* + *-ura* → *feiura*). Embora a afixação de *-ura* a bases substantivas não seja impossível (*beleza* + *-ura* → *belezura*), são formações incomuns e o rastreamento leva a outra explicação.

*Le Trésor de la Langue Française informatisé*<sup>24</sup> data *usure*<sub>2</sub> (equivalente francês) de 1530 e o descreve como derivado de *user* ‘usar’ + *-ure*. Inicialmente considerada, a hipótese de sobrevivência e evolução do sentido latino ‘uso’ teve de ser descartada porque nenhuma língua românica examinada (nem o inglês) apresentou tal sentido antes da forma francesa e porque essa lexia é condizente com os mecanismos derivacionais em francês, onde *-ure* forma substantivos deverbais (funcionando distintamente de *-ura* em português). Isso é constatado no fr *chausser* ‘calçar’ + *-ure* → *chaussure* ‘sapato’ (compare-se com o port *calçado*) e fr *procéder* ‘proceder’ + *-ure* → *procédure* ‘processo’ (compare-se com o port *procedimento*). *Usura*<sub>2</sub> foi identificada em outras línguas românicas apenas entre finais do século XIX e início do XX, como o it *usura*<sub>2</sub> e o port *usura*<sub>2</sub>, época em que o francês ainda era a principal língua de prestígio internacional. Tudo isso corrobora a hipótese de que *usura*<sub>2</sub> seja empréstimo francês e tenha por étimo fr *usure*<sub>2</sub>, não sendo, portanto, formação vernácula. Acresça-se aqui ser essa informação também fornecida por dicionários italianos (cf. verbete no *Dizionario italiano De Mauro*).<sup>25</sup>

<sup>24</sup> Disponível em: <https://www.cnrtl.fr/definition/usure>. Acesso em: 18 ago. 2020.

<sup>25</sup> Disponível em: [https://dizionario.internazionale.it/parola/usura\\_2](https://dizionario.internazionale.it/parola/usura_2). Acesso em: 21 ago. 2020.

*Usura* faz parte do léxico herdado do latim, é derivada pela conversão de um particípio futuro latino, caracterizando-se como nome departicipial. Um resumo de sua etimologia está no Quadro 4.

#### QUADRO 4 – Etimologia de *usura*

port <i>usura</i> 1264-84 < <i>ūsūram</i> ~ lat <i>ūsūra</i> 191 a.C. ← <i>ūsūra</i> ~ <i>ūsūrus</i> <small>PART FUT</small> de <i>ūtī</i> 200 a.C.
---

Fonte: Elaboração própria.

### 4 *Usurar*

O verbo *usurare* é atestado para o latim medieval através do *Glossarium Ad Scriptores Mediae et Infimae Latinitatis* de Charles du Fresne du Cange. O material, que glosa vozes do latim medieval, foi consultado através do *Corpus corporum*, ali datado de 1710 (tomo 1). Esse glossário recebeu sucessivos acréscimos ao longo dos séculos XVIII e XIX<sup>26</sup> e, em adendo de 1766, é fornecida a datação de 1188 para o verbo *usurare*. A princípio, não se havia localizado este verbo em latim, e supunha-se um étimo reconstruído *\*ūsūrāre*, considerando-se a existência da forma verbal em todas as línguas românicas rastreadas, com exceção do romeno e do galego (cf. QUADRO 2). Houaiss e Villar (2009) fornecem como *terminus a quo* para o port *usurar* o século XIV, indicando a existência da lexia ainda em galego-português. Como apresentado acima, os verbos correspondentes nas outras línguas românicas, que ocorrem ainda na Idade Média, foram atestados nesta ordem: francês (sXIII), italiano (sXIV) e catalão (sXV). Além da atestação da própria forma latina, a antiguidade dessas lexias e sua difusão entre o grupo românico suportam, nesse caso, um étimo latino medieval, *usurare*, definido por Du Cange *et al.* (1883-1887) como ‘fazer usuras’ (*usuras producere*). Transcreve-se a seguir parte da abonação fornecida no verbete:

(...) *quamdiu debitor erit in peregrinatione, non Usuret.*<sup>27</sup>

(Bened. abb. Petroburg. de Gest. Henr. II. reg. Angl. ad ann. 1188.  
tom. 2. edit. Hearn. p. 498)

<sup>26</sup> O material pode ser conferido em: <http://ducange.enc.sorbonne.fr>. Acesso em: 12 jan. 2021.

<sup>27</sup> Em tradução livre: “enquanto o devedor estiver em peregrinação, não usure.”

O lat *usurare* seria, então, um verbo denominativo derivado a partir da conversão do lat *ūsūra* 191 a.C. A interpretação da base derivacional como nominal é sustentada por Houaiss e Villar (2009), e também pelos verbetes do *Tesoro della Lingua Italiana delle Origini*<sup>28</sup> e de *Le Trésor de la Langue Française Informatisé*,<sup>29</sup> porém os dados não apontam para a formação em cada vernáculo, como indicam esses materiais, e sim no próprio latim.

A prática usurária foi um tema jurídico e religioso durante a Idade Média, condenada em vários documentos (cf. Seção 3), ao que se pode supor que o verbo foi transmitido pela escrita, sendo um empréstimo erudito às línguas românicas. Os *corpora* medievais galego-portugueses em que se fez o rastreio não deram nenhum resultado para *usurar* e não se conseguiu, desta forma, confirmar em outros textos ou recuar o *terminus a quo* fornecido por Houaiss e Villar (2009), mas uma forma alternativa, a locução verbal *dar a usura* (que não é exclusiva ao português, cf. sardo *donau a usura*), se registra nas *Cantigas de Santa Maria* (1264-84; fonte: CIPM):

*Ca muit' é cousa sen guisa de fazeren avolezas  
os que creen ena Virgen, que é Sennor de nobrezas,  
que mais ama limpidõe que avarento requezas,  
e piadad' e mercee ca judeu dar [a] usura.  
(Afonso X, CSM312, 1264-84)*

O texto mais antigo em que o vocábulo foi localizado em português, nesta pesquisa, é o *Supplement au nouveau dictionnaire des langues françoise et portugaise*, de Joseph Marques, de 1775 (Google Livros). A abonação é o verbo a seguir:

*Agioter, v. a. Usurar, commerciar com usuras intoleraveis, e illicitas, fazer usuras.  
(Supplement au Nouveau Dictionnaire..., 1775, p. 19)*

Aqui pode ser observada a especialização e intensificação do sentido que seria prototípico a esse verbo, 'praticar usura, fazer empréstimos a juros', em que *usurar* é especificamente 'emprestar a juros abusivos' ou 'vender a preços exagerados', o que caracteriza ainda prática criminosa. O sentido sem essa especialização é abonado por um verbo

<sup>28</sup> Cf. <http://tlio.oiv.cnr.it/TLIO/>. Acesso em: 21 ago. 2020.

<sup>29</sup> Cf. <http://www.le-tresor-de-la-langue.fr/>. Acesso em: 21 ago. 2020.

do segundo tomo do *Diccionario da Lingua Portuguesa*, de António Moraes da Silva (1813), também consultado através do Google Livros:

**USURÁR**, v. n. *Dar dinheiro á usura, ou ao ganho.*  
(*Diccionario da Lingua Portuguesa*, Tomo Segundo, p. 824)

Ambos os significados permanecem em português e são comuns às outras línguas românicas. A relativa uniformidade semântica e o fato de as abonações dessa forma virem sobretudo de dicionários parece indicar que não teve ou teve pouca difusão popular, estando restrita a espaços jurídicos e ao debate econômico.

Pela proposição acima exposta, *usurar* seria um cultismo latino emprestado às línguas românicas na Idade Média, sendo um verbo denominal resultante da conversão de um nome de particípio. Resumida, sua etimologia é:

#### QUADRO 5 – Etimologia de *usurar*

port <i>usurar</i> sXIV <sup>HOUAIS</sup> < lat med <i>usurare</i> 1188 ← lat <i>ūsūra</i> 191 a.C.
---

Fonte: Elaboração própria.

### 5 *Usureiro/usurário*

A razão pela qual se preferiu reunir *usureiro* e *usurário* em uma mesma seção é o fato de serem resultados divergentes do mesmo étimo, situação comum a pares corradicais com os sufixos *-eir(o)* e *-ári(o)*. Para uma descrição pormenorizada do desenvolvimento histórico desses dois sufixos desde o latim até as línguas românicas e do espectro semântico que recobrem nesse grupo linguístico, consulte-se o estudo de Simões Neto (2020). As evidências mais antigas do lat *ūsūrārīus* são do século II a.C., ainda no latim arcaico (cf. Seção 3). É novamente em Plauto, na peça *Curculio*, que se encontra o exemplo mais antigo. O sentido primevo seria ‘que serve ou é adequado ao uso, de que se tem uso ou proveito’<sup>30</sup> (LEWIS; SHORT, 1879), o que tornaria a palavra latina semelhante ao port *utilitário*. A abonação é do *Corpus corporum*:

<sup>30</sup> Trad. nossa do original “That serves or is fit for use, of which one has the use or enjoyment”.

*cupio aliquem emere puerum, qui **usurarius** nunc mihi quaeratur.*<sup>31</sup>  
(Plauto, *Curculio*, 3, 1; 5. sII a.C.)

Como derivado em *-ārīus* (lat *ūsūrārīus* ← lat *ūsūra* 191 a.C. + *-ārīus*), *ūsūrārīus* apresenta na averbação o sentido que seria prototípico a formas com esse sufixo. Em outra obra atribuída também a Plauto, *Truculentus*, do mesmo século, é possível observar uma especialização semântica, denotando o derivado o sentido de ‘que diz respeito a juros, de que se cobra juros’<sup>32</sup> (LEWIS; SHORT, 1879). Tal constatação interessa muito à pesquisa etimológica, pois revela que os dois significados de *ūsūrārīus* já conviviam no século II a.C., o que implica ainda que *ūsūra* provavelmente teria um sentido especializado, relativo a *juros*, muito antes da primeira atestação dessa acepção, no século I a.C., e que ele conviveu por muito tempo com o sentido relativo a *uso*. As palavras, diz Viaro (2009b, p. 142), “já nascem polissêmicas”. O exemplo, novamente, é tomado do *Corpus corporum*:

*quos quidem quam ad rem dicam in argentariis referre habere, nisi pro tabulis, nescio, ubi aera prescribantur **usuraria**: accepta, dico, expensa ne qui censeat.*<sup>33</sup>  
(Plauto, *Truculentus*, 1, 1; 8. sII a.C.)

O primeiro significado latino foi perdido em português e nas línguas românicas, porém o segundo é atestado nas duas formas. O exemplo de *usurário* é do *Corpus do Português* (DAVIES, 2006-2018) e o de *usureiro* está no Livro IV da *Collecção da legislação antiga e moderna do Reino de Portugal* (1786), documento acessado pelo Google Livros. Esse livro transcreve as *Ordenações do Senhor Rey Dom Affonso V*, o que aponta (junto com a ortografia) para um original medieval, que teria sido escrito no reinado desse monarca, no século XV. Serão necessárias, porém, a expansão dos *corpora* e uma nova pesquisa para precisar a datação do texto.

<sup>31</sup> Em tradução livre: “desejo comprar alguém, um menino, que sirva ao uso que agora me é desejado”.

<sup>32</sup> Trad. nossa do original “Of or belonging to interest or usury, that pays interest”.

<sup>33</sup> Tradução: “Eu não sei dizer com que finalidade eles permanecem junto dos banqueiros, a não ser para servir de tabuinhas onde se fazem os registros dos dinheiros relativos a juros – refiro-me aos dinheiros recebidos, não vá alguém pensar em dinheiros gastos.” (PLAUTO, Trad. Adriano M. Cordeiro, 2010)

*Com ganho **usurário** recompensarei os esmaltes da minha descendência*  
(Amaro Roboredo, *Centúrias*, 1619-21)

*E aquelle que o contraíro fizer, e ouver de receber gaança algũa do dito  
contrauto, perca todo o principal, que deu, por aver a dita gaança [...]*  
*e per aqui entendemos, que poderá contrauto **usureiro** tam ilicito*  
*da nossa terra, e Senhorio seer esquivado.*  
(*Collecção da legislação antiga...* Livro IV. 1786, p. 94-95)

A transformação que viria a gerar um sufixo agentivo *-eir(o)* em português (e nas outras línguas românicas) se manifesta ainda em latim (cf. SIMÕES NETO, 2020; VIARO, 2011a), e é o que se pôde observar nas buscas para o artigo, com este exemplo de Ambrósio, Bispo de Milão (*Corpus corporum*):

*Usurarius est egenus, cogentibus vobis, habet quod reddat:*  
*quod impendat, non habet.*<sup>34</sup>  
(Ambrósio de Milão, *De Tobia*, 14. 0763B. sIV d.C.)

O que se identifica é uma metonimização. Da coisa *usurária*, passa-se à pessoa *usurária* e, finalmente, ao *usureiro/usurário*, o que pode ser representado pela sequência de paráfrases: ‘(a coisa) de que se cobra juros’ >> ‘(a pessoa) que cobra juros’ >> ‘a pessoa que cobra juros’, em que primeiro a palavra ganha um sentido ativo, mas ainda é dependente do núcleo de um sintagma nominal (a pessoa, o profissional *que faz X*), e depois, por omissão desse núcleo, torna-se ela mesma o elemento central do sintagma nominal (*o que faz X*, cf. VIARO, 2011a). Ambos os derivados possuem, em português, o significado agentivo. O primeiro exemplo é das *Cantigas de Santa Maria* (CIPM), que é também o *terminus a quo* de *usureiro*, e o segundo do livro *Maria rosa mystica* (1686), do Padre António Vieira, consultado via Google Livros:

*Ca ben como se lle ouvesse dito Santa Maria: “vai, e dar-ch-ey*  
*quito teu fillo do **usureiro** maldito”*  
(Afonso X, CSM062, 1264-84)

<sup>34</sup> Em tradução livre: “o usurário é pobre, recolhendo de vós, tem o que dá: o que gasta, não tem”.

[...] & com tudo Christo Senhor nosso diz que tendo hum destes  
**usurarios** dous devores, hum que lhe devia cincoenta dinheiros,  
 & outro quinhentos, a ambos perdoou a divida.  
 (Pe. António Vieira, *Maria rosa mystica*, v. 1, p. 473, 1686.)

Por uma intensificação e especialização do sentido, primeiro com *usureiro* (e depois com *usurário*), a referência que se faz é a alguém que ‘empresta a juros abusivos, vende a preços excessivos, um agiota’. O exemplo de *usureiro* (*vsureyros*) a seguir é de Machado Filho (2019, p. 693), e o de *usurário* é extraído do *Corpus do Português* (DAVIES, 2006-2018). Machado Filho (2019) traz também a definição ‘avarento’ para o mesmo verbete, porém esse sentido não é tão claro com a única abonação fornecida:

*confessese se tomou esmola dos publicos roubadores ou publicos vsureyros  
 ou dos desfazedores dos pobres ou dos que estan en mal querença.*  
 (Tratado dos sacramentos da ley antiga e NOVA, 1399. Fólio 64, reto, coluna 1)

*Porque o dinheiro que tem por idolo, & a quem em tudo obedece lhe manda  
 que jure falso, seja **usurario**, & venda por mais do justo preço, inda  
 que Deos vivo lho defenda.*  
 (Amador Arrais, *Diálogos*, 1589)

Com a abonação retirada de Davies (2006-2018), há um recuo no *terminus a quo* de *usurário* fornecido por Houaiss e Villar (2009) e Cunha (2010), já que ambos datam a forma de 1614, mas a ocorrência identificada nesta pesquisa é de 1589. *Usurário*, como forma divergente de *usureiro* a partir do mesmo étimo, transmite-se inicialmente por via erudita, tomado de empréstimo diretamente do latim escrito, mas ganha terreno onde se usava apenas *usureiro*, de transmissão popular. Quanto à conexão entre o *usureiro* e a característica de ser *avarento*, mesmo sem evidência textual definitiva de que existia já na Idade Média a intercambialidade ou sinonímia parcial das formas (p.ex., *velho avarento* = *velho usureiro*), é visível a associação entre a profissão e a *avareza*, como expresso nas *Cantigas de Santa Maria* (CIPM):

*Ena vila u foi esto avia un **usureiro**  
 mui riqu’ e mui’orgullos’ e sobervi’ e tortiçeiro;  
 e por Deus nen por sa Madre non dava sol nen dinneiro,  
 e de seu corpo pensava mui’ e de sa alma nada.*  
 (Afonso X, CSM075, 1264-84)

A conceptualização do *usureiro* como alguém *avarento* é recorrente na Idade Média (cf. LE GOFF, 2004 e a Seção 3 deste artigo), e não é difícil visualizar o desenvolvimento pelo qual, com atuação da metonímia, as duas formas se tornaram sinônimas, seja no Medievo (cf. MACHADO FILHO, 2019), seja em momento posterior. Note-se o paralelismo da evolução semântica de *usureiro* e *usurário*, intercambiáveis em português moderno. Houaiss e Villar (2009), por exemplo, limitam-se a indicar, na definição do verbete *usureiro*, a sinonímia com *usurário* e, outrossim, o faz Ferreira *et al.* (2004), acrescentando uma remissão a *avaro*. A próxima abonação de *usurário* (do Google Livros) exemplifica o significado *avarento* e a concretização da metonímia (que supõe ainda uma generalização) ‘é comum que *usurários* sejam *avarentos*’ >> ‘todo *usurário* é *avarento*’ >> ‘*usurário* = *avarento*’:

*E nem o lastimoso espectaculo de experimentarem vigorosamente as tres mayores perseguições de peste, fome, & guerra abrandava os animos dos **usurarios** & ambiciosos para deyxarem de perseguir com avareza & malicioso engano aos q̃ não haviam chegado à ultima miseria.*

(Luis de Menezes, *Historia de Portugal restaurado*, 1679, p. 874)

*Usureiro* não foi localizado em *corpora* dialetais do português brasileiro e, muito provavelmente, é uma palavra desusada; já *usurário* ‘avarento’ é comum no APFB e, no restante do Nordeste, está presente com o mesmo significado em Sergipe e na Paraíba (ARAGÃO, 2014; ARAGÃO; MENEZES, 1984), além de ser amplamente dicionarizada com essa acepção.

Como exposto acima, *usureiro* compõe o léxico herdado e *usurário* é um empréstimo latino do século XVI (talvez do séc. XV). Ambos se formaram por derivação sufixal a partir de um nome de participial (*usura*) e funcionam tanto como substantivos como adjetivos. O Quadro 6 resume a etimologia do par:

#### QUADRO 6 – Etimologia de *usureiro*/*usurário*

port <i>usureiro</i> 1264-84 < (* <i>usurairo</i> <) lat <i>ūsūrārĭum</i> ~ <i>ūsūrārĭus</i> sII a.C. (← lat <i>ūsūra</i> 191 a.C. + <i>-ārĭus</i> , a, um) > port <i>usurário</i> 1589
--

Fonte: Elaboração própria.

## 6 Usurável

*Usurável*, que no APFB apresenta os significados de ‘avarento’ e ‘invejoso’, é uma forma para a qual não se encontrou atestação em latim. O rastreamento permitiu identificar formas equivalentes em francês, em castelhano, em catalão, em italiano e em inglês, mas esse é um número de línguas insuficiente para uma reconstrução. Tanto o português como o catalão, além disso, poderiam ter recebido a forma como empréstimo de qualquer uma das outras, o que enfraqueceria ainda mais qualquer tentativa de reconstruir um étimo latino. A maior pista que se tem sobre a transmissão de *usurável* são as datações que a busca revela:

QUADRO 7 – *Termini a quo* de *usurável* e de equivalentes em outras línguas

francês	italiano	inglês	castelhano	catalão	português
<i>usurable</i> 1336	<i>usurabile</i> 1850	<i>usurable</i> 1879	<i>usurable</i> 1895	<i>usurable</i> 1913	<i>usurável</i> 1963
DMF (2015)	Google Livros	Google Livros	Pesquisa avançada do Google	Google Livros	APFB (1963)

Fonte: Elaboração própria.

É perceptível a enorme diferença entre a datação do francês *usurable*, 1336 (do *Dictionnaire du Moyen Français*<sup>35</sup> — DMF, 2015), e todas as demais. Mesmo que esses *termini a quo* possam ser recuados com o rastreio em novos *corpora*, não há, por enquanto, evidências de que uma forma como essa tenha existido em latim, e a proximidade das datações em inglês, italiano, castelhano e catalão torna improvável a formação independente da palavra em todas essas línguas, favorecendo o caso de um empréstimo francês. Não se deve defender verdades categóricas em questões de Etimologia, uma vez que tais verdades inexistem e as línguas, não raro, surpreendem a cada nova investigação. Sendo assim, toda descrição etimológica é uma proposta de caráter provisório, de acordo com o melhor que se tem em mãos, pronta para ser aperfeiçoada ou mesmo rejeitada à luz de novos dados. Dito isso, e com base não só nas datações, mas reconhecendo o francês como importante fonte de empréstimos para a língua portuguesa, o melhor que se pode propor é que

<sup>35</sup> Disponível em: <http://www.atilf.fr/dmf>. Acesso em: 21 ago. 2020.

*usurável* seja um galicismo, tendo por étimo o fr *usurable* ‘que se usura, usurário’, formado do verbo fr *usurer* ‘usurar’ + *-(a)ble*. A abonação do étimo a seguir é de Godefroy (1880-1895, p. 124):

*Il estoit tenus et obligies, en pluseurs gries debtes **usurables**, lesquelles paier ne pooit se, par droite necessité, aucuns de ses hiretages n'estoient vendus. (C'est dou pooir que Jehans de Tournay..., 1336. S.-Brice, Arch. Tournai.)*

O mesmo significado (não mais encontrado no francês, em corpora recentes) se encontra em português, como se observa no exemplo a seguir, extraído do portal do Senado Federal:<sup>36</sup>

*Por exemplo, enquanto o sistema financeiro, tomando dinheiro ao preço de uma inflação que, felizmente, mal alcança 1%, chega a emprestar esse dinheiro a até 10% ao mês, ou dez vezes a inflação, como no caso de cheques especiais e cartões de créditos, num País de cidadãos endividados, contribuindo de forma **usurável** com a estabilização da moeda [...]*  
(Pronunciamento de José Alves em 15/01/1999)

Nota-se, em português, mais uma vez por metonímia, que a paráfrase ‘que se usura, de que se cobra usura’ também pode ser interpretada como ‘que usura, que cobra usura’, em outras palavras, não só caracteriza um *paciente*, mas ainda um *agente*, e disso resulta transformação análoga à de *usurário* e *usureiro*, já que a associação ao sujeito da prática usurária gera a sinonímia de *usurável* e *avarento*, o que é documentado no APFB (ROSSI *et al.*, 1963):

*[usurável] gente ruim, que tem pena de dar de comer a uma pessoa.*  
(APFB, Carta 104, P-16, inf. A)

Mesmo que os textos não ajudem tanto na definição de uma cronologia, o desenvolvimento de *usurável* encontra muitos paralelos em *usurário/usureiro* e *usura*, e é pela história da base lexical que pode ser explicada a polissemia de *usurável*. Pela sinonímia *cobiça* = *ganância* = *usura*, pode-se supor metonímia e pejoratividade atuando sobre o significado ‘que cobra usura’, o que geraria o sentido ‘cobiçoso, ganancioso’ para essa palavra. Em outros termos: o sujeito que dá

<sup>36</sup> Disponível em: <https://www25.senado.leg.br/web/atividade/pronunciamentos>. Acesso em: 21 ago. 2020.

empréstimos usurários, o faz porque tem cobiça ou ganância, logo *usurável* é ‘que(m) cobiça’. Por fim, uma especialização restringiria essa cobiça ao que é alheio, e *usurável* seria ‘que(m) cobiça (especificamente) o alheio (os bens, as características, a posição etc.)’. Essas observações são apontadas em comentários dos informantes do APFB (em notas):

[*usurável*] quando tem usura pelo que é dos outros.

(APFB, 1963, Carta 104, P-10, inf. A)

[*usurável*] pessoa que não pode ver outra com nada, olho grosso, invejoso.

(APFB, 1963, Carta 104, P-7, inf. A)

[*usurável*] invejoso, tem usura pelo que é dos outros.

(APFB, 1963, Carta 104, P-29, inf. A)

Quanto às suas realizações fonéticas no APFB, a forma *usurável* ocorre com aférese em dois pontos: [zu' ravi] (ponto 40) e [zu' ravi] (ponto 30). Essas variantes (junto com [zu' rētu], para *usurento*, e [zu' raru], para *usurário*), como se verá nas seções 9 e 10, foram importantes para a explicação da formação de *zura* e *zuraco*. Há ainda uma realização que mostra flexão de gênero: [uzu' rava] (ponto 48); o que é interessante, visto que adjetivos em *-vel* não costumam ser flexionados. Uma explicação possível seria a redução do segmento final e sua realização como [vu], como em [uzu' ravu] (pontos 23 e 25), que permitiria a analogia com outras formas em *-o*, que geralmente admitem flexão.

Cardoso e Ferreira (2000), em *O léxico rural*, registram a existência de *usurável* com o significado de ‘avarento’ em Sergipe, que, junto com a Bahia, pertence à área dialetal baiana (NASCENTES, 1955). Essa forma não foi encontrada em nenhum outro *corpus* dialetal consultado, o que parece indicar uma distribuição diatópica, sendo *usurável* mais comum nos falares baianos. Deve-se mencionar, por fim, a identificação de um homônimo, *usurável*<sub>2</sub>, ‘que pode sofrer desgaste ou corrosão, fácil de se corroer’, pelo Google Livros. Sua etimologia ainda não é suficientemente conhecida, mas não teria relação com *usurável*<sub>1</sub> ou *usura*<sub>1</sub>. Por seu sentido, há uma clara associação com *usura*<sub>2</sub> (cf. Seção 3). Equivalentes estão presentes em francês, inglês e italiano, porém os dicionários só registram o it *usurabile*. Nessas línguas, tais formas ocorrem em lojas virtuais, nas indicações de garantia dos produtos, p.ex. “as partes usuráveis (i.e., desgastáveis) não são cobertas pela garantia”.

Conforme a proposta que aqui se expôs, *usurável* seria um empréstimo francês, embora a datação de sua entrada em português ainda não esteja bem esclarecida. É um adjetivo deverbal, formado por derivação sufixal do verbo fr *usurer*, sendo sua etimologia resumida a seguinte:

QUADRO 8 – Etimologia de *usurável*

port <i>usurável</i> 1963 < fr <i>usable</i> 1336 ← fr <i>usurer</i> 1300-50 + -(a)ble
--

Fonte: Elaboração própria.

## 7 *Usuroso*

*Usuroso* não teve equivalente identificado nos textos latinos, e é uma forma que oferece certo número de complicações. Primeiro, só conta com uma realização no APFB, [auzu'rozu] (Carta 105, P-17, inf. B), e poderia ser apenas uma instância de criatividade lexical, em que o falante acede ao *léxico virtual* (BASILIO, 2004); segundo, a coexistência de duas bases nas línguas românicas, *usura* e equivalentes (presentes em todas as línguas românicas e no inglês) e outra que teria por étimo o latim medieval *usuria* (que foi encontrada como palavra em galego, castelhano, francês, sardo e inglês e somente como base de derivados no catalão e no italiano), cria divisões na preferência por uma ou outra e impede o estabelecimento de relações diretas necessárias à tentativa de reconstrução do étimo. E, finalmente, as datações não permitem o estabelecimento de uma cronologia mais ou menos regular.

Até o momento de finalização do rastreamento (julho/2020), formas em *-os(o)* a partir de bases equivalentes ao português *usura* foram encontradas em castelhano, italiano e inglês. De todos os materiais utilizados na busca, *usuroso* só gerou resultados em português no APFB, em *O léxico rural* (CARDOSO; FERREIRA, 2000) e na pesquisa avançada do Google. *O léxico rural*, no entanto, glosa as lexias do APFB e do *Atlas Linguístico de Sergipe* (FERREIRA et al., 1987), e a ocorrência da forma ali deve-se ao APFB. Como o resultado mais antigo fornecido pela pesquisa avançada do Google é de março de 2020, a datação mais recuada que se tem é a do APFB, 1963. Em inglês, o Google Livros permite datar *usurous* de 1716, mas essa forma é definida no *Collins*

*English Dictionary*<sup>37</sup> como ‘uma variante obsoleta de *usurious*’,<sup>38</sup> sendo o único dicionário consultado que a registra. Em castelhano e italiano (cf. QUADRO 2), a pesquisa avançada do Google encaminhava a publicações em blogs de 2012 e 2016, respectivamente. Nenhuma dessas datas é convincente como *terminus a quo*, ainda mais considerando que as formas românicas em *-oso* partindo do lat *usura* começam a aparecer no século XV, sendo o fr *usurieux* a primeira, em 1497 (DMF, 2015), e seria estranho o empréstimo inglês nesse caso, sendo a forma pouco frequente naquela língua e estruturalmente latino-românica.

Curioso é que os derivados de **usur-** em *-oso* estão quase inteiramente ausentes dos dicionários. Somente o ing *usurous*, como dito, está em um dicionário moderno, e o fr *usurieux* no DMF (2015), o que pode indicar a atuação pancrônica de um paradigma sufixal latino extremamente produtivo, *X-ōsus*, que se espalhou para fora dos limites românicos (o inglês é língua germânica) e que forma lexias transparentes, facilmente interpretadas através da paráfrase desse sufixo, ‘provido de X’. Sendo assim, se o conhecimento de língua do informante no APFB põe o par *usura* e *avareza* como sinônimos, quem tem *avareza* é o *usuroso* assim como quem tem ganância é o *ganancioso*. A mesma explicação seria válida para qualquer outro significado de *usura*. A real resposta, no entanto, é que ainda não se sabe o suficiente, e a etimologia é incerta. Sabe-se que a base é *usura* ou cognato em outra língua e o sufixo é *-os(o)* ou equivalente estrangeiro, mas a palavra realmente surgiu independentemente em todas essas línguas? Quando surgiu? Transmitiu-se já formada de uma língua às outras? Nada disso é conhecido ainda.

Uma consulta ao *Twitter* (que não gera resultados pelo Google) em junho de 2020 identificou que *usuroso* tem certa frequência em publicações em língua espanhola e italiana na plataforma, porém isso é outra pesquisa em si mesma e não poderia ser tratado neste artigo. Como abonação de *usuroso* em português, o trecho abaixo é da seção de comentários do arquivo de um canal excluído do *YouTube*, encontrado através da pesquisa avançada do Google, em que há clara depreciação no uso da palavra:

<sup>37</sup> Disponível em: <https://www.collinsdictionary.com/dictionary/english/usurous>. Acesso em: 21 ago. 2020.

<sup>38</sup> Trad. nossa do original “an obsolete variant of *usurious*”.

[...] a maioria dos cristão dessa geração é **usuroso**, vigarista, mentiroso, ganancioso, individualistas, enroladas e caloteiros. Apenas usa o termo e a igreja para fazer nome, se popularizar, galgar cargos criar mecanismos corporativismo para interesses particulares.

(Abraão Ferreira O Realista.<sup>39</sup> Seção de comentários, março de 2020)

Em resumo, *usuroso* é um adjetivo denominal de etimologia incerta, formado por derivação sufixal a partir de um substantivo departicipial.

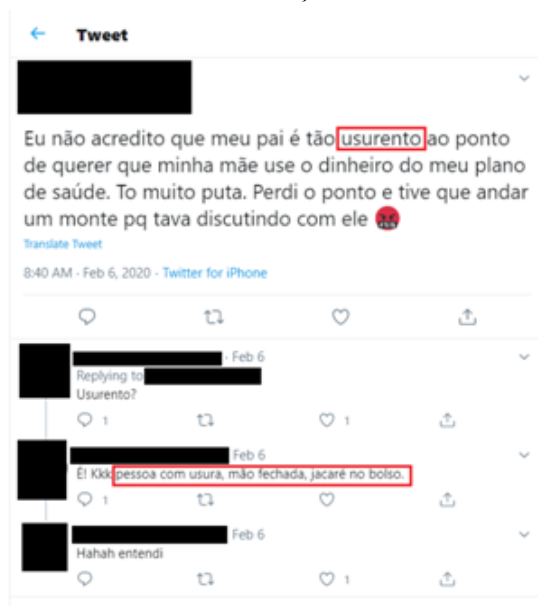
## 8 *Usurento*

Apesar de mais comum no APFB que *usuroso*, *usurento* (4 ocorrências) também não teve um equivalente latino identificado, e tampouco foi observada a ocorrência de formas equivalentes nas outras línguas, com exceção apenas do castelhano. O cast *usuriento* foi identificado em uma publicação de blog de 2012, através da pesquisa avançada do Google (não se sabe ainda que peso a ausência do galego nas opções desse mecanismo teria tido aqui). A oposição entre português e castelhano e as outras línguas românicas pesquisadas pode apontar para uma conexão ibero-românica, uma vez que o sufixo *-ent(o)* é produtivo nas duas línguas. Apenas com dados do espanhol e do português e sem evidências claras de empréstimo, a explicação de que o derivado foi formado independentemente em ambas seria mais provável que a reconstrução de um étimo latino em *X-entus*. Essa é uma das dificuldades da comparação entre línguas muito próximas: suas estruturas são semelhantes, e, onde convergem, é difícil atribuir a origem a um ou a outro sem evidência textual. Acrescentar dados de outras línguas ibero-românicas, como o leonês e o aragonês, poderia ajudar, mas não sem levantar o mesmo problema. Melhor seria, de fato, uma atestação em latim, ou fora do grupo ibero-românico para confirmar que a forma é herdada, o que parece não ser o caso.

A forma pode ser ainda um hispanismo em português, ou um lusitanismo em espanhol, ou ser uma formação independente nas duas línguas. Como no caso de *usuroso*, não se sabe o suficiente para uma resposta segura. Se interpretada através da paráfrase do sufixo *-ent(o)*,

<sup>39</sup> Disponível em: <https://armemory.info/users/7hw3qJxXCG4sBZuMP7IEaA.html>. Acesso em: 21 ago. 2020.

‘que tem X’ ou ‘que faz X’, comumente com sentido pejorativo e frequentativo (cf. port. *briguento*, *catarrento* etc.), é uma formação transparente e derivaria da base *usura* ou o equivalente em latim, espanhol ou em outra língua ibero-romance, com o sufixo supracitado. Essa derivação pode ter ou não sofrido analogia formal de *avarento*. Uma consulta ao *Twitter*<sup>40</sup> (08/06/2020) também identificou *usurento* na plataforma, e essa é a fonte do seguinte exemplo:

FIGURA 1 – Abonação de *usurento*

Fonte: *Twitter*.

Recapitulando, *usurento* tem etimologia ainda incerta. É formado por derivação sufixal a partir de um substantivo de participial, *usura* ou equivalente em outra língua, sendo, portanto, um adjetivo denominal.

## 9 Zura

A identificação de *zura* (e também de *zuraco*) com a família lexical estudada foi estabelecida por duas razões: 1) a indicação dos dicionários

<sup>40</sup> Disponível em: <https://twitter.com/>. Acesso em: 08 jun. 2020.

(FERREIRA *et al.*, 2004; HOUAISS; VILLAR, 2009; Priberam; Caldas Aulete; *Infopédia*) de que essa forma derivaria de *usurário*; e 2) as realizações aferéticas de *usurário*, *usurável* e *usurento* no APFB ([z] *uraro*, [z] *urave* e [z] *urento*). Mesmo que *zura* não ocorra nesse atlas, há certo padrão fonético em todas essas formas e seu significado ‘avarento, sovina’ cria inegável conexão entre ela e as outras lexias populares. De acordo com o *Dicionário Gaúcho* (OLIVEIRA, 2002), acessado pelo Google Livros, *zura* seria regionalismo do Rio Grande do Sul, mas se sabe pouco sobre sua distribuição diatópica para confirmar isso.

Admitida a hipótese dos dicionários, tratar-se-ia de um *truncamento* (GONÇALVES, 2016), com a supressão de partes da palavra para gerar um derivado. Dado que a forma não possui *u*-inicial, não derivaria diretamente de *usurário*, mas de *zuraro*. Assim se teria *zuraro* → *zura*, desenvolvimento comparável aos exemplos de Gonçalves (2016): *português* → *portuga*, *comunista* → *comuna*. O maior problema seria a datação, pois *zuraro* é atestado, nos materiais desta pesquisa, em 1963, e *zura* pôde ser datado de 1954, pelo Google Livros. Essa diferença é, contudo, pequena, não rechaçando completamente a possibilidade de *zuraro* ser mais antiga, e a fonética parece fortalecer a proposta da variante aferética de *usurário* como base derivacional, já que a aférese atuou e continua atuando na história no português, p.ex., lat *episcōpum* > port *bispo* e port *aguentar* > [gwẽ'ta] (VIARO, 2014), corroborando a ideia de que *usurário* > *zuraro* anteceda a formação de *zura*.

O texto mais antigo em que *zura* foi encontrado é um periódico identificado pelo Google Livros como *Anhembi* (1954, v. 14), sem identificação de autoria. A forma está na página 371,<sup>41</sup> no trecho transcrito a seguir:

*Escuta, galego zura, você já juntou dinheiro que dá prá pagá um navio prá trazê os peitos dela, e outro prá o rabo dela, também, não é?*  
(*Anhembi*, v. 14, p. 371, 1954)

Há claro sentido depreciativo e vulgar no uso acima, o que seria também uma concordância semântica com alguns exemplos de truncamento (*comuna*, *vagaba*). ‘Avarento’ é o único sentido encontrado

<sup>41</sup> Disponível em: <https://books.google.com.br/books?id=TEUMAQAIAAJ&dq=editions:STANFORD36105128997256&hl=pt-BR&lr=>. Acesso em: 21 ago. 2020.

para essa forma. Todos os dicionários de língua portuguesa consultados apontam *zura* como brasileirismo, e nenhum equivalente foi encontrado nas outras línguas românicas.

*Zura* muito provavelmente é derivado de *zuraro*, variante fonética de *usurário*, e possui conotações pejorativas. É um adjetivo deadjetival, tendo sido formado por truncamento da base derivacional. O quadro seguinte resume sua etimologia:

#### QUADRO 9 – Etimologia de *zura*

port <i>zura</i> 1954 ← port <i>zuraro</i> ≈ <i>usurário</i> 1589 (< lat <i>ūsūrārius</i> sII a.C.)
---

Fonte: Elaboração própria.

### 10 *Zuraco*

Como *zura*, *zuraco* também é registrado nos cinco dicionários de língua portuguesa consultados. Não foi encontrado em outras línguas, e seria exclusivo ao português. Ferreira *et al.* (2004) e Houaiss e Villar (2009) apontam-no como regionalismo do Sul. O *Dicionário Gaúcho* (OLIVEIRA, 2002) e o *Dicionário de cearensês: a cultura do povo cearense* (CAVALCANTE, 2012) apontam *zuraco* como regionalismo gaúcho e cearense, respectivamente, o que na verdade implica que a *lexia* tem distribuição mais ampla no Brasil. Todos a definem como ‘avarento’, mas poderia ser usada em outros sentidos de *usurário*, conforme revela a abonação seguinte (Google Livros):

***Zuraco***, adj. *Usurario*. -« Este **zuraco** empresta dinheiro a vinte por cento ao anno. »

(*Revista do Instituto Histórico e Geográfico de São Paulo*, v. 16, p. 223, 1914)

O que os dicionários fornecem como etimologia para *zuraco* é que teria sido derivado de *zura* + *-aco*, tendo o sufixo uma carga pejorativa (cf. *-aco*, em Cunha (2010)). A datação identificada nesta pesquisa impõe, todavia, um problema. É verdade que o *terminus a quo* encontrado para *zura* (1954) poderia ser recuado com uma expansão dos *corpora*, mas um fator importante deve ser acrescido à equação: os dicionários ignoram ou desconhecem as três formas com aférese encontradas no APFB (*zuraro*, *zurave* e *zurento*), logo a única palavra que lhes resta para comparação é

*zura*. Considerando a frequência da aférese de vogais iniciais (cf. VIARO, 2014), a maior antiguidade de *zuraco* dentro dos limites deste artigo, a produtividade da base *usura* e a existência de um paradigma aferético nessa família léxica, não é implausível propor que *zuraco* (1914) seja na verdade derivado de *\*zura* (< *usura*) + *-aco* ou ainda resultado da queda do [u] inicial em *\*usuraco* (← *usura* + *-aco*). Não se pode dar uma resposta definitiva, e uma explicação dupla é possível nesse caso, reforçada pelos dados do português popular.

*Zuraco* é um adjetivo denominal, formado por derivação sufixal. O afixo *-aco*, que é depreensível nessa palavra, tem carga semântica pejorativa. A pesquisa aponta para uma explicação etimológica dúplice, como se resume no Quadro 10:

QUADRO 10 – Etimologia de *zuraco*

port <i>usura</i> 1264-84 > port <i>*zura</i> ‘usura’ + <i>-aco</i> → port <i>zuraco</i> 1914 < port <i>*usuraco</i> ← port <i>usura</i> 1264-84 + <i>-aco</i>
---

Fonte: Elaboração própria.

## 11 *Usurariamente*

*Usurariamente* é um derivado encontrado principalmente em dicionários (Priberam, *Infopédia* e Caldas Aulete) e tem constituição morfológica transparente, como a maioria das palavras em *-mente*. Não parece ter difusão fora da linguagem escrita e formal. A pesquisa identificou cognatos em francês, castelhano, catalão e italiano. A forma francesa, *usurairement*, datada de 1448, é a mais antiga e é o étimo mais provável para as outras formas românicas. O maior indício disso é que a atestação mais recuada desse advérbio em italiano (1717), castelhano (1744) e português (1758) está em dicionários bilíngues de francês (consultados pelo *Google Livros*). A própria ocorrência no dicionário da Porto Editora é como tradução do fr *usurairement*. A forma catalã, *usurariament*, é atestada em um dicionário multilíngue de 1839 (*Diccionari Catalá-Castellá-Llatí-Frances-Italià*, 1839, de autoria não identificada), e é temporalmente mais distante dos outros. Considerada, porém, a posição do francês como língua de prestígio na maioria das cortes europeias no século XVIII, a hipótese mais provável é o empréstimo ou decalque a partir do francês, dada a natureza dos

documentos onde são atestadas (dicionários bilíngues), não obstante serem possíveis como formações vernáculas. O fr *usurairement*, por sua vez, é derivado de *usuraire* 1340 (DMF, 2015) + *-ment* e sua primeira atestação seria a seguinte (GODEFROY, 1895-1902):

*Ont contracté usurairement avec gens de tous estats.*  
(Mai 1448, Ord., XIV, 20.)

Tem sentido prototípico, considerada sua base e o sufixo *-ment*, também de significação aparentemente uniforme. Esse é o mesmo sentido encontrado em português, e não foram identificadas transformações semânticas. A primeira atestação é no *Nouveau Dictionnaire des Langues Françoise et Portugaise* (1758),<sup>42</sup> de Joseph Marques (Google Livros):

*Usurairement, d'une manière usuraire. Usurariamente, por via de usura.*  
(*Nouveau Dictionnaire Des Langues Françoise Et Portugaise...* v. 1. 1758, p. 668)

*Usurariamente* é um galicismo. É um advérbio deadjetival, formado em francês por derivação sufixal em *-ment* a partir do fr *usuraire*. Sua etimologia pode ser resumida no Quadro 11.

#### QUADRO 11 – Etimologia de *usurariamente*

port <i>usurariamente</i> 1758 < fr <i>usurairement</i> 1448 ← fr <i>usuraire</i> 1340 + <i>-ment</i>
---

Fonte: Elaboração própria.

## 12 *Usurador*: confusão ou criação neológica?

Os primeiros rastreamentos não identificaram *usurador* em português. Uma das dificuldades para encontrá-lo foi a confusão que os mecanismos de pesquisa fazem entre essa forma e *usurpador*, que aparece como resultado nas buscas mesmo que entre as opções se selecione exatamente a primeira e se exclua a segunda de maneira explícita. Não raro, como se aludiu na seção de metodologia, a palavra estava em duas linhas (linha 1 *usur-* e linha 2 *pador*) e a sequência era identificada como *usurador*. Mas usos inequívocos dessa *forma*, seja ela considerada um

<sup>42</sup> Disponível em: <https://books.google.com.br/books?id=3cRKAAAACAAJ&printsec=frontcover&hl=pt-BR#v=onepage&q&f=false>. Acesso em: 21 ago. 2020.

membro da família lexical explorada neste estudo, uma variante de *usurpador* ou mesmo uma amálgama das duas coisas, acabaram sendo localizados em revisões posteriores do trabalho. Ausente em todos os *corpora* de sincronias pretéritas, nos dicionários de língua e dicionários etimológicos de português, só foi identificada com a busca avançada do Google, em publicações na internet.

Dentro do conjunto de línguas examinadas, derivados em *-dor* ou equivalentes nessa família léxica foram encontrados em latim, galego, castelhano, catalão, francês e inglês (cf. QUADRO 2). O lat *usurator* ‘usureiro’ foi identificado em um vocabulário medieval recolhido por Thomas Wright (WRIGHT, 1884)<sup>43</sup> e é datado do século XV. A identificação em galego deu-se de forma indireta, pois está no título de um *e-book* anunciado pelo site da Amazon em espanhol, *Usurador do pavo real* (WURFEL, 2020),<sup>44</sup> sobre o qual se aponta: “edição galega”. A contração da preposição com o artigo definido (*de + o* → *do*) também impossibilita que o título esteja em castelhano. As ocorrências desta palavra para o castelhano e para o catalão foram encontradas pelo Google Livros. As atestações mais antigas, considerados os limites da pesquisa, estão no mesmo documento, o volume 5 da *Enciclopedia moderna catalana*, de Joseph Fiter (1913). Ela nos fornece o cat *usurador* como tradução para o fr *usuraire* e traduz essa mesma palavra com o cast *usurador* e *usurero*. A forma francesa, *usureur*, já existia no francês antigo (séc. XII) como *usureor* e é assim registrada em Godefroy (1880-1895). O rastreamento deste trabalho também identificou a palavra em francês moderno, em publicações na *web* (sobre a evolução de derivados *X-tōr* em francês cf. Nyrop, 1903 e o étimo lat *peccātōrem* > fr *pêcheur* ‘pecador’). A forma inglesa, *usuror*, por sua vez, pôde ser datada de 1686 através do *Google Livros*. Seria prematuro atribuir como étimo das demais lexias a forma do latim medieval, dada a diferença nas datações e o fato do vocábulo francês apresentar-se com um resultado popular (*usureur* e não \**usurateur*), o que faria do derivado latino em *X-tōr* muito mais antigo

<sup>43</sup> Acedeu-se ao material através do *Corpus of Middle English Prose and Verse*, mantido pela Universidade de Michigan. O texto pode ser consultado em: <http://name.umdl.umich.edu/CME00034>. Acesso em: 21 ago. 2020.

<sup>44</sup> Disponível em: [https://www.amazon.es/Usurador-do-pavo-real-Galician-ebook/dp/B08415KB5G/ref=sr\\_1\\_1?\\_encoding=UTF8&dchild=1&qid=1597159107&refinements=p\\_27%3A+Dennis%5C+Wurfel&s=digital-text&sr=1-1](https://www.amazon.es/Usurador-do-pavo-real-Galician-ebook/dp/B08415KB5G/ref=sr_1_1?_encoding=UTF8&dchild=1&qid=1597159107&refinements=p_27%3A+Dennis%5C+Wurfel&s=digital-text&sr=1-1). Acesso em: 21 ago. 2020.

que a atestação do século XV. Curiosamente, entre as línguas românicas, essas formas se manifestaram (ou se preservaram) apenas em línguas da România Ocidental, embora os dados não deem ainda segurança a respeito de sua inexistência na România Oriental.

Em português, a atestação mais antiga obtida pela pesquisa avançada do Google é de uma publicação de 2013, em um fórum sobre futebol:<sup>45</sup>

*Portanto vendo Palandrê e o Irmão do Sócrates, questionarem o **usurador** de medalhas da CBF, vejo que estão batendo em prego enterrado [...].*  
(Meu Timão. Profissionais do futebol, são muito engraçados mesmo. 30/08/2013)

Há argumentos que podem sustentar a hipótese de que a forma não resulta de uma confusão com *usurpador*. O primeiro é que nessa situação específica *usurpador* não se encaixa muito bem. Embora tenha contiguidade de sentido com formas em **usur-**, por exemplo, *usurário*, no que se refere ao apoderamento injusto e abusivo de algo, *usurpador* não está necessariamente associado ao acúmulo, à avarizia, como está a série de derivados discutidos anteriormente. O contexto não corrobora a interpretação da usurpação de medalhas, e sim de seu acúmulo e/ou da cobiça a elas, o que geraria uma paráfrase plausível para uma forma *usurar* + *-dor*. O segundo é que tal formação é possível (existindo em línguas muito íntimas ao português), seguiria os padrões derivacionais comuns e se valeria de um esquema estabelecido e produtivo, *X-dor* (*armador*, *pegador* etc.). A estes um terceiro argumento poderia ser acrescido: palavras em **usurp-** estão mais circunscritas a poucos contextos, enquanto a família lexical de *usura* não só é relativamente mais produtiva dentro e fora da România (como o rastreamento feito constata), como também está presente na fala popular (nos atlas linguísticos) e em contextos informais na *web*.

Todavia, são poucos os resultados fornecidos pela busca avançada do Google e fez-se necessário, para esta lexia específica, recorrer a outros materiais para o contraste dos exemplos e o resgate de novas abonações que fortaleçam (ou enfraqueçam) a hipótese de *usurador* como membro da família lexical estudada em português. Foi-se então

---

<sup>45</sup> Disponível em: [https://www.meutimao.com.br/forum-do-corinthians/off-topic/37949/profissionais\\_do\\_futebol\\_sao\\_muito\\_engracados\\_mesmo?pag=6](https://www.meutimao.com.br/forum-do-corinthians/off-topic/37949/profissionais_do_futebol_sao_muito_engracados_mesmo?pag=6). Acesso em: 21 ago. 2020.

ao *Twitter*, cujo mecanismo de busca permite restrições semelhantes à ferramenta avançada do Google. A consulta, feita em 29/07/2020, teve como resultados onze publicações. A primeira delas, de 2016, faz referência ao ex-presidente Michel Temer (2016-2018):

FIGURA 2 – Abonação de *usurador*



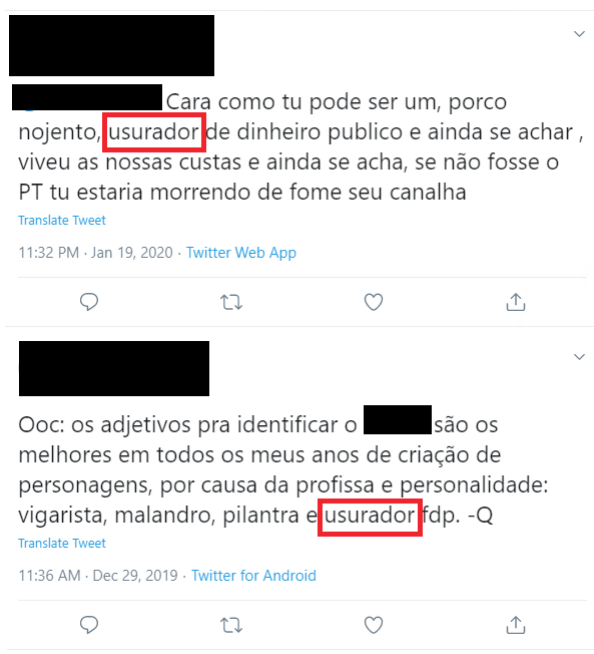
Fonte: *Twitter*.

Tendo assumido o governo durante o controverso processo de *impeachment* que levou à cassação do mandato da então presidenta Dilma Rousseff e sido classificado como golpista pelos apoiadores de sua antiga companheira de chapa, é fácil enxergar no exemplo uma referência a Temer como *usurpador*. Tratar-se-ia, como se poderia argumentar, de uma confusão com um MLB mais comum (**usur-** por **usurp-**) ou ainda um erro de digitação. Verdade seja dita, essa não é uma explicação impossível, mas ela desconsidera a existência da forma em outras línguas e a possibilidade de uma manifestação independente de *usurador* em português, que congregaria naquele uso a carga pejorativa comum a seus corradicais e sofreria, reconheça-se, certa analogia semântica com *usurpador*, de significação contígua.

Situação análoga é identificada em espanhol, com referências no *Twitter* a Nicolás Maduro, presidente da Venezuela, como cast *usurador*. Neste caso, o uso é muito mais frequente e certamente não se trata de mero equívoco ou confusão. A primeira ocorrência do cast *usurador* nessa rede social é de 2010, mas a forma se torna mais frequente depois do início do

governo Maduro, em 2013. Impõe-se aqui uma questão: seria *usurador* um empréstimo do espanhol venezuelano? Não necessariamente. Todas as razões já expostas e o primeiro exemplo trazido apontam para certa independência de *usurador* em relação a *usurpador*, porém a intercambialidade (i.e., sinonímia) entre as duas formas, verificável nas referências a Temer, talvez tenha sido influenciada pelo uso venezuelano. Se houve empréstimo da *forma*, ele não pode ser relacionado diretamente ao espanhol venezuelano pelas datações, e a plausibilidade da criação neológica enfraquece a hipótese de hispanismo. Os exemplos abaixo, em português, mostram usos menos ligados a *usurpador*:

FIGURA 3 – Abonações de *usurador* no Twitter



Fonte: *Twitter*.

Embora não sejam casos em que a substituição por *usurpador* é inadmissível, seria igualmente aceitável a troca por *usurário* ou outro sinônimo corradical. Apesar da existência de uma forma latina, o vácuo temporal que existe entre esta e a forma francesa e as formas castelhana e catalã e a recência da atestação galega e dos registros em português

parecem não sustentar a proposição de um étimo latino, pelo menos não para todas essas línguas. É possível que o latim vulgar do norte Gália tenha tido um derivado *\*usuratore*, mas se essa forma teve difusão fora da área onde os dialetos de *langue d'oïl* se formariam, poderia ter se arcaizado ou se perdido. O fr *usureur* sXII poderia ser o étimo do ing *usuror* 1686, mas sua esporadicidade levanta dúvidas. Menos provável ainda seria a palavra francesa, que não é dicionarizada, ter sido decalcada e dicionarizada no castelhano e no catalão no início do século XX como *usurador*, e o mesmo pode ser dito das formas galega e portuguesa: embora antigo, o vocábulo francês teve poucas atestações ao longo do rastreamento (que englobou sincronias diferentes) e o caso é distinto do aparecimento de *usurável* (que entra em várias línguas num curto espaço de tempo, 500 anos depois da atestação em francês) ou de *usurariamente* (cujas abonações mais antigas estão em dicionários bilíngues de francês). O fato de não se ter localizado atestações de formas em *X-dor* ou equivalentes em *corpora* medievais de outras línguas românicas de igual maneira dificulta a proposição de um étimo no latim vulgar, ainda que isso permaneça no campo da possibilidade. Por outro lado, a ocorrência aparentemente pontilhada no tempo das formas românicas talvez aponte, como no caso de *usuroso*, para a manifestação não de uma palavra, mas de um esquema derivacional pancrônico herdado por todas elas.

*Usurador* tem etimologia incerta, mas integra o impressionante grupo de oito sinônimos corradicais identificados nesta pesquisa. Sua constituição mórfica é transparente, com o tema de um verbo, *usurar* ou equivalente, e um sufixo derivacional, *-dor* ou equivalente. Tal seria válido nas várias hipóteses delineadas: criação neológica, empréstimo de outra língua românica ou ainda étimo no latim vulgar. É, assim, um substantivo deverbal.

### 13 Sinônimos corradicais em *usur-/zur-*: convivência, frequência, variação e mudança

A observação panorâmica da família lexical de *usura* conduz, inevitavelmente, à constatação da manifesta representatividade quantitativa das formas nominais sufixadas geradas em seu interior: *usurário*, *usureiro*, *usurável*, *usuroso*, *usurento*, *usurador* e *zuraco*. Das onze formas que compõem a totalidade do conjunto, sete são vozes nominais/adjetivais sufixadas (cerca de 2/3 dos elementos desse

paradigma léxico-semântico), todas correspondendo, junto com a forma truncada *zura*, ao sentido de ‘pessoa sovina’.

Se, de um lado, o emprego, numa mesma área dialetal (a dos falares baianos, captados pelo APFB), de pelo menos quatro dessas oito formas derivadas veiculadoras de um mesmo sentido comum (*usurário*, *usurável*, *usuroso* e *usurento*)<sup>46</sup> evidencia a sinonímia entre elas, podendo ser tomadas como um caso de *variação morfológica* (BAGNO, 2007; SIMÕES NETO, 2018), como *doublets lexicais* (SOLEDADE, 2004) ou *variantes derivacionais* (CAMBRAIA, 2010), desponta, de outro, como uma clara demonstração de que o chamado efeito de bloqueio (ARONOFF, 1976) não tem um alcance plenipotenciário sobre a emergência de lexias neológicas da língua.

A princípio, a ideia propulsada por Aronoff (1976), da existência de um efeito de bloqueio (*blocking*), é, sem dúvida, interessante e convincente. Consistiria, essencialmente, “[...] na não ocorrência de uma forma devido à simples existência de outra.” (ARONOFF, 1976, p. 43),<sup>47</sup> com o consequente impedimento da geração indiscriminada de sinônimos operados sobre uma mesma forma-base (ARONOFF, 1976). Como explica Viaro (2011b, p. 48), “[...] uma palavra formada com radical *x* e sufixo *y* é bloqueada se já houver, de antemão, uma outra com o mesmo radical *x* e sufixo *z*, sendo  $z \neq y$  do ponto de vista do significante, mas  $z = y$  do ponto de vista do significado.” Assim, ao falante lhe seria bloqueada a criação da forma *\*inconfiança* por já existir uma forma derivada que veicula o mesmo sentido visado: *desconfiança*. O bloqueio serviria, portanto, para explicar algumas lacunas existentes em processos morfológicos produtivos (ABREU, 2006; ARONOFF; ANSHEN, 1998), ou seja, a não aceitabilidade ou restrita frequência de palavras consideradas possíveis segundo os esquemas formativos da língua (DÍAZ HORMIGO, 2004-2005), o que não se confunde com o fenômeno da agramaticalidade. Sobre a natureza das restrições operadas pelo bloqueio, é interessante o que aponta Díaz Hormigo (2004-2005): o filtro atuante no não aparecimento, na modesta difusão ou na não

---

<sup>46</sup> A flutuação entre formas nominais corradicais sufixadas de MLB *usur-* não é exclusiva ao português, sendo atestada em outras línguas românicas, o que reforça o caráter não necessariamente fortuito e esporádico de tais derivados. Vide os Quadros 2 e 3.

<sup>47</sup> Trad. nossa. No original: “Blocking is the non-occurrence of one form due to the simple existence of another.”

consagração pela norma de formações virtualmente possíveis desde o ponto de vista do sistema pode dever-se a fatores de distinta ordem, desde fonológicos e morfológicos, até sintáticos, semânticos e pragmáticos.

Embora essa força de contenção, o bloqueio, ative-se exitosamente no impedimento da produção de algumas formas, não parece ser suficientemente robusta para barrar a emergência de inúmeras outras formas sinonímicas, dentre as quais estão os quatro adjetivos sufixados da família de *usura* supramencionados. Formas com o mesmo sentido central, usadas no mesmo contexto e na mesma área dialetal, compartilhando o mesmo MLB, e com uma única diferenciação: o elemento sufixal que contém. Vê-se, assim, que as forças geradoras de inovação, nesses casos, conseguem prevalecer sobre as correntes de bloqueio, propiciando a manifestação do curioso fenômeno de coexistência de vocábulos corradicais sinonímicos.

Ao se considerar o uso estendido nas línguas de formas sinonímicas (corradicais ou não), vê-se que o bloqueio não tem um alcance absoluto (SOLEDADE, 2004) ou sistemático (DÍAZ HORMIGO, 2004-2005), mas sim, condicionado, ocorrendo de modo atrelado a outros aspectos, como a frequência de uso de dada estrutura vocabular (PLANK, 1981; RAINER, 1988), a sua acessibilidade lexical e armazenamento mnemônico (RODRIGUES, 2016) e a sua natureza semântica (grau de transparência, de composicionalidade, de prototipia/nuclearidade, de homonímia, de polissemia etc.). Assim, pode-se detectar o verdadeiro potencial do fenômeno do bloqueio: não se trata de uma força geral de impedimento para a geração de formas equivalentes/sinonímicas, mas sim, de um filtro, que se traduz no léxico em uma preferência por formas mais facilmente acessíveis, em detrimento de outras menos funcionais. Tal filtro, como perceptível, não tem um alcance plenipotenciário, permitindo, por exemplo, a coocorrência/concorrência de formas corradicais, mas quase sempre sinalizando o alçamento preferencial de uma delas sobre as demais; ou, mais ainda, sinalizando a diferença entre, de um lado, as possibilidades léxicas no plano simbólico ou virtual e, do outro, a concretude efetiva de usos no plano discursivo.

São inúmeros os casos de *doublets* lexicais na língua, de natureza variada, fazendo da variação corradical um fenômeno de amplíssima extensão, perpassando vários subdomínios do léxico (PONCE DE LEÓN, 2010). Além dos pertencentes a esquemas sufixais (*descaração* ~ *descaramento*, *pegajoso* ~ *peguento*), há também os gerados via

prefixação (*antecessor* ~ *predecessor*, *desobediência* ~ *inobediência*), por oposição formal — mas equivalência semântica — entre formas com e sem prefixo (*assoprar* ~ *soprar*, *dependurar* ~ *pendurar*), ou mesmo por aplicação de diferentes esquemas morfolexicais, como na geração de verbos (parassíntese X mera verbalização: *embranquecer* ~ *branquear*, *amadurecer* ~ *maturar*).

A existência factual de todos esses exemplos, com diferença morfológica, mas com equivalência semântica, põe em xeque a validade plena do princípio do bloqueio, pelo menos como concebido em sua formulação mais geral. Na verdade, ainda que se tomem como base explanações teórico-analíticas mais minudenciadas sobre o fenômeno, o princípio não parece se sustentar completamente, ao ser confrontado com dados de língua. Assim, se Gonçalves, Yakovenko e Costa (1998), ao discutirem a criação de formas corradicais em *-eiro* para formas preexistentes em *-ista* (*sambista/sambeiro*, *romancista/romanceiro*), alegam que não há violação do princípio mencionado por haver certa diferença de conotação emotiva para as respectivas formas em *-eiro*, poderiam ser arrolados inúmeros outros casos em que sequer se notam diferenças semânticas dessa natureza, sendo absolutamente intercambiáveis, como as lexias reproduzidas anteriormente.

Com isso, não se quer defender aqui que a aplicação de operações morfolexicais para a criação vocabular não se subordine a pautas de contenção. Há, de fato, restrições diversas: i) fonológicas (*\*conredator*, *\*taxiizinho*); ii) morfológicas (*\*pensativamente*, *\*reantes*); iii) semânticas (*\*desreavaliar*, *\*alegrudo*); iv) lexicais (*\*infeio*, *\*criançologista*); v) morfossintáticas (*\*Aquele alegremente garoto*; *\*Enriquecimentos propuseram impensável*); vi) e mesmo etimológicas (*lamber* > *lambão*, mas *digitar* > *\*digitão*).<sup>48</sup> Tais restrições, porém, pelo que aponta a língua, não se aplicam como num jogo de tudo-ou-nada, e sim sob uma escala gradiente de aceitabilidade/rejeição, mormente ao que diz respeito às pautas semânticas e lexicais. Assim, por exemplo, *\*infeio* parece menos provável de ser materializado que *\*ileal*; do mesmo modo, *\*lindante* e *\*persistitório* parecem ser menos prováveis de ocorrer que *\*usurante* e *\*usuratório*. Esse entendimento, que é respaldado pela observação dos dados empíricos, coaduna-se àquele exposto e preconizado por Soledade

<sup>48</sup> Para maior detalhamento sobre os diversos tipos de restrições que atuam na morfologia derivacional, cf. Rodrigues (2016) e Díaz Hormigo (2004-2005).

(2004) e Simões Neto (2018), com a percepção do bloqueio como uma restrição relativa, não absoluta.

Ao que parece, a geração das unidades do léxico (operada pelo falante, em sua mente, materializada no uso e na interação e, consequentemente, incrustada em dado contexto comunicativo-situacional), encontra-se imersa numa tensão entre forças centrífugas, de expansão – a criatividade lexical do falante, o potencial gerador do léxico e a analogia, por exemplo – e forças centrípetas, de contenção e controle – os bloqueios, quer lexicais, quer morfológicos, quer semânticos, quer de outra natureza, a tradicionalização morfológica<sup>49</sup> (RIO-TORTO; LOPES, 2019) etc. –. A difícil questão a ser deslindada, portanto, seria a determinação da justa medida da potência e do alcance de dois feixes de forças em permanente tensionamento no léxico de uma língua: de um lado, as forças de geração de novas palavras; do outro, as forças de contenção de tais criações.

Sabendo que as forças de bloqueio serviriam para uma economia cognitiva e de expressão, a questão que permanece é a seguinte: por que o falante se lança à criação de novos construtos morfolexicais (com igual MLB, mas com afixos diferentes) para a expressão de sentidos para os quais já existem vocábulos correspondentes em seu léxico mental e em uso em seu ambiente interacional? Em outras palavras: por que se dá ao trabalho de gerar uma ou mais palavras para expressão de um sentido já materializado em uma lexia preexistente, como no caso dos adjetivos em **usur-**, em que, em vez de se contentar com *usurário*, cria no mínimo outras sete (*usureiro*, *usurável*, *usuroso*, *usurento*, *usurador*, *zuraco* e *zura*) para denotar igual sentido, o de ‘avarento’?

Várias razões podem ser aventadas para se tentar explicar o fenômeno de emergência e coexistência de corradicais sinonímicos, quer levando em conta o fator diacrônico, quer se eximindo de suas interferências. Como perpassa este artigo a consideração dos efeitos histórico-diacrônicos na configuração da língua, é natural que sejam considerados numa análise do fenômeno da coocorrência de vocábulos corradicais convergentes no plano do significado.

No caso do conjunto dos nomes sufixados de base **usur-**, é crível pensar que a coexistência de quatro deles (*usurário*, *usurável*, *usuroso* e

---

<sup>49</sup> Fomentada pela escolarização, pela normatização gramatical ou pela frequência/vitalidade do uso, entre outros fatores.

*usurento*) num mesmo recorte temporal e numa mesma zona dialetal dá-se por influência de fatores intra e extralinguísticos, como (i) o sinmorfismo sufixal<sup>50</sup> (os sufixos empregados possuem flagrante contiguidade semântica, orbitando em volta de um sentido mais ou menos parafraseável por ‘aquele que tem X’, não raramente com um matiz semântico adicional de avaliação depreciativa); (ii) a vitalidade e frequência dos sufixos empregados (todos são produtivos e frequentes no léxico geral) ou mesmo a recorrência/antiguidade de seu intercâmbio em outras famílias lexicais;<sup>51</sup> (iii) a interferência analógica (de uma forma muito recorrente e regular, cria-se outra, como pode ser conjecturado para a criação de *usuroso*, por influência de *ganancioso*; *usurento*, por influência de *avarento*); (iv) a criatividade lexical do falante e (v) uma menor influência de fatores que contribuiriam para uma aplicação rigorosa e imediata do travamento das inovações via tradicionalização morfológica, entre elas a escolarização, a urbanização e o acesso aos meios de comunicação em massa (todos esses pouco ou nada presentes na realidade social dos utentes da língua que produziram as formas em questão).

Seriam pelo menos esses cinco fatores os que atuariam como um conjunto de forças propulsoras da inovação lexical e da proliferação de formas equipolentes, comprometendo, por assim dizer, o potencial de interferência e de aplicação dos diversos bloqueios, das forças de contenção e controle da criatividade lexical e do potencial neológico da língua.

Se a coocorrência numa mesma área dialetal das quatro lexias nominais sufixadas em **usur-** é comprovada pela observação de dados empíricos, há de se dizer que também o é a diferença no peso e representatividade de cada uma delas. No APFB, dominam a cena as vozes *usurário* e *usurável*, enquanto as ocorrências e o espraio territorial para as outras duas lexias do grupo (*usuroso* e *usurento*) são bastante minguados, o que demonstra que embora *potencialmente* intercambiáveis, são de alguma forma dessemelhantes quanto a seu emprego na comunicação *efetiva*, ao menos quanto à frequência de sua materialização.

<sup>50</sup> Termo cunhado por Soledade (2004) e que nada mais é que a sinonímia entre afixos (nesse caso, entre formantes sufixais).

<sup>51</sup> Vários sufixos em situação de sinmorfismo no português contemporâneo apresentam-se em flutuação desde o período arcaico da língua, como é comprovado por Soledade (2004).

Ao considerar o léxico comum brasileiro, os dados dialetais e a própria história da língua portuguesa, constata-se que de todas as sete formas nominais corradicais sufixadas da família lexical de *usura* que comprovadamente existiram em sua diacronia (*usurário*, *usureiro*, *usurável*, *usuroso*, *usurento*, *zuraco* e *usurador*), a única cujo uso hodierno é relativamente estendido no país é *usurário*, tendo as demais ou um emprego muito limitado (a certas áreas ou como formações *ad hoc*) ou mesmo já tendo sofrido um processo de arcaizamento (como parece ter ocorrido com *usureiro*, que não foi detectado em nenhum *corpus* dialetal para o português brasileiro). Esse processo de desuso de umas formas a favor do incremento na frequência de outra(s) certamente está ligado, no caso dessas variantes morfológicas da família lexical de *usura*, a uma questão de viés diafásico: a avaliação dos derivados *usureiro*, *usurável*, *usuroso*, *usurento*, *zuraco* e *usurador* como mais coloquiais e periféricos e a avaliação de *usurário* como mais formal e mais prototípico.

Desse panorama relativo aos corradicais nominais sufixados em **usur-**, chega-se a algumas conclusões: primeiro, que a flutuação/intercâmbio entre formas pode não se traduzir numa verdadeira proporcionalidade equilibrada quanto ao seu emprego na fala real pelos utentes; segundo, que, não raramente, além de se encontrarem imersos num quadro de variação, também se encontram atingidos por efeitos de mudança linguística, com o alçamento (quantitativo ou qualitativo) de uma ou mais variantes sobre as demais;<sup>52</sup> terceiro, que, embora possa ocorrer (e comumente ocorre) um processo de abandono de uma ou várias formas do conjunto de lexias em coocorrência, em geral as possibilidades de sua reativação permanecem latentes, não havendo travas absolutas que as impeçam de ressurgir em dado falante, em dado grupo linguístico ou em dada época da língua, inclusive influenciando a entrada inovadora de afixos anteriormente alheios à família lexical (como o caso de *-dor*, em *usurador*, cujo emprego parece ser muito recente); por fim, que a frequência de uso de uma variante (cf. RODRIGUES, 2016) e sua antiguidade podem ser fatores decisivos para o seu alçamento sobre as demais com as quais coocorre/concorre (como se deu com *usurário* frente às suas congêneres).

---

<sup>52</sup> Não obstante, variantes morfolexicais podem permanecer em uso na língua por séculos a fio, sem despontar um quadro de mudança linguística e dissolução da coocorrência, como afirma Soledade (2012).

## 14 Considerações finais

A investigação da constituição e funcionamento de paradigmas léxico-morfológicos mostra-se como um campo propício para o entrelaçamento entre o labor etimológico, a Linguística Histórica e a Dialetoлогия, em franco diálogo com aspectos históricos, geográficos, culturais e sociais. Trata-se, portanto, de um objeto de estudo notadamente interfacial, que propicia uma apreciação holística e que flui entre a morfologia e o léxico, movendo-se sobre o eixo da semântica, do significado, da significação.

Não obstante a patente relevância da perquirição das famílias lexicais, não obstante haver vários estudos sobre campos lexicais e campos semânticos para o português, são raras as pesquisas e publicações voltadas ao rastreamento e descrição de famílias de palavras nessa língua (sobretudo diacronicamente), o que constitui uma considerável lacuna na produção científica atinente às ciências do léxico e à morfologia lexical vernaculares.

Visando a oferecer algum contributo ao tema em tela, apresentou-se, neste artigo, de modo sumário, um conjunto de informações, análises e achegas concernentes à família léxica de *usura*, um paradigma morfolexical bem restrito e bem pouco expressivo quanto ao número de elementos constituintes, mas dotado de indiscutível valor pela antiguidade de sua manifestação (no latim e no vernáculo), pelo seu espraioamento panromânico (não raras vezes processado mediante empréstimos) e pelo vigoroso sinmorfismo sufixal detectável em várias de suas formas nominais. Em outras palavras, uma família lexical quantitativamente modesta, mas qualitativamente relevante.

De uma observação de dados dialetais da área que compreende os denominados *falares baianos*, chegou-se à atestação da coexistência de formas corradicais sinonímicas portadoras de diferentes sufixos e à detecção dos demais constituintes do paradigma lexical de *usura*, no português (tal como exposto neste artigo) e em várias outras línguas românicas e no inglês (que, como dito, ficaram para outra publicação). Concomitantemente ao delineamento da evolução formal dos vocábulos de tal família léxica, houve a apreciação da evolução semântica dessas mesmas formas, atentando-se também para as interinfluências ocorridas no interior desse paradigma morfolexical e, inclusive, para as relações processadas com outros (como o de *usurpar*, por exemplo). Assim,

constatou-se na prática quão importantes são os dados extraídos de atlas linguísticos e de outros materiais de natureza dialetal para as investigações de morfologia lexical, lexicologia, lexicografia, semântica e etimologia.

O conjunto tipologicamente diversificado de materiais empíricos que subsidiaram a investigação, constituindo o seu *corpus* específico de análise, abarcando de atlas linguísticos a obras lexicográficas, de obras raras digitalizadas a registros de redes sociais, sem contar os inúmeros *corpora* textuais verificados, sinaliza a complexidade do trabalho histórico-diacrônico (e comparativo) atilado sobre famílias lexicais, com as suas inúmeras demandas e as exigências de uma rota metodológica cuidadosamente delineada, quanto aos dados empíricos, quanto às ferramentas de busca desses mesmos dados e quanto aos procedimentos para um tratamento eficiente e seguro destes.

A indicação (ou confirmação) das primeiras ocorrências detectáveis das formas constituintes do paradigma lexical de *usura* a partir de um vasto conjunto de fontes textuais, lexicográficas e digitais, bem como o traçado histórico do desenvolvimento semântico desses vocábulos (quase sempre conectado a fluxos de metonimização, generalização e analogia), constituem os contributos principais da investigação realizada, sintetizada nestas laudas. Espera-se, sem maiores pretensões, que as achegas metodológicas e os resultados obtidos sirvam em algo para o surgimento de novos estudos sobre outras famílias lexicais do português e das demais línguas românicas, mormente sob um viés histórico-diacrônico.

### Agradecimentos

Os autores deste artigo são gratos aos pareceristas anônimos pelas observações e sugestões apontadas, que contribuíram efetivamente para o aperfeiçoamento do texto em sua versão final para publicação. Agradecem também à Profa. Silvana Ribeiro (UFBA) por ter proposto o trabalho de análise das cartas ‘avarento’ do APFB, que serviu de inspiração à pesquisa apresentada neste artigo, e à Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) pelo subsídio vigente através de bolsa de iniciação científica.

## Contribuição dos autores

O artigo é fruto do trabalho conjunto de seus dois autores, em todas as fases de elaboração e revisão do texto. Matheus Pinto foi especialmente responsável pelo levantamento de *corpora*, pelo rastreamento das formas da família lexical de *usura* em português e em todas as demais línguas selecionadas, pela varredura e tratamento dos dados para a comparação interlinguística, pela redação dos resumos, da Introdução, da seção de metodologia e pelas seções referentes às análises etimológicas e morfossemânticas das lexias estudadas, assim como pela elaboração dos quadros e recolha das figuras. Por sua vez, Mailson Lopes foi especialmente responsável pela segmentação das seções do artigo, pela revisão completa da análise dos dados, pela proposição de aspectos concernentes à evolução semântica dos derivados, pelo rastreamento das lexias em cerca de 30 atlas linguísticos brasileiros e em outros estudos dialetais, pela elaboração da seção de análise dos corradicais sinonímicos e pela redação das Considerações finais. A versão final do artigo, revisada à luz dos comentários dos pareceristas anônimos, resultou igualmente de um trabalho realizado de modo conjunto pelos dois autores.

## Referências

ABREU, K. Focalizando a morfologia improdutiva: um estudo sobre siglas. *SIGNUM*, Londrina, n. 9, v. 2, p. 9-26, 2006. DOI: <https://doi.org/10.5433/2237-4876.2006v9n2p9>. Disponível em: <http://www.uel.br/revistas/uel/index.php/signum/article/view/3478>. Acesso em: 21 ago. 2020.

ÁLVAREZ, R. (org.). *Tesouro do léxico patrimonial galego e português*. Santiago de Compostela: Instituto da Língua Galega, 2014-2020. Disponível em: <http://ilg.usc.es/Tesouro>. Acesso em: 21 ago. 2020.

ARAGÃO, M. S. S. Sinônimos e parassinônimos em capitais do Nordeste Brasileiro: dados do ALiB. *Acta Semiótica et Linguística*, João Pessoa, v. 19, n. 1, p. 7-20, 2014. Disponível em: <https://periodicos.ufpb.br/index.php/actas/article/view/23372>. Acesso em: 19 jul. 2020.

ARAGÃO, M. S. S.; MENEZES, C. P. B. *Atlas Linguístico da Paraíba*. Brasília: UFPB; CNPq, Coordenação Editorial, 1984. v. 1-2.

ARONOFF, M. *Word Formation in Generative Grammar*. Cambridge, MA; London: MIT, 1976.

ARONOFF, M.; ANSHEN, F. Morphology and the Lexicon: Lexicalization and Productivity. In: SPENCER, A.; ZWICKY, A. (ed.). *The Handbook of Morphology*. Oxford: Blackwell, 1998. p. 236-247.

BAGNO, M. *Nada na língua é por acaso: por uma pedagogia da variação linguística*. São Paulo: Parábola, 2007.

BASILIO, M. *Formação e classes de palavras no português do Brasil*. São Paulo: Contexto, 2004.

BASSETTO, B. F. *Elementos de Filologia Românica: história externa das línguas românicas*. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2005. v. 1

CAMBRAIA, C. N. *Dulceça, dulçor, dulçura e dulcidom: um estudo de caso de variantes derivacionais no português medieval*. *Estudos de Lingüística Galega*, Santiago de Compostela, n. 2, p. 37-56, 2010. DOI: <https://doi.org/10.3309/1989-578X-10-2>. Disponível em: <https://revistas.usc.gal/index.php/elg/article/view/1507>. Acesso em: 8 jul. 2020.

CARDOSO, S. A. M.; FERREIRA, C. S. *O léxico rural*. Glossário. Comentários. Salvador: Universidade Federal da Bahia, 2000.

CAVALCANTE, R. *Dicionário de cearensês: a cultura do povo cearense*. 3. ed. Fortaleza: Edição do Autor, 2012.

CENTRE NATIONAL DE RESSOURCES TEXTUELLES ET LEXICALES. Disponível em: <https://www.cnrtl.fr>. Acesso em: 18 ago. 2020.

CLACKSON, J.; HORROCKS, G. *The Blackwell History of the Latin Language*. Malden: Blackwell Publishing, 2007.

COLLINS ENGLISH DICTIONARY. Disponível em: <https://www.collinsdictionary.com/dictionary/english>. Acesso em: 21 ago. 2020.

CUNHA, A. G. *Dicionário etimológico da língua portuguesa*. 4. ed. Rio de Janeiro: Lexikon, 2010.

DAVIES, M. *O corpus do português*. [S.l]: [S.n], 2006-2018. Disponível em: <https://www.corpusdoportugues.org/>. Acesso em: 21 ago. 2020.

DÍAZ HORMIGO, M. T. Restricciones del sistema y restricciones de la norma en la formación de palabras. *Linred: Lingüística en la Red*, Madrid, n. 2, p. 1-26, 2004-2005. Disponível em: <https://ebuah.uah.es/dspace/handle/10017/24638>. Acesso em: 2 ago. 2020.

DICIONÁRIO DE LATIM-PORTUGUÊS PORTUGUÊS-LATIM. Porto: Porto Editora, 2014.

DICIONÁRIO ONLINE CALDAS AULETE. [S.l.]: Lexikon Editora Digital. Disponível em: <http://www.aulete.com.br/>. Acesso em: 21 ago. 2020.

DICIONÁRIO PRIBERAM DA LÍNGUA PORTUGUESA [on-line]. [S.l.]: [S.n.], 2008-2020. Disponível em: <https://dicionario.priberam.org/>. Acesso em: 21 ago. 2020.

DIZIONARIO ITALIANO DE MAURO: Vocabolario online della lingua italiana. Disponível em: <https://dizionario.internazionale.it/>. Acesso em: 21 ago. 2020.

DMF – Dictionnaire du Moyen Français, version 2015 (DMF 2015). Lorraine: ATILF/CNRS; Université de Lorraine. Disponível em: <http://www.atilf.fr/dmf>. Acesso em: 21 ago. 2020.

DU CANGE, C. F. *et al.* *Glossarium mediae et infimae latinitatis*. Niort: L. Favre, 1883-1887.

DUBOIS, J. *et al.* *Dicionário de linguística*. 2. ed. São Paulo: Cultrix, 2014.

FERREIRA, A. B. H.; FERREIRA, M. B.; SILVEIRA, A. M. *Novo dicionário Aurélio da língua portuguesa*. 4. ed. Curitiba: Positivo, 2004.

FERREIRA, C. *et al.* *Atlas Linguístico de Sergipe*. Salvador: UFBA – Instituto de Letras; Fundação Estadual de Cultura de Sergipe, 1987.

FITER, J. *Enciclopedia moderna catalana*. Barcelona: Joseph Gallach, 1913. v. 5.

GODEFROY, F. *Complément du dictionnaire de l'ancienne langue française et de tous ses dialectes du XI<sup>e</sup> au XV<sup>e</sup> siècle*. [S.l.]: [S.n.], 1895-1902. Disponível em: <http://micmap.org/dicfro/introduction/complement-godefroy>. Acesso em: 18 ago. 2020.

GODEFROY, F. *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IXe au XVe siècle*. [S.l.]: [S.n.], 1880-1895. Disponível em: <http://micmap.org/dicfro/introduction/dictionnaire-godefroy>. Acesso em: 18 ago. 2020.

GONÇALVES, C. A. *Atuais tendências em formação de palavras*. São Paulo: Contexto, 2016.

GONÇALVES, C. A. V.; YAKOVENCO, L. C.; COSTA, R. G. R. Condições de produtividade e condições de produção: uma análise das formas X-eiro no português do Brasil. *Alfa*, São Paulo, v. 42, p. 33-61, 1998.

GOOGLE LIVROS. Disponível em: <https://books.google.com.br/>. Acesso em: 31 jul. 2020.

HARRISON, S. (ed.). *A Companion to Latin Literature*. Malden: Blackwell Publishing, 2005. DOI: <https://doi.org/10.1002/9780470996683>

HASPELMATH, M. *Understanding Morphology*. London: Arnold, 2002.

HASPELMATH, M.; SIMS, A. *Understanding Morphology*. 2. ed. London: Hodder Education, 2010.

HERNÁNDEZ AROCHA, H. ¿Son las familias de palabras un subproducto de la morfología o es la morfología un subproducto de las familias de palabras? *Revista de Filología*, San Cristóbal de La Laguna, n. 40, p. 69-103, 2020. DOI: <https://doi.org/10.25145/j.refiull.2020.40.05>. Disponível em: <https://www.ull.es/revistas/index.php/filologia/article/view/992>. Acesso em: 28 dez. 2020.

HOUAISS, A.; VILLAR, M. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

ILARI, R. *Linguística românica*. 2. ed. São Paulo: Contexto, 2018.

INFOPÉDIA – *Dicionários Porto Editora*. Porto: Porto Editora, 2003-2020. Disponível em: <https://www.infopedia.pt/>. Acesso em: 21 ago. 2020.

INSTITUT D'ESTUDIS CATALANS. Disponível em: <https://www.iec.cat/activitats/entrada.asp>. Acesso em: 21 ago. 2020.

INSTITUTO HISTÓRICO E GEOGRÁFICO DE SÃO PAULO. *Revista do Instituto Histórico e Geográfico de São Paulo*, São Paulo, v. 16, 1914.

JESPERSEN, O. *Language: Its Nature, Development and Origin*. Londres: George Allen & Unwin, 1954.

LE GOFF, J. *A bolsa e a vida: economia e religião na Idade Média*. Tradução de Rogério Silveira Muio. São Paulo: Brasiliense, 2004.

LE TRESOR DE LA LANGUE FRANÇAISE INFORMATISE. Disponível em: <https://www.le-tresor-de-la-langue.fr/>. Acesso em: 21 ago. 2020.

LEWIS, C. T.; SHORT, C. *A Latin Dictionary*. Oxford: Clarendon Press, 1879. Disponível em: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059>. Acesso em: 21 ago. 2020.

MACHADO FILHO, A. V. L. *Novo dicionário do português arcaico ou medieval*. [S.l.]: Independent Edition, 2019.

MATTOS E SILVA, R. V. *Caminhos da linguística histórica: ouvir o inaudível*. São Paulo: Parábola, 2008.

MORERA, M. Familia de palabras vs. campo semántico. Los casos particulares de las familias *punt-*, *punz-* y *pinch-*. *Revista de lexicografía*, A Coruña, n. 8, p. 149-222, 2002. DOI: <https://doi.org/10.17979/rlex.2002.8.0.5588>

NASCENTES, A. Divisão dialectológica do território brasileiro. *Revista Brasileira de Geografia*, Rio de Janeiro, v. 17, n. 2, p. 213-219, 1955.

NYROP, K. *Grammaire historique de la langue française*. Copenhague: Imprimerie Nielsen & Lydiche, 1903. Tome 2ème

OLIVEIRA, A. J. *Dicionário Gaúcho: termos, expressões, adágios, ditados e outras barbaridades*. 3. ed. Porto Alegre: Age, 2002.

PENA, J.; CAMPOS SOUTO, M. Propuesta metodológica para el establecimiento de familias léxicas en una consideración histórica: el caso de *hacer*. *Cuadernos del Instituto Historia de la Lengua*, San Millán de la Cogolla, n. 2, p. 21-51, 2009. Disponível em: [https://www.cilengua.es/sites/cilengua.es/files/page/docs/02\\_propuesta\\_metodologica\\_para\\_el\\_establecimiento\\_de\\_familias\\_lexicas\\_en\\_una\\_consideracion\\_historica\\_el\\_caso\\_de\\_hacer.pdf](https://www.cilengua.es/sites/cilengua.es/files/page/docs/02_propuesta_metodologica_para_el_establecimiento_de_familias_lexicas_en_una_consideracion_historica_el_caso_de_hacer.pdf). Acesso em: 10 jul. 2020.

PESQUISA avançada do Google. Disponível em: [https://www.google.com/advanced\\_search](https://www.google.com/advanced_search). Acesso em: 21 ago. 2020.

PLANK, F. *Morphologische (Ir-)Regularitäten: Aspekte der Wortstrukturtheorie*. Tübingen: Günter Narr, 1981.

PLAUTO. *O Truculento*. Tradução do latim, introdução e notas de Adriano Milho Cordeiro. Coimbra: Universidade de Coimbra, 2010. DOI: <https://doi.org/10.14195/978-989-8281-63-0>

PONCE DE LEÓN, R. F. Z. de. Esquemas rivales en la formación de palabras en español. *ONOMÁZEIN*, Santiago de Chile, n. 22, p. 59-82, 2010. Disponível em: [http://onomazein.letras.uc.cl/Articulos/22/3\\_Zacarias.pdf](http://onomazein.letras.uc.cl/Articulos/22/3_Zacarias.pdf). Acesso em: 5 ago. 2020.

RAINER, F. Towards a Theory of Blocking. In: BOOIJ, G.; VAN MARLE, J. (ed.). *Yearbook of Morphology*. Dordrecht: Foris, 1988. p. 155-185.

REAL ACADEMIA ESPAÑOLA (RAE); ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (AALE). *Nueva gramática de la lengua española – NGLE*. Madrid: Espasa, 2009. v. 1

RIO-TORTO, G. et al. (org.). *Gramática derivacional do português*. 2. ed. Coimbra: Imprensa da Universidade de Coimbra, 2016. Disponível em: [https://digitalis.uc.pt/pt-pt/livro/gram%C3%A1tica\\_derivacional\\_do\\_portugu%C3%AAs\\_0](https://digitalis.uc.pt/pt-pt/livro/gram%C3%A1tica_derivacional_do_portugu%C3%AAs_0). Acesso em: 26 dez. 2020.

RIO-TORTO, G. M.; LOPES, M. Fluctuación prefijal en el gallego-portugués y en el castellano medievales. *Estudos de Lingüística Galega*, Santiago de Compostela, v. 11, p. 103-136, 2019. DOI: <https://doi.org/10.15304/elg.11.5105>. Disponível em: <https://revistas.usc.gal/index.php/elg/article/view/5105>. Acesso em: 6 ago. 2020.

ROCHA, L. C. da S. *Estruturas morfológicas do português*. Belo Horizonte: Editora da UFMG, 1998.

RODRIGUES, A. S. Noções basilares sobre a morfologia e o léxico. In: RIO-TORTO, G. et al. (org.). *Gramática derivacional do português*. 2. ed. Coimbra: Imprensa da Universidade de Coimbra, 2016. p. 35-133. DOI: [https://doi.org/10.14195/978-989-26-0864-8\\_1](https://doi.org/10.14195/978-989-26-0864-8_1). Disponível em: [https://digitalis.uc.pt/pt-pt/livro/gram%C3%A1tica\\_derivacional\\_do\\_portugu%C3%AAs\\_0](https://digitalis.uc.pt/pt-pt/livro/gram%C3%A1tica_derivacional_do_portugu%C3%AAs_0). Acesso em: 6 ago. 2020.

ROELLI, P. (org.). *Corpus corporum*. Repositorium operum Latinorum apud universitatem Turicensem. Zürich: Universität Zürich, 2016. Disponível em: <http://www.mlat.uzh.ch/MLS/index.php?lang=0>. Acesso em: 15 jul. 2020.

ROSSI, N. et al. *Atlas prévio dos falares baianos*. Rio de Janeiro: MEC; INL, 1963.

SÁNCHEZ MARTÍN, F. J. *Estudio del léxico de la geometría aplicada a la técnica en el Renacimiento hispano*. 2008. 484f. Tese (Doctorado en Filología Hispánica) – Facultad de Filología, Universidad de Salamanca, Salamanca, 2008. Disponível em: <https://gredos.usal.es/handle/10366/22435>. Acesso em: 26 dez. 2020.

SILVA, A. M. *Diccionario da Lingua Portuguesa*. Tomo Segundo. Lisboa: Lacerdina, 1813.

SIMÕES NETO, N. A. *O esquema X-ari- do latim às línguas românicas: um estudo comparativo, cognitivo e construcional*. 2020. 5v. 4297f. Tese (Doutorado em Língua e Cultura) – Instituto de Letras, Universidade Federal da Bahia, Salvador, 2020.

SIMÕES NETO, N. A. Variação morfológica: aproximações entre dialetologia e diacronia. *Miguilim – Revista Eletrônica do Netlli*, Crato, v. 7, n. 1, p. 39-54, 2018. Disponível em: <http://periodicos.urca.br/ojs/index.php/MigREN/article/view/1552>. Acesso em: 4 ago. 2020.

SIMÕES NETO, N. A.; SOLEDADE, J. Túnel morfológico: polissemia, alomorfia, sinmorfismo e doublets no português arcaico e no português brasileiro. *Estudos Linguísticos e Literários*, Salvador, n. 47, p. 105-126, 2013. DOI: <https://doi.org/10.9771/2176-4794ell.v1i47.14456>. Disponível em: <http://periodicos.urca.br/ojs/index.php/MigREN/article/view/1552>. Acesso em: 20 dez. 2020.

SOLEDADE, J. O *sinmorfismo* e os *doublets* no português arcaico. In: MATTOS E SILVA, R. V.; OLIVEIRA, K.; AMARANTE, J. (org.). *Várias navegações: português arcaico, português brasileiro, cultura escrita no Brasil, outros estudos*. Salvador: EDUFBA, 2012. p. 45-65.

SOLEDADE, J. *Semântica morfolexical: contribuições para a descrição do paradigma sufixal do português arcaico*. 2004. 2v. 575f. Tese (Doutorado em Letras e Linguística) — Instituto de Letras, Universidade Federal da Bahia, Salvador, 2004.

THE PACKARD HUMANITIES INSTITUTE. Disponível em: <https://latin.packhum.org/index>. Acesso em: 31 jul. 2020.

TWITTER. Disponível em: <https://twitter.com/>. Acesso em: 31 jul. 2020.

VAAN, M. *Etymological Dictionary of Latin and the other Italic Languages*. Leiden; Boston: Brill, 2008.

VÄÄNÄNEN, V. *Introducción al latín vulgar*. Tradução de Manuel Carrión. Madrid: Gredos, 1968.

VIARO, M. E. A formação do significado agentivo de -eiro. In: CONGRESO INTERNACIONAL DE LA ALFAL, XVI., 2011. Alcalá de Henares. *Actas [...]*. Alcalá de Henares: Universidad de Alcalá, 2011a. p. 2671-2679.

VIARO, M. E. *A derivação sufixal do português: elementos para uma investigação semântico-histórica*. 2011. 220f. Tese (Livre-docência) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011b.

VIARO, M. E. *Etimologia*. São Paulo: Contexto, 2014.

VIARO, M. E. *Manual de etimologia do português*. 2. ed. São Paulo: Globo Livros, 2013.

VIARO, M. E. Proposta de um método de análise para derivações sufixais. *Domínios de Linguagem*, Uberlândia, v. 3, n. 2, p. 140-165, 2009b. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/11511>. Acesso em: 31 jul. 2020.

VIARO, M. E. Uma nova metodologia para dados etimológicos e diacrônicos: o problema da datação dos fenômenos. In: TORRES MORAIS, M. A. C. R.; ANDRADE, M. L. C. V. de O. (org.). *História do português paulista*. Campinas: Editora da Unicamp, 2009a. p. 445-463.

WRIGHT, T. *Anglo-Saxon and Old English vocabularies*. Editado por Richard Paul Wülcker. Londres: Trübner, 1884.

WURFEL, D. *Usurador do pavo real* (e-book). [S.l]: [S.n], 2020.

XAVIER, M. F. (org.). *Corpus Informatizado do Português Medieval – CIPM*. Disponível em: <https://cipm.fcsh.unl.pt/>. Acesso em: 21 ago. 2020.



## **“Tô notando, sô!” A partícula vocativa *sô* e a interface sintático-pragmática**

### ***“Tô notando, sô!” The vocative particle *sô* and the syntactic-pragmatic interface***

Juliana Costa Moreira

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG),  
Belo Horizonte, Minas Gerais / Brasil

julianaichs@yahoo.com.br

<http://orcid.org/0000-0003-4329-7142>

**Resumo:** Moreira (2005, 2008) e Moreira e Alkmim (2013) observam que a forma *senhor* e suas variantes de gênero e número são frequentemente utilizadas como vocativo em construções extraídas de peças teatrais escritas por autores mineiros nos séculos XIX e XX. Considerando os trabalhos citados, a nossa hipótese é a de que a incidência do uso da forma *senhor* como vocativo em posição final resultou em um processo de gramaticalização, que originou as formas reduzidas desse item: *sinhô*, *seu*, *siô*, *sôr* e *sô*. Neste artigo, a partir de dados de intuição e da avaliação de falantes nativos, objetiva-se descrever e analisar as construções em que o item *sô* se realiza. Seguimos Hill (2007, 2014) que situa o vocativo na estrutura sintática, considerando a função pragmática que este item desempenha. Como resultado, tem-se as configurações arbóreas correspondentes às construções em que o item *sô* se realiza e considerações sobre o vocativo na estrutura sintática do dialeto mineiro.

**Palavras-chave:** *senhor*; *sô*; vocativo; estrutura sintática; função pragmática.

**Abstract:** Moreira (2005, 2008) and Moreira and Alkmim (2013) observe that the form *senhor* and number and gender and number variants are frequently used as a vocative in constructions taken from plays written by authors from Minas Gerais in the 19th and 20th centuries. Considering the works cited above, our hypothesis is that the incidence of the use of the form *senhor* as a vocative in final position resulted in a process of grammaticalization that originated the reduced forms of this item: *sinhô*, *seu*, *siô*, *sôr* and *sô*. In this article, based on intuition and evaluation data from

native speakers, we aim to describe and analyze the constructions in which the item *sô* is realized. We follow Hill (2007, 2014), who places the vocative in the syntactic structure, considering the pragmatic function that this item performs. As a result, we present tree configurations corresponding to the constructions in which the item *sô* is realized, as well as considerations about the vocative in the syntactic structure of the dialect of Minas Gerais.

**Keywords:** *senhor*; *sô*; vocative; syntactic structure; pragmatic function.

Recebido em 31 de agosto de 2020

Aceito em 18 de novembro de 2020

## 1 Introdução

Moreira (2005, 2008) e Moreira e Alkmim (2013, p. 85) observam que a forma *senhor*, e suas variantes de gênero e número, são frequentemente utilizadas como vocativo, respectivamente, em *corpus* constituído por peças teatrais escritas por autores mineiros e, em um segundo *corpus*, representativo do Português Brasileiro dos séculos XIX e XX.<sup>1</sup> São exemplos do *corpus* de Moreira (2005):

- (1) a. Senhora, já estais sciente dos meus intentos, conforme acabei de ouvir. (REZENDE, 1882, s.p.)  
b. Oh, senhores, custa-me a acreditar em tamanha felicidade (PAIVA, 1893, s.p.)
- (2) Eu, Snr<sup>es</sup>, não sou nenhum vaqueiro, Que viva de guardar alheio gado... (PAIVA, 1893, s.p.)
- (3) Não há mais barreira, senhor. (WERNECK, 1900, p. 120.)

Em (1a), os vocativos *senhora* e *senhor*, respectivamente estão situados à esquerda da oração. Observe-se, no entanto que em (1b), há a precedência da interjeição *oh*. Já em (2), uma abreviatura de *senhor*, *Snr<sup>es</sup>*, vocativo, situa-se em uma posição intermediária, entre o sujeito e o verbo e; em (3), observe-se que o termo *senhor*, ao exercer também a função sintática de vocativo, está alocado à direita da oração.

---

<sup>1</sup> Alguns exemplos citados neste artigo pertencem ao *corpus* de Moreira (2005) para o desenvolvimento de estudo sobre o vocativo na língua coloquial de Minas Gerais.

Moreira (2005, p.37) observa, em relação ao dialeto mineiro, e também Moreira (2008, p.59), tratando-se do Português Brasileiro, que, no século XIX, o vocativo é utilizado com mais frequência em posição inicial [Voc + Oração], em relação à ocorrência deste constituinte em posição medial [Oração + Voc + Oração] e em posição final [Oração + Voc].

Ao investigar a ordem do vocativo na oração do Português Brasileiro, Moreira (2008) obtém os resultados apresentados na Tabela 1.

TABELA 1 – Frequência de uso de vocativo em função do tempo

<i>Estruturas</i>	<i>T1</i>			<i>T2</i>			<i>T3</i>			<i>T4</i>			<i>Total</i>
	Nº	%	PR	Nº	%	PR	Nº	%	PR	Nº	%	PR	
<b>[Voc + Or]</b>	160	43/	.61	114	46	.56	131	35	.43	106	36	.39	511
<b>[Or +Voc]</b>	209	57	.38	133	54	.43	245	65	.57	189	64	.60	776
<b>Total</b>	369	100	-	247	100	-	376	100	-	295	100	-	1287

Fonte: Moreira (2008, p. 60.)

De acordo com a Tabela 1, na primeira metade do século XIX (T1), o vocativo é utilizado com mais frequência em posição inicial (.61) e, no século XX, o índice de ocorrência do vocativo nesta posição decresce e o que se obtém, na 2ª metade do século XX (T4) é o peso relativo de .39. Por outro lado, construções com vocativo em posição final são menos frequentes ao início do século XIX (T1), apresentando o peso relativo de .38 e se tornam mais recorrentes no século XX, de modo que atingem na última faixa temporal analisada o peso relativo .60. O perfil descrito é característico de um processo de mudança de ordem do vocativo na oração do Português Brasileiro.<sup>2</sup>

Devido à baixa frequência de ocorrência de vocativo em posição intermediária (9%), optou-se, neste estudo, por descartar esses enunciados, no que se refere à análise quantitativa realizada, já que esta é feita a partir de uma variável binária.

Moreira (2008) investiga os fatores linguísticos que podem estar levando a mudança à frente no Português Brasileiro, utilizando o Programa Goldvarb 2001. Ao detalhar um desses fatores, qual seja, o material que

<sup>2</sup> Esses resultados foram obtidos a partir da submissão dos dados ao Programa Goldvarb 2001.

compõe os vocativos, verifica o peso relativo consideravelmente alto desse constituinte, em posição final, quando corresponde a um apelido (.79), a uma expressão nominal de tratamento (.69)<sup>3</sup> e quando se trata de epíteto (.57).<sup>4</sup>

Todavia, há de se considerar que é baixo o número de apelidos no total do *corpus*: há apenas sete contruções em que o vocativo corresponde a um apelido, os quais se encontram na peça “Dois perdidos numa noite suja”, de Plínio Marcos, obra em que se observa uma linguagem coloquial. Diante dessa observação, tratando-se dos constituintes que podem representar sintaticamente o vocativo, verifica-se que há maior índice de ocorrência de vocativo como expressão nominal de tratamento em posição final. Esse resultado é esperado, pois, em contextos em que não se tem conhecimento do nome de alguém, um pronome, uma expressão nominal de tratamento ou um vocativo profissional, por exemplo, pode ser utilizado, uma vez que são mais genéricos do que aqueles representados morfofonologicamente por um nome próprio.

Partindo esses resultados, Moreira e Alkmim (2013, p. 87) tratam amudnça de posição do vocativo na oração como hipótese para explicar a redução de *senhor* > *sô* quando à direita da oração. Ao levar em conta essa hipótese, postulam que o sintagma que o sintagma que constitui o vocativo no caso aqui analisado com a mudança deposição na oração, da esquerda para a direita, torna-se mais gramatical. Desse modo, a posição à direita poderia ser considerado um ambiente propício para a ocorrência da forma reduzida *sô*. A partir desta hipótese, Ramos (2011, p.75) argumenta:

embora as autoras não justifiquem teoricamente esta correlação nem apresentem um estudo quantitativo da forma reduzida conforme a posição do vocativo no eixo do tempo e se restrinjam

---

<sup>3</sup> De acordo com a Profa. Jânia Ramos (comunicação pessoal), há várias razões para se considerar o item *senhor* como uma expressão nominal de tratamento e não como um pronome: o fato de ora vir com artigo (*O senhor aceita um café?*), ora sem artigo (*Senhor, aceita um café?*) é uma dessas razões. Comparando com *a gente*, pode-se ver claramente a diferença. No termo *a gente*, o item “a” não pode mais ser descrito como um artigo por ter perdido a capacidade de indicar o traço [+ definido] do nome.

<sup>4</sup> “Epíteto é a palavra ou expressão que se associa a um nome ou a um pronome para qualificá-lo; qualificação elogiosa ou injuriosa dada a alguém; alcunha, qualificativo”. (HOUAISS; VILLAR, 2001, p.70).

à apresentação de um estudo quantitativo da distribuição de vocativos conforme sua posição na sentença, a hipótese formulada é interessante, se analisada sob um tratamento formal de gramaticalização. Se tivermos em conta que o vocativo constitui o *locus* em que nomes e pronomes ocorrem sem determinante, as autoras ao chamar atenção para uma configuração de sentença especial [oração + vocativo], acabam por ressaltar uma função sintática relevante. (RAMOS, 2011, p. 76).

Partindo da hipótese apresentada, a autora considera que a ocorrência de *senhor* como vocativo à direita, à esquerda ou no meio da oração teria sido o gatilho para a reanálise do nome como pronome, como explicitado na *cline* abaixo:

(4) *senhor* > ... > *sinhô* > ... > *sô*

Em vista disso, apresenta um estudo sobre as ocorrências de *senhor* em vocativos, com o objetivo de mostrar que *senhor* e *sô* são pronomes, sendo o primeiro uma forma homônima de um nome. A partir da análise de três *corpora*, um diacrônico e dois sincrônicos, conclui que esse processo se desenvolveu nos últimos dois séculos e que a ocorrência do item na posição de vocativo favoreceu a reanálise sintagma nominal > núcleo pronominal, chamando a atenção para a importância dos vocativos no processo de pronominalização, o que resultou na realização das formas reduzidas *sinhô*, *seu*, *siô* e *sô*. São exemplos de construções contendo as formas reduzidas da expressão nominal de tratamento *senhor*:

(5) Não sei, não *sinhô*, foi um branco que me deu. (PAIVA, 1897, s.p.)

(6) a. Seu *dotô*, não há esse que não tenha um *podre* na vida. (BRAGA, 1915, s.p.)

b. Oia, seu *dotô*, uma lembrança dela. (BRAGA, 1915, s.p.)

(7) A *onça*, *Sôr* Porteiro, a *onça* que cá está. (PAIVA, 1897, s.p.)

(8) Ah *siô* Belchior! é um *estirão* do *Porqueiro* até aqui. (REZENDE, 1906, s.p.)

(9) Tô notando, *sô*, e aperciando e hei de vortá aqui mais *veis*... (PAIVA, 1897, s.p.)

No exemplo (5), temos a realização da forma *sinhô*, antecedida por *não*, formando uma resposta negativa;<sup>5</sup> em (6a) e (6b) observa-se que a forma “seu” ocorre seguida por um nome *doto*.<sup>6</sup> No *corpus* de Moreira (2005), encontramos também a forma *sôr*, exemplificada em (7). À semelhança de *seu*, as formas *sôr*, em (7) e *siô*, em (8), ocorrem seguidas de um nome, como em *Sôr Porteiro* e *siô Belchior*. Observe-se também em (8) que a forma *siô*, em posição inicial, é precedida pela interjeição “ah”. Já no exemplo (9), a forma *sô se* realiza ao final da primeira oração do período. No entanto, encontramos também a ocorrência do vocativo *sô* em posição inicial:

(10) Uai, *sô*, que baruío é esse ... (REZENDE, 1906, s.p.)

(11) Uê *sô*! Vance ta com vergonha de dizê que bebe cachaça? (REZENDE, 1906, s.p.)

No exemplo (10), a forma *sô* ocorre em posição inicial, precedida pela interjeição *uai*, ao passo que em (11), a interjeição *uê* a precede.

Como visto, nos exemplos do dialeto mineiro, a forma *sô* pode ocorrer ao final da oração ou ao início, se precedida de interjeição. Tendo em vista o exposto, propomos descrever a ocorrência de *sô* no dialeto mineiro contemporâneo, utilizando dados de intuição e também a avaliação de falantes nativos, na seção 3. Com base na descrição proposta, é também objetivo deste artigo representar sintaticamente as construções contendo a forma *sô*. Partimos do pressuposto de que esta forma ocorre na mesma categoria sintática do que o vocativo, considerando-se que esta posição é o *locus* do processo de pronominalização.

Partimos da caracterização das funções pragmáticas do vocativo como chamamento e destinatário, conforme Moreira (2008), levando em consideração trechos de peças de teatrais, embora a autora não utilize esses termos, ao declarar que as posições de colocação do vocativo na sentença podem gerar diferentes interpretações. Em detalhes, observe-

---

<sup>5</sup> Moreira (2005) observa que esse tipo de construção (*sim/não* + *senhor*), à semelhança de alguns vocativos, têm em sua composição um pronome (*senhor*, *senhora*) e vêm separadas por vírgula do restante da oração.

<sup>6</sup> A forma *seu* não ocorre isolada e, pelo que se observa, geralmente, é seguida por nome, como é o caso de *doto*, nos exemplos em (6). Ressalte-se que pode se tratar também de um nome próprio *Seu Belchior*, por exemplo.

se que, quando o falante está próximo do ouvinte ou aguardando uma resposta deste no contexto comunicativo, não havendo disputa de atenção daquele, o vocativo pode se situar à direita da oração, como no trecho:

(12) Tonho: (...) Eu estudei. Uma briga com o negrão não acaba nunca. Se eu acerto ele hoje, ele me pega de faca amanhã. Se eu escapo amanhã, ele me pega depois. Só acaba com a morte.

Paco: Mata ele.

Tonho: Eu estudei, meu chapa. Não estou a fim de apodrecer na cadeia por causa de um desgraçado qualquer.

(MARCOS, 1978, p. 38 *apud* MOREIRA, 2008, p. 91)

No trecho acima, *Tonho* não disputa a atenção de *Paco* com outro(s) interlocutor(es). O diálogo se dá entre os dois personagens que estão próximos e, portanto, a fala do último é esperada pelo primeiro.

Conforme Moreira (2008), o vocativo pode ser o primeiro constituinte da oração quando o enunciador está distante do ouvinte ou quando este deseja chamar a atenção daquele, principalmente quando há, no contexto do ato de fala, disputa da atenção deste último, como ilustrado no trecho:

(13) Jorge: Vizinha, vizinha, o que é? O que foi? Não vejo ninguém...

Florência: Quem está aqui?

Jorge: Vizinha, somos nós.

Emília, *dentro*: Minha mãe, minha mãe!

Florência: Ah, é o vizinho Jorge! E estes senhores! (*Levantando-se ajudada por Jorge*).

Emília: Minha mãe, o que foi? (MARTINS PENA, 1956, p. 329.)

Neste exemplo, embora o enunciador esteja próximo do ouvinte, dirige-se a ele com espanto, o que nos leva a crer que a intenção do primeiro é chamar a atenção do último. Podemos observar, ainda, que há disputa da atenção do ouvinte: tanto *Jorge*, o vizinho, como a filha *Emília*, requerem a atenção de *Florência*.

O vocativo pode também ocorrer dentro do enunciado quando o falante opta por enfatizar outro constituinte interno a oração, como no exemplo abaixo:

(14) Florência: O rapaz não tem inclinação nenhuma para ser frade.

(...)

Mestre: O dia que o Sr. Carlos sair do convento será para mim um dia de descanso. (...) Não se passa só um dia em que se não tenha de lamentar alguma travessura desse moço (...).

Florência: Foi sempre assim, desde pequeno.

Ambrósio: E seria uma crueldade violentar-lhe o gênio.

Mestre: E se o conheciam, senhores, para que o obrigaram a entrar no convento, a seguir uma vida em que se requer tranquilidade de gênio?

Florência: Ah, não foi por meu gosto, meu marido é que persuadiu-me.

Ambrósio, *com hipocrisia*: Julguei assim fazer um serviço agradável a Deus.

Mestre: Deus, senhores, não se compraz com sacrifícios alheios.

(MARTINS PENA, 1956, p. 316.)

No exemplo acima, o sujeito da oração *Deus* é topicalizado e se situa à esquerda do vocativo *senhores*. Observe-se que, também neste caso, o falante (o *Mestre*) está próximo dos seus interlocutores (*Florência* e *Ambrósio*) e não há, ainda, disputa de atenção destes.

Em resumo, no contexto exemplificado em (12), em que o vocativo se situa em posição inicial, a função do vocativo é chamar a atenção do interlocutor ou evocá-lo, principalmente quando, no contexto, há disputa de atenção entre os interlocutores. Já nos trechos em (13) e (14), com vocativo em posição final e dentro do enunciado, respectivamente, a intenção do falante parece ser manter o contato já estabelecido.<sup>7</sup>

Hill (2007, 2014) também diferencia também os chamamentos dos destinatários, no que diz respeito à interpretação do vocativo e sua distribuição sintática. Em termos gerais, se o vocativo se situa em posição inicial, exerce função de chamamento, enquanto que, ao exercer função de destinatário, pode ocorrer em outras posições. Assim, o vocativo pode ser interpretado como chamamento (*call*) ou destinatário (*address*), conforme a posição na oração, como nos exemplos a seguir:

---

<sup>7</sup> A observação do contexto pragmático em que os vocativos ocorrem e a relação entre função sintática, papel pragmático e diferentes interpretações dos vocativos é o que motiva assumirmos a postulação dos traços da periferia pragmática suas checagens e movimentos, o que será apresentado em seções subsequentes a esta.

- |         |                               |              |
|---------|-------------------------------|--------------|
| (15) a. | Maria, tira a roupa do varal. | Chamamento   |
| b.      | Tira a roupa do varal, Maria. | Destinatário |

É atribuída ao vocativo “Maria”, em posição inicial, em (15a), a leitura de chamamento ou interpelação. Porém, esta mesma leitura não pode ser atribuída aos vocativos em posição final, como em (15b). A interpretação que é atribuída a estes últimos é a de destinatário.

Considerando as possíveis funções pragmáticas e a posição do vocativo na oração, Hill (2007, 2014) apresenta uma proposta para a posição do vocativo na estrutura sintática, a qual será apresentada na próxima seção.

Nossa discussão, neste texto, está organizada basicamente em quatro seções. Na seção 2, apresentamos o referencial teórico utilizado para a análise das construções que contêm o item *senhor* ou da forma reduzida deste item, qual seja, *sô*. Na seção 3, descrevemos o comportamento sintático da forma *sô* no dialeto mineiro. Na seção 4, apresentamos a análise dessa partícula na interface sintático-pragmática. Por fim, na seção 5, constam as nossas considerações finais a respeito do fenômeno aqui abordado.

## 2 Referencial teórico: a interface sintático-pragmática

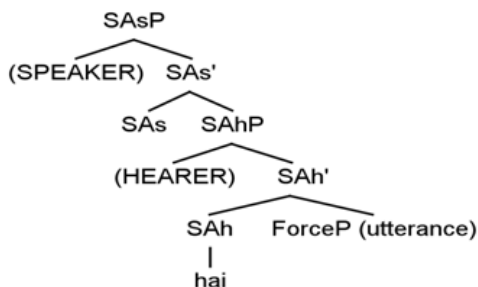
Conforme mencionado anteriormente, seguindo Ramos (2011, p. 81), consideramos que a posição de vocativo é o *locus* do processo de pronominalização da forma *senhor*. Com base em Speas e Tenny (2003), Hill (2007, 2014) considera que este constituinte, que representa o ouvinte, bem como constituintes que manifestam a opinião do falante, como as interjeições, situam-se na interface sintático-pragmática, em uma projeção denominada *Speech Act Phrase* (SAP).<sup>8</sup>

A projeção SAP é localizada acima de ForceP, projeção mais alta do domínio CP (RIZZI, 1997, p. 297), como expresso em (16):<sup>9</sup>

<sup>8</sup> De acordo com Hill (2007, 2014), esta projeção se assemelha à concha (vP/VP) que representa sintaticamente as construções bitransitivas (e.g. *I gave the book to Mary*).

<sup>9</sup> Hill (2014, p. 147) propõe essa configuração ao analisar dados do romeno, do búlgaro e do umbundu.

(16)



Fonte: Hill (2014, p.147).

Hill (2007, p. 2100) propõe que os vocativos são situados em uma categoria SAhP, na área de ouvinte. As interjeições, por sua vez, são situadas na área de falante, mais especificamente em uma categoria que a autora denomina SAs.

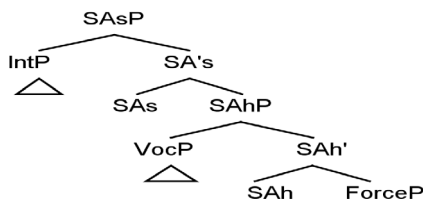
Observe-se a ocorrência da partícula *hai*, nesse núcleo, correspondente ao traço [V] realizado lexicalmente no romeno. Esse traço seleciona um enunciado, ou seja, ForceP, e se relaciona com as projeções de C/T/V. Assim, compreendemos o fato de o vocativo poder apresentar os mesmos traços *phi* de um constituinte argumental, isto é, a possível relação de SAhP com um constituinte argumental.

Segundo Hill (2007, p. 2092), a partícula *hai* apresenta uma terminação morfológica que pode variar de acordo com o tipo sentencial, a saber, a forma *haideti*, utilizada no modo imperativo, e a forma *haidem*, que ocorre no modo subjuntivo. É, assim, evidente que há uma relação entre o traço do ato de fala com a categoria ForceP, categoria do domínio de CP, na qual é codificado o tipo sentencial.

Seguindo uma linha de raciocínio cartográfica, consideramos a existência da categoria VocP, situada no especificador de SAhP e, equiparadamente, consideramos que as interjeições podem ser representadas por uma categoria IntP (*interjection phrase*) na posição de especificador de SAsP.<sup>10</sup> É o que mostramos a seguir:

<sup>10</sup> Na próxima seção, apresentaremos evidências para a existência da categoria VocP; no entanto, não exploraremos, neste artigo, essa categoria IntP, uma vez que não está dentre os nossos objetivos.

(17)



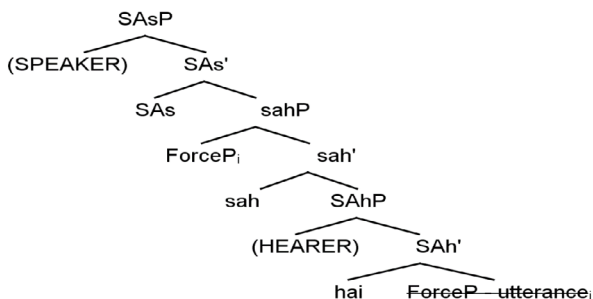
Fonte: Elaboração própria.

Pressupõe-se que existe uma hierarquia para a atribuição de Caso e papel temático/pragmático no domínio de SAP, tal qual ocorre na concha larsoniana: (i) o Caso Exclamativo e o papel pragmático “falante” são atribuídos ao constituinte que ocupa a posição de argumento mais alta, geralmente uma interjeição, e (ii) o Caso Vocativo e o papel pragmático “ouvinte” são atribuídos ao constituinte que esteja situado na posição intermediária de argumento, podendo ou não envolver uma marca morfológica. Em Latim o nome vocativo apresenta uma marca morfológica de Caso (por exemplo, Lat. *lupus* ‘lobo’ NOM *versus* *lupe* ‘lobo VOCC’.<sup>11</sup> Ressalte-se que, como é sabido, o fato de haver desinências casuais visíveis não implica dizer que, na sua ausência em outras línguas, não exista a atribuição de Caso abstrato. O caso vocativo morfológicamente realizado é opcional nos nomes, visto que mesmo sem essas marcas um termo pode ser interpretado como vocativo. Quando não há marca morfológica, o que indica que um nome seja interpretado como vocativo é a entoação com que é pronunciado e, além disso, a função do substantivo de chamamento ou de destinatário no diálogo.

Além das posições argumentais, Hill (2014, p.171) estipula a existência de uma categoria discursiva no domínio de SAP, como ilustrado a seguir:

<sup>11</sup> Como pontua Hill (2014, p. 4), no Latim e no Grego Antigo, nomes da segunda declinação no singular são terminados com -e, marca morfológica de vocativo. Esta marca remonta ao proto-indoeuropeu e ainda é encontrada no grego, no búlgaro e no romeno. A abordagem da origem, os tipos e a distribuição dos morfemas do Caso Vocativo se enquadram no campo da Linguística Histórica tradicional. Não estenderemos esta discussão, uma vez que a discussão sobre a dimensão histórica dos vocativos não se enquadra dentre os objetivos deste artigo.

(18)



Fonte: Hill (2014, p. 171).

Em (18), SAh seleciona ForceP como complemento e o vocativo como especificador. Observe-se a existência de uma categoria discursiva sah, localizada na projeção mais alta da área de ouvinte, a qual é licenciada pela presença de um traço discursivo, a saber, o traço [ênfase] – ou, nos termos da autora, [emphatic]. Este traço desencadeia o movimento de ForceP para o especificador de sahP, movimento esse que gera uma alteração na interpretação da sentença, uma vez que o vocativo, que representa o ouvinte, passa a desempenhar a função pragmática de destinatário, ao invés de chamamento. Assim, é a existência do traço [ênfase], que determina a interpretação da sentença, como injuntiva ou avaliativa.<sup>12</sup>

Se houver movimento de ForceP, temos a seguinte ordem de palavras:

(19) Só encontrei esse livro na Fnac, João.

Só encontrei esse livro na Fnac – FORCEP/sahP, João – VOCP

Se é outro constituinte que se movimenta, temos:

(20) Esse livro, João, só encontrei \_ na Fnac.

Esse livro – sahP, João – VOCP, só encontrei\_ na Fnac – FORCEP

Se houver topicalização de um outro constituinte como um DP, a ordem de palavras é a seguinte:

<sup>12</sup> A autora usa o termo *bonding reading* para se referir a um contexto avaliativo, em contrapartida de um contexto injuntivo, em que temos o modo imperativo.

- (21) Esse livro, João, eu só encontrei \_ na Fnac.  
 Esse livro – TOP, João – VOC, eu só encontrei \_ na Fnac – FORCEP

Continuando o mapeamento da interface sintaxe-pragmática, utilizando-se de dados do flamengo ocidental, Hill (2014, p.176) investigou o comportamento sintático de partículas vocativas, *zé* e *zè*. Estas partículas são derivadas do verbo *zien* (ver). De acordo com Hill (2014), a partícula *zé*, pronunciada com entonação mais forte, geralmente ocorre em posição inicial (exemplo de Hill (2014, p. 176)):

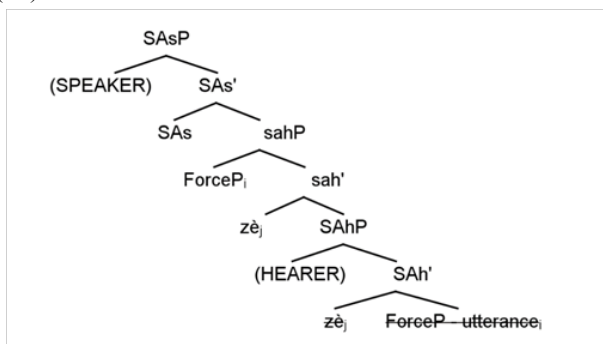
- (22) *Zé*, Valère is doa!  
*Zé* Valère está ali  
 ‘Veja, Valère está ali!’

Por sua vez, *zè* é pronunciado com entonação mais fraca e ocorre em posição final, como no seguinte exemplo de Hill (2014, p. 176):

- (23) Valère is doa *zè*  
 Valère está ali *zè*  
 ‘Valère está ali, como você pode ver.’

A partir dessa descrição, a autora postula a existência de um outro traço no núcleo de sahP, o traço [atenção]. Este traço está presente no núcleo de Sah e quando esta posição é preenchida, é valorado, de modo que se obtém [+atenção]. Este é o caso da construção em (22). Por outro lado, quando o traço [ênfase] está presente e há movimento de ForceP para o especificador de sahP, temos também o movimento da partícula *zè* para o núcleo de sahP, de forma que obtemos [-atenção]. O movimento de ForceP condiciona o movimento da partícula, como se pode observar na configuração abaixo (HILL, 2014, p.179):

(24)



Fonte: Hill (2014, p. 179.)

Na configuração acima, há também o movimento de ForceP para o especificador dessa categoria devido à atuação do traço [ênfase]. Portanto, em (24), o constituinte que se move para o especificador de sahP é alvo do traço [ênfase], mas não do traço [atenção]. O último é verificado pela partícula *zè* a partir do movimento núcleo a núcleo.<sup>13</sup>

De acordo com Hill (2014), a partícula vocativa pode estar associada a traços funcionais: um traço interpessoal e o traço de 2ª pessoa, nos termos da autora, temos, respectivamente, [i-p] e [2nd]. O traço [i-p] codifica a relação interpessoal entre o falante e o ouvinte, ao passo que o traço de 2ª pessoa, ao qual nos referiremos, doravante, como [2ªp], que atribui ao nome o estatuto de ouvinte.

Como veremos na próxima seção, considerando o percurso de *senhor* de adjunto a núcleo, Ramos (2011), uma evidência de que vocativos são licenciados por um traço de 2ª pessoa é a de que não são antecidos por artigos definidos e, portanto, são sintagmas nominais (*Noun Phrases*). Por outro lado, se um nome é precedido por um artigo definido, temos sintagmas determinantes (*Determiner Phrases*). Assim, o núcleo do sintagma determinante, D°, valora nomes de 3ª pessoa, o que não é o caso dos vocativos. Pode-se notar, portanto, uma análise dos vocativos como VocPs, ao invés de NP/DP, é mais adequada, na medida em que a interpretação de ouvinte tem origem em uma configuração sintática em

<sup>13</sup> Hill (2014, p.177) ressalta que o romeno e o flamengo ocidental exibem as mesmas possibilidades para relacionar a distribuição e a interpretação na estrutura de SAhP, uma vez que a leitura de atenção está associada ao valor [+atenção] e à posição inicial.

que VocP seleciona um NP e valora um traço de 2ª pessoa. Dessa forma, este traço de 2ª pessoa de ouvinte também é mapeado na sintaxe do nome, o qual interage na sintaxe com [i-p] na derivação de VocP.

Nesta seção, foi apresentada a proposta de Hill (2007, 2014) para a posição do vocativo e também para a posição da interjeição na interface sintático-pragmática ou domínio de SAP. Em resumo, de acordo com a autora:

- (i) o domínio de SAP é dividido em duas áreas, a saber, a área de ouvinte, SahP, que é o *locus* do vocativo; e a área de falante SASP, onde se situam as interjeições;
- (ii) no domínio de SAP, Caso Vocativo e Papel pragmático “ouvinte” são atribuídos ao vocativo, ao passo que Caso Exclamativo e Papel pragmático “falante” são atribuídos às interjeições;
- (iii) a observação do comportamento sintático da partícula “hai” do romeno que se realiza com duas outras terminações (haideti/ haidem) evidencia a relação entre o núcleo de SahP (o núcleo da área de ouvinte) e ForceP, visto que a terminação da partícula varia de acordo com o tipo sentencial.
- (iv) na área de ouvinte da interface sintático-pragmática, há a atuação de dois traços [ênfase] e [atenção],
- (v) a existência da categoria VocP, situada no especificador de SahP, é evidenciada por um traço de 2ª pessoa que atribui o estatuto de ouvinte ao nome que desempenha função sintática de vocativo;
- (vi) há também outro traço presente em Vocº, qual seja [i-p], o qual codifica a relação interpessoal entre falante e ouvinte;
- (vii) a posição sintática do vocativo está atrelada à função pragmática deste constituinte, de forma que condiciona também a interpretação.

Após a apresentação dos pressupostos teóricos que norteiam este artigo, na próxima seção, descreveremos construções com a partícula “sô”, realizada sintaticamente como vocativo.

### 3 As formas vocativas: uma descrição

Segundo Hill (2014), as formas vocativas podem ser diretas ou indiretas. As primeiras identificam o ouvinte, pois nelas, um vocativo, muitas das vezes constituído pelo próprio nome do interlocutor, sempre

está presente e, portanto, constituem um apelo direto a quem a forma de chamamento é dirigida. Já as formas de chamamento indiretas são representadas pelas partículas vocativas indiretas desacompanhadas de vocativo que podem atuar como chamamento. São exemplos dessas últimas, citados pela autora, em português: *Ei, Oh. Ah* etc. e outras expressões, como *ok, obrigado, olá, oi*, etc.<sup>14</sup>

De acordo com a autora, as formas vocativas podem conter um vocativo, como em (25), uma partícula de chamamento indireto ou uma interjeição<sup>15</sup> seguida por um vocativo, como em (26), uma partícula de chamamento indireto, como em (27), conforme se vê, abaixo:

(25) Mamãe, não está vendo que estou falando com você? PB

(26) Oh, Maria, o que houve? PB

(27) Oh, o que houve? PB

No exemplo (25), a forma de chamamento é composta pelo vocativo *mamãe*; em (26), temos uma forma de chamamento composta pela interjeição “*Oh*” e o vocativo *Maria* e, em (27), é a partícula de chamamento direto *Oh* que configura uma forma de chamamento.

As formas de chamamento ocorrem à esquerda da oração, o que está em concordância com a constatação da existência de dois tipos de vocativos, chamamentos e destinatários, e que a função que desempenham está relacionada com a posição sintática: os chamamentos ocorrem à esquerda da oração, ao passo que os destinatários ocorrem à direita de um constituinte ou da oração.<sup>16</sup> A autora observa ainda que formas vocativas diretas apresentam-se reduzidas em várias línguas, resultando em partículas vocativas, como *(a)be/bre, +le e Ma*, no Búlgaro, *Vre e (mo)re*, no Grego; *Ó e pá*; no Português Europeu; *zé e zê*, no Flamengo Ocidental, dentre outros exemplos, de outras línguas. À semelhança dessas

<sup>14</sup> As partículas vocativas são geralmente tratadas como interjeições nas gramáticas tradicionais.

<sup>15</sup> Seguindo Hill (2014, p. 157), assumimos a existência de dois tipos de interjeições: as propriamente ditas, que expressam o estado mental, as emoções e sentimentos do falante, e as partículas de chamamento indireto, que podem atuar como um chamamento ou como um cumprimento. Para obter mais detalhes sobre diferentes tipos de interjeições, consulte Moreira (2017).

<sup>16</sup> Esta hipótese foi apresentada na Introdução deste texto.

partículas, identificamos, no dialeto mineiro, a partícula *sô*. Exploraremos as possibilidades de ocorrência dessa partícula na seção 3.1.

### 3.1 A partícula vocativa *sô* no dialeto mineiro

Observamos que a partícula *sô* ocorre em diferentes posições na oração:

- (28) a. Uai *sô*, o negócio tá feio.  
       b. Ô *sô*, o negócio tá feio.  
       c. \**Sô*, o negócio tá feio.
- (29) a. Pára com isso, *sô*.  
       b. ?Pára com isso,(#) ô *sô*.  
       c. ?Pára com isso, (#) uai *sô*.
- (30) a. Ô, pára com isso, *sô*.  
       b. Uai, pára com isso, *sô*.
- (31) a. Cadê os prego que eu deixei aqui?  
       b. Os prego, *sô*, sei cadê não.

Observe-se que a realização da partícula *sô* é possível em posição inicial se houver precedência de interjeição. Em (28a), a partícula *sô* é precedida pela interjeição *uai*; enquanto que em (28b) é a partícula de chamamento indireto *Ô* que a precede. Por sua vez, a construção em (28c) é agramatical, visto que não há nenhum item lexical à esquerda da partícula vocativa. Já em posição final, como exemplificado em (29), não é usual que essa partícula seja precedida por interjeição ou partícula de chamamento indireto, a não ser que haja uma pausa enfática entre a oração e um desses constituintes, o que pode indicar que se tem duas construções, estando uma delas antes e outra após esta pausa. Há ainda a possibilidade de uma partícula de chamamento indireto, como *Ô*, ou uma interjeição, como *Uai*, ocorrer ao início da oração, ao passo que a partícula vocativa *sô* ocorre ao final, como em (30a) e (33b), respectivamente. Ressalte-se, ainda, que na resposta, em (31), a partícula vocativa *sô* está alocada à direita de um constituinte topicalizado.

Segundo Ramos (2011, p. 81), a ocorrência do item *sô* na posição sintática de vocativo permitiu identificar que o processo de

pronominalização se desenvolveu nos últimos dois séculos no português e que a ocorrência do item na posição sintática de vocativo favoreceu a reanálise. Esse processo se deu ao final da oração e depois o seu uso foi expandido para outras posições disponíveis para o vocativo. Conforme assinala Head (1976, *apud* RAMOS, 2011, p. 71), não é incomum nas línguas que nomes e expressões nominais adquiram características de pronomes: nome > nome pronominalizado > pronome > pronome pessoal. Para Ramos (2011 p. 70), como já dissemos, *senhor* e *sô* são pronomes, sendo o primeiro uma forma homônima de um nome.

Sobre esse processo, nos termos da gramática gerativa, de acordo com Vitral (2002), a perda de substância fônica, em um processo de gramaticalização/ pronominalização/ cliticização, envolve a mudança de estrutura, isto é, a mudança de projeção máxima para núcleo pode ser vista como uma consequência da atração exercida pelas categorias gramaticais.<sup>17</sup>

Roberts e Rousseau (2003, p.45) exemplificam casos em que um DP<sup>18</sup> em posição de especificador é reanalisado como núcleo. Gelderen (2006, p.4) implementa essa proposta e argumenta que a gramaticalização constitui fonte de inovação linguística prevista por princípios da Gramática Universal, na medida em que três desses princípios, que implicam a noção de economia, são relevantes: (a) *Head Preference Principle* (HPP): quando possível seja núcleo, e não um sintagma (GELDEREN, 2006, p. 6); (b) *Late Merge Principle* (LMP): juntar o mais tarde possível (GELDEREN, 2006, p. 10) e (c) *Specifier Incorporation Principle* (SIP): quando possível, seja um especificador e não um adjunto (GELDEREN, 2006, p.15); e quando possível seja um núcleo e não um sintagma.

Levando em conta esse princípio, Ramos (2011, p.80) postula que o percurso de *senhor* poderia ser descrito assim:

(32) Adjunto > especificador > núcleo nominal > núcleo de determinante

---

<sup>17</sup> Para autor, isso se deve ao fato de que são as categorias gramaticais que atraem itens pertencentes a outras categorias gramaticais e também pertencentes às categorias lexicais, que passam a se alocar, através de movimento ou inserção, nas posições previstas pelo domínio de atração das categorias gramaticais “atratoras”. Conforme o autor, ao que parece há uma cooptação de itens, exercida pelas categorias gramaticais, que provoca a redução fonética e a alteração da natureza do significado (ou esvaziamento semântico), captadas por meio da noção de processo de gramaticalização.

<sup>18</sup> *Determiner Phrase*.

Em (33), *senhor* (flexionado no feminino) se realiza como adjunto:

(33) A *senhora* duquesa deseja alguma coisa? (RAMOS, 2011, p.80.)

Como observa a autora também que a inserção de um adjetivo não torna o sintagma malformado, como no exemplo abaixo:

(34) a. A gentil e honesta *senhora* duquesa deseja alguma coisa?

b. A bela e gentil *senhora* duquesa.

c. A *senhora* e bela duquesa. (RAMOS, 2011, p.80.)

O núcleo do sintagma nominal seria o nome *duquesa*; *senhora* seria o aposto. Conforme a autora, no eixo do tempo, os sintagmas contendo *senhor* deixam de ocorrer com outros elementos. A relação entre o artigo e o item *senhor* se reestrutura, e o sintagma nominal deixa de permitir adjetivos, como em (35b):

(35) A: - Posso me assentar ali?

B': - Sim, o senhor pode se assentar ali.

B'': - \*Sim, um senhor pode se assentar ali.

B''': - \* O bom senhor pode se assentar ali. (RAMOS, 2011, p.71)

Nas respostas em (35), o item *senhor* refere-se àquele que faz a pergunta e apenas o contexto extralinguístico permite identificar que, ele é. Veja-se que em (B'') *senhor* vem precedido de um artigo definido e em (B''') de um adjunto. O resultado, nesses exemplos, é a impossibilidade de *senhor* referir-se a quem faz a pergunta (A).

Além disso, o artigo deixa de poder ser substituído por demonstrativos. Essa rigidez pode ser analisada como resultante de uma reanálise, que fez com que *senhor* deixasse de ser adjunto e passasse a ser núcleo. Uma consequência dessa reanálise foi a realização de *sô*.

(36) A: - Posso me assentar ali?

B': - Sim, *sô* pode se assentar ali.

B'': - \*?Sim, o *sô* pode se assentar ali.

B''': - \*?Sim, um *sô* pode se assentar ali.

B''': - \*Sim, o bom senhor pode se assentar ali. (RAMOS, 2011, p.72.)

A comparação dos diálogos permite verificar que a forma reduzida não aceita artigo definido, enquanto a forma plena o aceita. Isso constitui uma indicação de que o item pleno seria estruturalmente mais complexo do que o item reduzido e aponto para o processo de gramaticalização sobre o qual discorremos.

A descrição do percurso de *senhor* apresentada por Ramos (2011) vai ao encontro do que pressupõe Hill (2014, p.179), ao situar a partícula vocativa *zé* e sua variante *zê*, do Flamengo Ocidental, no núcleo de SAhP ou no núcleo de sahP, como na configuração em (24).<sup>19</sup>

A ocorrência das construções com a partícula *sô* conduz-nos ao entendimento de que a pronominalização do item *sô* se deu ao final da oração e depois o seu uso foi expandido para outras posições disponíveis para o vocativo. Portanto, este é um caso em que um DP em posição de especificador é reanalisado como núcleo. Como afirma Gelderen (2006), essa seria uma etapa necessária. Como ela afirma, “nomes na posição argumental não são núcleos, pois sempre contêm determinante” e acrescenta que “nomes puros ocorrem como adverbiais, predicativos e vocativos” (GELDEREN, 2004, p.187).

Nesta seção, exploramos os contextos de ocorrência da partícula *sô* no dialeto mineiro, considerando a avaliação de falantes nativos, a fim de subsidiar a elaboração das configurações arbóreas correspondentes às sentenças que contém essa partícula, a serem apresentadas na seção 4. Observamos que a realização da partícula *sô* é possível em posição inicial se houver precedência de interjeição e, em posição final, não é precedida por interjeição ou partícula de chamamento indireto, exceto quando há uma pausa enfática entre a oração e um desses constituintes. A descrição dos dados leva-nos a conjecturar que ocorreu um processo de pronominalização da forma *sô*, de forma que *senhor* passou de adjunto a especificador e de especificador a núcleo, em concordância com Ramos (2011).

#### 4 A partícula *sô* na interface sintático-pragmática

Como visto, em posição inicial, a partícula *sô*, deve ser precedida por interjeição ou por uma partícula de chamamento indireto, como se vê nos exemplos abaixo:

---

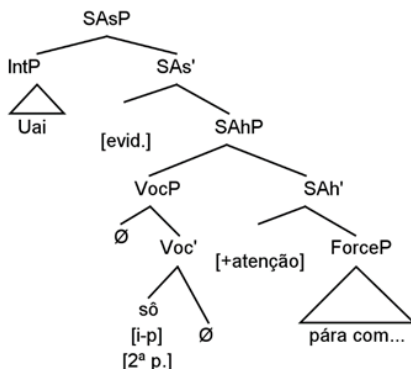
<sup>19</sup> Outro ponto a ser debatido é que, no processo de aquisição, a criança seja exposta a um amplo *input* de estruturas que provêm de evidências de que sintagmas plenos possam ser analisados como tal.

(37) a. Uai sô, pára com isso.

b. Ô sô, pára com isso.

Segue-se a configuração referente a construção em (37a):

(38)



Fonte: Elaboração própria.

Nota-se que há uma diferença do comportamento da partícula “sô” e dos vocativos que exercem função de chamamento. Estes últimos podem ocupar a posição inicial, se precedidos ou não de interjeição ou uma partícula de chamamento indireto. A ocorrência da partícula “sô” em posição inicial é, portanto, mais marcada do que a ocorrência de vocativos nesta posição. O traço [atenção] de SAh é valorado, de modo que se obtém [+atenção]. Ilustramos, ainda, a ocorrência de traços dentro do sintagma vocativo, VocP, a saber o traço de 2ª pessoa, que atribui ao nome que desempenha essa função sintática o estatuto de ouvinte, e o traço interpessoal [i-p], que codifica a relação entre falante e ouvinte.<sup>20</sup>

No que diz respeito à área de falante, a interjeição *uai*, presente no exemplo (37a) seja do tipo assertiva, nos termos Alonso-Cortès (1999). Segundo o autor, esse tipo de interjeição indica implicitamente a avaliação por parte do falante de algo que o interlocutor disse ou algo que aconteceu.

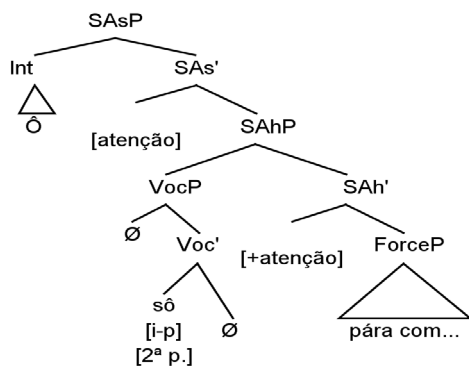
<sup>20</sup> É importante pontuar que não nos dedicamos, neste artigo, a uma análise detalhada das interjeições, entretanto, como estes constituintes obrigatoriamente precedem a partícula de chamamento direto “sô”, em posição inicial, precisamos esboçar uma análise destes constituintes.

Os outros dois tipos de interjeições consideradas pelo autor são: (i) expressivas, que indicam o que o falante está pensando: assombro, surpresa, admiração, dor, lamento, alegria, rejeição, asco etc.; e as diretivas, que acompanham o imperativo, indicando a força ilocutiva do ato instativo.<sup>21</sup>

Hill (2007) se refere a este último tipo de constituinte como partículas de chamamento indireto, diferenciando-os dos outros tipos de interjeições, que se tratam, nesta perspectiva das interjeições propriamente ditas. Partindo de estudo realizado anteriormente (MOREIRA, 2017), consideramos que as interjeições, como *Uai*, em (30) são licenciadas pelo traço [evidencialidade], enquanto que a derivação das partículas de chamamento direto, como *oi*, *olá*, *psiu*, *ô*, envolve o traço [atenção].

A seguir, apresentamos a configuração arbórea referente à construção em (37b):

(39)



Fonte: Elaboração própria.

Na representação sintática acima, a partícula de chamamento indireto “ô” é alocada no especificador de SAsP, na área de falante. Como vimos, este constituinte se enquadra dentre os tipos de interjeições elencados

Considerando a sintaxe interna de VocP, a partícula “sô” está presente em Voc°, o que verifica os traços [i-p] e [2ºp.]. Já em relação à posição do vocativo na frase, temos o preenchimento do especificador

<sup>21</sup> Para mais detalhes acerca da análise das interjeições e das partículas de chamamento indireto, consulte Moreira (2017).

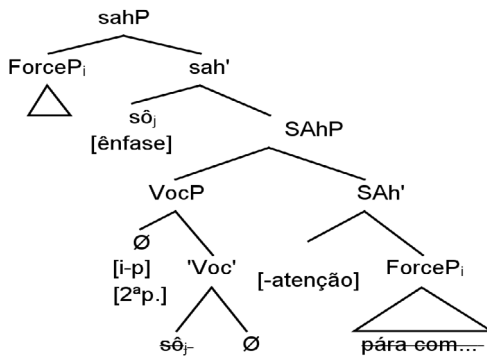
de SasP e, por consequência, a valoração do traço de Sah°, de forma que obtemos [+atenção].

Já em posição final, a partícula *sô* não ocorre precedida por interjeição ou partícula de chamamento indireto, como exemplificado em (29 a) e, novamente, abaixo:

(40) Pára com isso, sô.

A configuração arbórea correspondente é a seguinte:

(41)



Fonte: Elaboração própria.

Na configuração acima, há atuação do traço [ênfase] que desencadeia o movimento de ForceP para o especificador de sahP e também o movimento núcleo a núcleo da partícula *sô*, de forma que obtemos [-atenção] e na interpretação de destinatário.

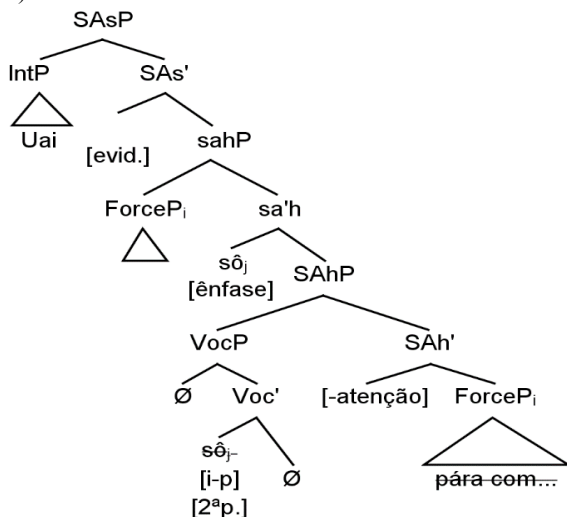
São exemplos apresentados em (37a) e (37b):

(42) a. Uai, pára com isso, sô.

b. Ô, pára com isso, sô.

A configuração referente à construção em (42a) segue abaixo:

(43)



Fonte: Elaboração própria.

Na área de falante, tem-se a ocorrência da interjeição *Uai*, a qual é licenciada pelo traço [evidencialidade]. Já na área de ouvinte, o traço [ênfase] desencadeia o movimento de ForceP para o especificador da categoria discursiva sahP. Apesar disso, a partícula “sô” permanece no núcleo de VocP, categoria que, por sua vez, está situada no especificador de SAhP, o que está em conformidade com a valoração de [-atenção].

Nesta seção, analisamos as ocorrências da partícula *sô* em diferentes posições na oração e a situamos em Voc°, núcleo do sintagma vocativo. Conforme foi observado: (i) a ocorrência da partícula “sô”, em posição inicial é mais marcada do que a dos vocativos nesta posição; (ii) ilustramos a realização de traços dentro e fora do sintagma vocativo, em perspectiva macro e microssintática; (iii) exploramos o comportamento sintático das interjeições que precedem o vocativo, em casos em que este sintagma é alocado em posição inicial, e identificamos aquelas que desempenham função de chamamento em oposição àquelas que exercem função de uma interjeição propriamente dita.

## 5 Considerações finais

Neste estudo, foi possível: (i) descrever as propriedades das construções contendo o item *sô*, que se realiza sintaticamente como vocativo, no dialeto mineiro; (ii) descrever as propriedades das construções contendo este item e a relação com as interjeições ou partículas de chamamento indireto; (iii) definir a posição da forma reduzida *sô* na interface sintático-pragmática e (iv) iniciar o estudo da sintaxe interna de VocP, tendo em vista que a análise da relação do traço de 2ª pessoa ao NP poderá ser explorada em trabalho futuro.

O estudo do comportamento sintático da partícula *sô* possibilitou-nos, ainda, certificar a posição dos chamamentos à esquerda dos destinatários na interface sintático-pragmática. Investigando os contextos de ocorrência de *sô*, no dialeto mineiro e a partir das representações sintáticas, corroboramos com a constatação de Ramos (2011) de que a partir de um processo de pronominalização, houve uma reanálise de *senhor*, que ao passar de nome a pronome, passou de especificador a núcleo. Uma questão emerge neste ponto e ficará cargo de uma nova pesquisa: *sô* pode estar a caminho de alcançar o estatuto de clítico?

Vale dizer, ainda, que a observação do comportamento das interjeições e das partículas de chamamento indireto trouxe-nos pistas de que é válida a distinção entre dois tipos de vocativos, a saber, chamamento e destinatário. Por fim, neste artigo, tentamos oferecer um contributo para o estudo da interface sintático-pragmática e para a descrição do dialeto mineiro.

## Agradecimento

Agradeço ao Prof. Lorenzo Vitral pelas sugestões e observações, que contribuíram para o desenvolvimento deste artigo.

## Referências

ALONSO-CORTES, A. Las construcciones exclamativas. La interjeccion y las expresiones vocativas. In: BOSQUE, DEMONTE (org.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa-Calpe, 1999. v. 3, p. 3993-4050.

BRAGA, T. *Terra ideal*. Revista local em 3 atos. Manuscrito. São João Del Rei: Acervo Clube Teatral Artur Azevedo; Biblioteca da Universidade Federal de São João Del-Rei, 1915.

GELDEREN, E. *Principles and Parameters in Change*. Amsterdam: John Benjamins, 2004.

GELDEREN, E. *Economy of Merge and Grammaticalization: Two Steps in the Evolution of Language*, 2006. Disponível em: <http://www.public.asu.edu/~gelder/elly.htm>. Acesso em: 29 jan. 2020.

HEAD, B. F. Respect Degrees in pronominal reference. In: GREENBERG, J. (ed.) *Universals of Human Language*. Cambridge: Cambridge University Press, 1976. v. 3. p. 151-211.

HILL, V. Vocatives and the Pragmatics-Syntax Interface. *Lingua*, [S.l.], v. 117, p. 2077-2106, 2007. DOI: 10.1016/j.lingua.2007.01.002

HILL, V. *Vocatives: How Syntax Meets with Pragmatics*. Leiden: Brill, 2014. DOI: <https://doi.org/10.1163/9789004261389>

HOUAISS, A.; VILLAR, M. de S. *Minidicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva, 2001. DOI: <https://doi.org/10.5628/rpcd.01.03.03>

MARCOS, P. *Dois perdidos numa noite suja*. São Paulo: Ed. Global, 1978.

MARTINS PENA, L. C. O Noviço. In: \_\_\_\_\_. *Teatro de Martins Pena*. Rio de Janeiro: Ediouro, 1956. p. 293-335.

MOREIRA, J. C. *O vocativo na língua coloquial de Minas Gerais nos séculos XIX e XX: uma abordagem variacionista*. 2005. 92f. Monografia (Bacharelado) – Instituto de Ciências Humanas e Sociais da Universidade Federal de Ouro Preto, Mariana, 2005.

MOREIRA, J. C. *O vocativo no Português Brasileiro nos séculos XIX e XX: um estudo de mudança linguística*. 2008. 108f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

MOREIRA, J. C. Interjeições e invocações: a ordem de constituintes exclamativos no Português Brasileiro. *Revista de Estudos Linguísticos e Literários*, Salvador, v.58, p. 61-82, 2017. DOI: <https://doi.org/10.9771/ell.v0i58.26806>

MOREIRA, J. C.; ALKMIM, M. G. R. de. Preenchedores de vocativo em peças teatrais. In: RAMOS, J. M.; COELHO, S. M. (org.). *Português brasileiro dialetal: temas gramaticais*. Campinas: Mercado das Letras, 2013. p. 73-90.

PAIVA, Modesto de. *Mudança de Capital*. Comédia em 1 ato. Manuscrito do Acervo do Clube Teatral Artur Azevedo. Biblioteca da Universidade Federal de São João Del Rei, São João Del Rei, 1893.

PAIVA, M. de. *A onça*. Revista local em 3 atos. Manuscrito. São João Del Rei: Acervo Clube Teatral A. Azevedo; Biblioteca da Universidade Federal de São João Del Rei, 1897.

RAMOS, J. M. De nome a pronome: um estudo sobre o item senhor. *Caligrama*, Belo Horizonte, v. 16, n. 2, p. 69-83, 2011. DOI: <https://doi.org/10.17851/2238-3824.16.2.69-84>

REZENDE, S. N. C. de. *A virgem Mártir de Santarém*. Drama em 4 atos. Manuscrito Acervo do Clube Teatral A. Azevedo. Biblioteca da Universidade Federal de São João del Rei, São João Del Rei, 1882.

REZENDE, S. N. C. de. *Santo Antônio nas águas*. Manuscrito. São João Del Rei: Acervo do Clube Teatral A. Azevedo; Biblioteca da Universidade Federal de São João del Rei, 1906.

RIZZI, L. The Fine Structure of the Left Periphery. In: HAEGMAN, L. (ed.). *Elements of Grammar*. Dordrecht: Kluwer, 1997. p. 281-337. DOI: [https://doi.org/10.1007/978-94-011-5420-8\\_7](https://doi.org/10.1007/978-94-011-5420-8_7)

ROBERTS, I.; ROUSSEAU, A. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: University Press, 2003. DOI: <https://doi.org/10.1017/CBO9780511486326>

SPEAS, M; TENNY, C. Configurational Properties of Point of View Roles. In: DI SCIULLO, A. M. (ed.). *Asymmetry in Grammar*. Amsterdam: John Benjamins, 2003. p. 315-344. DOI: <https://doi.org/10.1075/la.57.15spe>

VITRAL, L. T. A interpolação de “se” e as consequências para a teoria da cliticização. *Revista da Abralin*, Aracaju, v.1, n. 2. p. 161-197, 2002. DOI: <https://doi.org/10.5380/rabl.v1i2.52697>

WERNECK, A. *Lucrecia*. Belo Horizonte: Imprensa Oficial de Minas Gerais, 1900.



## Reeditar é preciso? As cartas oficiais norte-rio-grandenses e os corpora diacrônicos

### *Is it necessary to reedit? The Rio Grande do Norte official letters and the diachronic corpora*

Felipe Morais de Melo

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN),  
Natal, Rio Grande do Norte / Brasil

[felipemorais\\_m@yahoo.com.br](mailto:felipemorais_m@yahoo.com.br)

<https://orcid.org/0000-0003-1881-6262>

**Resumo:** São bastante estreitas e profícuas as relações entre a Linguística Histórica e a Linguística de *Corpus*, sobretudo a partir da década de 90, quando a primeira ganha fôlego renovador no Brasil, tendo na edição de *corpora* diacrônicos uma das motrizes basilares para seu progressivo desenvolvimento. Sob o signo desse entrelaçamento entre as duas áreas, este trabalho, através da análise de conteúdo, apresenta um caso de reedição de *corpus* diacrônico, as *cartas oficiais norte-rio-grandenses* (1713-1950), e, por meio de pesquisa bibliográfica, propõe algumas reflexões acerca do contato entre esses dois campos de estudo no Brasil e dos impactos, positivos e negativos, do meio digital nessa conexão. O artigo revela como alterações nas orientações de uma investigação podem justificar um processo frutífero de reedição e evidencia tanto riscos éticos quanto aprimoramentos funcionais a que esses *corpora* estão sujeitos devido a fatores como a tecnologia e a internet, propondo, ao final, que haja maior visibilidade das questões levantadas no estudo.

**Palavras-chave:** reedição; cartas oficiais; corpus diacrônico; linguística histórica.

**Abstract:** The relations between Historical Linguistics and Corpus Linguistics are quite close and fruitful, especially from the 90's, when the former gains renewed breath in Brazil, having in the edition of diachronic corpora one of the basic drivers for its progressive development. Under the sign of this intertwining between the two areas, this work, through content analysis, presents a case of reedition of a diachronic corpus, the *Rio Grande do Norte official letters* (1713-1950), and, through bibliographic research, proposes some reflections about the contact between these two fields of study in Brazil

and the impacts, positive and negative, of the digital medium in this connection. The paper reveals how changes in research guidelines can justify a fruitful process of reedition and highlights both ethical risks and functional improvements to which these corpora are subject due to factors such as technology and the internet, proposing, in the end, a greater visibility of the issues raised in the study.

**Keywords:** re-edition; official letters; diachronic corpus; historical linguistics.

Recebido em 13 de outubro de 2020

Aceito em 22 de dezembro de 2020

## 1 Considerações iniciais

La biblioteca es ilimitada y periódica. Si un eterno viajero la atravesara en cualquier dirección, comprobaría al cabo de los siglos que los mismos volúmenes se repiten en el mismo desorden (que, repetido, sería un orden: el Orden). Mi soledad se alegra con esa elegante esperanza<sup>1</sup>

(BORGES, 2009, p. 566).

editar é preciso

(CAMBRAIA, 1999, p. 14).

Enxergamos – na mecânica borgiana pela qual se antepõe ao espelho descoberto o próprio infinito e da qual extraímos essas linhas em que a engrenagem ficta do escritor argentino ressuma em corredores de estantes – uma alegoria do próprio fazimento de se (re)editar um *corpus*. É quase como se ouvíssemos uma tradução filosófica (claramente circunstancial e assaz pragmática dentre as tantas e tão profundas expressões que os cantos de Borges podem ativar e evocar) do que escreve Almeida Cabrejas: “es imposible en las arduas tareas de transcripción y edición no cometer errores; la única manera de reducir su aparición es la

---

<sup>1</sup> “A Biblioteca é ilimitada e periódica. Se um eterno viajante a atravessasse em qualquer direção, comprovaria ao fim dos séculos que os mesmos volumes se repetem na mesma desordem (que, reiterada, seria uma ordem: a Ordem). Minha solidão alegra-se com essa elegante esperança” (BORGES, 1998, p. 523).

revisión”<sup>2</sup> (ALMEIDA CABREJAS, 2013, p. 15). Ainda que revisitado, revisado, repassado, recuidado, no cabo dos dias, num próximo lance de olhos, para além do reconhecimento das melhorias, constatar-se-á a repetição da desordem: outros erros que ficaram, velhos deslizes que sobraram, e outros bruxuleios, inda que só de ilusórias imprecisões. Se a ciência não nos dá – a história sim – a paciência da eternidade como meta para a ordem, contamos com, pelo menos, mais um par de repetições da desordem, para, titubeando-a, aproximá-la do eixo:<sup>3</sup> na eventualidade de uma futura publicação, a sação para novas revisões.

Mesmo em face de possíveis errâncias, Cambraia (1999, p. 19) solta um peremptório aviso: “editar é preciso”. A primeira razão que embasa essa sua assertiva é o fato de o linguista nem sempre ter acesso direto aos manuscritos, seja por causa das dificuldades impostas pelos arquivos, seja pela distância geográfica desses centros, sendo-lhe de interesse, portanto, uma edição. A segunda diz respeito ao cariz que o manuscrito expõe: abreviaturas, legibilidade, a escrita, enfim, de outrora exige conhecimento para que possa ser lida ou paciência a fim de se obter a competência necessária para tal. Por fim, Cambraia (1999, p. 14) faz uma ressalva aos fac-símiles, fotografias e cópias xerográficas: “nem mesmo esses recursos são capazes de reproduzir com absoluta fidelidade as características de um original”. Daí por que, mesmo oferecendo-se um edição fac-símile – de grande relevância por permitir que o leitor faça, sempre que queira ou careça, um cotejo da transcrição com o documento original –, é importante que a edição, ainda que de forma superficial, descreva alguns feítios materiais dos documentos apresentados.

Este trabalho desenvolve-se precisamente sob esses auspícios: o da repetição/revisão e o da edição. Inicialmente, apresentaremos as *cartas oficiais norte-rio-grandenses*, um *corpus* diacrônico editado (2012), e reeditado (2018). Tratando-se de uma temática pouco explorada – porque irrelevante ou (quase) inexistente?! –, este artigo, após esta

---

<sup>2</sup> “é impossível, nas árduas tarefas de transcrição e edição, não cometer erros; a única maneira de reduzir sua aparição é a revisão” (tradução nossa, doravante TN).

<sup>3</sup> As remissões são fatais, tanto pelo mote quanto pela gravidade borgiana (cada barra vertical simples exprime mudança de verso; a dupla, de estrofe ou espaço entre título e texto): “Errância || Só porque | erro | encontro | o que não se procura || só porque | erro | invento | o labirinto || a busca | a coisa | a causa da | procura || só porque | erro | acerto: me | construo. || Margem de | erro: margem | de liberdade” (FONTELA, 2006, p. 202).

introdução (seção 1), concentra praticamente metade de seu texto à compreensão do que são essas cartas (seção 2), do que ensajou e de como se operacionalizou sua reedição (seção 3). A segunda metade (seção 4), mais conceitual (histórica e metodologicamente, embora amiúde entrecortada por exemplos ora das *cartas* ora da pesquisa grafemática que, com base nelas, se realizou), percorremos as confluências entre a Linguística Histórica e a Linguística de *Corpus* por duas sendas mui intimamente conectadas ao próprio (re)fazimento das *cartas oficiais*: a virada na Linguística Histórica brasileira que se engendrou a partir do final dos anos 90; e algumas consequências da integração dos *corpora* diacrônicos ao mundo digital, maiormente as referentes à divulgação dos dados e aos novos instrumentais tecnológicos.

## **2 As cartas oficiais norte-rio-grandenses (2012-2018)**

As *cartas oficiais norte-rio-grandenses* são um *corpus* diacrônico nascido no mestrado de Moraes de Melo (2012), constituindo um agrupamento de 107 documentos (da C1 até a C107)<sup>4</sup> que circularam, entre 1713 e 1931, no meio público com temáticas atinentes ao que hoje é o Rio Grande do Norte. As principais razões que motivaram o autor a levar a cabo essa empresa foram seu interesse, desde a época da licenciatura, pelos fenômenos da variação e mudança linguísticas e, conjunturalmente, a sua participação, de 2010 a 2012, na equipe do Para a História do Português Brasileiro no Rio Grande do Norte (PHPB-RN), projeto local cujos objetivos centrais eram estabelecer um *corpus* mínimo comum essencial para as pesquisas em Linguística Histórica desenvolvidas no Rio Grande do Norte e contribuir na construção de um banco de dados para o programa nacional Para a História do Português Brasileiro (PHPB). Como consequência, um trabalho dessa sorte legaria material empírico para investigadores dos estudos da linguagem, auxiliando sobretudo as pesquisas diacrônicas acerca do português brasileiro. O autor, que tinha como plano de mestrado inicial efetuar uma análise de ofícios novecentistas potiguares, terminou, após a entrada no PHPB-RN, expandindo seu escopo e responsabilizando-se por uma das

---

<sup>4</sup> C1 deve ser lido “carta um”; C107, “carta cento e sete” e assim sucessivamente. Apresentamos, ainda neste tópico 2, uma lista com os intervalos de cartas que fazem parte de cada quarto de século, desde a primeira metade do XVIII até a segunda metade do XX.

categorias, as cartas oficiais, do “*corpus* mínimo comum – manuscrito” estabelecido pelo grupo nacional.

Ao lado da edição – que constou como apêndice da dissertação e foi incorporada ao conjunto de *corpora* reunidos no site do PHPB e disponibilizados para consulta de todos os interessados,<sup>5</sup> estava o corpo da pesquisa: uma caracterização das cartas, organizadas do ponto de vista da Paleografia, da Diplomática e mais pormenorizadamente das Tradições Discursivas (TD). Através desta última vertente, foram analisadas macroestruturas (no caso em tela, os gêneros textuais) e as microestruturas do tipo fórmula textual presentes no *corpus*.<sup>6</sup>

Na sua tese de doutorado, Morais de Melo (2018) procedeu a uma análise em que o texto cedesse seu protagonismo a outro ator, os usos (orto)gráficos sem transcendência fônica, passando as cartas a funcionar como palco para essa nova averiguação. Para tanto, deliberou revisar o labor editorial de 2012 para dirimir eventuais erros, conforme a lógica supracitada de Almeida Cabrejas (2013, p. 15), bem como excluir quaisquer textos não lavrados no Rio Grande do Norte a fim de situar o resultado final o máximo possível nas trilhas de uma dialetologia histórica mais controlada. As duas ações continham, em seu horizonte, a expectativa de engendrar um *corpus* que superasse o primeiro em rigor, fidedignidade e transparência de modo a atender melhor aos estudos científicos, mormente no campo da Linguística Histórica voltada para o português brasileiro. No tópico seguinte, “Por que reeditar?”, serão esmiuçadas melhor essas duas decisões e as demais que alteraram o primeiro *corpus*, gerando as *cartas oficiais norte-rio-grandenses* de 2018, um conjunto ordenado de 129 cartas (da C1até a C129) escritas no Rio Grande do Norte e trocadas por pelo menos um representante da administração pública no intervalo de 1713 a 1950.

O qualificativo “oficiais” é empregado para nomear o *corpus*<sup>7</sup> em consonância com a seguinte definição retirada do *Manual de redação da Presidência da República* (BRASIL, 2002, p. 4):

---

<sup>5</sup> Disponível em: <https://sites.google.com/site/corporaphpb/home/corpora-manuscritos/manuscritos-rio-grande-do-norte>. Acesso em: 12 jan. 2020.

<sup>6</sup> Um recorte dessas análises pode ser lido em Morais de Melo e Lima (2016a, 2016b) e Morais de Melo (2017).

<sup>7</sup> Uma explicação mais esmiuçada do porquê das expressões “cartas” e “oficiais” para batizar o *corpus* encontra-se em Morais de Melo (2012, p. 90-92).

As comunicações oficiais são necessariamente uniformes, pois há sempre um único comunicador (o Serviço Público) e o receptor dessas comunicações ou é o Serviço Público (no caso de expedientes dirigidos por um órgão a outro) – ou o conjunto dos cidadãos ou instituições tratadas de forma homogênea.

Contudo, esse paradigma é relativizado por Barbosa (2002, p. 428), para quem os “documentos da administração colonial em circulação pública” podem ser de duas naturezas: os de *caráter deliberativo oficial*, respeitando aos documentos emanados de órgãos ou autoridades do governo, do Serviço Público, em concordância com os conceitos focalizados acima; e os de *requerimento pessoal*, que são endereçados a representantes ou entidades do serviço público, mas não são deste necessariamente provenientes, podendo terem sido escritos por cidadãos externos à administração pública. É precisamente essa definição que nos brinda Barbosa (2002) a que melhor convém às *cartas oficiais*.

Inclui-se nesse círculo de correspondências públicas a esfera jurídica, que não participava do primeiro *corpus*, de 2012, uma vez que, para a reedição de 2018, grande parte das cartas obtidas no Arquivo Público Estadual do Rio Grande do Norte provinha de transações judiciais. Embora estranho à conformação original, essa esfera indiscutivelmente integra a administração pública, o que se comprova por qualquer busca na internet em que se encontra, a modo de ilustração, que o judiciário é um dos três poderes do Estado (público, portanto, não privado).

A grande maioria dos documentos de nosso *corpus*<sup>8</sup> (102 dos 129, isto é, 79,06%) tem como missivistas pessoas que representam o Governo, sejam capitães-mores (C1, C2 e C3, por exemplo), provedores (C5 e C6, v.g.), escrivães da Fazenda Real (C8 e C31, e.g.), presidente de uma província (C62, C67), secretários do governo (C72, C80) ou juízes (C108, C109). As demais 27 cartas (20,93%) são enviadas para (mas não por) membros da administração pública, sendo a maioria delas (20 documentos) requerimentos, como a C32, de Francisco de Paulo Moreira, solicitando confirmação de seu posto de cirurgião; ou a C119, de um 3º sargento ao presidente do Supremo Tribunal de Justiça.

---

<sup>8</sup> Neste artigo, o vocábulo *corpus* empregado de modo absoluto dirá respeito sempre às *cartas oficiais norte-rio-grandenses* reeditadas em 2018. A referência à primeira edição sempre virá acompanhada de algum especificador do tipo “o *corpus* de 2012”, “o primeiro *corpus*” ou soluções afins.

Somente em duas dessas 27 cartas não é nenhum dos interlocutores um membro estrito da administração pública, na C42 e na C111: a primeira é uma cópia de um ofício enviado por José Barbosa e João Luís Pereira a membros do clero com o intuito de que estes intercedessem à Rainha para a obtenção de verba em prol da construção de um hospício de Santo Antônio na cidade de Natal; e a segunda, um ofício enviado por Arthur Napoleão Soares de Macedo ao coronel José Soares Filgueira Sobrinho questionando o militar sobre a compra de terras. De toda forma, além de, na C42, haver uma remissão final à Rainha, aparecem, como uma das pontas da correspondência, em ambos os casos, componentes que, a seu modo, se incluíam no *modus operandi* da administração brasileira. Afinal, não é gratuita a inclusão, na obra *A administração no Brasil Colonial*, organizada por Salgado (1985), dos quatro seguintes eixos organizacionais: “Estrutura judicial”, “Administração fazendária”, “Organização militar” (a exemplo da C111, escrita já fora dos lindes coloniais, mas num contexto em que essa organização exercia papel de destaque na administração pública) e “Administração eclesiástica” (no caso da C42). Sobre essa última instância, de modo especial, comprovando seu papel integrante na dinâmica administrativa (pelo menos na colonial, da qual a C42 faz parte), lemos:

A expressão ‘funcionários eclesiásticos’ dá bem a medida de como a Igreja nascente nas terras americanas dependia do Estado português, situação que se prolongou por todo o período colonial e durante o Império, cujas raízes, conforme já dissemos, estavam no direito do padroado. Além da integração político-religiosa, a Coroa se beneficiou, e muito, da sua condição de administradora dos dízimos eclesiásticos, em muitas regiões talvez a principal fonte da renda colonial (SALGADO, 1985, p. 115).

Afora todas essas justificativas, encetadas pela concepção de Barbosa (2002) sobre documentos da administração pública e aliadas às de Salgado (1985), essas cartas que não são lavradas por pessoas que assumem, no sentido mais estreito, função na máquina pública alinham-se às demais por uma nota que não se pode perder de vista: todos os 129 textos que integralizam as *cartas* representam uma minoria letrada e com razoável ou alto nível de instrução numa sociedade em que rareavam as condições de oferta e acesso à educação.

Além dessas questões por trás de “oficiais”, o uso de “cartas oficiais” é também uma declarada evocação às cartas oficiais da Paraíba

organizadas por Fonseca (2003, p. 120-121) e à decisão nomenclatória da autora:

Dos documentos selecionados, encontram-se neste *corpus*, de acordo com a classificação feita por Martinheira: *carta de lei*, *carta régia*, *aviso dos secretários*, *ofícios*. Dentre esses, os *ofícios* serão focalizados com a denominação geral de cartas por serem maioria absoluta e por apresentarem maior heterogeneidade.

Embora na nova edição os *ofícios* não representem a maioria absoluta (no antigo, ocupavam 70% do *corpus*), ainda é o gênero mais expressivo, uma vez que das 129 cartas (denominação geral), 63 (48,83%) são *ofícios* que formam, somados às 25 cartas,<sup>9</sup> 68,2% do total. A maior diferença entre as primeiras *cartas* e a reedição de 2018 foi o crescimento relevante de requerimentos, que constituem agora o segundo gênero mais recorrente, com 26 exemplares (20,15%). Os três juntos – *ofícios*, *requerimentos* e *cartas* – expressam 88,35% das 129 *cartas oficiais norte-rio-grandenses*.

Outra elucidação é sobre o porquê da marcação do termo com *itálico*. Decidimos fazê-lo para ressaltar a identidade do *corpus* organizado e editado nesta tese, e evitar, pelo registro de cartas oficiais norte-rio-grandenses, assim, sem marcas, a abertura para se ler, nas menções que fazemos da expressão, a ideia, lassa e equívoca para este estudo, de estarmos tratando de qualquer carta escrita ou no Rio Grande do Norte ou por norte-rio-grandenses e em qualquer período, a despeito do perímetro bem delimitado com o qual trabalhamos e no qual as análises aqui promovidas e suas conclusões estão estritamente circunscritas.

---

<sup>9</sup> O *ofício* é um gênero de estrutura bastante semelhante à da carta e da carta régia, diferenciando-se, contudo, por uma razão de ordem sociofuncional: são cartas oficiais trocadas, em sua maioria, entre membros da administração pública, sob a condição de nenhum dos coenunciadores, o remetente ou o destinatário, ser o Rei. A tradição da carta é, formalmente, muito símile à do *ofício*, mas é ativada pela necessidade de comunicar algum assunto relativo à administração pública ao Rei, daí todas elas trazerem como *inscriptio* apenas o pronome de tratamento “Senhor” que, conforme explica Fonseca (2003, p. 150), é exclusivo para autoridade real. Devido à fixidez de seu endereço, o Rei, a carta goza sempre de um caráter ascendente, isto é, é escrita por alguém hierarquicamente inferior à pessoa a quem o documento é dirigido. *Inscripção* é um termo da diplomática o qual podemos encontrar na obra de Belloto (2002) e indica uma das partes que pode ocorrer no protocolo inicial (equivalente à seção pré-textual) de um documento.

Do ponto de vista metodológico, decidimos fixar um mínimo de 2.500 palavras por quarto de século. Embora o trabalho de 2018 já não estivesse atrelado a nenhuma ação do PHPB – ao contrário do empreendimento de 2012, que esteve formalmente envolvido com o projeto –, resolvemos seguir, conquanto de forma adaptada,<sup>10</sup> a orientação dada pelo programa nacional de obter 5.000 palavras para cada metade de centúria. Vejamos abaixo as cartas de cada período seguidas por uma descrição geral sobre os documentos de cada século.

- XVIII.1.1<sup>11</sup> – 13 cartas, da C1 (1713) até a C13 (1725);
- XVIII.1.2<sup>12</sup> – 10 cartas, da C14 (1726\*)<sup>13</sup> até a C23 (1747);
- XVIII.2.1 – 12 cartas, da C24 (1756) até a C35 (1772\*);
- XVIII.2.2 – 10 cartas, da C36 (1777) até a C45 (1798);
- XIX.1.1 – 14 cartas, da C46 (1806) até a C59 (1822\*);
- XIX.1.2 – 16 cartas, da C60 (1833) até a C75 (1849);
- XIX.2.1 – 13 cartas, da C76 (1852) até a C88 (1874);
- XIX.2.2 – 16 cartas, da C89 (1876) até a C104 (1891);
- XX.1.1 – 14 cartas, da C105 (1913) até a C118 (1919);
- XX.1.2 – 11 cartas, da C119 (1931) até a C129 (1950).

---

<sup>10</sup> A adaptação concerne a trabalharmos com quarteis, e não com metades de século, como estabelece o PHPB e como foi acatado em Moraes de Melo (2012). Essa alteração deu-se durante a pesquisa doutoral para que se pudesse controlar melhor eventuais mudanças nos usos gráficos – objeto de análise de Moraes de Melo (2018) – no eixo diacrônico investigado (1713-1950).

<sup>11</sup> O primeiro algarismo arábico após a indicação do século alude a alguma metade do século (1 representa a primeira e 2, a segunda) e a segunda cifra arábica indica os primeiros ou os últimos 25 anos dessa metade. Assim, de cada centenário correspondente, 1.1 expressa o intervalo que vai do ano 01 ao ano 25; 1.2, de 26 até 50; 2.1, de 51 a 75; e 2.2, de 76 a 00. Nessa linha, XVIII.2.1 engloba um hiato que vai de 1751 a 1775. Nas *cartas oficiais* de 2018, a primeira carta desse período é a C24, de 1756, e a última é a C35, escrita provavelmente em 1772.

<sup>12</sup> Agradecemos à professora Carmen Alveal, do Departamento de História da UFRN, e a seus bolsistas, por terem cedido, em 2011, a transcrição de quase 2.000 palavras da primeira metade do século XVIII para nosso *corpus*.

<sup>13</sup> O asterisco informa que a data não é exata (não consta no documento), mas sugerida por Lopes (2000).

**Textos do século XVIII:** somando 45 cartas oficiais (19 cartas, 10 requerimentos, 6 ofícios, 3 cartas de registro, 3 certidões, 1 provisão, 1 carta de sesmaria, 1 carta régia e 1 carta patente), os documentos setecentistas de nosso *corpus* foram retirados de um conjunto de textos do Arquivo Ultramarino de Lisboa (AHU) relacionados ao Rio Grande do Norte. Esse conjunto foi catalogado por Lopes (2000), professora aposentada do departamento de História da UFRN, como parte do projeto Resgate Barão do Rio Branco, e disponibilizado pela equipe de História da referida universidade. Esses documentos do AHU inventariados por Lopes (2000) abarcam dois séculos de cartas oficiais, datadas as mais antigas de 1623 até as de 1823. Interessaram-nos apenas aquelas escritas a partir de 1701, conforme nosso recorte cronológico. Mesmo comungando da afirmação de Acioli (1994, p. 62), para quem “a dificuldade de leitura não reside obrigatoriamente no retrocesso cronológico”, porquanto houve, de fato, cartas setecentistas fáceis de serem transcritas, vimos nessa fonte os textos mais difíceis para serem transcritos, principalmente pela ilegibilidade de alguns e pela considerável presença de documentos com trechos em escrita que julgamos ser processada,<sup>14</sup> ambas evitadas em nossa seleção.

**Textos do século XIX:** contendo 59 cartas oficiais (46 ofícios, 6 cartas, 5 requerimentos e 2 relatórios), os textos oitocentistas mais antigos das *cartas oficiais* ainda foram obtidos dos documentos do AHU catalogados por Lopes (2000). Mas cerca de 50% da primeira metade e todos os da segunda metade da centúria em pauta são ofícios localizados no Instituto Histórico e Geográfico do Rio Grande do Norte (IHGRN), preservados em oito caixas de arquivo, havendo em cada uma delas subpastas organizadas por data. O ofício mais antigo oriundo do Instituto de que dispusemos em nosso *corpus* data de 1812, e o mais recente, de 1891. As imagens desses documentos não estavam captadas, como os do século XVIII, exigindo um trabalho de campo para conhecer o acervo, selecionar as cartas e fotografá-las a fim de transcrevê-las.

---

<sup>14</sup> Segundo Acioli (1994, p. 42), foi uma escrita que promovia “uma rapidez de traçado extraordinária”, tornando-se “a preferida dos escrivães”. É uma escrita cursiva e de traços deturpados (degeneração da Cortesã, consoante a autora), caracterizada pela separação irregular das palavras, confusão no traçado das letras e grande quantidade de traços supérfluos.

**Textos do século XX:** com 25 cartas oficiais (11 ofícios, 11 requerimentos, 1 auto, 1 termo e 1 certidão) – número reduzido, quando comparado aos dos séculos anteriores, por nosso *corpus* cobrir apenas a primeira metade dos Novecentos –, os documentos foram conseguidos graças a visitas ao Arquivo Público Estadual do Rio Grande do Norte para seleção, fotografiação e posterior transcrição para a edição. Esse intervalo de 50 anos é marcado pela presença majoritária de documentos do circuito jurídico, com cartas expedidas, *v. g.*, por juízes, promotores públicos e vários pedidos/ordens de habeas-corpus (categorizados como requerimentos), o que não se verificava nas *cartas oficiais* de 2012.

Duas últimas considerações sobre as *cartas* de um modo geral, antes de entrarmos nas razões que motivaram sua reedição. A primeira tange à sua ordenação. A numeração que vemos, de 1 a 129, atende quase absolutamente à cronologia das cartas, com exceção da C111 e C112. Devido à desatenção, só percebemos, em fase bastante avançada da tese, que a posição deveria ser invertida, posto que a C111 data de 26 e 27 de agosto de 1918, ao passo que a C112 foi escrita em 10 de agosto do mesmo ano, devendo, por conseguinte, ser a C111. Devido aos vários desdobramentos que essa mudança implicaria e diante do pouco tempo que nos restava quando avistamos o erro, preferimos deixá-lo.

O segundo ponto tem a ver exatamente com essa dupla datação da C111. São duas datas porque, na realidade, existem dois ofícios escritos, não obstante no mesmo fôlio. O primeiro é remetido por Arthur Napoleão Soares de Macedo ao coronel José Soares Filgueira Sobrinho, perguntando ao militar se ele comprara quinhentos braços de terra à firma Severo & Irmão. Ao pé deste documento, conforme solicitado por Arthur Soares (<Preciso que VS., abem da verdade, me | responda ao pé desta>), há um segundo ofício com a resposta do coronel. Se optamos por evitar a correção na numeração das C111 e C112, pois, ainda que se tratasse de apenas uma permuta, reclamaria uma cadeia de pequenos ajustes, imagine-se a inclusão de uma outra carta: afetaria a numeração das 19 cartas que sucedem a C111, gerando a necessidade de um sem-fim de reparos. Mantivemos, assim, uma carta dupla, que, desmembrada, faria com que esta nova reedição contabilizasse não 129, mas 130 documentos.

### 3 Por que reeditar?

Nesta seção, apresentaremos as razões que subjazem a decisão de revisitar e reeditar as *cartas oficiais norte-rio-grandenses* (MORAIS DE MELO, 2012) através de cinco eixos: a exclusão de cartas, a revisão das transcrições, a procura de novas cartas, a transcrição desses novos documentos e a nova apresentação desse *corpus* diacrônico.

#### 3.1 Exclusão de cartas

A primeira edição das *cartas* constitui um conjunto de 107 missivas escritas na administração pública no intervalo de 1713 a 1931 em que questões vinculadas ao Rio Grande do Norte são abordadas.

Quantitativamente, 82, 24% do *corpus* de 2012 é redigido no Rio Grande do Norte, o que representa 88 das 107 cartas, ao que se somam 5 (4,6% do total) escritas no Rio de Janeiro, 1 (0,9%) em Sergipe, 4 (3,7%) em Lisboa e outras 9 sem identificação tópica precisa. Daí pode-se afirmar que 87,8% (a soma das cartas escritas no Rio Grande do Norte, no Rio de Janeiro e em Sergipe) representam o que Barbosa (2002) chama de português *no* Brasil, isto é, textos lavrados em solo brasileiro, sem terem sido emanados, necessariamente, de um punho brasileiro (entenda-se uma pessoa que, mesmo não tendo nascido no Brasil, tenha vivido a maior parte de sua vida, principalmente seus anos de formação, em terras brasileiras).

Embora nem todas as cartas tivessem sido escritas no Rio Grande do Norte, decidimos, à época, cunhá-las *cartas oficiais norte-rio-grandenses* no rasto de Acioli (1994, p. 55, grifo nosso), para quem, diante da influência ibérica impressa patentemente nos manuscritos brasileiros, avisa:

*Entendeu-se que brasileiros seriam todos os manuscritos relacionados com o nosso país, quer oriundos do Brasil, quer de Portugal. Assim sendo, tanto cartas remetidas da colônia quanto documentos régios ou consultas do Conselho Ultramarino, despachados da metrópole, foram considerados brasileiros quando o assunto em questão descreve problemas desta possessão portuguesa na América.*

Após vários diálogos no período da tese, especialmente com os orientadores, decidimos excluir, no processo de reedição, todos os documentos do primeiro *corpus* que não tivessem sido escritos no

Estado. Muito embora o contorno analítico levado a cabo em Morais de Melo (2018) – o exame dos usos gráficos sem transcendência fônica, um plaiño, portanto, sumamente gráfico – não estivesse a princípio atrelado a variáveis diatópicas, a escrita, em sua latência, pode revelar liames com o geográfico, máxime no que toca aos virtuais rastros da oralidade deixados nos traços visuais.<sup>15</sup> E como aspirávamos a que o novo *corpus* não estivesse a serviço apenas da fatia investigativa executada dentro das balizas doutorais, mas que se constituísse em um *corpus* útil à mais ampla gama de finalidades dentro das lindas que oferece, resolvemos proceder a essa talha.

---

<sup>15</sup> Numa disciplina sobre metodologia de pesquisa ofertada por José Luis Ramírez Luengo aos alunos de mestrado e doutorado da Universidad Autónoma de Querétaro (México), em 2015, a que o autor deste artigo assistiu na condição de ouvinte, comentou-se mais de uma vez sobre a existência, entre os historiadores da língua do domínio hispânico que se interessam por questões grafemáticas, de certo entendimento segundo o qual a variável geográfica não acarreta influxo nas realizações gráficas sem transcendência fônica. Eximiam-se, porém, dessa aparente incolumidade as possíveis marcas da oralidade decalcadas na escrita e alguns usos gráficos relativos ao período de formação das escritas romances, na Idade Média. No que tange a esse último ponto, Sánchez-Prieto Borja (2006), por exemplo, mostra como as tradições de escrita medievais também eram criadas no seio de povoados, geralmente pela fonte irradiadora de monastérios, chancelarias e câmaras régias, como se pode depreender de trechos como: “Claro que la valoración de las [sic] usos díspares que muestran los documentos también puede hacerse en clave sociolingüística, teniendo en cuenta posibles diferencias entre un núcleo de población en el que los notarios públicos adoptaban usos que si no habían logrado una estandarización sí estaban extendidos más allá del primitivo reino de Castilla, mientras que la tradición monástica de Vega de Espinaredo se revela más arcaizante, o quizá simplemente más acorde con la lengua hablada” [“Claro que a valoração dos usos díspares que os documentos mostram também pode se tornar chave sociolingüística, levando em conta possíveis diferenças entre um núcleo de povoação no qual os notários públicos adotavam usos que, se não haviam alcançado uma standardização, estavam, sim, estendidos para além do reino de Castela, enquanto a tradição monástica de Vega de Espinaredo se revela mais arcaizante, ou talvez simplesmente mais de acordo com a língua falada” (TN)] (SÁNCHEZ-PRIETO BORJA, 2006, p. 240) e “Desde años me ha llamado la atención las marcadas diferencias dentro del llamado ‘castellano alfonsí’ entre los grandes códices de la Cámara Régia y los documentos salidos de la Cancillería” [“Desde muitos anos tem me chamado a atenção as diferenças marcadas dentro do chamado ‘castelhano alfonsino’ entre os grandes códices da Câmara Real e os documentos saídos da Chancelaria” (TN)] (SÁNCHEZ-PRIETO BORJA, 2006, p. 241).

Além dessas 19 cartas eliminadas por critérios diatópicos, também ficou decidida a eliminação de cópias que não fossem de época. Destarte, apenas mantivemos, das antigas cartas, a C23 (atual C27), C31 (atual C42) e C66 (atual C85). Por outro lado, todas as 21 cartas oficiais que compunham a primeira metade do século XX foram descartadas por se tratar de códices diplomáticos em registros,<sup>16</sup> todos transcritos por uma mesma mão (registradas talvez num mesmo dia) com a intenção de arquivar em livro os ofícios, não tendo sido efetivamente, as referidas cópias, enviadas. Ao final desse processo de filtro, 35 das 107 cartas de 2012 tinham sido rejeitadas para inteirar o *corpus* de 2018.

### 3.2 Revisão das transcrições

Silva Neto (1956, p. 27) discorre sobre uma lista de erros “mais comumente observados na leitura de manuscritos medievais”. A despeito de ser obra escrita na década de 50 e estar pensada para documentos medievais, o que nos relata o grande filólogo é um rol de alertas bastante atuais e de extrema importância para qualquer pesquisador que, no seu afã investigativo, tenha de praticar o exercício de transcrição textual, independentemente da época do documento a ser copiado.<sup>17</sup> São eles os erros devidos à má compreensão de letras, a desconhecimentos de fatos linguísticos, à ignorância de siglas e abreviaturas, os saltos-borrões<sup>18</sup> e os erros devidos à má separação das sílabas. Muito embora, na nossa labuta do mestrado, tenhamos efetuado uma transcrição cuidadosa, atenta sobremaneira a esses quesitos, e que foi revisada pelo autor, pelos orientadores e talvez pela equipe do PHPB nacional, uma vez que foi

---

<sup>16</sup> Os códices diplomáticos estão agrupados em livro. Quando esses códices são do tipo registro, as cópias dos documentos exarados são organizadas por sequência cronológica.

<sup>17</sup> Uma prova explícita da validade desse elenco que, na década de 50 fichou Silva Neto (1956), podemos encontrá-la em Rumeu, Barbosa e Callou (2002). Os autores escolhem alguns exemplos de textos coloniais transcritos, a maioria deles ligados à Biblioteca Nacional (anais, acervos), em que houve ou equívoco na transcrição ou na interpretação feita da transcrição por causa de um dos cinco erros (todos os cinco são ordeira e ilustrativamente contemplados no artigo dos três pesquisadores) de modo a evidenciar as achegas do grande filólogo brasileiro.

<sup>18</sup> Assim inicia o tópico destinado ao problema dos saltos-bordões: “Muitas vezes, por falta de atenção, o copista saltava uma ou várias linhas” (SILVA NETO, 1956, p. 31).

disponibilizada no site de *corpora* do projeto,<sup>19</sup> achamos por bem repassar cada uma das 72 cartas mantidas daquele *corpus* de 2012. Descansados pelo distanciamento dos anos, os olhos – quiçá sazoados, pelo repouso, para o labor transcripcional – lograram decifrar muitas palavras (e até frases) que tinham sido assinaladas com [*inint.*] (ininteligíveis), bem como localizar alguns equívocos (efeitos ora de uma, ora de outra causa de deslize alistada por SILVA NETO, 1956), que foram devidamente consertados.

Algumas cartas, no tempo do mestrado, foram transcritas no próprio IHGRN, tendo sido suas fotos tiradas, para o trabalho de edição, dos documentos arquivados nas caixas de ofícios que nos foram disponibilizadas.

IMAGEM 1 – Fotos com caixas de ofícios disponibilizadas pelo IHGRN entre 2010 e 2012



Fonte: Elaboração própria.

Algumas das fotos tiradas na época, culpa de nosso descuido, não ficaram tão nítidas como deveriam. Durante o processo de revisão, vimos a necessidade de sanar algumas dúvidas voltando ao Instituto. Infelizmente, porém, ao longo de todo o período do doutorado (2014-2018), o mais importante arquivo público do Estado permaneceu com suas portas fechadas para pesquisa devido a reformas que acabaram sendo embargadas num processo que parecia sem fim. Em mais de uma ocasião, explicando a responsáveis pela casa nossa necessidade de consumir apenas algumas verificações para um trabalho que gozou de todo o apoio da própria equipe do IHGRN entre 2010 e 2012, tentamos, sem êxito, ter acesso às caixas de ofícios (mostradas nas fotos acima)

<sup>19</sup> <https://sites.google.com/site/corporaphpb/home/corpora-manuscritos/manuscritos-rio-grande-do-norte>.

nas quais estavam guardadas essas cartas oficiais. Em março de 2017, o Instituto reabriu suas portas, mas apenas de modo parcial: para exposições e eventos. Continua fechado para pesquisa, impedindo a consecução dessa nova fotografiação.<sup>20</sup>

No final das contas, mesmo diante de persistentes percalços, Morais de Melo (2018) oferece, com absoluta certeza, um *corpus* com transcrições mais atiladas, com menos lacunas e com menos erros. Não perfeito, porém. Tão pertinente e irrepreensível é o comentário de Almeida Cabrejas (2013) acerca do trabalho de transcrição e tão bem resume o nosso sentimento ao final da reedição que reiteramos, malgrado já possa ser lido na introdução deste artigo, uma parte dele: “incluso donde no hay dudas, es imposible en las arduas tareas de transcripción y edición no cometer errores; la única manera de reducir su aparición es la revisión”<sup>21</sup> (ALMEIDA CABREJAS, 2013, p. 15). Sem dúvida, caso venhamos a publicar em livro as *cartas oficiais norte-rio-grandenses* (e esperamos que, para este então, já tenha retomado seu pleno funcionamento o IHGRN, de modo que possamos rever alguns documentos e precisar algumas outras informações), procederemos a mais uma revisão com o intuito de identificar e solucionar o máximo possível de problemas remanescentes, aproximando, assim, nosso ofício transcricional de um estado ótimo.

### 3.3 Procura de novas cartas

Quando elaboramos a edição de 2012, tínhamos como meta suprir as demandas do PHPB, isto é, obter 5.000 palavras para cada metade de século da primeira metade do século XVIII até a segunda do XX. Como estávamos responsáveis por textos manuscritos, abandonamos o intervalo dos últimos 50 anos (1950-2000) por não encontrarmos, nas caixas de documentos a que tivemos acesso, textos escritos à mão após 1950. Todos já estavam tipografados. Após a supressão dos 35 documentos desenredada no tópico acima, terminamos com uma quantidade bastante irregular de palavras para cada corte cronológico. Ainda que nossa tese não estivesse atrelada a nenhum projeto que assentasse uma quantidade

<sup>20</sup> Em outubro de 2020, lê-se “Temporariamente, não está aberto para pesquisa em seus acervos”, na página principal de seu *site*: <http://ihgrn.org.br/instituicao/instituto>.

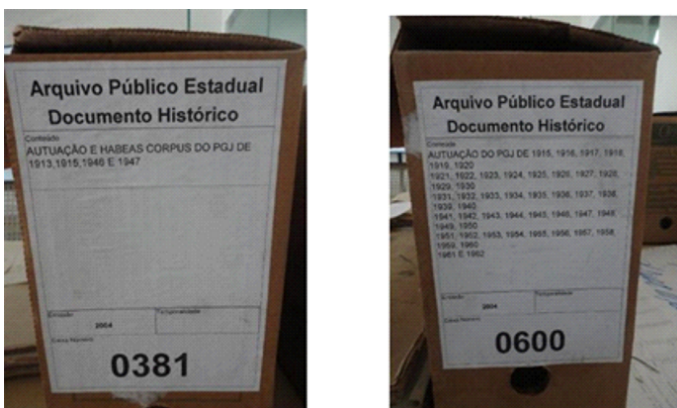
<sup>21</sup> “Inclusive onde não há dúvida, é impossível nas árduas tarefas de transcrição e edição não cometer erros; a única maneira de reduzir sua aparição é a revisão” (TN).

mínima de palavras e mesmo sabendo da possibilidade de se trabalhar com quantidades desiguais (efetuando-se os cálculos de forma proporcional ao total de cada lapso de tempo delimitado, neutralizando-se, dessarte, por meio de um equilíbrio relativo, as somas conflitantes de palavras), optamos por reestabelecer um equilíbrio absoluto entre os intervalos temporais.

Para se conseguir preencher os vãos abertos após a exclusão das 35 cartas, foi preciso voltar à lide primeira: ir a arquivos, procurar textos, selecioná-los, trazer à luz velhos espíritos em escrita sobrevividos. Parafraseando o efusivo mote que nos presenteia Tarallo (1990, p. 175), *picaretas em punho*, voltamos a *cavar*. Para todo o século XVIII até 1823, ainda podíamos contar com a farta base de dados do Arquivo Ultramarino de Lisboa catalogada por Lopes (2000) e a cujo conjunto integral de fotos temos acesso. O problema maior era recauchutar as baixas, que não foram muitas, do século XIX após 1823 e, de modo crítico, todo a primeira metade do século XX. Malogrados os intentos de conseguir aceder ao IHGRN, vimos no Arquivo Público do Estado nossa nova manancial de documentos.

Revivemos todas os estágios por que passamos entre 2010 e 2012: fizemos algumas visitas ao Arquivo, quando, guiados pelo crivo do olho para reconhecer os textos que fossem aparentemente mais legíveis e assegurados de que tinham sido escritos em Natal e circulados na administração pública, fotografamos uma série de cartas para posterior sondagem e eventual transcrição e inclusão como parcela documental do *corpus*.

IMAGEM 2 – Algumas das caixas cedidas para pesquisa pelo Arquivo Público



Fonte: Elaboração própria.

A maioria das missivas preservadas nas caixas que nos foram cedidas era relativa a trâmites da burocracia jurídica. Não nos agradou, a princípio, incluirmos sob o rótulo de “cartas oficiais” textos de ordem jurídica. Todavia, uma vez que quase todos os textos se afinavam com os gêneros com que já vínhamos trabalhando (ofício e requerimento,<sup>22</sup> majoritariamente) e que o poder judiciário se incorpora à administração pública, sofreamos a inquietude e aceitamo-los como “cartas oficiais”.<sup>23</sup> A seguir, elencaremos, para cada quarto de século, a quantidade e tipo de documentos na configuração final após o trabalho de revisão das cartas remanescentes, seleção e transcrição das novas. Entre parênteses, temos a quantidade de textos inéditos (num total de 57), isto é, que não faziam parte de Moraes de Melo (2012). Assim, “7 cartas (2)”, na primeira linha, significa que há 7 cartas, duas das quais foram obtidas durante os afazeres doutorais; e “8 requerimentos (8)”, na última linha, revela que dos 8 requerimentos, todos são “novos”.

---

<sup>22</sup> Cinco dos documentos que classificamos como requerimentos a rigor são pedidos/ordem de *habeas corpus*. Analisando, não obstante, sua estrutura com base no estudo prévio que realizamos sobre as tradições discursivas dos gêneros textuais/espécies documentais que compunham o *corpus* de 2012 (MORAIS DE MELO, 2012), verificamos que se tratava da mesma macroestrutura, contendo, ainda por cima, algumas das mesmas expressões formulaicas que mapeamos para esse gênero no referido estudo. Além disso, após conversa com advogados a quem mostramos esses cinco documentos, eles afirmaram que, antes de serem pedidos/ordens de *habeas corpus*, todos são com efeito requerimentos.

<sup>23</sup> Inclusive se olharmos a tabela que pautou as agendas de trabalho do PHPB disponível em Hora e Silva (2010, p. 423), para a macrocategoria “corpus mínimo comum - manuscritos”, notaremos que, dentre as seis tipologias definidas (a saber: testamento, processos-crime, atas da câmara, cartas particulares, cartas da administração privada e cartas oficiais), é na de cartas oficiais que esses textos que circularam no meio jurídico melhor se inserem.

QUADRO 1 – Relação com intervalo de cartas, quantia de cada gênero e, entre parênteses, número de exemplares inéditos para cada quartel de século

SÉCULO XVIII	XVIII.1.1	C1 a C13 – 7 cartas (2), 3 cartas de registro, 1 requerimento (1), 1 certidão e 1 ofício. TOTAL: 13 cartas, sendo 3 inéditas.
	XVIII.1.2	C14 a C23 – 8 cartas (5), 1 carta de sesmaria (1) e 1 requerimento. TOTAL: 10 cartas, sendo 6 inéditas.
	XVIII.2.1	C24 a C35 – 4 cartas (1), 2 certidões (1), 1 carta régia, 1 carta patente (1), 3 requerimentos e 1 ofício. TOTAL: 12 cartas, sendo 3 inéditas.
	XVIII.2.2	C36 a C45 – 1 provisão, 5 requerimentos (4) e 4 ofícios. TOTAL: 10 cartas, sendo 4 inéditas.
SÉCULO XIX	XIX.1.1	C46 a C59 – 6 cartas (6), 4 requerimentos (4), 4 ofícios (3). TOTAL: 14 cartas, sendo 13 inéditas.
	XIX.1.2	C60 a C75 – 15 ofícios e 1 requerimento. TOTAL: 16 cartas.
	XIX.2.1	C76 a C88 – 7 ofícios, 2 relatórios (2) e 4 ofícios (1). TOTAL: 13 cartas, sendo 3 inéditas.
	XIX.2.2	C89 a C104 – 16 ofícios. TOTAL: 16 ofícios.
SÉCULO XX	XX.1.1	C105 a C118 – 11 ofícios (11) e 3 requerimentos (3). TOTAL: 14 cartas, todas inéditas.
	XX.1.2	C119 a C129 – 8 requerimentos (8), 1 auto (1), 1 certidão (1) e 1 termo (1). TOTAL: 11 cartas, todas inéditas.

Fonte: Elaboração própria.

### 3.4 A transcrição das cartas

Para a transcrição, continuamos seguindo grande parte das normas propostas pelo PHPB. O modelo de transcrição foi publicado por Mattos e Silva (2001), intitulado “Normas para transcrição de documentos manuscritos para a história do português brasileiro”. Em 2010, elas foram repassadas para todas as equipes regionais do projeto, dentre as quais para o PHPB-RN, com algumas atualizações, chamando-se “Normas de transcrição de documentos manuscritos e impressos”. O modelo direciona uma edição semidiplomática dos documentos, que Spina (1977, p. 79) também chama de diplomático-interpretativa, uma vez que “vai mais longe na interpretação do original, pois já representa uma tentativa de melhoramento do texto, com a divisão das palavras, o desdobramento das abreviaturas (trazendo as letras, que não figuram no original, colocadas entre parênteses) e às vezes até com pontuação”.

Cambraia indigita esse tipo de transcrição como presumivelmente a melhor para uma edição que tenha em mente os linguistas:

Quando se tem em mente como principal público-alvo (mas não o único) lingüistas, o tipo mais adequado parece ser a **edição semidiplomática**, pois esse tipo de edição, em uma versão um pouco mais conservadora do que como definida por Spina (1994), tem como vantagem respeitar ao máximo as características do original, fazendo-se, no entanto, pequenas intervenções (sempre assinaladas!) com o objetivo de viabilizar a leitura de seu público. Embora voltada para um público em especial, isto não significa que estudiosos de outras áreas não possam também utilizá-la: este tipo de edição serve também, por exemplo, a pesquisadores de literatura ou historiadores que, com um pequeno esforço inicial para se habituarem ao sistema de transcrição adotado, certamente não encontrarão maiores dificuldades na leitura do texto (CAMBRAIA, 1999, p. 16).

No caso dos critérios propostos pelo PHPB, não há inclusão de sinais de pontuação e as palavras que estão unidas por razões linguísticas, como a natureza clítica de algum termo, são mantidas, sendo separadas, contudo, aquelas unidas devido à morfologia das letras do original manuscrito, que, por apresentarem geralmente as extremidades alongadas, alcançam a palavra seguinte e a ela se ligam por seus ornatos. Numa transcrição diplomática, por outro lado, ocorre uma “reprodução tipográfica do original manuscrito, como se fosse completa e perfeita cópia do mesmo” (SPINA, 1977, p. 78).

### 3.5 Lição justalinear e facsímile

Desde a época do mestrado, não julgamos o arranjo dos documentos transcritos proposto pelo PHPB como o melhor no sentido de promover uma justa correspondência visual entre a transcrição e o documento original. A funcionalidade do formato se via ainda mais comprometida à medida que pensávamos numa edição facsimilar, pois invalidava parcialmente o potencial de localização e comparação que o facsímile compreende. Trazemos, à continuação, um recorte da C61 seguido pela transcrição antiga acompanhada pela atual. Através delas, podemos contrastar seus efeitos, perceptíveis sobremodo quando vistos em cotejo com a foto.

Aproveitemos, antes, para, a partir desse microexemplo que virá, detalhar alguns poucos parâmetros que, embora talvez já sejam dados por certos numa lição justalinear, julgamos merecedores de menção.

Um dos vários trabalhos que efetuam esse tipo de registro nos moldes que seguimos, conquanto não reserve nenhuma parte do livro para dar, como o fazemos (inda que assaz sumariamente), um esclarecimento sobre os critérios seguidos, é Megale, Toledo Neto e Fachin (2009, p. 29-61). Entendemos que a bússola básica desse tipo de apresentação dita: preservar o máximo possível da disposição visual que caracteriza o documento original. Dessa maneira, um texto que esteja mais à direita será transcrito mais à direita (vejamos, no quadro abaixo, na coluna reservada à lição justalinear, a localização da assinatura <Bazilio Quaresma Torreão>, colacionando-a com a foto), as proporções (veja que <Bazilio> está alinhado verticalmente a <Natal> de forma parelha ao original), os espaços (note-se a sangria em <Deus Guarde>) tentarão ser mantidas etc.

Esse rigor, sem embargo, não é total, máxime no que respeita aos espaçamentos. Ao atentarmos para o fragmento de foto, divisamos que o espaço entre a linha que termina em <1833> e a assinatura de Basílio Torreão é maior do que o que se emprega entre a assinatura e a linha seguinte (que começa com <Ill.<sup>mos</sup> Senhores>). Na transcrição, contudo, as duas foram registradas como um espaço simples (uma linha em branco apenas). Há outros casos, como o da C1, em que a assinatura, no manuscrito, se dá no final da página, ao passo que o texto imediatamente anterior (o final da seção textual) se encontra na parte de cima. Seriam necessários, caso se quisesse ater, na transcrição, para esse aspecto, pelo menos 20 ou 25 linhas em branco para se emular o que figura o original. Deixamos, entretanto, uma apenas.

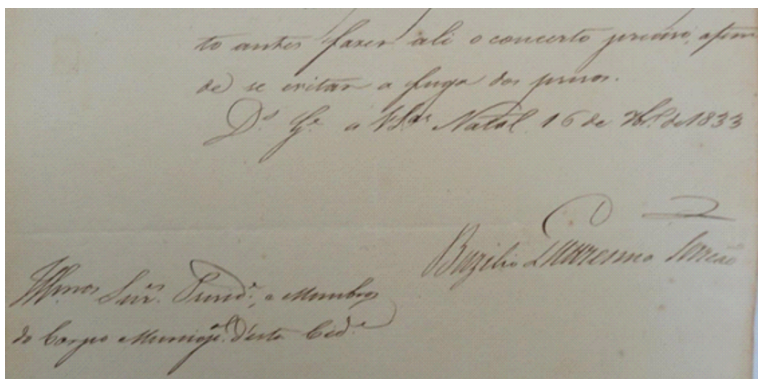
Duas últimas convenções devem vir à luz antes de irmos à mostra. A primeira diz respeito a uma situação em que a transcrição não honrará a biunivocidade linha a linha ansiada pela lição justalinear. Haverá casos esporádicos em que a transcrição de uma linha do documento original, devido à quantidade de abreviaturas desenvolvidas, não pôde ser comportada numa única linha do Word.

Nesses eventos casuais, lançaremos mão do colchete aberto ( [ ) para indicar que aquela linha introduzida por esse sinal não corresponde a uma linha do original, mas ao final da linha anterior. Por fim, uma decisão que diz respeito à disposição da edição nas páginas da tese e que foi tomada em vista de se tentar reduzir um pouco o volume do trabalho. Sempre que a carta for suficientemente pequena de modo que caibam, numa única página, tanto a foto do fôlio quanto sua transcrição,

elas serão justapostas nessa mesma folha. Do contrário (a maioria dos casos), surgirá, de cada fôlio e nesta ordem, a transcrição e a imagem cada qual numa página.

Casos explanados, aduzimos o exemplo.

IMAGEM 3 – Fragmento da C61



Fonte: Elaboração própria.

QUADRO 2 – Comparação entre a disposição da transcrição seguida em Morais de Melo (2012) e Morais de Melo (2018)

Morais de Melo (2012) seguindo o modelo do PHPB	Lição justalinear de Morais de Melo (2018)
<p>to antes fazer ali o concerto preciso, afim   de se evitar a fuga dos presos.    Deus Guarde a VS.<sup>as</sup> Natal 16 de abril de 1833   Bazilio Quaresma Torreão    Ill.<sup>mos</sup> Senhores Presidente, e Membros   do Corpo Municipal d'esta Cidade</p>	<p>to antes fazer ali o concerto preciso, afim de se evitar a fuga dos presos. Deus Guarde a VS.<sup>as</sup> Natal 16 de abril de 1833 Bazilio Quaresma Torreão Ill.<sup>mos</sup> Senhores Presidente, e Membros do Corpo Municipal d'esta Cidade</p>

Fonte: Elaboração própria.

#### 4 Linguística de *Corpus* e Linguística Histórica

Depois de apresentada a reedição das *cartas oficiais norte-rio-grandenses*, desenvolvemos, como última parte deste artigo, uma

discussão mais geral acerca da relação entre a Linguística de *Corpus* e a Linguística Histórica nos estudos da linguagem. Essa conexão possui um escopo assaz vasto e poderia, portanto, ser abordada segundo uma miríade de perspectivas diversas. A que trazemos nas páginas seguintes é bastante parcial, na medida em que trata de um recorte espacial (a guinada dos estudos históricos no Brasil pós-década de 90) e temático (a disponibilização de *corpora* diacrônicos na rede mundial de computadores) que estão vinculadas diretamente a reflexões, inquietações, vivências e roteiros que formaram a travessia do autor deste artigo pelas *cartas oficiais* ao longo de 8 anos (2010, início do mestrado – 2018, final do doutorado), período no qual foi sujeito-(agente e paciente)-pesquisador sob a constelação desse entrelaçamento.

A Linguística Histórica é o grande campo dos estudos da linguagem que trata de estudar as mudanças por que passa, com o transcurso do tempo, a língua em seus mais diversos níveis de análise. Mattos e Silva (2008) distingue uma linguística histórica no sentido *lato* e outra no *stricto sensu*. A primeira corresponde a qualquer tipo de linguística que trabalha com *corpora* datados e localizados; a segunda, apenas com dados provenientes de sincronias passadas. Esta última, a linguística histórica *stricto sensu*, por sua vez, ainda se subdivide, para a autora, em uma linguística histórica de caráter sócio-histórico, por considerar fatores extralinguísticos na análise de fenômenos da língua; e uma linguística diacrônica, que possui uma natureza associal, considerando, sobretudo, os fatores imanentes à estrutura. Como não efetivamos qualquer tipo de controle sobre a biografia dos escreventes (uma tarefa, por si só, já bastante complicada, o que se complexifica ainda mais quando se busca desvendar a real mão que escreveu a carta),<sup>24</sup> a análise promovida por Moraes de Melo (2018) sobre os usos gráficos nas *cartas oficiais* 8 não teve qualquer cacife para, com propriedade, imiscuir-se por variáveis extralinguísticas, fazendo, portanto, para usar dos termos lançados por Mattos e Silva (2008), uma linguística diacrônica de caráter grafemático.

---

<sup>24</sup> Sobre as dificuldades implicadas na identificação da mão que lavra um documento escrito em sincronias pretéritas, *vide* Gonçalves e Ferreira (2001, p. 485-486). Muito valioso igualmente é o testemunho de Mattos e Silva (2002) sobre os óbices enfrentados por vários trabalhos ligados ao PHPB quanto à constituição de *corpora* diacrônicos.

Os estudos históricos da língua,<sup>25</sup> que hoje se albergam preponderantemente sob o título destrinchado pela autora baiana, dimanam de uma tradição filológica.<sup>26</sup> Concentrando nossa atenção em modelos clássicos de obras que faziam filologia e mesmo História da Língua no século XX, notamos que a fonte primária para os manuais e artigos/ensaios investigativos eram, de modo semelhante ao que se faz hoje, manuscritos ou, mais comumente, edições feitas desses manuscritos... desde que se tratasse do período medieval. Para os demais séculos, a fonte básica eram geralmente os textos literários ou, às vezes, obras metalinguísticas da época a ser descrita. Podemos confirmar este panorama novecentista através de três marcos consagrados: para o início do século, as *Lições de Filologia portuguesa*, escritas nos primeiros anos da década de 10 por Vasconcelos (s/d); para o meio, os *Ensaios de Filologia portuguesa*, de Silva Neto (1954 [primeira edição]); e o *I seminário de Filologia e língua portuguesa*, ocorrido em 1997 (RODRIGUES; ALVES; GOLDSTEIN, 1999). Um passar de olhos sobre os índices ou uma folheada nessas obras é suficiente para que se comprove o diagnóstico feito acima. Isso justifica o porquê de, em 1996, Castro escrever:

Em relação ao português clássico, em especial dos séculos XVII e XVIII (para não falar do desconhecidíssimo XIX), quem o quiser estudar tem de se resignar a fazer de cabouqueiro, desenterrando penosamente os seus documentos, peneirando os

---

<sup>25</sup> Aqui estamos falando do contexto lusófono, e mais especificamente do brasileiro, porque é sobre o qual temos maior conhecimento, muito embora a mudança no *zeitgeist* metodológico marcada na década de 90 que será comentada na sequência do parágrafo do qual esta nota deriva, pelas discussões que tivemos na disciplina dada por José Luis Ramírez Luengo já anteriormente referida em nota de rodapé, também tiveram lugar no quadro hispânico, inclusive um pouco antes do que se dá no Brasil, na década de 80 pelo menos. Vejam-se, por exemplo, as alusões que Fontanella de Weinberg (1998, p. 95) faz a seus trabalhos anteriores.

<sup>26</sup> Sobre esse sulco filológico, recomendamos a bela – e mais extensiva, haja vista trazer alguns exemplos que extrapolam os muros da realidade brasileira, dentro dos quais se situam nossos apontamentos – discussão que promove Maia (2012). Principiando seu artigo com a contextualização da guinada (cujo expoente nas terras brasileiras foi o PHPB) que inseriu, no final do século XX, a Linguística Histórica na agenda dos estudos da linguagem, concentra-se, então, na relação entre Filologia e Linguística Histórica, enfocando, primeiramente, o peso daquela nesta e, logo, a importância desta (e da Linguística moderna como um todo) para aquela.

dados, organizando uma taxionomia inexistente e, se ainda tiver coragem e tempo de vida, formulando hipóteses interpretativas que ficarão à espera de um debate crítico só possível se outros investigadores se transviarem pelos mesmos terrenos (CASTRO, 1996, p. 137).

O registro de uma transição radical vemo-lo um ano após a citação de Castro e no mesmo ano em que sucede o supra aludido Seminário de Filologia: em 1997, quando ocorreu o I Seminário para a História do Português Brasileiro, cujos textos frutos do evento foram publicados, em livro, no ano seguinte (CASTILHO, 1998). A grande mudança que as discussões levantadas nesse seminário traziam, no bojo dos compromissos a que se propunha o PHPB em formação, era – ao lado das preocupações com a sintaxe histórica e a história social, que despontam desde a publicação de Castilho (1998) e às quais se unem, em obras futuras do projeto, outras, como as TD – a constituição de um *corpus* diacrônico do português brasileiro. Reflexo disso é uma série de artigos em que se vão relatando depoimentos acerca de processos de formulação de *corpora*, alguns dos quais mencionamos: Carneiro e Almeida (1998) para documentos do século XVIII ao XX de Feira de Santa, na Bahia; uma primeira sondagem dos textos disponíveis na Biblioteca Nacional, por Barbosa (1998); os primeiros passos para a formação de um banco de documentos paranaenses, por Cyrino, Barrichello e Paula (2002); um *corpus* quase concluído com cartas de homens “ilustres” na Bahia, por Carneiro e Almeida (2002); dois *corpora* (um do século XVIII e o outro, do XIX) com cartas públicas e privadas paranaenses (o último texto da série de livros *Para a história do português brasileiro* cujo cerne está no testemunho acerca da confecção de *corpora*), por Rumeu (2006).

Os filólogos e linguistas interessados pela História da Língua viam-se envolvidos numa maratona que demandava idas a arquivos, manuseio de textos (manuscritos em seu grosso), seleção, transcrição (nos próprios arquivos ou, caso permitida fotografiação ou microfilmagem, em outros ambientes) e edição de documentos. Numa aplicação cabal do que delinea Berber Sardinha (2000, p. 325) sobre um de seus campos de interesse – “A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” –, os novos linguistas históricos estavam

fazendo linguística de *corpus*, e em completude: não só explorando os conjuntos ordenados de dados, mas coletando-os, conformando-os.<sup>27</sup>

Muito rapidamente essa caminhada de labor arquivístico viu-se, pelas premissas da modernidade, interpelada pelo mundo digital, o que fazia jus à própria esteira constitutiva da Linguística de *Corpus*, cuja história, segundo Berber Sardinha (2000, p. 329), está “intimamente ligada à disponibilidade de corpora eletrônicos”. Se desde o primeiro da série *Para a história do português brasileiro* viamos, a cada livro, uma seção devotada a “estudos sobre a organização do *corpus* diacrônico” (ALKMIN, 2002, p. 7, por exemplo), somos surpreendidos, no quarto volume da série (DUARTE; CALLOU, 2002), com uma produção congruente e judiciosa em que é cogitada, por prismas diversos (o ético, inclusive), a vinculação desses *corpora* à rede mundial de computadores: trata-se do artigo de Barbosa, Lopes e Callou (2002).

Os autores, após definirem o termo *corpus diacrônico* como um “conjunto de materiais de diferentes sincronias passadas” (BARBOSA; LOPES; CALLOU, 2002, p. 30), alertam que a consolidação desse agrupamento de dados organizado não reside apenas nas precisões tecnológicas das ferramentas de busca, mas na “formação do *corpus* de textos sobre o qual os instrumentos incidem”. Daí o necessário esmero que tivemos na confecção das *cartas oficiais norte-rio-grandenses*, e em sua reedição, no que toca à observância de fatores como o recorte temporal, local de produção e natureza do documento. Dada essa definição de *corpus diacrônico*, os autores anunciam:

A equipe carioca [do PHPB], que tem suas raízes no Projeto NURC-Rio, deixa disponível, desde já, na *internet*, seu *corpus* diacrônico: transcrições de impressos do século XIX e edições diplomático-interpretativas de manuscritos dos séculos XVIII e XIX. *Essa ação pioneira* contribui para que a comunidade acadêmica venha a trabalhar com dados mais seguros na reconstrução da história do português brasileiro, a modalidade linguística de mais de 160 milhões de pessoas (BARBOSA; LOPES; CALLOU, 2002, p. 30, grifo em “Essa ação pioneira” nosso).

<sup>27</sup> Bawarshi e Reiff (2013, p. 55-59) dedicam um item de sua obra para refletir sobre as contribuições que a “linguística histórica/de *corpus*” (assim, em quase todas as aparições, grafada) tem para dar “com a pesquisa e ensino de gêneros ao dar conta da natureza das tipologias e da mudança linguística”.

Esse passo de integração de uma, poderíamos cunhá-la deste modo, Linguística de *Corpus* Histórica à rede virtual, realizada, neste novo cenário do fazer diacronia linguística surgido na década de 90 em terras *brasilis*, pela primeira vez (ao menos segundo a autoclassificação de pioneira lida na citação acima) pela equipe do PHPB-RJ viria a se tornar a recomendação repassada pelo PHPB (a meta era a constituição desses *corpora*, havendo a orientação de que fossem, após concluídos, disponibilizados em rede) a todas as equipes locais. Efeito disso foi a criação da plataforma <https://sites.google.com/site/corporaphpb/>. Tornou-se, assim, acessível ao grande público uma série de *corpora*, ordenados em três grandes grupos: *corpora* impressos, *corpora* manuscritos e *corpus* diferencial. Em cada um deles, tem-se via para materiais produzidos pelas equipes regionais de diversos estados e resultantes, como afirma a página principal, “de teses, dissertações, livros, CDs, relatórios, enfim, não apenas diferentes resultados acadêmicos de trabalhos concebidos e produzidos sob a chancela do PHPB, mas também de colaboradores externos”. Aí estão as *cartas oficiais* fruto de Moraes de Melo (2012).

Amplamente promissora, portanto, a investida de Barbosa, Lopes e Callou (2002). O artigo segue, no entanto, com algumas vigilâncias a que esse tipo de associação – *a priori*, tão prolífica, salutar e, como muito bem asseveram os autores, legítima/necessária, mormente quando esse trabalho de coleta e formação de *corpora* contém, em seus passos, financiamento público<sup>28</sup> – deve atentar. Alguns deles estão relacionados com o crédito dado ao autor do trabalho. Os autores exibem um cabeçalho de sua edição eletrônica para fazer-nos ver que, a despeito de haver um trabalho conjunto, deve ficar discriminado o autor exato que se responsabilizou por cada documento. Com isso, prosseguem, “os indivíduos não verão sua produção acadêmica dispersa e escamoteada no trabalho coletivo, mas, pelo contrário, ela será contabilizada na edição eletrônica a ser citada pelas pesquisas de outrem” (BARBOSA; LOPES; CALLOU, 2002, p. 31). Isso atenua, mas não extingue, os riscos de um uso – quiçá oriundo de uma

---

<sup>28</sup> “Nossa política de divulgação assume que o acesso virtual deva ser gratuito, uma vez que o ônus da execução foi, direta ou indiretamente, do dinheiro público. Somente faz sentido cobrar pela confecção de versões em CDs em função dos gastos materiais e de prestação de serviço posteriores à publicação da página da rede. A cobrança por *softwares* deveria seguir a mesma lógica, caso não tenham sido financiadas pelo setor privado” (BARBOSA; LOPES; CALLOU, 2002, p. 31).

compreensão (insonte ou não) de que a disponibilização de um material em linha, ao permitir o “domínio” público (escorreito seria dizer “acesso” público), é de todos, é da rede e, por conseguinte, não tem autoria – do acervo aberto sem as devidas referências.<sup>29</sup>

Outro cuidado (e com ele fechamos nosso bosquejo sobre esse artigo que aporta questões extremamente relevantes sobre o fazer linguística de *corpus* histórica nos tempos de hoje, questões muitas vezes dadas por certas e, por isso, talvez, não trazidas à baila, não ditas) ao qual se deve atender afeta ao ineditismo que se queira preservado por parte do autor: “É primordial que se resguardem edições que sejam objeto de teses, dissertações ou de trabalhos que os editores julguem de maior relevo antes de serem incluídos no acervo de cada equipe” (BARBOSA; LOPES; CALLOU, 2002, p. 32).

Dando prosseguimento a essa pauta e, ao mesmo tempo, cosendo o remate deste item 4, que, por sua vez, faz despontar o fecho do próprio artigo, seguimos a trilha desenhada por alguns dos artigos que compõem o livro organizado por Kabatek (2016a). Abrindo o livro, deparamo-nos com um estudo do próprio organizador em que, partindo da constatação de que existem diferentes “disciplinas” contidas na denominação Linguística de *Corpus*, o autor distingue três vertentes: uma primeira, destinada à elaboração de *corpus*, englobando suas etapas, que vão desde a coleta até a apresentação dos dados organizados; uma segunda, à primeira estreitamente interligada, que respeita ao tratamento dos dados por meio de etiquetagem, contagens e estatísticas; e uma terceira – que o linguista considera a mais presente nos estudos reunidos na obra – que faz um linguística “com *corpus*”, em que se focalizam fenômenos específicos da história da língua com base nos dados que comporta determinado *corpus*

---

<sup>29</sup> Um exemplo que ilustra esse risco é o artigo de Braga e Bispo (2016, p. 3) em que se recorre às *cartas* que editamos em 2012 e que foram divulgadas na plataforma do PHPB (onde há, não obstante, a discriminação dos organizadores: o autor e seus dois orientadores de mestrado), referenciando-as apenas como dados “extraídos das amostras de 86 cartas oficiais, provenientes da plataforma de *corpora* do projeto Para a História do Português Brasileiro (PHPB), que circularam no estado do Rio Grande do Norte entre os séculos XVIII e XIX”. Uma citação que traz um nítido detrimento autoral em favor de um agente usurpador: a plataforma praticamente “se passa” (!) por autor. O escamoteamento de que falam Barbosa, Lopes e Callou (2002) se dá aqui não (apenas) em nome de um trabalho coletivo, mas (também) em nome do suporte que veicula a produção acadêmica.

(cf. KABATEK, 2016b, cap. 1, par. 4).<sup>30</sup> O exercício praticado para a confecção das *cartas oficiais* mobilizou todas as instâncias implicadas por essas três vertentes, inclusive porque, como assume o autor, “entre las tres vertentes hay, obviamente, una estrecha relación y, en tanto que un corpus no es un fin en sí, sino que se crea *para algo*, se necesita un intercambio continuo del creador del corpus con los usuarios que

---

<sup>30</sup> Alguns dos livros utilizados nesta tese só foram obtidos em sua versão eletrônica, os *e-books* (não confundir com livros em pdf), disponibilizados, após compra, em plataformas de leitura, como o Kindle e o Kobo. Não há consenso – às vezes, na realidade, sequer orientação – sobre como se devem fixar as citações dessas obras, já que, na maioria das vezes (ao menos com relação aos livros de que dispomos através desses leitores digitais), não está disponível a informação da página correspondente na edição impressa junto aos números das localizações (*location numbers*) que aparecem nos aplicativos. Na falta de uma resolução, decidimos seguir, ainda que não ortodoxamente, as orientações dadas pelas Bibliotecas Bodleian (Bodleian Libraries) da Universidade de Oxford (Oxford University), por julgarmos que elas atendem efetivamente à meta de orientar o leitor a encontrar (seja na edição impressa, seja na virtual) a citação em apreço. Essas recomendações podem ser lidas no link <https://libguides.bodleian.ox.ac.uk/e-books/citing>. Uma vez que, como salienta o próprio site, os números de localização (os *location numbers*), encontrados ao menos em todos os e-books compatíveis com Kindle ou Kobo, não fornecem uma saída estável, haja vista se alterarem conforme tipo de fonte ou tamanho (zoom) da tela de leitura, as Bodleian Libraries sugerem, em caso de não haver a informação da página (correlata à edição impressa), que se indique primeiramente o capítulo (que abreviaremos “cap.”) seguido pelo parágrafo (que transformaremos em “par.”) no qual aparece o excerto. Mesmo cômicos de que em alguns livros essa solução pode gerar caminhos bastante áridos para o autor que dela se valer (imaginemos um livro sem capítulos), é, na ausência de modelos pré-estabelecidos institucionalmente, o rumo que melhor nos convenceu. Talvez não fosse necessário esclarecer – fazemo-lo, contudo, para evitar desencontros quando de um possível confronto em busca de determinada passagem que nos chegou via e-book e também por serem os livros eletrônicos uma realidade recente (que o diga o quase desconcerto dos profissionais especializados, os bibliotecários, a que recorremos ao serem questionados sobre possíveis vias a seguir) – que citações deslocadas, por sua extensão (por conterem mais de 3 linhas, por exemplo, para evocar a ABNT), à semelhança de um parágrafo, não serão contabilizadas como tal, contando apenas os parágrafos no sentido mais consabido do termo, ainda que, por razões editoriais ou de apresentação cibernética, alguns (os primeiros de um capítulo, e.g.) não apresentem sangria. Essa contagem será feita a partir do início do capítulo (desconsiderando-se conteúdos supratextuais, como a epígrafe e as notas de rodapé, e organizacionais, como o título e os subtítulos). O resultado de todos esses passos pode-se ver em “(cf. KABATEK, 2016b, cap. 1, par. 4)”, reproduzido aqui em bis ilustrativo.

lo utilizan para un estudio concreto”.<sup>31</sup> Em nosso caso, neste primeiro momento, criador e usuário coincidiram numa mesma pessoa.

O capítulo escrito por Enrique-Arias (2016), por sua vez, abre com uma certificação – a qual, em que pesem os problemas levantados por Barbosa, Lopes e Callou (2002), é aplicável perfeitamente ao que se vê nos últimos anos, quando *corpora*, a exemplo dos criados, vêm sendo lançados na internet no Brasil – acerca dos efeitos e/ou impactos da divulgação de *corpora* na rede mundial de computadores: “La investigación en diacronía del español se ha visto beneficiada en los tiempos recientes por la disponibilidad de grandes bases de datos textuales de uso libre en la red”<sup>32</sup> (ENRIQUE-ARIAS, 2016, cap. 2, par. 1).

Entendemos que uma das grandes vantagens que introduz o acesso a esses *corpora* em linha é efetivamente a possibilidade de utilizar ferramentas de busca – sejam as mais simples, como o Ctrl L do Microsoft Word ou Ctrl F do Adobe PDF, até programas desenvolvidos para agilizar ou até determinar a eficácia no trabalho de coleta – sobre o conjunto ordenado de dados. No caso das *cartas oficiais* reeditadas, para o controle dos usos de grafemas sem transcendência fônica, valemo-nos de um *software* criado para a identificação de palavras contendo os 23 ambientes gráficos definidos para análise em Moraes de Melo (2018). Recorremos, contudo, no momento de coletar os dados a serem escrutinados, também a esses mecanismos já dados, como o Ctrl L, para contornar algumas vicissitudes que eventualmente escapassem às malhas do programa.

Dando continuidade à sua reflexão sobre as vantagens e os desafios que o *corpus* eletrônico ou disponível digitalmente carrega, Enrique-Arias (2016, cap. 2, par. 9) aponta que numa leitura linear, isto é, numa leitura feita em material impresso na sua versão física, o leitor tem de processar o texto em sua ordem de aparição ao passo que “en el texto electrónico es mucho más habitual acceder al texto a través de una concordancia generada por una máquina de búsqueda”.<sup>33</sup> De fato, ao se

---

<sup>31</sup> “Entre as três vertentes há, obviamente, uma relação estreita e, posto que um corpus não é um fim em si, mas que se cria *para algo*, é necessário um intercâmbio contínuo do criador do corpus com os usuários que o utilizam para um estudo concreto” (TN).

<sup>32</sup> “A pesquisa em diacronia do espanhol tem-se visto beneficiada nos tempos recentes pela disponibilidade de grandes bases de dados textuais de uso livre na rede” (TN).

<sup>33</sup> “no texto eletrônico, é muito mais habitual acessar o texto através de um padrão gerado por uma ferramenta de busca” (TN).

ter acesso a um *corpus* impresso apenas em versão física, sem sequer um redobro digitalizado ou disponível virtualmente, o pesquisador está atado ao dever de averiguar o material em sua inteireza, linha por linha e palavra por palavra, e ir rascunhando e agrupando os elementos que lhe sejam de interesse. Essa labuta é consideravelmente mitigada quando se pode ir direto – um direto passível de ser torto segundo os limites do dispositivo de busca – ao que se almeja examinar.

Esse contraste entre a leitura linear e a leitura de um *corpus* digital é sumária e assaz polidamente esquematizado pelo autor: “en la lectura lineal se accede a las estructuras lingüísticas en el orden *contexto* > *forma*, mientras que en el corpus informatizado se accede en el orden contrario, es decir, *forma* > *contexto*”<sup>34</sup> (ENRIQUE-ARIAS, 2016, cap. 2, par. 9). Conquanto assinala os benefícios que o *corpus* digital contém, o pesquisador não deixa de sublinhar um possível complicador. Ao se lidar com textos de sincronias passadas – em que há uma manifesta variação, com relação ao que se tornou padrão hoje, na forma das palavras, além de em várias outras camadas: estratégias retóricas, encadeamento textual etc. –, o investigador vê-se obrigado, para poder contar com as benesses do texto informatizado, a “conocer de antemano, a partir de gramáticas históricas, diccionarios o estudios previos, cuáles son las formas utilizadas para expresar la función que el investigador se propone reastrear en el corpus”<sup>35</sup> (par. 9). Daí, progride o autor, ser crucial o modo de acesso aos dados, a que se pode conferir uma parcela dos eventuais limites da Linguística de *Corpus* (cf. ENRIQUE-ARIAS, 2016, cap. 2, par. 9).

Destaca-se, nessa enseada e dimensionando a questão para nossa realidade, o papel da tabela matriz, uma das estratégias desenvolvidas por Moraes de Melo (2018), a fim de operacionalizar a análise dos usos de grafemas sem transcendência fônica nas *cartas oficiais*. Para cada palavra não abreviada identificada nas *cartas* – a qual ficava disposta, com as demais variantes gráficas que porventura surgissem, em uma linha –, registrava-se, numa coluna criada à extrema esquerda da tabela, o vocábulo

---

<sup>34</sup> “na leitura linear, acessa-se às estruturas linguísticas na ordem *contexto* > *forma*, ao passo que, no *corpus* informatizado, o acesso se dá na ordem contrária, ou seja, *forma* > *contexto*” (TN).

<sup>35</sup> “conhecer de antemão, a partir de gramáticas históricas, dicionários ou estudos prévios, quais são as formas utilizadas para expressar a função que o pesquisador se propõe a rastrear no *corpus*” (TN).

conforme sua forma ortográfica atual. Estavam as realizações, assim, lematizadas.<sup>36</sup> Foi graças a essa estaca referencial que se puderam efetuar com sucesso, por meio do *software*, a busca e a apreensão das palavras segundo os padrões pré-estabelecidos e nos moldes que idealizamos. A palavra “aceitação”, por exemplo, foi materializada em nosso *corpus* apenas como <aseitação> na C42 e <acceitação> na C74; nenhuma em forma compatível com a da ortografia hodierna. Se não houvesse essa entrada estándar, o lema, complicar-se-ia em altíssima potência o trabalho de obtenção das palavras através da rotação do *software* sobre as 23 notações (cada uma corresponde a um contexto grafemático de análise) instituídas, inviabilizando-o até, na verdade. É evidente que estamos relatando aqui um processamento que não recai de modo direto nas *cartas oficiais*, mas nessa captura que delas fizemos, a tabela matriz, a qual, sem embargo, pode ser perfeitamente compreendida como uma extensão de acesso do próprio *corpus* linguístico. É igualmente óbvio, por sua vez, que esse acesso é um acesso parcial ao *corpus*: a tabela contém todas as palavras não abreviadas que puderam ser transcritas das *cartas*, o que

---

<sup>36</sup> Vaamonde (2015) aponta a falta de lematização como uma das limitações dos grandes *corpora* do espanhol, nas figuras do *Corpus Diacrónico del Español*, o CORDE, concebido pela RAE, e do *Corpus del Español*, o CdE, criado por Mark Davis. Os outros limites são: 1º o uso de textos escritos como fonte de dados, uma “limitación forzosa y obvia de todo corpus diacrónico” [“limitação forçosa e óbvia de todo *corpus* diacrônico” (TN)] (VAAMONDE, 2015, p. 2), diante da inexistência de fontes orais dos períodos pretéritos da língua; 2º níveis de acesso ao documento, pois o autor defende que, para que possa o *corpus* atender às mais plurais expectativas, deveria ser apresentado em distintas edições: semidiplomática, crítica e fac-símiles; 3º anotação linguística detalhada, com, para além da lematização, as etiquetas morfológicas e sintáticas; e 4º gerenciamento dos fatores extratextuais que vão além do cronológico e do local de produção, como a caracterização social do autor, sua procedência dialetal e propósito comunicativo do texto (cf. p. 4). Saltando o primeiro ponto, por razões evidentes, ao encarar nosso próprio labor e premeditando a possibilidade de uma futura publicação (linear e informatizada) das *cartas*, gostaríamos de oferecer, ao lado das atuais edições semidiplomática e fac-similar (no caso desta, com a revisão de algumas fotos, logo que o IHGRN reabra suas portas para pesquisa), uma edição crítica/modernizada, que faria às vezes de um hiperlema textual, bem como de tentar uma supervisão de outros critérios extralinguísticos, como os que cita Vaamonde (2015). A princípio, por outro lado, não prevemos nenhuma investida em etiquetagens morfossintáticas. Na sequência do artigo, o autor nos traz um exemplo de um *corpus* diacrônico em que se supera grande parte dessas limitações assinaladas por ele, o Post Scriptum.

permite uma série de estudos relativos aos grafemas alfabéticos, mas, claramente, não serve para perscrutações focadas em outros aspectos grafemáticos (abreviações e pontuação, por exemplo) ou em demais estratos de organização da língua (sintático, textual, discursivo).

Outro texto compilado em Kabatek (2016a) que lança luz – pela apresentação de um competente modelo – sobre as confluências fecundantes e necessárias entre Linguística de *Corpus* e Linguística Histórica é o artigo de Carvalheiro *et al.* (2016). Os autores relatam um pouco do Post Scriptum, *corpus* linguístico desenvolvido pelo Centro de Linguística da Universidade de Lisboa que promove um acervo regulado de cartas privadas – merecendo já os devidos louros, pois são um dos gêneros mais difíceis de serem angariados, precisamente por circularem na esfera privada – escritas entre o século XVI e o início do século XIX em Portugal e na Espanha com a meta (hoje já atingida) de oferecer 2 milhões de palavras (1 milhão oriundo de cartas privadas em português e o outro milhão, em espanhol). Reportamo-nos a este capítulo pelo horizonte de expectativa que enxergamos no comprometimento e seriedade com que a equipe trouxe à luz um *corpus* médio-grande<sup>37</sup> tão esmerado em cumprir uma série de requisitos que são agudamente complicados (controle biográfico, zona dialetal, notação morfossintática) e, por isso, muitas vezes deixados de lado.<sup>38</sup> Encerram o artigo dando uma amostragem do potencial do *corpus* com uma análise do marcador discursivo “pois”, em português, e de dois mecanismos anafóricos, o “cuja” do português e os pronomes oblíquos de terceira pessoa “le/la/lo” do espanhol. Por meio dessas amostragens analíticas, os pesquisadores ilustram dois argumentos que pretendem defender:

---

<sup>37</sup> Consoante Berber Sardinha (2000, p. 346), para um *corpus* de abordagem histórica, existem cinco classificações no que toca à sua extensão: pequena (com menos de 80 mil palavras, sendo, portanto, a categoria na qual se incluem nossas *cartas oficiais*), pequeno-médio (entre 80 e 250 mil palavras), médio (entre 250 mil e 1 milhão de palavras), médio-grande (entre 1 e 10 milhões, onde entra o Post Scriptum) e grande (com 10 milhões ou mais de palavras).

<sup>38</sup> O exemplo mais próximo de Post Scriptum no Brasil seja, quiçá, o Projeto Tycho Brahe, conduzido por uma equipe da Unicamp, em que as anotações linguísticas, tanto morfológicas (em 44 textos, num total de 1.962.176 palavras, segundo a página oficial do projeto) quanto sintáticas (em 27 textos, atingindo 1.234.323 palavras) são operacionalizadas. O site desse *corpus* histórico é o <http://www.tycho.iel.unicamp.br/corpus/index.html>.

i) o de que o conceito de *desvio linguístico* é facilmente desmontável com a demonstração da antiguidade de processos que, enquanto <<desvios>>, costumam ser estigmatizados pelo público leigo; ii) o de que os atlas histórico-dialetais são um recurso linguístico que se pode realizar, hoje em dia, com um grau de incerteza cada vez mais insignificante (CARVALHEIRO *et al.*, 2016, cap. 9, par. 5).

O último estudo de que nos valem, encerrando, com ele, este artigo, é o de Nieuwenhuijsen (2016). Abre seu texto a autora com um fragmento que, à guisa de ciclo, retoma – deitando-o agora em solo hodierno – a tônica com que encetamos este item 4.3 relativa à conjunção entre as duas rotas da linguística aqui consideradas: “Hoy en día es un hecho que la lingüística histórica como disciplina académica está cambiada *profunda y definitivamente*, debido al desarrollo de la lingüística de corpus y las nuevas metodologías ofrecidas y hasta impuestas por los corpus diacrónicos digitales”<sup>39</sup> (NIEUWENHUIJSEN, 2016, cap. 11, par. 1, grifo nosso). Essa mudança acarreta uma latente perda da empatia (termo que a autora toma de uma expressão cunhada por Kabatek, “linguística empática”) com a matéria maior do *corpus*, na medida em que o pesquisador perde a obrigação de palmilhar os textos em sua plenitude; já pode transitar fácil e diretamente, sem desvios, do objetivo ao objeto. O que vemos nessa corretíssima consideração é quase uma metáfrase do debate que Enrique-Arias (2016), acima glosado, faz ao diferenciar as duas formas de acesso às estruturas linguísticas: da ordem *contexto* > *forma* à ordem *forma* > *contexto*.

A pesquisadora elenca alguns impasses que podem ameaçar a busca de palavras e, consequentemente, azar algumas inconsistências nas análises. Foi muito grato ver que nossa faina, mesmo com suas eivas, oferece condições para que se driblem esses fatores de embaraço. Tomando as mesmas medidas de análise de que se vale Vaamonde (2015), o Corpus Diacrónico del Español (CORDE) e o Corpus del Español (CdE), o primeiro óbice que Nieuwenhuijsen aponta são as buscas de elementos clíticos que, por muitas vezes virem aglutinados a palavras pesadas (nacionais, substantivas, v. g.), ou são de difícil obtenção pelas

<sup>39</sup> “Hoje em dia é um fato que a Linguística Histórica como disciplina académica está mudada profunda e definitivamente devido ao desenvolvimento da linguística de corpus e às novas metodologias oferecidas até impostas pelos *corpora* diacrónicos digitais” (TN).

ferramentas de busca oferecidas por esses *corpora* ou são tomadas por outras (NIEUWENHUIJSEN, 2016, cap. 11, par. 3).<sup>40</sup> Um segundo risco com um *corpus* digital de que fala a autora é a categorização errônea de algumas palavras (ela traz o exemplo de homônimos: uma realização de “fuera” [“fora”], advérbio de lugar, coletado como “fuera” [“fosse”], pretérito imperfeito do subjuntivo de “ser”). Como a separação e organização de cada uma das cerca de 26.000 palavras da tabela matriz medraram do trabalho manual de Moraes de Melo (2018), e não de uma equipe ou de um programa, acreditamos que erros – desse tipo, ao menos – não ocorram ou, na pior das hipóteses, sejam raros. Se para o primeiro tipo de problema, a linguista não enxerga um remédio, para o segundo ela aventa a revisão manual de todos os exemplos. Assim, propõe como solução o que fizemos como missão, o que não nos isenta, *a posteriori*, do dever da reinspeção.

Por fim, a autora demonstra como a aplicação de um teste de regressão logística binário pode precisar e esclarecer melhor o comportamento das formas linguísticas em análise. Embora o exame dos usos gráficos nas *cartas* tenha sido efetuado, em Moraes de Melo (2018), sobre os cálculos dos valores brutos e percentuais, e não por meio de um teste logístico binário, confiamos ter obtido um quadro suficientemente eficaz para pensar e acompanhar com rigor as tendências dos usos gráficos nos documentos potiguaros em tela. De mais a mais, o modo como arquitetamos a apresentação dos resultados (que chamamos de “perfis de saída” na tese) lega a transparência oportuna para que se possa, aquando das análises, transitar descomplicada e fertilmente de interpretações de base mais quantitativa às de cunho mais qualitativo. Mesmo sugerindo o teste de regressão logística, Nieuwenhuijsen (2016, cap. 11, par. 51) fecha seu artigo com um belo remate no qual abona a experiência que enarramos neste parágrafo. Com ele, destarte, fechamos o tópico:

A pesar del gran valor de los test estadísticos para la lingüística histórica, no queremos abogar aquí por la supresión de los análisis tradicionales y la sustitución completa de los análisis tradicionales

---

<sup>40</sup> Repetimos que a tabela matriz, embora desfaça esse obstáculo, se revela exígua – devido às convenções tomadas em sua criação (cf. MORAIS DE MELO, 2018, p. 148-155) – para o reconhecimento de outras formas, quais sejam a colocação do til, maiúscula e minúsculas e pontuação. Para esses casos, é necessária uma leitura linear da transcrição completa, invalidando-se, a princípio, a utilização de acessórios de busca.

por las pruebas estadísticas. El análisis cuantitativo sigue siendo imprescindible para formarse una idea global de la frecuencia y desarrollo de una forma o construcción sintáctica. Asimismo, el análisis cualitativo, es decir el detenido estudio de ejemplos específicos en su contexto, permite identificar posibles factores que hayan influido en la evolución del cambio lingüístico. El tratamiento estadístico, en cambio, constituye una herramienta complementaria muy potente, que sirve para comprobar la validez de las conclusiones sacadas en ambos tipos de análisis y para medirla posible influencia de los distintos factores identificados en el material estudiado.<sup>41</sup>

## 5 Considerações finais

O movimento renovador pelo qual a Linguística Histórica brasileira passou na década de 90 – num contínuo progressivo que transformou essa área de estudos em uma das mais produtivas atualmente dentro das ciências da linguagem no Brasil – fez da edição de *corpora* seu fito basilar, estando, desde então, sua consecução em perene fabrico. Se, por um lado, nesta seara dos estudos diacrônicos, editar textos goza de um protagonismo notório, o que se tem dito sobre reeditá-los? Este trabalho propôs-se, então, a apresentar as *cartas oficiais norte-rio-grandenses*, um conjunto de documentos oficiais organizados por primeira vez em 2012 como parte das ações locais vinculadas ao PHPB. Devido a novas orientações e a alterações nos objetivos de pesquisa, as *cartas* foram reeditadas, afã finalizado em 2018, quando veio à luz um *corpus* composto por 129 missivas consignadas entre 1713 e 1950 no estado do Rio Grande do Norte.

---

<sup>41</sup> “Apesar do grande valor dos testes estatísticos para a Linguística Histórica, não queremos advogar aqui pela supressão das análises tradicionais e pela substituição completa das análises tradicionais pelas provas estatísticas. A análise quantitativa continua sendo imprescindível para se formar uma ideia global da frequência e desenvolvimento de uma forma ou construção sintática. Da mesma forma, a análise qualitativa, isto é, o estudo detido de exemplos específicos em seu contexto, permite identificar possíveis fatores que tenham influenciado na evolução da mudança linguística. O tratamento estatístico, por outro lado, constitui uma ferramenta complementar muito potente que serve para comprovar a validade das conclusões tiradas de ambos os tipos de análise e para medir a possível influência de distintos fatores identificados no material estudado” (TN).

O processo de reedição exigiu uma série de ingerências no *corpus* de 2012. O saldo final foi positivo: afinou-se o feixe ordenado de cartas a novos critérios e necessidades surgidos ao longo do percurso doutoral e realizou-se a revisão das cartas antigas, logrando-se, inclusive, a decifração de muitas passagens ininteligíveis aos olhos que no começo desta década iniciaram a missão transcricional dos documentos. Porquanto tanto a tarefa de 2012 quanto a de 2018 nascem do cruzamento entre Linguística Histórica e Linguística de *Corpus*, resolvemos dedicar a segunda metade deste artigo para refletir sobre algumas repercussões desse encontro. Destacamos dois pontos: a disponibilização de *corpora* diacrônicos na internet facilita muito o acesso dos dados aos pesquisadores interessados, mas precisa atentar para a preservação dos direitos autorais; as novas tecnologias potencializam as abordagens e os filtros aplicados aos *corpora*, mas não invalidam estratégias menos avançadas.

Esperamos que as experiências compartilhadas neste artigo e os argumentos levantados possam dar maior visibilidade a estas facetas que certamente, ao menos em alguma medida, participam da engrenagem dos estudos históricos da língua: a reedição de *corpus* linguístico, as vantagens e riscos da liberação dos *corpora* em meio virtual e os horizontes abertos pelas novas tecnologias para aproximações cada vez mais ricas dos dados.

### **Agradecimento**

A Hozanete Lima, mais do que orientadora, uma querida amiga, por ter me incentivado a querer publicar.

### **Referências**

ACIOLI, V. L. C. *A escrita no Brasil colônia: um guia para a leitura de documentos manuscritos*. Recife: FUNDAJ; Massangana; Ed. Universitária da UFPE, 1994.

ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3.

ALMEIDA CABREJAS, B. Escuchar los textos: el análisis de los textos en el estudio de la fonética y fonología de épocas pasadas. *Linguística en la Red*, Alcalá de Henares, n. 11, p. 1-18, 2013.

BARBOSA, A. G. O português escrito no século XVIII: fontes reunidas na Biblioteca Nacional do Rio de Janeiro. In: CASTILHO, A. T. (org.). *Para a história do português brasileiro: primeiras idéias*. São Paulo: Humanitas, 1998. v. 1. p. 229-238.

BARBOSA, A. G. O contexto dos textos coloniais. In: ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3. p. 421-431.

BARBOSA, A.; LOPES, C. R. S.; CALLOU, D. Organização dos *corpora* diacrônicos do PHPB-RJ na rede mundial de computadores. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ/FAPERJ, 2002. v. 4. p. 29-37.

BAWARSHI, A. S.; REIFF, M. J. *Gênero: história, teoria, pesquisa, ensino*. São Paulo: Parábola, 2013.

BELLOTO, H. L. *Como fazer análise diplomática e análise tipológica de documentos de arquivo*. São Paulo: Arquivo do Estado e Imprensa Oficial do Estado, 2002.

BERBER SARDINHA, T. Linguística de corpus: história e problemática. *Delta*, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>.

BORGES, J. L. *Obras completas I: 1923-1949*. Buenos Aires: Emecé, 2009.

BORGES, J. L. *Obras completas*. São Paulo: Globo, 1998. v. 1.

BRAGA, A. P. A.; BISPO, E. B. Estratégias de relativização em cartas oficiais norte-rio-grandenses dos séculos XVIII e XIX. *Odisseia*, Natal, v. 1, n. 2, p. 3-16, 2016.

BRASIL. Presidência da República. *Manual de redação da Presidência da República*. Brasil: Presidência da República, 2002.

CAMBRAIA, C. N. Subsídios para uma proposta de normas de edição de textos antigos para estudos lingüísticos. In: RODRIGUES, A. C. S.; ALVES, I. M.; GOLDSTEIN, N. S. (org.). *I Seminário de Filologia e Língua Portuguesa*. São Paulo: Humanitas, 1999. p. 13-23.

CARNEIRO, Z. O. N.; ALMEIDA, N. L. F. Documentos dos séculos XVIII-XX para a constituição de um banco de dados do português. In: CASTILHO, A. T. (org.). *Para a história do português brasileiro: primeiras idéias*. v. 1. São Paulo: Humanitas, 1998. p. 197-210.

CARNEIRO, Z. O. N.; ALMEIDA, N. L. F. Informes sobre *corpus* em fase de conclusão: cartas de homens “ilustres” do século XIX (PB/Bahia). In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ/FAPERJ, 2002. v. 4. p. 61-75.

CARVALHEIRO, C. *et al.* A idade dos <<desvios>>: diacronia, variação social e lingüística de *corpus*. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book] p. 175-196. DOI: <https://doi.org/10.1515/9783110462357-009>

CASTILHO, A. T. (org.). *Para a história do português brasileiro*. São Paulo: Humanitas; FAPESP, 1998.

CASTRO, I. Para uma história do português clássico. In: CONGRESSO INTERNACIONAL SOBRE O PORTUGUÊS, 1996, Lisboa. *Actas [...]*. Lisboa: Associação Portuguesa de Linguística; Edições Colibri, 1996. p. 135-150. DOI: <https://doi.org/10.2307/3980045>

CYRINO, S. M. L.; BARRICHELO, J.; PAULA, F. F. Formação de um banco de documentos paranaenses: primeiros resultados. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ; FAPERJ, 2002. v. 4. p. 77-85.

DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do Português Brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: Faculdade de Letras da UFRJ; FAPERJ, 2002. v. 4.

ENRIQUE-ARIAS, A. Sobre la noción de perspectiva en lingüística de corpus: algunas ventajas de los corpus paralelos. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book]. p. 21-39. DOI: <https://doi.org/10.1515/9783110462357-002>

FONSECA, M. C. A. P. *Caracterização lingüística de cartas oficiais da Paraíba dos séculos XVIII e XIX*. 2003. 499f. Tese (Doutorado em Linguística) – Centro de Artes e Comunicação, Universidade Federal de Pernambuco, 2003.

FONTANELLA DE WEINBERG, M. B. La variable sexo y las grafías de los hablantes bonaerenses en los siglos XVIII y XIX. In: BLECUA, J. M.; GUTIÉRREZ, J.; SALA, L. (org.). *Estudios de grafemática en el dominio hispánico*. Bogotá: Ediciones Universidad Salamanca; Instituto Caro y Cuervo, 1998. p. 83-95.

FONTELA, O. *Poesia reunida [1969-1996]*. São Paulo: Cosac Naify; Rio de Janeiro: 7 Letras, 2006.

GONÇALVES, U. S.; FERREIRA, P. S. Aventura no reino das traças: contribuindo para uma história lingüística da Bahia. In: MATTOS E SILVA, R. V. (org.). *Para a história do português brasileiro: primeiros estudos*. São Paulo: Humanitas; FAPESP, 2001. v. 2. p. 483-504.

HORA, D.; SILVA, C. R. (org.). *Para a história do português brasileiro: abordagens e perspectivas*. João Pessoa: Ideia, 2010. v. 8.

KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016a. [E-book]. DOI: <https://doi.org/10.1515/9783110462357>

KABATEK, J. Un nuevo capítulo en la lingüística histórica iberorrománica: el trabajo crítico con los corpus. Introducción a este volumen. In: \_\_\_\_\_. (ed). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016b. [E-book]. p. 1-19. DOI: <https://doi.org/10.1515/9783110462357-001>

LOPES, F. M. (org.). *Catálogo de manuscritos avulsos da capitania do Rio Grande do Norte (1623-1823)*. Natal: EDUFRN, 2000.

MAIA, C. Lingüística Histórica e Filologia. In: LOBO, T. et al. (org.). *ROSAE: lingüística histórica, história das línguas e outras histórias*. Salvador: EDUFBA, 2012. p. 533-542.

MATTOS E SILVA, R. V. (org.). *Para a história do português brasileiro: Primeiros estudos*. São Paulo: Humanitas; FAPESP, 2001. v. 2.

MATTOS E SILVA, R. V. Reflexões e questionamentos sobre a constituição de *corpora* para o Projeto *Para história do português brasileiro*. In: DUARTE, M. E. L.; CALLOU, D. (org.). *Para a história do português brasileiro: notícias de corpora e outros estudos*. Rio de Janeiro: FAPERJ, 2002. v. 4. p. 17-28.

MATTOS E SILVA, R. V. *Caminhos da Linguística Histórica: ouvir o inaudível*. São Paulo: Parábola Editorial, 2008.

MEGALE, H.; TOLEDO NETO, S. A.; FACHIN, P. R. M. (org.). *Caminhando mato dentro: documentos do ouro do século XVIII*. São Paulo: Espaço Editorial, 2009.

MORAIS DE MELO, F. *Cartas oficiais norte-riograndenses dos séculos XVIII, XIX e XX: constituição e caracterização de um corpus diacrônico*. 2012. 329f. Dissertação (Mestrado em Estudos da Linguagem) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, 2012.

MORAIS DE MELO, F. As fórmulas textuais das “Cartas oficiais norte-rio-grandenses” (1713-1931). In: NEGRO ROMERO, M.; ÁLVAREZ, R.; MOSCOSO MATO, E. (org.). *Gallaecia*. Estudos de lingüística portuguesa e galega. Santiago de Compostela: Universidade de Santiago de Compostela; Servizo de Publicacións e Intercambio Científico, 2017. p. 465-482.

MORAIS DE MELO, F. *Nas trilhas da escrita: reedição e análise grafemática das cartas oficiais norte-rio-grandenses (1713-1950)*. 2018. 961f. Tese (Doutorado em Estudos da Linguagem) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, 2018.

MORAIS DE MELO, F.; LIMA, M. H. A. Uma microanálise de cartas oficiais norte-rio-grandenses. *Alfa: revista de Linguística da UNESP*, São José do Rio Preto, v. 60, p. 61-77, 2016a. DOI: <https://doi.org/10.1590/1981-5794-1604-3>. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/7438>. Acesso em: 12 jan. 2020.

MORAIS DE MELO, F.; LIMA, M. H. A. Four Analyses on the Official Letters of Rio Grande do Norte (Brazil). *Scriptum digital*, [S.l.], n. 5, p. 25-43, 2016b. Disponível em: [http://www.scriptumdigital.org/documents/2-Melo\\_y\\_Lima\\_Wok.pdf](http://www.scriptumdigital.org/documents/2-Melo_y_Lima_Wok.pdf). Acesso em: 12 jan. 2020.

NIEUWENHUIJSEN, D. Notas sobre la aportación del análisis estadístico a la lingüística de corpus. In: KABATEK, J. (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlin: De Gruyter, 2016. [E-book]. p. 215-237. DOI: <https://doi.org/10.1515/9783110462357-011>

RODRIGUES, A. C. S.; ALVES, I. M.; GOLDSTEIN, N. S. (org.). *I Seminário de Filologia e Língua Portuguesa*. São Paulo: Humanitas, 1999.

RUMEU, M.; BARBOSA, A.; CALLOU, D. Textos coloniais na América Portuguesa e seus problemas. In: ALKMIM, T. M. (org.). *Para a história do português brasileiro: novos estudos*. São Paulo: Humanitas, 2002. v. 3. p. 433-442.

RUMEU, M. C. B. Para uma história do português no Brasil: edição de cartas setecentistas e oitocentistas. In: LOBO, T. et al. (org.). *Para a história do português brasileiro: novos dados, novas análises*. Salvador: EDUFBA, 2006. v. 6. p. 819-844.

SALGADO, G. (org.). *Fiscais e meirinhos: a administração no Brasil Colonial*. Arquivo Nacional. Rio de Janeiro: Nova Fronteira, 1985.

SÁNCHEZ-PRIETO BORJA, P. Interpretación fonemática de las grafías medievales. In: CONGRESO INTERNACIONAL DE HISTORIA DE LA LENGUA ESPAÑOLA, VI., 2003, Madrid. *Actas [...]*. Madrid: Arco; Libros, 2006. v. 1. p. 219-260. Disponível em: <https://ebuah.uah.es/dspace/bitstream/handle/10017/7223/Interpretación%20Fonemática.pdf?sequence=1>. Acesso em: 11 abr. 2016.

SILVA NETO, S. *Ensaio de Filologia Portuguesa*. São Paulo: Companhia Editora Nacional, 1954.

SILVA NETO, S. *Textos medievais portugueses e seus problemas*. Rio de Janeiro: Casa de Rui Barbosa, 1956.

SPINA, S. *Introdução à edótica*. São Paulo: Cultrix, 1977.

TARALLO, F. *Tempos lingüísticos: itinerário histórico da língua portuguesa*. São Paulo: Ática, 1990.

VAAMONDE, G. Limitaciones en el uso de corpus diacrónicos del español. Nuevas aportaciones desde el proyecto de investigación Post Scriptum. *E-AESLA*, Madrid, n. 1, p. 1-10, 2015. Disponível em: <https://cvc.cervantes.es/lengua/eaesla/pdf/01/60.pdf>. Acesso em: 03 jun. 2017.

VASCONCELOS, C. M. V. *Lições de Filologia Portuguesa: segundo as preleções feitas aos cursos de 1911/12 e de 1912/13 seguidas das Lições Práticas de Português Arcaico*. Lisboa: Martins Fontes, [s./d.].



## **Vivendo à margem da lei: histórias de brasileiros em situação irregular no contexto europeu**

### ***Living on the fringes of the law: stories of Brazilians in an irregular situation in the European context***

Glaucia Muniz Proença Lara

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil  
gmplara@gmail.com

<https://orcid.org/0000-0003-3813-1850>

**Resumo:** O presente artigo examina e compara, à luz da Análise do Discurso Francesa (ADF), quatro narrativas de vida – coletadas por meio de entrevistas – de migrantes brasileiros que vivem em situação irregular na França ou na Inglaterra. O principal objetivo é apreender as representações discursivas de si, dos outros, do mundo que tais sujeitos constroem nas ditas narrativas e, ao mesmo tempo, verificar se e como a situação de irregularidade afeta seu cotidiano no novo país. Para analisar e cotejar essas quatro histórias, foram utilizadas categorias que integram a Semântica Global de Maingueneau (2005): temas, vocabulário, dêixis enunciativa e modo de enunciação. Se os resultados obtidos revelam diferenças na forma de contar e avaliar a experiência migratória, eles permitem também apreender aspectos comuns, tais como a motivação prioritariamente econômica para a migração, as dificuldades (culturais, linguísticas etc.) enfrentadas, sobretudo por ocasião da chegada ao país de destino, e as condições precárias de trabalho a que os migrantes brasileiros se submetem por estarem em situação irregular. Embora dois dos sujeitos entrevistados tenham preferido silenciar sobre essa questão, suas opiniões podem ser recuperadas nas entrelinhas do discurso.

**Palavras-chave:** migração; brasileiros; Europa; situação irregular; narrativas de vida.

**Abstract:** This paper examines and compares, in the light of French Discourse Analysis, four life stories – collected through interviews – that were produced by Brazilian migrants that live illegally in France or in England. The main objective is to find out the discursive representations (of themselves, of the others, of the world) constructed by such individuals in their narratives and, at the same time, verify if and how their irregular situation interferes with their routine in the new country. In order to analyze

and compare the four stories, some categories that integrate Maingueneau's Global Semantics (2005) were used: themes, vocabulary, enunciative deixis and enunciation mode. If the results reveal differences in the way of telling and evaluating the migratory experience, they also allow us to apprehend similarities, such as the primarily economic motivation to migrate, the difficulties (cultural, linguistic etc.) faced especially when arriving in the destination country and the precarious working conditions to which Brazilian migrants are submitted due to being in an irregular situation. Even though two of the interviewed subjects have preferred to be silent on this issue, their opinions can be implicitly identified.

**Keywords:** migration; Brazilians; Europe; irregular situation; life stories.

Recebido em 26 de outubro de 2020

Aceito em 23 de dezembro de 2020

## 1 Problematização

O que leva alguns brasileiros a deixar seu país natal, onde, pelo menos em tese, gozam de plenos direitos como cidadãos, para viver à margem da lei num país estrangeiro? Buscando responder a essa pergunta, selecionamos, entre 30 narrativas de vida que foram coletadas ao longo de uma pesquisa maior com migrantes<sup>1</sup> brasileiros que vivem atualmente na Europa, quatro narrativas de sujeitos que, apesar de residirem e trabalharem na França ou na Inglaterra, não têm autorização de permanência nesses países.<sup>2</sup> Suas histórias, assim como as dos demais participantes da pesquisa, foram obtidas por meio de entrevistas que eles

---

<sup>1</sup> A exemplo da posição que assumimos em trabalhos mais recentes (ver LARA, 2019), utilizaremos, neste artigo, migração (e seu correlato migrante) por se tratar de um termo relativamente neutro, que descreve simplesmente um processo de mobilidade (cf. CALABRESE; VENIARD, 2018, p. 11). Manteremos, porém, imigrante/imigração (ou emigrante/emigração) em citações de textos que empregam tais termos.

<sup>2</sup> A referida pesquisa (pós-doutorado): “Emigrantes brasileiros no contexto europeu” foi desenvolvida de agosto de 2019 a julho de 2020, no LAEL/PUC-SP, com a supervisão da Prof<sup>a</sup> Beth Brait. No período de outubro de 2019 a março de 2020 (seis meses), contamos com uma bolsa de Professor Visitante Sênior (PRINT-CAPE/UFMG) na França, com a (co)supervisão do Prof. Dominique Ducard (Université Paris-Est Créteil – UPEC). A pesquisa incluiu França, Inglaterra e Portugal. Não encontramos, porém, migrantes em situação irregular nesse último país.

nos concederam e nos autorizaram a utilizar em publicações e eventos, desde que fosse mantido o anonimato e eliminado qualquer dado que pudesse levar à sua identificação.<sup>3</sup>

Antes, porém, de apresentar e analisar as quatro narrativas de vida selecionadas, julgamos oportuno discutir questões referentes à terminologia utilizada correntemente para designar tais migrantes (aqueles que não dispõem de visto de permanência), bem como à realidade vivenciada por esses sujeitos no país de chegada, em decorrência, sobretudo, dessa condição de irregularidade. Para tanto, apoiamo-nos na ampla literatura que existe atualmente sobre essa temática em domínios como o dos Direitos Humanos e o do Multiculturalismo, fazendo-os “dialogar” com a Análise do Discurso francesa (ADF) – nossa teoria de base –, dado o caráter constitutivamente interdisciplinar que perpassa a ADF.

## 2 Em busca da definição de termos e de categorias

Falar dos fluxos migratórios contemporâneos é sempre um desafio. Além de se tratar de um tema sensível, o(a) analista se depara com uma profusão de termos para designar os sujeitos deslocados ou em deslocamento: (i)migrantes, refugiados, solicitantes de asilo, clandestinos, sem documento, apátridas... (CLOCHARD, 2007; BLANCHARD *et al.*, 2016), que, a rigor, implicam estatutos e direitos distintos, revelando uma multiplicidade de situações que impactam diretamente as condições de vida de cada uma dessas categorias no novo país.

Sem ter a pretensão de deslindar essa “trama designativa”, voltaremos nosso olhar para o que se costuma chamar de “migração clandestina” ou “migração sem documentos”. Bartram *et al.* (2014, p. 144) definem *undocumented (illegal) migration* (“migração (ilegal) sem documentos”) como “o resultado de uma entrada clandestina ou, mais comumente, de [o sujeito] ultrapassar o tempo de permanência do seu visto e/ou engajar-se em atividades (por exemplo, emprego) não autorizadas pelo seu visto”.<sup>4</sup>

---

<sup>3</sup> Esclarecemos que todos os participantes da pesquisa assinaram o Termo de Compromisso Livre e Esclarecido (TCLE), conforme orientações do CEP (Comitê de Ética em Pesquisa)/UFMG.

<sup>4</sup> Tradução livre de: “[...] it results from clandestine entry or (more commonly) from overstaying one’s visa and/or engaging in activities (e.g. employment) not authorized by one’s visa.”

Os autores ressaltam que a forma como a expressão “imigração ilegal” circula na maioria dos debates públicos (por exemplo, no meio político e na instância midiática) é inadequada, pois se refere àqueles que entram/permanecem sem autorização num país estrangeiro como criminosos, ainda que eles constituam uma importante parcela do mercado de trabalho e da economia informal. Diante disso, atribui-se ao governo a responsabilidade de fazer algo para enfrentar a “ameaça” que tais indivíduos representam: seja no que diz respeito aos danos que eles alegadamente trazem para a segurança ou para a identidade nacional, seja pelo que significam em termos de perda de controle do Estado sobre suas fronteiras, entre outros problemas.

Nesse cenário, uma das principais dificuldades que surgem relaciona-se ao uso da terminologia. Se os governantes e certos grupos contentam-se com “imigração ilegal”, parcialmente porque tal termo legitima a lei e a ordem que esses grupos representam, outros julgam inaceitável descrever as pessoas como ilegais. Um termo alternativo seria *undocumented immigration* (“imigração sem documentos”), sugerindo que alguns imigrantes apenas não dispõem de todos os documentos exigidos para uma autorização legal de permanência (BARTRAM *et al.*, 2014, p. 145).

Nessa mesma direção, mas agora no contexto francês, Blanchard *et al.* (2016, p. 44) pontuam que o que se entende por imigração clandestina é tanto a entrada ilegal de um estrangeiro num dado país, quanto o fato de as pessoas se tornarem ilegais como resultado de uma recusa em partir, mesmo diante da não renovação de um visto ou da não obtenção de uma posição favorável relativa a um pedido de asilo. Explicam, porém, que o fato de tais sujeitos serem chamados, frequentemente, de *sans papiers* não significa que eles não disponham de documentos de identidade (que podem simplesmente ter expirado), tampouco que sejam necessariamente clandestinos, pois muitos entraram legalmente no país e, portanto, são conhecidos das autoridades.

Quanto aos pesquisadores, o que se constata é que eles prefeririam usar termos neutros. No entanto, não fica claro se há termos que sejam genuinamente neutros (BARTRAM *et al.*, 2014, p. 145). De fato, a escolha de uma palavra, em detrimento de outra(s), não raramente marca uma posição política que acaba por influenciar o próprio sentido dessa palavra. Isso implica que as palavras ganham seus sentidos nos usos que delas são feitas pelos locutores, o que está imbricado no

“duplo movimento” que existe entre linguagem (discurso) e sociedade (CALABRESE; VENIARD, 2018, p. 22).

Assim, embora “irregular” ou “não autorizado” também passem por conotações políticas e normativas – como vimos a neutralidade não existe já que as palavras estão mergulhadas no social que as envolve (e vice-versa) – parece-nos que esses termos são menos fortes do que “ilegal” ou “sem documentos”, razão que nos leva a utilizá-los preferencialmente a outras denominações no presente artigo. Mas, afinal, como vivem os migrantes em situação irregular? Que consequências eles experimentam no seu cotidiano pelo fato de não estarem legalmente autorizados a permanecer num dado país? É o que abordaremos a seguir.

### 3 Migrantes em situação irregular: dificuldades e obstáculos

Com o aumento e a diversificação dos fluxos migratórios para a Europa, a partir da década de 1990, assistiu-se também a um acréscimo no número dos ditos “ilegais”, mesmo que seja extremamente difícil obter estimativas confiáveis e exatas sobre esses sujeitos, principalmente em função de suas diferentes trajetórias no que se refere ao direito (ou não) de permanência no novo país: vistos de curta duração, de estudante ou outros tipos de visto que tenham perdido a validade, não obtenção do direito de asilo, entre outras situações (HORTA, 2013; BERNARDOT, 2019).

No caso específico da França, entre as pessoas estrangeiras em situação irregular que permanecem no país, existem, inclusive, aquelas que são qualificadas pela administração como “*ni-ni*” [“nem-nem”], ou seja, “*nem regularizáveis* (segundo os critérios definidos pelo ministério do interior, que mudam de acordo com os governos), *nem expulsáveis* (segundo a regulamentação internacional e francesa)”<sup>5</sup> (BERNARDOT, 2019, p. 18; grifos do original). Nesse último caso, enquadram-se, por exemplo, indivíduos cujos filhos tenham nascido na França, os que tenham laços familiares unicamente nesse país ou ainda os que estejam doentes. Entre os dispositivos que impedem a expulsão/recondução ao

---

<sup>5</sup> Tradução livre de: “[...] *ni régularisables* (selon les critères définis par le ministère de l’intérieur, qui évoluent selon les gouvernements), *ni expulsables* (selon la réglementation internationale et française)”.

país de origem, encontra-se, por exemplo, a Convenção Europeia dos Direitos do Homem (*Convention européenne des droits de l'homme*) que, pautada em critérios éticos, protege o direito à vida em família e proíbe o Estado de adotar procedimentos de distanciamento entre familiares.

De qualquer forma, como destaca Peixoto (2013, p. 165) em relação ao contexto europeu, em geral, e ao português, em particular, os migrantes encontram-se entre os segmentos mais vulneráveis do mercado de trabalho: eles, com frequência, desempenham tarefas abaixo do seu nível de qualificação, assumem contratos temporários, com remuneração inferior à dos nativos e jornadas mais extensas, além de estarem mais sujeitos ao desemprego.

Esse quadro se agrava no caso dos migrantes em situação irregular: muitos trabalham sem ser declarados por seus empregadores e sem se beneficiarem de nenhuma proteção ou direito. Submetem-se, porém, a essas condições laborais precárias – tornando-se “presas” fáceis e atrativas para certos patrões – a fim de evitar a visibilidade sobre si mesmos. Afinal, eles não estão dispostos a reclamar os direitos que teriam formalmente, temendo que o contato com as autoridades levará à sua deportação. Mesmo assumindo esse e outros cuidados, eles convivem cotidianamente com a insegurança e com o medo da expulsão, já que sempre existe a possibilidade de denúncia ou de eles se depararem com algum tipo de fiscalização mais rigorosa (BARTRAM *et al.*, 2014; CASTRO *et al.*, 2015; BERNARDOT, 2019).

Os brasileiros, via de regra, entram legalmente na Europa, embora o façam na posse de documentos não condizentes com a decisão de lá permanecerem. Em outras palavras, eles se beneficiam do chamado “visto de turismo” – na verdade e paradoxalmente, uma ausência de visto – que permite a turistas e outras categorias (decorrentes de negócios, cobertura jornalística ou missão cultural) a estadia em certos países europeus, como é caso de Portugal (e – acrescentamos – da França), por um período de 90 dias<sup>6</sup>, sem a necessidade de documentos adicionais (autorização de residência, visto de trabalho etc.). Assim, findo o período

---

<sup>6</sup> Quanto ao Reino Unido, o prazo máximo para permanência sem visto é de 180 dias, no caso de turismo ou de negócios, segundo informação do Itamaraty. Disponível em: [http://cglondres.itamaraty.gov.br/pt-br/viagem\\_ao\\_reino\\_unido.xml#Visto](http://cglondres.itamaraty.gov.br/pt-br/viagem_ao_reino_unido.xml#Visto). Acesso em: 27 set. 2020.

de 90 dias, que prescinde de visto, eles simplesmente não retornam ao Brasil e, com isso, caem numa situação de ilegalidade (MARQUES; GÓIS, 2015, p. 110).

Cabe mencionar ainda que, em geral, os migrantes em situação irregular – e os brasileiros não fogem à regra – têm na motivação econômica a principal razão para seu deslocamento: eles buscam, em geral, melhorias na qualidade de vida. Muitos querem juntar dinheiro para, posteriormente, retornar ao Brasil e adquirir um imóvel ou abrir um negócio (CASTRO *et al.*, 2015; BERNARDOT, 2019).

#### **4 Relatos de brasileiros em situação irregular na Europa: questões teóricas e metodológicas**

A partir do breve panorama apresentado nas seções anteriores, o objetivo do presente artigo, como já foi dito, é o de analisar narrativas de vida de brasileiros que vivem em situação irregular no contexto europeu. Por meio da noção de “narrativa de vida” (BERTAUX, 2005), utilizada originalmente no âmbito da etnossociologia – e que fazemos aqui “dialogar” com a ADF –, buscamos (tanto no projeto maior quanto no presente artigo) dar a palavra a esses sujeitos para que eles próprios contem suas experiências de vida, ampliando seus espaços de fala para além da esfera privada. Isso porque, conforme constatamos em outros trabalhos (ver LARA, 2018, 2019), os debates públicos sobre as migrações contemporâneas limitam-se, na maioria das vezes, a mencionar números, gráficos e porcentagens ou a valorizar o que dizem os especialistas, os agentes governamentais e/ou os jornalistas, tomados como porta-vozes dos migrantes, sem que estes tenham a oportunidade de se manifestarem sobre o que é ser/viver como migrante num país estrangeiro, com tudo o que isso implica: dificuldades e obstáculos, mas também avanços e conquistas.

Esclarecemos que a expressão “narrativa de vida”, tradução de *récit de vie*, foi introduzida na França, em 1976, pelo sociólogo Daniel Bertaux. Para ele, há narrativa de vida sempre que um sujeito conta a outro (pesquisador ou não) um episódio qualquer de sua experiência de vida. O verbo “contar” (“fazer o relato de”) mostra-se, nesse caso, fundamental para sinalizar que a produção discursiva do sujeito assumiu a forma narrativa (BERTAUX, 2005, p. 36).

Trata-se, pois, de um gênero por meio do qual um ser do mundo fala de si para os outros, ou seja, relata a outro sujeito certos acontecimentos que protagonizou ao longo de seu percurso de vida. Esse *ser-que-se-conta* representa a si mesmo e aos outros, devendo, pois, responder a questões como: *Quem eu sou?*; *Como me represento?* Isso implica que o *eu* que escreve ou fala, na presente instância de enunciação, o *eu* do *aqui* e do *agora*, (re)cria, por meio da linguagem, um *outro*, o do *lá* e do *outrora*, dando, assim, por meio desse movimento de (re) contar-se, um melhor contorno a suas experiências de vida (MACHADO; LESSA, 2013, p. 105).

Parafraseando Arfuch (2010, p. 114), diremos que há um deslizamento da *pessoa* ao *personagem*, como se o sujeito construísse uma nova versão de si mesmo. Isso implica um embate entre realidade e ficção, pois, como afirma Charaudeau (1992, p. 712-713), contar é uma atividade posterior à existência de uma “realidade passada” (inventada ou não), o que implica, simultaneamente, o nascimento de um outro universo, o “universo contado”. Nessa perspectiva, nada garante que uma dada narrativa possa ser “o reflexo fiel de uma realidade passada”, ainda que ela tenha sido vivida pelo sujeito que (se) conta, uma vez que as memórias de alguém são sempre reconstruções e, por isso, são histórias que oscilam entre efeitos de real e efeitos de ficção. Moreira (2018, p. 140), por sua vez, admite que nesse exercício de se dar a conhecer ao outro, o sujeito empenha-se em buscar na memória lembranças e reminiscências, mas tem que lidar, inevitavelmente, com porosidades e lacunas. O resultado, portanto, é a produção de uma história complexa e heterogênea, em que vale mais o “dizer verdadeiro” do que a verdade ontológica – pelo menos no escopo da análise do discurso.

Para a coleta de dados, tomamos por base os procedimentos da entrevista narrativa (BERTAUX, 2005). Inicialmente, a partir de um roteiro prévio, registramos os relatos dos colaboradores (migrantes brasileiros na Europa), por meio do aplicativo “Gravador de voz avançado”, instalado em aparelho celular. As entrevistas tiveram a duração média de 15 a 20 minutos e foram realizadas em locais, dias e horários previamente combinados com os colaboradores nas capitais dos três países selecionados (Lisboa, Paris e Londres), no período de novembro de 2020 a fevereiro de 2021.

O referido roteiro propunha uma questão geral: “Conte-me como você vivia antes no Brasil e como vive atualmente no país de destino.”,

desdobrada em cinco perguntas mais específicas: 1) quais foram suas motivações para migrar para a Europa e, particularmente, para o país escolhido?; 2) na sua opinião, quais foram/são os pontos positivos e negativos da mudança?; 3) como você avalia o olhar do nativo em relação ao migrante, sobretudo o migrante brasileiro?; 4) como é o seu contato com nativos, brasileiros e outros estrangeiros no novo país?; 5) você tem algum projeto de retorno ao Brasil? De acordo com Bertaux (2005, p. 149), a proposição de um roteiro prévio permite que o entrevistador interfira o mínimo possível na narrativa, deixando-a fluir naturalmente, e, ao mesmo tempo, impede que o entrevistado se afaste da temática focalizada.

Além dessas perguntas, se, no momento do registro de dados (idade, escolaridade, estado civil, situação no novo país etc.), o entrevistado mencionasse não ter visto de permanência, solicitávamos que falasse um pouco dessa questão ao responder às perguntas propostas, esclarecendo, porém, que, caso ele se recusasse, seu silêncio seria respeitado. Foi assim que chegamos aos quatro relatos selecionados para este artigo.

O passo seguinte foi transcrever as entrevistas, de acordo com as normas do Laboratório ICAR da Universidade de Lyon (CALABRESE; VENIARD, 2018, p. 28). Depois dessa etapa, nós as editamos para efeitos de análise, tendo em vista que nosso interesse maior é o conteúdo dos textos.<sup>7</sup>

Ora, se a metodologia para a coleta de dados e a constituição do *corpus* seguiu, em linhas gerais, os procedimentos da entrevista narrativa (BERTAUX, 2005), a análise posterior dos textos registrados foi feita com categorias da ADF. Em linhas gerais, trabalhamos com alguns planos propostos por Maingueneau (2005, p. 79-102), no âmbito de sua Semântica Global, entendida como o sistema de restrições que incide, de forma integrada, sobre os vários planos do discurso, tanto

---

<sup>7</sup> Na transcrição, a difícil questão de como reproduzir, na modalidade escrita, a oralidade fornecida pelas entrevistas se impõe ao pesquisador. Para facilitar a leitura dos excertos apresentados neste trabalho, optamos por editá-los, introduzindo sinais de pontuação e eliminando ocorrências como pausas, hesitações e autocorreções. Mantivemos, porém, certas marcas de oralidade (*né, entendeu, tá* etc.) e as inadequações relativas ao uso do português padrão (problemas de concordância, regência, entre outros), considerando, sobretudo, a informalidade da situação.

na ordem do enunciado quanto na ordem da enunciação. É o caso do *vocabulário* (palavras-chave, índices de avaliação, nominalizações), dos *temas* (impostos ou específicos), da *dêixis enunciativa* (categorias de pessoa, tempo e espaço) e do *modo de enunciação* (o “tom” do discurso, que remete à construção do éthos). Além desses quatro planos – que constituem, para nós, os mais produtivos no exame das narrativas de vida, há a *intertextualidade*, o *estatuto do enunciador e do destinatário* e o *modo de coesão*, totalizando sete planos. Cabe esclarecer que utilizaremos os planos escolhidos de maneira mais abrangente do que faz o autor. Não vemos, porém, incompatibilidades entre o que ele propõe e a nossa “releitura” desses planos.<sup>8</sup>

Com o dispositivo teórico-metodológico brevemente descrito acima, acreditamos poder apreender, em grande medida, que representações discursivas<sup>9</sup> (de si, dos outros/nativos, dos países de partida e de chegada, entre outras) os migrantes brasileiros que vivem do outro lado do Atlântico constroem por meio das histórias que contam. E, além disso, se e como (em que medida) a situação de irregularidade interfere no seu dia a dia.

## 5 Com a palavra os migrantes brasileiros

No Quadro 1, a seguir, apresentamos informações mais detalhadas (idade, estado civil, escolaridade, profissão no Brasil e no novo país, tempo de permanência como migrante) sobre os quatro “narradores”, na ordem em que foram entrevistados. Quanto ao seu estatuto jurídico, lembramos que todos eles se encontram em situação irregular nos respectivos países (alguns, como Sofia e Magdalena, há mais de dez anos). Lembramos também que, de acordo com as orientações do CEP (Comitê de Ética em Pesquisa)/UFMG, utilizamos nomes fictícios para designá-los.

---

<sup>8</sup> Para maiores detalhes sobre a escolha desses planos (temas, vocabulário, dêixis enunciativa e modo de enunciação), remetemos o(a) leitor(a) a nossos trabalhos anteriores (LARA, 2018, 2019).

<sup>9</sup> Cabe esclarecer que, neste artigo, tomamos o termo “representações (socio)discursivas” na acepção de Charaudeau (2007, 2015). Muito resumidamente, diremos que se trata de formas de ver e julgar o mundo que se manifestam por meio do discurso.

QUADRO 1 – Informações sobre os sujeitos da pesquisa

Nome fictício	Novo país / tempo de permanência <sup>10</sup>	Idade	E. Civil / Filhos	Escolaridade	Profissão no Brasil / atual
Flávia	França / 8 meses	42	Casada (2 filhos no Brasil)	Ens. Médio / Curso Técnico em Radiologia	Recepcionista de prédio/ faxineira
Sofia	França / 11 anos	44	Solteira	Ens. Médio	Vendedora de loja/ faxineira e manicure
Magdalena	Inglaterra / 12 anos	41	Solteira	Ens. Médio	Atendente de telemarketing/ faxineira
Antônio	França / 1 ano	33	Casado (2 filhos na França)	Ens. Fundamental	Marceneiro / marceneiro

Fonte: elaboração própria.

No que tange ao estado brasileiro de onde vieram os entrevistados, cabe dizer que Flávia nasceu no Maranhão, mas viveu a maior parte de sua vida em Goiás; Antônio, natural da Bahia, morava no Espírito Santo antes de se mudar para Paris. Tanto Magdalena quanto Sofia vieram de Minas Gerais, onde nasceram.<sup>11</sup>

Dito isso, passemos à análise do *corpus* que, como comentamos, seguirá, em linhas gerais, quatro planos da Semântica Global de Maingueneau (2005): os temas, o vocabulário, a dêixis enunciativa e o modo de enunciação (éthos). Começando pela análise dos temas, esclarecemos que as narrativas de vida serão abordadas a partir de três eixos temáticos, “recortados” das perguntas do roteiro prévio: 1) motivações para a migração e para um possível retorno; 2) aspectos positivos e negativos da mudança para o novo país (incluindo a questão do estatuto jurídico irregular); 3) relações com o *outro*: nativos, brasileiros e demais estrangeiros (com destaque para o olhar do nativo sobre o

<sup>10</sup> Mínimo de 6 meses, contados a partir do dia da entrevista.

<sup>11</sup> Como já foi dito (vide nota 2), não encontramos, ao longo da pesquisa, nenhum sujeito em situação irregular que tenha migrado para Portugal, embora saibamos que os brasileiros constituem hoje a maior nacionalidade estrangeira naquele país. Porém, como constatarem os pesquisadores do projeto “Vagas Atlânticas: a Imigração Brasileira em Portugal”, transformado, posteriormente, em livro (PEIXOTO *et al.*, 2015), foram poucos os imigrantes irregulares encontrados, provavelmente em função das várias possibilidades de legalização (algumas atribuídas apenas a brasileiros) oferecidas por Portugal.

migrante, sobretudo o brasileiro). São esses três eixos que conduzirão a análise.

Para Maingueneau (2005, p. 88), os temas estão integrados semanticamente a um dado discurso por meio do sistema de restrições que o rege, podendo ser de dois tipos: a) os temas impostos, que são obrigatórios para que um discurso seja bem aceito; b) os temas específicos, que são próprios a um dado discurso. No caso deste artigo (e evidentemente da pesquisa que o originou), consideramos, em consonância com o autor, que, num discurso que se proponha a falar da experiência migratória de um sujeito, é imprescindível a presença de temas como aqueles que foram contemplados no roteiro prévio da entrevista e que “recortamos” nos três eixos mencionados no parágrafo anterior. São, portanto, a nosso ver, temas impostos. Outros temas, porém, decorrentes ou não dos temas impostos, vão surgindo em certos textos – mas não em outros – respondendo pelos temas específicos na “grande história” da migração de brasileiros para o continente europeu.

Esclarecemos também que não vemos como trabalhar com os temas (impostos ou específicos), sem analisar paralelamente o vocabulário. Se, para Maingueneau (2005, p. 83-84), “a palavra em si mesma não constitui uma unidade de análise pertinente”, não podemos deixar de observar como, em função de seus usos, as palavras se comportam (dialogam, polemizam, complementam-se, chamam umas às outras) nos discursos que integram.

No eixo temático 1 (motivações para a migração e para um possível retorno), constatamos que dos quatro entrevistados, três tiveram, principalmente, motivação econômica. Segundo Bernardot (2019, p. 34), dados coletados no âmbito dos países da OCDE (*Organisation de coopération et de développement économiques*) revelam quatro grandes motivos para a migração: 1) a migração econômica; 2) a migração familiar; 3) a migração humanitária; e 4) a migração para os estudos. A autora comenta, porém, que as motivações são, no geral, complexas e múltiplas.

Assim, é que, embora não possamos descartar o peso dos fatores econômicos na decisão de migrar, as falas dos três entrevistados, apresentadas a seguir, nos revelam outras perspectivas: como o desejo de estudar (fazer cursos, aprender uma nova língua) e de se reunir com parentes já migrados: Flávia tinha a sogra, Sofia, um irmão e Antônio, primos que já viviam (e trabalhavam) em solo francês. Foram esses sujeitos que os incentivaram ao deslocamento e/ou deram o apoio inicial

para que eles se estabelecessem num país que lhes era completamente desconhecido até então. Vejamos como cada um deles relata suas motivações:<sup>12</sup>

T1: Eu morava em Goiânia, trabalhava, era recepcionista em um prédio, trabalhei doze anos no mesmo local e eu sempre tive muita vontade de mudar, de vivenciar outras culturas e tudo e aí, eu e meu marido decidimos vir para a França porque minha sogra já mora aqui há uns 10 anos e aí ela nos convidou e nós resolvemos aceitar, com o propósito de avançar, mudar de vida, aprender outra língua, mas também mudar de vida financeiramente. Nós viemos mais em busca de trabalho [...] E assim, o que nos trouxe mais aqui mesmo é o fato de que a gente no Brasil paga aluguel e nós queríamos juntar, trabalhar para juntar um dinheiro e comprar uma casa lá porque nós não queríamos comprar financiada. Então, a gente queria comprar à vista. (Flávia)

T2: Bom, eu trabalhava no Brasil com vendas e caixa de loja, e eu quis ser autônoma e resolvi comprar uma máquina para fazer batatas, tipo *ruffles*, para vender na praia. Só que a máquina quebrou e eu perdi todo o meu investimento. Então, eu comecei a buscar uma ideia para recomeçar. E fiz um curso de depilação para fazer uma cabine de depilação porque se gasta pouco e já começa a desenvolver o negócio. Mas, quando eu fiz o curso, abriu o leque, e eu quis entrar nessa área de estética. Mas eu precisaria investir em mais cursos, e aí veio a ideia de vir pra França. Eu já tinha um irmão aqui [...] Então, ele me buscou no aeroporto e eu fiquei na casa dele durante 3 meses, depois eu fui morar com uma francesa que eu conheci através dele e aí eu comecei a trabalhar, a fazer minhas coisas. (Sofia)

T3: Bom, então, a minha motivação, né, de tá vindo pra Europa foi com relação à família, parentes, porque eu tenho parentes aqui na França já há muito tempo, eu tenho primos que moram aqui já onze, doze anos, né?, e um belo dia eles me fizeram o convite de tá vindo, né?, trabalhar aqui na região e tal, que eles tavam gostando e que seria uma oportunidade pra mim, pra minha família, e daí, então, eu fui motivado através deles, me motivei através dos meus familiares, meus parentes. (Antônio)

---

<sup>12</sup> Para facilitar os comentários e remissões, os trechos citados serão numerados: T1, T2 e assim sucessivamente.

Já as motivações de Magdalena para se mudar para a Inglaterra foram um pouco diferentes: ela conta que estava meio perdida com a morte da mãe e infeliz com o emprego de atendente de telemarketing, do qual acabou sendo demitida. Sem o apoio materno e sem perspectivas no Brasil, ela resolveu aceitar o insistente convite de uma irmã que morava em Londres. A motivação familiar parece ter sido, portanto, a principal razão que a levou ao deslocamento, embora possamos considerar que a motivação econômica, relacionada à demissão do emprego no Brasil, também tenha contribuído para a mudança, o que reafirma o comentário de Bernardot (2019) sobre o caráter complexo e múltiplo das motivações para a migração. Diz Magdalena:

T4: Antes de vir pra cá, eu tinha perdido a minha mãe, eu tava assim meio deslocada, eu não sabia bem o que fazer com a minha vida. Eu tava trabalhando nessa empresa de telemarketing, mas era uma coisa que eu não tava muito feliz [...] eu tava me sentindo assim um pouco perdida, não sabia o que eu queria da minha vida, o que eu fazia, aí eu fui demitida dessa empresa. Na verdade, eu já tava querendo sair, tava me estressando, tava me fazendo mal psicologicamente, fisicamente e a minha irmã tava aqui já tinha um tempo e ela sempre queria me trazer pra cá, me chamava pra vir pra cá, mas eu até aquele momento não tinha decidido. Eu sempre quis aprender outras línguas, eu sempre gostei do inglês, sempre tive interesse [...] Depois que eu saí desse trabalho e tudo, aí ela me chamou e falou: “Vai ser a última vez que eu te chamo.” Aí, eu não tinha dinheiro, não tinha nada. E eu falei “Tudo bem”. Pra falar a verdade, eu não tinha perspectiva nenhuma. Não tinha. Aí eu falei assim: “Não, tudo bem. Eu vou.” Aí ela pagou minha passagem e tal e eu vim. Mas sem perspectiva.

Dois aspectos chamam a atenção na fala de Magdalena em T4. Primeiramente, do ponto de vista do vocabulário, o uso de palavras e frases de cunho negativo para descrever sua vida no Brasil (*deslocada, perdida, não sabia o que fazer com a vida, não muito feliz, não sabia o que queria, estressando, fazendo mal psicologicamente e fisicamente*), além da repetição (na negativa) do termo “perspectiva” como forma de destacar a falta de convicção que a levou Londres: “eu não tinha perspectiva nenhuma. Não tinha.”, “mas sem perspectiva” (o que faz eco com a condição meio “sem rumo” que ela experimentava naquele momento, como mostram os índices de avaliação *meio deslocada e um*

*pouco perdida*, entre outros). Em segundo lugar, a presença do discurso direto, por meio do qual ela simula um diálogo com a irmã, possivelmente para frisar a insistência do convite feito por esta, que soa, inclusive, como um ultimato. O discurso direto, sendo uma espécie de teatralização de uma enunciação anterior, “autentifica” os enunciados relatados, criando um efeito de sentido de realidade, de que as coisas foram ditas exatamente daquela forma (MAINGUENEAU, 1991, p. 134).

Mobilizando os recursos linguístico-discursivos descritos, Magdalena age como se a decisão de migrar tivesse dependido mais da insistência da irmã do que de sua própria vontade (mesmo que ela admita que sempre quis aprender outras línguas e sempre gostou do inglês). Nesse ponto, Magdalena destoa de Flávia, Sofia e Antônio, que parecem ter tido objetivos mais claros na/para a mudança de país. Mesmo assim, um tema específico que emerge na fala de Antônio (e que não é mencionado diretamente pelas demais entrevistadas) é o fato de que, por mais que o sujeito esteja determinado, a decisão de migrar não deixa de ser um processo doloroso de “desenraizamento”, que envolve múltiplos fatores: familiares, culturais, geográficos, econômicos (BERNARD, 2002; MOREIRA, 2018). Diz ele: “[...] toda pessoa quando vai imigrar, quando vai tomar uma decisão dessa, porque não é uma decisão fácil, é uma decisão muito difícil de você sair da sua parentela, sair do seu país, sair da sua cultura pra você poder tentar a vida em outro lugar, né?”.

Quanto à motivação para retornar (ou não) ao Brasil, os sujeitos se dividem – e essa decisão, curiosamente, independe do tempo de permanência de cada um no novo país. Assim, no momento da entrevista, Flávia, que estava há oito meses em Paris, e Magdalena, há doze anos em Londres, declararam que têm planos de voltar ao país natal, enquanto Sofia e Antônio com, respectivamente, onze anos e um ano em Paris disseram preferir – pelo menos, quando conversaram conosco – permanecer no exterior. É importante observar, porém, que todos eles modalizam suas falas – Antônio em maior grau – no sentido de admitir que podem mudar de ideia, como se vê nos trechos de T5 a T8 (grifos nossos). Afinal, a decisão de voltar pode ser tão difícil quanto a decisão de migrar:

T5: A nossa pretensão aqui é um período de uns cinco, seis anos. A gente não quer ficar mais tempo do que isso aqui e nós queremos voltar no Brasil [...] Então, assim, eu por mim, ficaria mais tempo, mas definitivo, não; a gente não tem esse pensamento. *Pode ser*

*que algum dia mude*, porque tá tudo muito recente, então, nós ainda tamos na fase de adaptação. Mas, por enquanto, nosso intuito é o mesmo de quando a gente saiu do Brasil: de vir, mas de voltar num determinado tempo, o qual nós estipulamos. (Flávia)

T6: Ainda não voltei no Brasil, então, não tenho um parâmetro, assim, para falar se eu quero voltar, uma ideia para saber se quero voltar ou não. Mas *ano que vem quando eu for, eu posso ter essa ideia* [de ficar no Brasil] [...]. Mas até então, a ideia é ficar aqui. Eu tô feliz aqui e é aqui que eu quero ficar. (Sofia)

T7: Eu não pretendo ficar morando aqui a vida inteira, *a não ser que eu mude*, tenho muita saudade do Brasil, muita vontade de voltar e a minha vontade é conseguir o objetivo, né? Meu foco é esse, é voltar pro Brasil. No momento, eu tô pensando em voltar. Já tem doze anos que eu não vou e tô pensando em voltar e ver como eu vou me adaptar. (Magdalena)

T8: Eu cheguei com o pensamento de voltar pro Brasil [...] você vai com o pensamento: “Eu vou e eu volto”, né? Mas, ao chegar aqui, esse pensamento ele vem mudando. Então, hoje *eu ainda não sei exatamente dizer assim*: “Ah, eu volto pro Brasil.” Hoje a ideia tá amadurecendo de permanecer aqui na França, né? *Não sei*, vamos vivendo e *vamos ver o que vai acontecer*. A intenção atual é de ficar, *vamos ver como vai ser mais pra frente*. (Antônio)

Ao que tudo indica, a decisão de permanecer na Europa ou de retornar ao Brasil está ligada à avaliação que o sujeito faz da experiência migratória, o que o leva a descrever o país de chegada em seus aspectos positivos e negativos e, não raro, a compará-lo com o Brasil. Isso nos conduz ao segundo eixo temático. Nesse eixo, interessa-nos saber ainda se – e em que medida – o estatuto jurídico irregular interfere no cotidiano dos quatro entrevistados.

Flávia, que, à época da entrevista, estava ainda em processo de adaptação, lembra a expectativa que ela e o marido tinham sobre a França e a realidade que encontraram (descrita pelo índice de avaliação: *totalmente diferente* em termos de cultura e de língua). O adjetivo *difícil*, intensificado pelo advérbio *muito*, é repetido várias vezes ao longo da sua fala para sinalizar o momento de chegada e os primeiros oito meses em Paris. Constitui, assim, uma palavra-chave ou, como diria Maingueneau (2005, p. 84), um “ponto de cristalização semântica” no/do discurso. Isso

significa que o lexema *difficil* assume, de forma privilegiada no discurso de Flávia (e de tantos outros migrantes na mesma situação), toda a carga negativa que envolve(u) cada etapa, cada aspecto de (sobre)vivência num outro país. Destacamos, nesse sentido, as condições precárias de trabalho a que Flávia se submete por ser migrante e, mais do que isso, por estar numa situação irregular.

T9: Nós viemos mais em busca de trabalho e nós tínhamos uma expectativa. Quando nós chegamos aqui, nós vimos que é uma cultura totalmente diferente, um povo totalmente diferente, nos sentimos meio que deslocados devido à língua ser *très difficile, muito difficil*. E, então, assim, a princípio [foi] *muito difficil*. Muito mesmo. Teve dias de querer voltar, ir embora, deixar tudo, independente do que já tinha gasto, da motivação que nos trouxe aqui [...] A vida aqui é *muito difficil*. Eu acordo 3h45 da manhã todos os dias, saio para trabalhar 4 horas da manhã. A gente trabalha muito, não tem um lugar fixo, você trabalha hoje aqui, amanhã você trabalha em outro lugar e assim vai sucessivamente. Então, assim, é *muito difficil*. (grifos nossos)

Apesar disso, ela reconhece o “lado bom” de estar na França. Na sua opinião, Paris é uma “cidade belíssima”, com uma cultura e uma gastronomia ricas. Ressalta, além disso, a oportunidade de estar estudando francês, o que vê como um investimento para o futuro.

Quanto à situação irregular de migrante, já sinalizada mais acima, Flávia é uma das que mais fala sobre a questão, reconhecendo os obstáculos que isso implica e descrevendo o medo que já sentiu em certas ocasiões, bem como a tristeza que decorre dessa sua condição, como mostra o trecho reproduzido a seguir.

T10: Não estou legal aqui. O que dificulta mais ainda a situação é não estar legal num país que não é o seu e você fica ilegal, você tem medo. Às vezes, você tá num no trem ou num metrô e entra a fiscalização. Às vezes é a fiscalização do Navigo, do passe do metrô, você já fica meio assim, achando que é a imigração. Então assim é muito complicado para arrumar trabalho, devido a você não estar legal no país, isso dificulta mais ainda e às vezes entristece mais ainda a gente [...] Quando você chega aqui você vê que é totalmente diferente e tem momentos que o seu mundo assim desaba. Será que eu fiz a coisa certa? Será que aqui é o meu lugar? Será que não tá na hora, será que eu não tenho que voltar?

Chama a atenção, no final de T10, o conjunto de perguntas retóricas que Flávia – e muitos migrantes – se faz(em), revelando dúvida, incerteza no tocante à decisão de migrar, o que remete, de certa forma, ao difícil processo de “desenraizamento” presente na fala de Antônio mais acima. Afinal, como vimos, por mais determinado que esteja, o sujeito, ao partir, defronta-se com inevitáveis rupturas relativas à língua, à cultura, às relações afetivas (familiares, amigos) que mantinha no seu país natal, o que pode afetá-lo em maior ou menor grau, levando-o não raras vezes a questionar a decisão tomada. Migrar não é, portanto – e nunca será – uma questão simples na vida de quem o faz. Como afirma Laacher (2012, p. 46), partir tem um custo para o sujeito: é um ato que jamais ocorrerá sem prejuízo simbólico e social. Para ele, apenas as gerações subsequentes serão capazes de transformar em lembranças as perdas que a experiência migratória representa.

Sofia, por sua vez, já há onze anos em Paris, também descreve sua experiência migratória inicial como *difícil*, em função, sobretudo, da língua francesa, que ela não dominava, e da falta de contatos com outras pessoas no novo país, problemas que o tempo ajudou a superar.

T11: A dificuldade que a gente tem é a língua porque eu, particularmente, cheguei sem falar nem um “oi”, mas hoje eu consigo desenvolver minha vida. [...] Olha, graças a Deus, eu não posso reclamar. Eu nunca fiquei sem trabalho [...] Então, pelo fato de fazer unha, você tem a possibilidade de conhecer muitas pessoas e perguntar, né?, ser um contato, ter um contato e ser um contato. Então, graças a Deus, nunca me faltou trabalho.

A exemplo de Flávia, Sofia se pergunta se, ao migrar, teria feito a coisa certa. Sua dúvida, porém, se refere a um ponto específico (e não à migração como um todo, que é o caso de Flávia). Vejamos:

T12: Olha, eu gostaria de estudar. Meu sonho é fazer Psicologia. Então, às vezes, por eu ter uma vida muito corrida, muito agitada, porque eu moro sozinha, então, acredito que é mais difícil, eu posso pensar que, se eu tivesse ficado no Brasil, será se eu teria feito a Psicologia? Não sei. É o único questionamento que eu tenho às vezes é esse. Aqui você tem a facilidade do preço, mas você não tem o tempo; é difícil você ter um tempo pra você estudar.

Quando se trata de abordar sua situação irregular na França, Sofia se cala. Podemos depreender, porém, nas entrelinhas do discurso, que isso afetou – afeta ainda – negativamente seu cotidiano. Diz ela: “Então, quando você chega, você faz o que vier, porque primeiro você não fala bem o francês ou não fala nada, então, você tem que fazer o que tiver pra fazer.”. Apesar de Sofia atribuir a necessidade de “fazer o que vier” apenas ao não domínio da língua estrangeira (no caso, o francês), não podemos deixar de associar essa impossibilidade de escolha também ao seu estatuto jurídico, o que a impede, por exemplo, de ter um contrato formal de trabalho, com horário determinado e salário fixo.

Hoje, falando francês e com um maior número de contatos, ela admite que “consegue, às vezes, até escolher um trabalho”, o que minimiza, mas não resolve o problema vivenciado pelo migrante, sobretudo aquele em situação irregular, de ter que se sujeitar ao que aparece em termos de serviços e de horários, mesmo que ela admita que nunca lhe faltou trabalho. Isso torna sua rotina “muito corrida, muito agitada e complicada” (índices de avaliação), o que a leva a concluir: “eu não tenho uma rotina, na verdade, né? [...] porque meus horários são aleatórios”. Essas considerações de Sofia acabam por corroborar nossa interpretação ainda que, como foi dito, ela não as relacione explicitamente a seu estatuto não autorizado de permanência na França.

Magdalena também relembra os problemas que teve quando migrou para Londres doze anos atrás – o que nos leva a reafirmar o vocábulo *difícil* (e variantes) como um ponto de cristalização semântica no discurso do migrante, em geral, sobretudo no período de adaptação ao novo país. Nessa perspectiva, chamam a atenção, em T13, verbos como *desistir* e *ir embora*, que assim como aqueles empregados por Flávia em T9: *querer voltar*, *deixar tudo*, revelam o estado de fragilidade do migrante num país em que, com frequência, ele não se sente bem acolhido:

T13: E o primeiro ano aqui foi *muito difícil*, eu quase desisti. Problema de trabalhar, de dinheiro, de saber o que fazer porque quando a gente chega aqui a gente não imagina trabalhar como limpar casa. Eu nunca, no Brasil, assim, eu fazia isso na casa da [minha] vó, na casa da família, mas nunca gostei muito desse tipo de trabalho. Assim, o primeiro ano foi *muito difícil* [se emociona], eu quase desisti [chora]. Se não fosse por ajuda, eu teria desistido, eu teria ido embora. (grifos nossos).

Como vemos em T13, de atendente de telemarketing no Brasil Magdalena se tornou faxineira em Londres, o que remete à afirmação de Peixoto (2013) de que os migrantes encontram-se entre os segmentos mais vulneráveis do mercado de trabalho, desempenhando, com frequência, tarefas abaixo do seu nível de qualificação. Essa condição pode ser atribuída também a Flávia e a Sofia, como mostra o quadro 1, e parece ser agravada pelo estatuto jurídico (irregular) das três nos respectivos países. Nesse sentido, é importante destacar que apenas Antônio continua exercendo a mesma profissão que tinha no Brasil – a de marceneiro. Ele não especifica, porém, qual é a sua situação na empresa onde trabalha em Paris: por exemplo, se tem algum contrato ou acordo, mesmo que informal; se isso lhe traz algum tipo de benefício etc., o que nos permite pensar que ele talvez se submeta a condições tão precárias quanto as das outras entrevistadas.

De qualquer forma, a exemplo de Flávia, Magdalena não se furta a abordar, mesmo que brevemente, sua situação irregular como migrante. Ela, porém, não menciona sentimentos de medo ou de tristeza como Flávia, mas fala da preocupação de ter algum problema mais sério, como ficar doente, e de um certo incômodo de não poder transitar livremente, o que se comprova no T14, a seguir:

T14: Acho o ponto negativo que eu sinto de tá aqui é por eu não tá legal aqui, eu não poder transitar, não poder viajar. Isso é uma coisa difícil porque eu gosto de viajar e tudo. E me preocupa muito, por exemplo, se eu ficar doente, alguma coisa acontecer comigo e eu não ter a segurança de poder ficar um tempo sem poder trabalhar.

Outro aspecto negativo apontado por Magdalena, agora em relação à Inglaterra – e a Londres, particularmente – é o tempo frio. Ela confessa sentir falta do sol do Brasil, mas, em contrapartida, destaca a segurança que tem em Londres, a liberdade de poder sair sozinha à noite, sem sofrer nenhum tipo de violência (diferentemente do que acontece no Brasil), o que é um aspecto bastante positivo da Europa, em geral (destacado, aliás, pela maioria dos migrantes brasileiros que entrevistamos, embora não tenha sido mencionado nem por Flávia, nem por Sofia). Segundo Magdalena:

T15: Mas aí tem a parte positiva, que é a parte de se sentir *segura*. Igual eu tava comentando hoje com o pessoal, eu posso sair à noite sozinha e me sentir *segura*, que é uma coisa que a gente não sente no Brasil. Igual antigamente assim aqui eu saía, saía à noite, saía pra divertir e tudo e voltava pra casa sozinha de madrugada e me sentia *segura*. Eu sempre me senti *segura* aqui nunca tive problema, nunca tive problema nem de violência, de nada. Eu sei que existe, mas não é igual a gente sente no Brasil. Essa *liberdade*, eu gosto dessa *liberdade*. (grifos nossos)

Apesar da segurança e da liberdade que diz sentir em Londres – e que reafirma, em T15, pela repetição –, Magdalena assume que lá não tem qualidade de vida: faltam-lhe ar puro e contato com a natureza, porque, segundo ela, “aqui a vida da gente é trabalhar”. Some-se a isso o fato de ela estar impedida de viajar, de circular livremente, devido à sua condição não autorizada de permanência no país. É, principalmente, por essa razão que ela pensa em voltar definitivamente para o Brasil e viver numa cidade pequena do interior, onde possa ter o que lhe falta em Londres, ou seja, ar puro e contato com a natureza. Caso não se (re) adapte ao Brasil, ela pensa em se mudar para outro país.

Já Antônio, a exemplo de Sofia, silencia sobre seu estatuto jurídico ao longo da entrevista. Tece, porém, um bom número de comparações entre a França e o Brasil. Destaca como ponto positivo da/na França, a condição de vida mais favorável pela facilidade de aquisição de bens e serviços (tema específico). Na sua opinião, consegue-se “viver melhor [na França] porque você consegue adquirir, com mais facilidade. [...] No Brasil, pra você conseguir adquirir um bem, um automóvel, uma mobília é muito mais complicado.”.

Assim como Magdalena no tocante à Inglaterra, aborda a questão da segurança que a França oferece, como comprova o trecho a seguir (T16). Nele, chama a atenção, do ponto de vista do vocabulário, as muitas palavras/expressões que Antônio utiliza para se referir a situações de violência que acontecem frequentemente no Brasil, mas são bastante raras na França:

T16: Você não tem a violência, né?, como naturalmente acontece principalmente no Brasil. Então, você pode sair sem se preocupar com nada; você pode sair com o celular, você pode sair com dinheiro [...]. Acontece, mas você raramente vai ouvir dizer que aconteceu, né?, um *latrocínio*, aconteceu um *roubo seguido de*

*morte* porque isso aqui é minoria, entendeu? É muito raro. Então, não existe esse negócio de *assalto à mão armada*, não existe esse negócio da pessoa te *forçar a entregar alguma coisa*. (grifos nossos).

Os pontos negativos da vida na França – e em Paris, especificamente – são, sobretudo, o clima, o fuso-horário (que, segundo Antônio, deixa o sujeito um pouco desorientado) e a distância dos familiares e amigos que ficaram no Brasil, o que gera a saudade. Porém, na sua opinião, os benefícios citados – com destaque para a segurança – compensam as dificuldades, principalmente as que teve/está tendo inicialmente, como o não domínio da língua francesa e o dia a dia num “país totalmente diferente”, com uma “cultura totalmente diferente”, avaliação que o aproxima de Flávia. Lembremos que, assim como ela, Antônio está há pouco tempo em Paris (um ano à época da entrevista) e, portanto, em pleno processo de adaptação. Ao contrário da maranhense, porém, Antônio, como vimos, considera a possibilidade de permanecer na França, embora em nenhum momento da entrevista ele mencione o que fará para resolver sua atual situação de irregularidade. Vejamos:

T17: É uma decisão muito difícil de você sair da sua parentela, sair do seu país, sair da sua cultura pra poder tentar a vida em outro lugar, né?, em um outro país totalmente diferente, uma cultura totalmente diferente, com a outra língua que você não conhece, não é fácil, é difícil. Apesar das dificuldades, né?, nós temos vários benefícios [...]. Então, por esse motivo, vale a pena todo, digamos assim entre aspas, o sofrimento [...]. Você pode viver melhor, você não tem a violência [...] Então, essa é a diferença da segurança que você tem de você poder chegar num parque com seus filhos e você deixar seus filhos no parque e poder ficar tranquilo com a sua família, sem se preocupar com nada. Então, assim, a segurança aqui ela não tem comparação. Segurança, saúde, não tem comparação com a que, infelizmente, nós temos no Brasil.

Avaliar a experiência migratória implica também considerar a relação com os nativos de um dado país e mesmo com brasileiros e outros estrangeiros que nele residem. Isso nos leva ao terceiro e último eixo temático, em que inevitavelmente surge a questão do preconceito e da

discriminação<sup>13</sup> contra os migrantes, em geral, e os migrantes brasileiros, em particular. A esse respeito, todos os entrevistados afirmaram nunca ter sido alvo direto de situações discriminatórias ou preconceituosas. Vejamos, mais detalhadamente, como cada um se manifesta sobre suas relações com os outros (nativos, brasileiros e estrangeiros).

Flávia descreve os franceses como “reservados” e “frios”, quando comparados aos brasileiros, por contraste, um povo “quente”, “que abraça” (índices de avaliação). Comenta ainda que, em Paris, é “cada um por si” e que mesmo os brasileiros que lá vivem adotam esse comportamento de desunião. Ela nada diz sobre suas amizades, talvez porque estivesse há apenas oito meses em Paris quando foi entrevistada.

Já Sofia afirma que, depois de onze anos de migração, fez muitos contatos, devido, principalmente, à profissão de manicure. Embora tivesse mencionado inicialmente que também faz faxinas, ela não retoma essa informação ao longo da entrevista, talvez por considerá-la uma profissão “menor” em relação à de manicure. Afirma que nunca sofreu nenhum tipo de preconceito ou discriminação, mas conta que os franceses, às vezes, riem dos brasileiros: “Só assim algumas palavras quando a gente, sei lá, fala alguma coisa que é muito engraçada. Eu acho que às vezes nem é por maldade, mas porque é engraçado mesmo”. Ressalta que isso também “pode acontecer [...] em relação à língua portuguesa, quando um francês fala”. Mesmo que ela admita que não tem nada a reclamar quanto a essa questão, fazer chacota ou zombar do modo como o estrangeiro se expressa em língua francesa (ou em outra língua qualquer) não deixa de ser uma forma sutil – porque velada – de preconceito, como atesta Bueno (2020) em relação aos migrantes japoneses que vivem no Brasil.

Por seu turno, Magdalena relata que, quando chegou a Londres, trabalhava de 2<sup>a.</sup> a 6<sup>a.</sup> feira e ia “pra farra” no final de semana, preferencialmente “pra lugar brasileiro”. Conheceu, assim, muita gente (muitos brasileiros), mas revela que amigos mesmo tem poucos,

---

<sup>13</sup> Carvalho (2020, p. 39-40) esclarece que “o preconceito tem a ver com disposições interiores, estando no plano do pensamento e da percepção, enquanto a discriminação seria a concretização dessas disposições interiores; seriam comportamentos, ações concretas”. A autora dá um exemplo bastante esclarecedor: “João, dono de uma padaria [...] considera que os negros são baderneiros, desorganizados e que os amarelos são muito pouco sociáveis. Entretanto, na hora de contratar para prestação de serviços em seu estabelecimento, João aceita os amarelos, mas recusa os negros. Logo, pode-se afirmar que João tem preconceito contra negros e amarelos, mas discrimina apenas os negros.”.

sobretudo na atualidade, quando, tendo abandonado a “farra” dos primeiros tempos, sai mais frequentemente para programas culturais, como ir ao cinema. Mantém amizade com uma família de brasileiros há doze anos e comenta que teve um amigo inglês, com quem, no entanto, perdeu contato. Por força de seu trabalho como faxineira, conhece(u) muitos estrangeiros e alguns ingleses, mas mantém (ou manteve) com eles uma relação estritamente profissional. Avalia os ingleses como pessoas “mais abertas” talvez por serem “mais estudadas”. Admite, porém, que eles têm preconceito “quando você não quer se integrar à cultura deles”. No seu caso, Magdalena assume que o fato de já ter algum conhecimento de inglês e de ter adquirido fluência na língua muito rapidamente (em seis meses) contribuiu para que ela se adaptasse mais facilmente ao novo país. Abrimos aqui um parêntese para falar, mesmo que de forma breve, sobre questões que envolvem a integração do migrante e o papel da língua nesse processo, já que tal tema (específico) é abordado por Magdalena.

Por integração, Bartram *et al.* (2014, p. 83) entendem “o processo por meio do qual imigrantes ganham filiação social e desenvolvem a capacidade de participar de instituições-chave no país de destino.”<sup>14</sup> Isso implica uma série de indicadores, como, por exemplo, na esfera econômica, integrar o mercado de trabalho em igualdade de condições com os nativos e, na esfera política, participar (votar, fazer campanha etc.) em padrões similares ou equivalentes aos dos nativos. Os autores destacam, porém, que a competência linguística é a capacidade-chave para a integração. Não é diferente a posição de Bernardot (2019, p. 99), para quem o domínio da língua (falada e escrita) num país facilita enormemente a inserção profissional, os laços sociais e de amizade com pessoas da sociedade de acolhida e mesmo a mobilidade (por exemplo, a orientação no transporte público) e a escolaridade dos filhos (por exemplo, a comunicação com a escola). Trata-se, pois, de um fator primordial para a integração.

Nessa perspectiva, é possível constatar que os quatro entrevistados se dão conta, de forma mais (ou menos) explícita, de que dominar a língua (e, evidentemente, a cultura) do país de destino implica não apenas ser aceito ou, pelo menos, não ser excluído pelo outro (nativo) – como disse

---

<sup>14</sup> Tradução livre de: “The process by which immigrants gain social membership and develop the ability to participate in key institutions in the destination country.”

Magdalena acima (ou como dirá Antônio mais adiante) –, mas também lidar com questões práticas do dia a dia. É o caso de Sofia, que chegou sem falar “nem um oi”, mas hoje consegue “desenvolver” sua vida porque aprendeu a se comunicar razoavelmente em francês.

Quanto a Antônio, o fato de a empresa onde trabalha ter, em sua maioria, funcionários portugueses (além de brasileiros) e prestar serviço para franceses, leva-o à seguinte comparação (tema específico): “os portugueses, eles são bem racistas [...] grande parte deles [...]. Agora com relação a francês, eu acho que, o pouco que eu convivo, o pouco que eu, né?, já trabalhei e encontrei, eu acho que é pouco, é bem menos que os portugueses”.

Mesmo se, na ótica de Antônio, os franceses são menos preconceituosos quando comparados aos portugueses, eles – na verdade, alguns deles, como Antônio faz questão de frisar – tratam, às vezes, a pessoa como se não fosse “digna [...] talvez por ser imigrante, por ser de fora, não sei”. O que mais chama a atenção, porém, ao longo de sua fala, é que ele, de certa forma, justifica o preconceito/a discriminação, seja quando o migrante não domina a língua do país de destino (no caso, o francês), seja quando ocupa postos no mercado de trabalho. Em suas palavras:

T18: Muitas das vezes, mesmo pra quem trabalha, no meu caso, em obra, esse negócio todo, é a questão de você estar se infiltrando no mercado de trabalho que, querendo ou não, não é só pra imigrante [...] diminui, né?, um pouco a mão de obra local. Então, eles, talvez alguns, não todos, mas alguns vê o imigrante como o que veio pra talvez diminuir a mão de obra local, entendeu? Às vezes eles acham, às vezes, tipo assim, por a gente tá imigrando, a gente trabalha com a mão de obra mais barata do que a mão de obra deles, então, a gente acaba, às vezes, tomando lugar [deles] [...] É como eu disse, eu acho que o francês é um pouco racista, né? [...] porque muitas das vezes nós podemos estipular que 99% dos imigrantes não falam a língua. Então, nem todo francês tem aquela paciência de entender que você está aqui sem falar a língua. E muitos, tipo assim, eu já ouvi isso várias vezes em obra, na qual a gente trabalha: “Poxa, como é que você está na França e não fala francês? Você tá na França tem que falar francês”. Então, isso aí, querendo ou não, eles estão dentro da casa deles, digamos assim,

né? e nós estamos na casa deles. Então, é uma questão que eu, particularmente, não tiro a total razão, entendeu?<sup>15</sup>

No que tange ao conteúdo de T18, destacamos dois aspectos: 1) no âmbito do vocabulário, a utilização de termos como *infiltrar-se*, *tomar o lugar (de)*, *estar dentro da casa deles (dos franceses)*; 2) o uso do discurso direto para reproduzir a fala de certos franceses (criando-se, com isso, um efeito de sentido de autenticidade): “Poxa, como é que você está na França e não fala francês? Você tá na França tem que falar francês”, em que se destaca a obrigatoriedade (expressa na modalidade deôntica *ter que*). É como se o migrante não tivesse o direito de estar ali, seja porque não domina a língua (e, portanto, deduz-se, não faz nenhum esforço para se integrar), seja porque ocupa um lugar que não é legitimamente seu.

No entanto, além de a integração ser uma questão complexa que, como vimos, envolve múltiplos fatores – não é por outra razão que muitos pesquisadores têm-se debruçado sobre as práticas e as políticas públicas, visando à efetiva integração de migrantes, como é o caso de Horta (2013) e Bernardot (2019) –, é preciso considerar que eles (os migrantes) constituem uma parcela importante da força de trabalho, assumindo, em geral, serviços pouco qualificados em setores mais carentes, ou seja, eles atuam mais em caráter complementar do que substitutivo à mão de obra nativa (PEIXOTO, 2013; BLANCHARD *et al.*, 2016).

Antônio admite, mais de uma vez ao longo de sua entrevista, que nunca foi alvo de preconceito ou de discriminação na França. Perguntamos, porém: será que, ao refletir longamente sobre o tema

---

<sup>15</sup> Um aspecto importante que pode observado na fala de Antônio e que “salta aos olhos” particularmente nesse trecho (T18) é que ele usa, com frequência (muito mais do Flávia, Sofia ou Magdalena), elementos como *né?*, *entendeu?*, *digamos assim*, por meio dos quais ele busca estabelecer uma espécie de convivência com a entrevistadora para as opiniões que manifesta. Segundo Authier-Revuz (1990), uma glosa como “digamos” funciona como uma injunção que instaura explicitamente uma enunciação conjunta, com tonalidade de desculpa, para que aqueles que “codizem” se contentem com um termo não muito satisfatório. Isso vai de par com os “torneios” que Antônio adota na hora de caracterizar os franceses como preconceituosos: “eles, talvez alguns, não todos, mas alguns”; “às vezes eles acham”, “um pouco racistas”. É como se, por estar “na casa deles”, o migrante perdesse o direito de qualificá-los negativamente, no estilo “não cuspir no prato [em] que comeu”.

(específico) da presença dos migrantes no mercado de trabalho francês, ou fazer alusão a comentários preconceituosos ouvidos na obra (e que parecem ter sido dirigidos a outrem),<sup>16</sup> Antônio não estaria, na verdade, falando de si, de sua experiência migratória, do fato de viver em situação irregular? Deixamos a questão em aberto para as considerações do(a) leitor(a) e passamos aos dois últimos planos da Semântica Global (MAINGUENEAU, 2005) que elegemos para o nosso dispositivo de análise: a dêixis enunciativa e o modo de enunciação.

Em linguística, o termo *dêixis* refere-se a palavras como “eu”, “aqui”, “agora” (ou seja, marcadores de pessoa, espaço e tempo) que só ganham sentido na situação de comunicação em que ocorrem. De forma mais ampla, cada discurso, em função do seu sistema de restrições semânticas, constrói uma determinada dêixis enunciativa espaciotemporal (e – acrescentamos – pessoal) para legitimar e autorizar sua enunciação (MAINGUENEAU, 2005, p. 93). Ora, conforme constatamos em outros trabalhos (LARA, 2018, 2019), as narrativas migratórias, em geral, constroem-se em dois grandes “momentos” espaciais/temporais: um *aqui-agora* no país estrangeiro e um *lá-então* no país natal, que, não raro, são confrontados quando se trata de justificar a migração e/ou a perspectiva (ou não) de um retorno. Como vimos, sobretudo no eixo temático 1, os relatos de nossos quatro entrevistados seguem essa “rotina”.

Já no que se refere à categoria de pessoa, falar em narrativa de vida implica, de imediato, considerar a predominância de um “eu” que (se) conta ao outro, como vimos na própria definição desse gênero de discurso. Isso pode ser facilmente constatado nos relatos de Flávia, Sofia, Magdalena e Antônio por meio dos excertos reproduzidos até aqui. Há, porém, outras semelhanças que julgamos oportuno destacar. Uma delas é o uso de **você** (genérico), ou seja, qualquer pessoa na mesma situação; outra é a utilização de **nós/a gente** (nós misto),<sup>17</sup> referindo-se seja aos brasileiros, seja aos migrantes em geral (nós, brasileiros / nós,

---

<sup>16</sup> Ao afirmar: “eu já ouvi isso várias vezes em obra, na qual a gente trabalha”, Antônio não deixa claro se o comentário estaria sendo dirigido a ele. Possivelmente, mas a forma como o enunciado foi construído não garante essa interpretação.

<sup>17</sup> Segundo Fiorin (2003, p. 165), a 1ª pessoa do plural pode assumir os seguintes significados: nós inclusivo (eu + você(s)); nós exclusivo (eu + ele(s)) ou nós misto (eu + você(s) + ele(s)).

migrantes) e contrapondo-se ao **eles** – nativos. Seguem exemplos (os grifos são nossos):

- a) Então, quando *você* chega, *você* faz o que vier, porque primeiro *você* não fala bem o francês ou não fala nada, então, *você* tem que fazer o que tiver pra fazer, e contato, *você* tem poucos contatos. (Sofia);
- b) E é assim que *você* tem que fazer, senão *você* tem que ir embora e pronto. Ou *você* desiste e vai embora ou *você* persiste e continua. (Magdalena);
- c) *A gente* trabalha com a mão de obra mais barata do que a mão de obra *deles*, então *a gente* acaba, às vezes, tomando o lugar [deles]. (Antônio).

Uma peculiaridade de Flávia, que distingue sua narrativa das demais, é que ela, frequentemente, faz uso do nós/a gente exclusivo (eu + meu marido), possivelmente para destacar que eles compartilham decisões, dificuldades e conquistas tanto na opção de migrar quanto na vivência do cotidiano parisiense. Esse emprego do nós exclusivo, que pode ser constatado nos dois exemplos que seguem, não aparece nas demais entrevistas, nem mesmo na de Antônio, que também é casado e está com a família em Paris:

- d) E assim, o que *nos* trouxe mais aqui mesmo é o fato de que *a gente* no Brasil paga aluguel e *nós* queríamos juntar, trabalhar para juntar um dinheiro e comprar uma casa lá porque *nós* não queríamos comprar financiada. Então, *a gente* queria comprar à vista. Então, foi isso que *nos* trouxe aqui. (grifos nossos)
- e) *Nós* tínhamos uma marcenaria de móveis planejados. Então, assim, *a gente* não vivia tão mal. [...] Mas *nós* vamos chegar ao nosso foco e atingir *nossas* metas. (grifos nossos)

Finalmente, quando ao modo de enunciação, Mainguenu (2005, p. 94-97) considera que todo texto, seja ele oral ou escrito, está associado a uma “maneira de dizer” específica (recuperada por meio de índices como o tom, o ritmo, a escolha das palavras e dos argumentos etc.) que remete a uma “maneira de ser”. Assim, a leitura (ou escuta) faz emergir uma “origem enunciativa”, uma “instância subjetiva encarnada” que funciona como garantia (fiador) do que é dito. Em trabalhos posteriores (ver MAINGUENEAU, 2006, 2008, entre outros), o autor associará o

modo de enunciação à noção aristotélica de *éthos*, definida, grosso modo, como a imagem de si que o orador constrói no/pelo discurso.

Nas narrativas de vida de Sofia e de Antônio predomina um tom assertivo que remete a um *éthos* de confiança e determinação. Não se percebe, em seus relatos, um tom de lamúria ou de nostalgia, mesmo nos momentos em que eles abordam as dificuldades e obstáculos que enfrentaram – enfrentam ainda – no novo país, como mostram T19 e T20, a seguir. Em T20, temos, inclusive, um toque de humor, quando o sujeito, no caso Antônio, fala de seu estranhamento em relação ao clima parisiense:

T19: Eu vim na intenção de ficar dois anos mais ou menos, dois, três anos, fazer uma certa quantia de dinheiro e voltar pro Brasil, pra montar alguma coisa no Brasil. E acabou que eu fui ficando ano, ano após ano e estou há onze anos. Em relação à língua, eu nunca fiz um curso formal, aprendi no dia a dia mesmo e consigo resolver a minha vida toda. Não preciso de ninguém para fazer nada para mim. É isso. (Sofia)

T20: Olha, o clima, a dificuldade maior é o frio [risos] pra quem tá no Brasil, principalmente, na região a qual eu habitava no Brasil, que é uma região turística, uma região de praia, né?, então, assim, o clima é sempre quente, é sempre sol [...] e aqui você se depara com clima de zero, -2, -3, então, é bem difícil de você conseguir superar, né, eu conheço pessoas que não aguentou o frio, foi embora por conta do tempo, por conta do clima. Então, assim, é difícil, né, mas dá pra aguentar [risos]. (Antônio)

Já Flávia e Magdalena assumem ora um tom esperançoso, como em T21 e T22, ora um tom de tristeza e/ou desânimo, o que (re)lembra a precariedade em que se encontram nos respectivos países em termos de trabalho (fazem faxina em escritórios e casas de família, apesar de terem o ensino médio completo) e mesmo de vida (estão distantes de pessoas queridas, que ficaram no Brasil; não têm muitos amigos; não podem se movimentar livremente; experimentam sentimentos negativos como o medo e a preocupação por não terem autorização de permanência). Nesses momentos, é o *éthos* de fragilidade que sobressai, como ocorre em T23 e T24 – e também em outros trechos já citados como T9 (no caso de Flávia) e T13 (no caso de Magdalena):

T21: Mas assim a gente tá perseverante [risos] tá crendo que vai dar tudo certo, tamos tendo fê, acreditando que os nossos objetivos vão se concluírem, que nós vamos atingir, chegar ao nosso foco, que é o da compra da casa, como eu tinha dito antes e assim tamos esperando (Flávia).

T22: Na verdade eu descobri tem pouco tempo, que eu demorei um pouco pra descobrir, descobri quem eu sou, o que eu quero, aí meus planos agora, assim, eu quero juntar um dinheiro e tô fazendo uns cursos, porque, quando eu voltar, eu tenho alguma coisa pra fazer e vou me dar um prazo de seis meses a um ano pra adaptar. Se eu não adaptar, eu tento ir pra outro lugar, porque eu descobri que o mundo é muito grande, e eu posso fazer o que eu quiser. (Magdalena)

T23: A nossa pretensão aqui é um período de uns cinco, seis anos. A gente não quer ficar mais tempo do que isso aqui e nós queremos voltar no Brasil, até mesmo porque eu deixei meus filhos e a falta deles é [se emociona] muito grande e é muito difícil ficar longe e aqui a vida é muito complicada. [Tenho] dois filhos, uma menina de vinte e um anos, tá no Brasil, faz Faculdade de Pedagogia, e um menino de dezoito anos. [...] E aí a gente olha para trás e vê tudo que deixou, a saudade, a vontade aumenta de voltar, mas aí você pensa no futuro que lá no Brasil você não vai conseguir o que você pode conseguir aqui com seu trabalho. E é isso que é mais difícil para mim é a falta dos filhos. (Flávia)

T24: Porque aqui a vida da gente é trabalhar [...] e, no meu caso assim, eu não posso viajar. Então, eu não sinto assim muita qualidade de vida aqui. Eu já pensei [em regularizar minha situação], mas o jeito que eu posso fazer é casando, ou é pagando pra casar, que eu não concordo muito, ou é casando por interesse, que eu não concordo muito, ou é casando por amor, que não aconteceu até hoje. Então eu também não sei se é a vida que eu quero, se é aqui que eu quero continuar (Magdalena).

A análise feita nesta seção procurou mostrar, por meio dos quatro planos da Semântica Global mobilizados, como os locutores das narrativas de vida se representam (discursivamente), representam os outros, representam, enfim, a experiência migratória, assumindo ou silenciando o fato de estarem em situação irregular nos respectivos países.

Ora, se cada sujeito que generosamente concordou em partilhar sua história conosco (e com os possíveis leitores) é único e sua maneira de ver – e viver – a experiência migratória, singular, não podemos perder de vista certos aspectos que apontam para a construção de um discurso comum. Entre eles, podemos citar a motivação econômica como razão (principal ou secundária) para a migração, bem como as condições precárias de trabalho enfrentadas pelo migrante em situação irregular (temas); o uso do vocábulo *difícil* (e similares) para descrever a vida no novo país, sobretudo, nos momentos iniciais; a predominância de um “eu” que (se) conta ao outro (prevista, aliás, pelo próprio gênero de discurso implicado); o movimento entre um *aqui/agora* – no país de chegada – e um *lá/antes* – no país de partida –, que são frequentemente comparados em seus pontos positivos e negativos; a presença tanto de um éthos otimista e determinado, que se projeta, em geral, no futuro, quanto de um éthos fragilizado, que, não raro, leva o sujeito a querer desistir, ir embora. Dito isso, resta-nos concluir o percurso que empreendemos ao longo destas linhas.

## 6 Considerações finais

Retomamos aqui a pergunta inicial, que julgamos poder responder pelo menos parcialmente, já que os resultados apontados neste artigo referem-se apenas às narrativas de vida selecionadas, embora, a nosso ver, eles possam ser estendidos a muitas e muitas outras. Assim, o que leva alguns brasileiros a deixar seu país natal, onde, pelo menos em tese, gozam de plenos direitos como cidadãos, para viver à margem da lei num país estrangeiro?

Ora, ainda que migrar implique, como vimos, inevitáveis – e, muitas vezes, dolorosas – rupturas com as “raízes”, os sujeitos são motivados pela vontade de ter uma melhor qualidade de vida, pelo sonho – nem sempre alcançado – de poder juntar dinheiro para comprar um imóvel ou montar um negócio no Brasil, ou simplesmente pela necessidade de buscar alternativas, diante da falta de oportunidades no país natal. Por outro lado, eles se aventuram num país que desconhecem, com uma língua que nem sempre dominam, porque já têm parentes migrados, os quais lhes fornecem o apoio necessário, sobretudo no momento da chegada.

E, se esses sujeitos vivem à margem da lei, no sentido de que não dispõem de visto de permanência e/ou outros documentos necessários para regularizar sua situação, há que se levar em conta que eles constituem

uma importante força de trabalho nos países de acolhida, realizando, em geral, serviços rejeitados pelos nativos, porque vistos como menos qualificados. Além disso, como também são consumidores, ajudam a movimentar a economia dos países onde se encontram.

Desse modo, ainda que se submetam a extensas jornadas de trabalho, com salário, em geral, mais baixo do que o dos nativos e sem um contrato formal de trabalho (que lhes garantiria certos direitos), e vivenciem cotidianamente sentimentos negativos, como o medo e a preocupação – embora, como vimos, nem todos se sintam confortáveis para abordar essas questões de forma explícita, o que demanda da parte do(a) analista disposição e capacidade de ler nas entrelinhas –, alguns vão ficando: um ano, dois anos, doze anos... e acabam conseguindo uma certa estabilidade e mesmo se regularizando no país, o que afasta, de vez, a possibilidade de um retorno definitivo ao Brasil. É o caso de Sofia, que nos informou estar iniciando seu processo de legalização.

Outros, porém, retornam ao país de origem por variadas razões: alcançaram a meta de juntar dinheiro, não conseguiram se adaptar ou simplesmente são movidos por um novo desejo de mudança. Como diz Magdalena, ao finalizar sua entrevista, “[...] porque a minha vida sempre foi mudar, mudar, mudar, às vezes eu tenho essa coisa de querer uma estabilidade, mas ao mesmo tempo eu quero essa mobilidade, eu tenho essa coisa de continuar tentando outras coisas”.

As narrativas de vida aqui apresentadas (e as análises que sobre elas empreendemos) não deixam dúvidas de que ouvir o que os próprios sujeitos deslocados têm a dizer de si, dos outros, do mundo é uma forma de restituir-lhes a palavra – confiscada pelos números e/ou pelos “portavozes autorizados” – e fazê-la circular para além do espaço privado, onde ela se constitui e, não raro, permanece.

## Referências

ARFUCH, L. *O espaço biográfico: dilemas da subjetividade contemporânea*. Trad. Paloma Vidal. Rio de Janeiro: Ed. UERJ, 2010.

AUTHIER-REVUZ, J. La non-coïncidence interlocutive et ses reflets méta-énonciatifs. In: BERRENDONER, A.; PARRET, H. (éd.). *L'interaction communicative*. Berne; Frankfurt; N.Y.; Paris: Peter Lang, 1990. p. 173-193.

BARTRAM, D. *et al. Key Concepts in Migration*. London: Sage, 2014. DOI: <https://doi.org/10.4135/9781473921061>

BERNARD, P. *Immigration: le défi mondial*. Paris: Gallimard, 2002.

BERNARDOT, M-J. *Étrangers, immigrés: (re)penser l'intégration*. Rennes: Presses de L'École des Hautes Études en Santé Publique, 2019.

BERTAUX, D. *Le récit de vie*. Paris: Armand Colin, 2005.

BLANCHARD, P. *et al. Atlas des immigrations en France*. Paris: Autrement, 2016.

BUENO, A. M. Imigrantes japoneses e a língua portuguesa: um caso de preconceito linguístico. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 28, n. 1, p. 455-478, 2020. DOI: <https://doi.org/10.17851/2237-2083.28.1.455-478>

CALABRESE, L; VENIARD, M. Mots, discours et migration, une relation dialectique. In: \_\_\_\_\_ (org.). *Penser les mots, dire la migration*. Bruxelles; Paris: Academia; L'Harmattan, 2018. p. 9-31.

CARVALHO, A. *Um estudo sociodiscursivo da temática do preconceito contra negros em sentenças de injúria racial*. 2020. 236f. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2020.

CASTRO, M. C. G.de *et al.* Contexto migratório de retorno. In: PEIXOTO, J. *et al.* (org.). *Vagas atlânticas: migrações entre Brasil e Portugal no início do século XXI*. Lisboa: Mundos Sociais, 2015. p. 159-176.

CHARAUDEAU, P. *Grammaire du sens et de l'expression*. Paris: Hachette, 1992.

CHARAUDEAU, P. Les stéréotypes, c'est bien. Les imaginaires, c'est mieux. In: BOYER, H. (org.). *Stéréotypage, stéréotypes: fonctionnements ordinaires et mises en scène*. Paris: L'Harmattan. 2007. [s.p.]. Disponível em: <http://www.patrick-charaudeau.com/Les-stereotypes-c-est-bien-Les,98.html>. Acesso em: 18 jun. 2020.

CHARAUDEAU, P. Os imaginários de verdade do discurso político. In: \_\_\_\_\_. *Discurso político*. São Paulo: Contexto, 2015. p. 185-245.

CLOCHARD, O. Les réfugiés dans le monde entre protection et illégalité. *EchoGéo*, [S.l.], v. 2, p. 1-8, 2007. Disponível em: <http://echogeo.revues.org/1696>. Acesso em: 30 set. 2019. DOI: <https://doi.org/10.4000/echogeo.1696>

FIORIN, J. L. Pragmática. In: \_\_\_\_\_. (org.). *Introdução à linguística II*. Princípios de análise. São Paulo: Contexto, 2003. p. 161-185

HORTA, A. P. B. A imigração em Portugal. Um contributo para o debate sobre as políticas e práticas de integração. In: FONSECA, M. L. et al. (org.). *Migrações na Europa e em Portugal*. Coimbra: Almedina, 2013. p. 227-250.

LAACHER, S. *Ce qu'immigrer veut dire: idées reçues sur l'immigration*. Paris: Le Cavalier Bleue, 2012.

LARA, G. M. P. A(s) voz(es) dos vulneráveis: narrativas de vida de imigrantes e refugiados à luz da análise do discurso. In: BARONAS, R. L. et al. (org.). *As ciências da linguagem e a(s) voz(es) e o(s) silenciamento(s) de vulneráveis*. Campinas: Pontes, 2018. p. 145-166.

LARA, G. M. P. De “Ouvrons les portes” a “Em casa no Brasil”: olhares contemporâneos sobre a migração. *Gláuks*, Viçosa, v. 19, n. 1, p. 79-100, 2019. DOI: <https://doi.org/10.47677/gluks.v19i1.151>. Disponível em: <https://www.revistaglauks.ufv.br/Glauks/issue/view/24/27>. Acesso em: 22 jun. 2020.

MACHADO, I. L.; LESSA, C. H. Reflexões sobre o gênero narrativa de vida do ponto de vista da análise do discurso. In: JESUS, S. N.; SILVA, S. M. R. da (org.). *O discurso & outras materialidades*. São Carlos: Pedro & João, 2013. p. 102-122.

MAINGUENEAU, D. *L'Analyse du discours: introduction aux lectures de l'archive*. Paris: Hachette, 1991.

MAINGUENEAU, D. *Gênese dos discursos*. Trad. Sírio Possenti. Curitiba: Criar, 2005.

MAINGUENEAU, D. *Cenas da enunciação*. Org. Sírio Possenti e Maria Cecília P. Souza-e-Silva. Curitiba: Criar, 2006.

MAINGUENEAU, D. A propósito do ethos. Trad. Luciana Salgado. In: MOTTA, A. R.; SALGADO, L. (org.). *Ethos discursivo*. São Paulo: Contexto, 2008. p.11-29.

MARQUES, J. C.; GÓIS, P. Processos de integração dos imigrantes brasileiros na sociedade portuguesa. In: PEIXOTO, J. *et al.* (org.). *Vagas atlânticas: migrações entre Brasil e Portugal no início do século XXI*. Lisboa: Mundos Sociais, 2015. p. 109-134.

MOREIRA, G. M. *Figures de migrants brésiliens en France: approche anthropologique et sociolinguistique*. 2018. 1495f. Thèse (Doctorat en Linguistique) – Université Paul Valéry – Montpellier III, Montpellier, 2018.

PEIXOTO, J. Imigração, emprego e mercado de trabalho em Portugal: os dilemas do crescimento e o impacto da recessão. In: FONSECA, M. L. *et al.* (org.). *Migrações na Europa e em Portugal*. Coimbra: Almedina, 2013. p.158-184.

PEIXOTO, J. *et al.* (org.). *Vagas atlânticas: migrações entre Brasil e Portugal no início do século XXI*. Lisboa: Mundos Sociais, 2015.





**“Isso tudo me traz de novo a vida que eu tinha”:  
a coconstrução de uma narrativa autobiográfica  
na Doença de Alzheimer**

***“It all brings me back to the life I had”: the co-construction  
of an autobiographical narrative in Alzheimer’s Disease***

Caio Mira

Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Rio  
Grande do Sul / Brasil

cmira@unisinos.br

<http://orcid.org/0000-0002-4858-1743>

Katiuscia Custodio

Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Rio  
Grande do Sul / Brasil

katiusciacustodio@gmail.com

<http://orcid.org/0000-0002-9735-6251>

**Resumo:** A Doença de Alzheimer é uma patologia neurodegenerativa que afeta, dentre outras funções cognitivas, a linguagem, causando dificuldades de acesso lexical, articulação fonológica, organização sintática e alterações de elementos pragmáticos da conversação que prejudicam as interações cotidianas. Tendo em vista esse contexto, o presente artigo analisa a interação face a face de uma pessoa acometida pela DA na fase moderada, com um pesquisador em uma situação de entrevista narrativa. A partir do pressuposto de que usar a linguagem é engajar-se em uma ação colaborativa, especificamente a produção de uma narrativa autobiográfica, o presente trabalho fundamenta-se no aparato teórico da Análise da Conversação, dos Estudos das Narrativas Oraís e da Linguística Textual. A narrativa autobiográfica que tomaremos em análise no presente artigo integra um *corpus* de interações gravadas com uma participante acometida pela DA. As análises demonstram que a participante mantém uma atitude colaborativa durante toda a interação, coconstruindo referentes com o interlocutor, inclusive quando surgem déficits ocasionados pela patologia. Além disso, a performance

narrativa da participante caracteriza-se por uma narrativa autobiográfica que emerge no contexto discursivo-interacional exercendo o papel de ressignificar sua experiência de vida frente à doença.

**Palavras-chave:** Doença de Alzheimer; narrativa autobiográfica; referência.

**Abstract:** Alzheimer's disease is a neurodegenerative pathology that affects, among other cognitive functions, language, causing difficulties in lexical access, phonological articulation, syntactic organization and changes in pragmatic elements of conversation that impair daily interactions. In view of this context, the present article analyzes the face-to-face interaction of a person affected by AD in the moderate phase, with a researcher in a situation of narrative interview. Based on the assumption that using language is to engage in a collaborative action, specifically the production of an autobiographical narrative, the present work is based on the theoretical apparatus of Conversation Analysis, Oral Narrative Studies and Textual Linguistics. The autobiographical narrative that we will analyze in this article includes a corpus of interactions recorded with a participant affected by AD. The analyzes show that the participant maintains a collaborative attitude throughout the interaction, co-building referents with the interlocutor, even when deficits caused by the pathology arise. In addition, the participant's narrative performance is characterized by an autobiographical narrative that emerges in the discursive-interactional context playing the role of reframing her life experience in the face of the disease.

**Keywords:** Alzheimer's disease; autobiographical narrative; reference.

Recebido em 27 de outubro de 2020

Aceito em 16 de dezembro de 2020

## Introdução

A linguagem é o elemento de colaboração de nossas ações no mundo social. Das grandes às pequenas tarefas cotidianas, a linguagem constitui o cenário básico em que as ações verbais são ações conjuntas, ou seja, onde usar a linguagem é sempre se engajar em alguma ação na qual a linguagem é o meio e o lugar onde a ação acontece necessariamente em coordenação com outras pessoas. O caráter colaborativo da linguagem consiste na ação conjunta que emerge quando falantes e ouvintes – ou escritores e leitores – desempenham duas ações individuais em coordenação, como um conjunto: “[...] são instâncias de uso da linguagem, atividades nas quais, com a linguagem, as pessoas fazem coisas” (CLARK, 1996, p. 3).

Uma das inúmeras formas de usar a linguagem para desempenhar uma ação conjunta pode ser expressa por uma atividade discursiva recorrente em nossas vidas: o ato de contar histórias. Apesar de parecer uma simples atividade cotidiana, narrar revela uma ação que fazemos em colaboração com o interlocutor, abarcando diferentes propósitos, seja para convencer alguém de algo, argumentar em favor de um ponto de vista, refutar uma ideia, defender uma causa, entreter ou até mesmo demonstrar quem de fato somos e construir nossas identidades.

Considerando a narrativa como uma ação conjunta, analisar esse tipo de produção discursiva implica investigar os recursos linguísticos e interacionais utilizados tanto na elaboração do enredo da história quanto no desempenho de papéis dos personagens/agentes que são construídos por quem narra e por quem ouve as histórias (DE FINNA; GEORGAKOPOLOU, 2012; MIRA, 2019). Em outras palavras, assumimos a narrativa como uma prática discursivo-interacional que é socialmente situada (BASTOS; BIAR, 2015). Portanto, é necessário considerar que narrar consiste em uma ação, isto é, uma performance em que construímos significados sobre quem somos, sobre quem são os outros e sobre as configurações do mundo social (KIM, 2016; MOITA LOPES, 2009).

As formas de desenvolvimento da narrativa são discursiva e textualmente formatadas pelo contexto interacional (NORRICK, 2007). Nesse sentido, Ochs e Capps (2001, p. 02) argumentam que as narrativas são moldadas durante as interações e que a ação de narrar constitui:

[...] uma ferramenta para refletir de forma colaborativa sobre situações específicas e seu lugar no esquema geral da vida. Em ensaios desse tipo, o conteúdo e a direção que os enquadramentos narrativos tomam são dependentes da contribuição narrativa de outros interlocutores, que fornecem, extraem, criticam, recusam e tiram inferências de facetas do relato que se desenrola. Nessas trocas, a narrativa torna-se uma conquista interacional e os interlocutores tornam-se coautores.<sup>1</sup>

---

<sup>1</sup> “Narrative activity becomes a tool for collaboratively reflecting upon specific situations and their place in the general scheme of life. In essays of this sort, the content and direction that narrative framings take are contingent upon the narrative input of other interlocutors, who provide, elicit, criticize, refuse, and draw inferences from facets of the unfolding account. In these exchanges, narrative becomes an interactional achievement and interlocutors become co-authors”.

Nessa prática de coconstrução entre os interlocutores, conhecimentos compartilhados são negociados e sentidos são construídos mutuamente à medida que a história se desenrola, incidindo diretamente sobre a relevância de algo estar sendo contado em determinado contexto. Como uma prática discursiva, a narrativa possibilita a análise linguística por duas perspectivas: a primeira focaliza o universo narrativo em si próprio (os eventos, os personagens e a textualidade desses elementos da história) e a segunda considera como ponto de partida de análise a interação na qual a narrativa é originada e o contexto em que ocorre a negociação de sentidos entre quem narra e quem ouve as histórias (FLANNERY, 2015; MIRA; CUSTODIO, 2019).

O presente trabalho tem como objetivo analisar uma narrativa autobiográfica produzida por uma pessoa que vive em uma fase moderada da Doença de Alzheimer (doravante DA). As características da linguagem da participante nessa fase são marcadas pelas dificuldades de acesso lexical, de articulação fonológica de alguns verbos e substantivos, repetições de itens lexicais e de tópicos durante a conversação. Além disso, são frequentes pequenos lapsos de memória e momentos de autocorreção dos enunciados. Apesar do progresso do quadro neurodegenerativo, que compromete algumas funções motoras e a visão, a participante consegue interagir de forma satisfatória nas interações com seus familiares e cuidadores e também nas situações de entrevista narrativa com o pesquisador. Dessa forma, consideramos, a partir da observação de nossos dados, que a linguagem da participante se manifesta no cenário interativo demonstrando as dimensões da narrativa e a coconstrução de referentes que se articulam no processo de coconstrução do sentido com o interlocutor, ressignificando as experiências de vida.

Aliada à concepção de narrativa no âmbito da perspectiva socioconstrucionista, o presente trabalho se baseia na perspectiva textual-interativa, desenvolvida no arcabouço teórico metodológico da Linguística Textual e da Análise da Conversação. Especificamente, o intuito dessa articulação teórica é compreender as estratégias de organização interativa e referencial envolvidas na constituição e manutenção dos elementos da narrativa autobiográfica produzida no contexto da DA, que é marcado por alterações linguísticas e cognitivas.

Com base na noção de referenciação, compreendida como uma atividade sociocognitiva de construção de sentidos, conforme proposto por Mondada e Dubois (2003), as estratégias linguísticas/textuais que

emergem em narrativas demonstram uma forma de produção discursiva em que os interlocutores constroem e interpretam colaborativamente os referentes das práticas interacionais e discursivas em que nos engajamos cotidianamente. Nessa perspectiva, Marcuschi (2004, p. 263) afirma que “o problema da significação não é resolver se às palavras corresponde algo no mundo externo e sim o que fazemos do ponto de vista semântico quando usamos as palavras para dizer algo”.

A referenciação consiste em um fenômeno textual e interativo que analisaremos a fim de demonstrar o caráter colaborativo próprio da linguagem na perspectiva da sociocognição, isto é, a construção conjunta de significados com base nos conhecimentos compartilhados entre os indivíduos, tendo como prática discursiva a narrativa autobiográfica, que explora as formas de construção e exibição de quem o narrador/personagem é, foi no passado ou projeta ser no futuro (HYDÉN, 2018). Nesse sentido, acreditamos que a contribuição seja evidenciar as estratégias colaborativas e compensatórias das pessoas que vivem com DA como um dos desafios enfrentados para se manterem como uma voz ativa nas interações em que contam a própria história ou (re)constróem eventos e fatos do cotidiano (HYDÉN, 2017; MIRA, 2019). Contar histórias sobre eventos passados é uma atividade conjunta em que os interlocutores se envolvem em uma ação que acontece necessariamente em coordenação com os outros sujeitos (CLARK, 1996; HYDÉN, 2017).

De acordo com Marcuschi (2001, p. 38), a construção da referência consiste “naquilo que, na atividade discursiva e no enquadre das relações interpessoais, é construído num comum acordo entre os atores sociais envolvidos numa dada tarefa comunicativa”. Ou seja, ao contar uma história, os interlocutores negociam objetos de discurso que exercem um papel fundamental na constituição da narrativa e na sua relevância no contexto interativo. Tais ações, muito comuns em situações triviais que desempenhamos nas interações cotidianas, são postas à prova quando nos colocamos em situações comunicativas nas quais a linguagem é afetada por uma patologia neurodegenerativa como a DA.

A DA é uma doença neurodegenerativa que acomete de 60% a 70% dos casos (SANTOS *et al.*, 2019) das neurodegenerências em idosos e que acomete funções cognitivas como memória e atenção, funções motoras como equilíbrio, força, flexibilidade e função aeróbica observando-se tais declínios já na fase inicial da doença. Trata-se, segundo Vieira *et al.* (2014), de uma doença ocasionada pelo acúmulo de

peptídeo *beta-amilóide* que se deposita nos vasos cerebrais e ocasiona a morte neuronal. Os impactos ocasionados pela DA incidem diretamente sobre as funções cognitivas, como memória, atenção e função executiva além de funções motoras, já nos estágios iniciais (ZIDAN *et al.*, 2012), o que, consequentemente, impede a pessoa acometida de realizar atividades cotidianas autonomamente.

Em relação à linguagem, na fase inicial da doença, “seriam identificados déficits na atividade de nomeação, repetições, circunlóquios, uso expressivo de dêiticos e de estruturas sintáticas consideradas ‘simples’, sem déficits expressivos no processamento fonológico” (MORATO, 2010, p. 104). Já na fase moderada, “os problemas mnésicos passam a ser incapacitantes, seguidos de crescente desorientação temporal-espacial e de problemas de linguagem mais crescentes e prontamente perceptíveis” (MORATO, 2016, p. 583-584). Por fim, na fase severa, a memória é afetada consideravelmente e a linguagem torna-se comprometida em todos os seus níveis (MORATO, 2016). A patologia também desencadeia impactos “nas formas de recepção social da doença (algo que inclui as práticas diagnósticas e a interação do doente com seus próximos), bem como de seu enfrentamento no plano psicossocial, médico-terapêutico e familiar” (MORATO, 2016, p. 584).

O contexto de declínios que envolve a DA afeta não somente questões relacionadas à saúde, mas a vida social do indivíduo em um âmbito maior, incluindo a sua participação social, impactando na compreensão que as pessoas não acometidas têm sobre a doença e como interagem com as pessoas acometidas. Diante desse quadro, o presente trabalho pretende trazer à pauta uma análise dos elementos linguísticos e interacionais que constituem uma cena interativa no contexto dessa doença neurodegenerativa. O nosso intuito é demonstrar como a participante ressignifica suas memórias e promove sua participação na interação a partir da constituição de uma narrativa autobiográfica que ocorre no contexto de entrevista narrativa com a presença de um pesquisador.

## **1 A narrativa**

A variedade de histórias que as pessoas contam, de acordo com o contexto interacional do qual estejam participando, motiva diferentes abordagens linguísticas a estudarem as narrativas. Passando de uma

abordagem inicial de estudos da narrativa, no campo da sociolinguística, mais voltada à forma de como se estrutura (LABOV; WALETZKY, 1967), para uma perspectiva interacionista de estudo da linguagem, as narrativas passam a ser vistas como práticas sociais e despertam o interesse das ciências humanas, ganhando cada vez mais espaço nesse campo de investigação.

Dada a ubiquidade do fenômeno narrativo, várias áreas da Linguística passaram também a se interessar por narrativas que emergem cotidianamente na cena interacional. Por meio da análise dos mecanismos discursivos utilizados pelos falantes, é possível compreender como as pessoas constroem imagens sobre si mesmas, deixando transparecer sua autopercepção por meio de suas histórias e de suas experiências de vida.

A narrativa autobiográfica é uma das abordagens de estudo que “consiste em um relato que auxilia na definição da identidade dos narradores, oferecendo uma versão coerente sobre por que e como eles se encontram nas posições atuais” (FLANNERY, 2015, p. 40). Em outras palavras, ela revela uma resignificação da experiência de vivida pelo “eu” que incide sobre como ele constrói sua identidade, ou como se autoavalia. Para De Finna (2015), a ação de contar histórias é um mecanismo cognitivo e psicológico diretamente relacionado com o autodesenvolvimento, propiciando um senso positivo de si mesmo e um artifício para lidar com momentos difíceis que enfrentamos.

De acordo com Linde (1993), esse tipo de narrativa tem como especificidade o fato de poder ser recontada várias vezes e não apenas em um único episódio. Além disso, as histórias de vida deixam transparecer uma avaliação permeada por valores socioculturais do meio em que vivemos. Em relação a esse aspecto da narrativa autobiográfica, vale destacar a sua importância no contexto em que situamos nosso estudo. Costumeiramente, as narrativas de pessoas acometidas por algum tipo de doença neurodegenerativa tendem a ser rotuladas como incompreensíveis, desorganizadas ou fragmentadas por apresentarem desvios do tópico, desordem temporal, serem repetitivas ou ainda apresentar trocas de personagens e de espaço sem uma preparação prévia ao interlocutor (HYDÉN, 2018). Entretanto, as narrativas têm um papel importante na rotina de quem vive com DA, pois:

Contar histórias ainda é uma atividade relevante para a pessoa com demência em todas as diferentes fases do processo da doença, pela simples razão de que tanto a pessoa com demência quanto

outros membros da família têm muito de sua identidade investida nas histórias do cotidiano, e todos continuam a contar histórias, mesmo quando a pessoa com demência tem graves problemas para animar as histórias. (HYDÉN, 2018, p. 5-6, tradução nossa).<sup>2</sup>

Em um cenário no qual a memória é desafiada, contar histórias de vida e ressignificar experiências adquire um novo sentido. As pessoas acometidas por patologias neurodegenerativas são pessoas que, como todas as outras, possuem “voz” e cuja identidade é coconstruída com os interlocutores por meio das histórias de experiências vividas e contadas conjuntamente. Pelas memórias e pelas histórias contadas por outras pessoas ou por si mesmas, as pessoas acometidas pela DA também podem reconstruir suas identidades reflexivamente na interação.

O sentido sobre nossa identidade é sustentado pela relação que estabelecemos com os outros na interação. Considerando tamanha importância da interação e das histórias de vida que contamos, “todas as pessoas, mas especialmente as pessoas com Doença de Alzheimer, dependem das interações sociais para a criação e manutenção de sua personalidade e senso de identidade” (SHENK, 2005, p. 6, tradução nossa).<sup>3</sup>

Conforme anteriormente discutido, o uso da linguagem, de acordo com Clark (1996), consiste em desempenhar uma ação conjunta, elaborada a partir de ações individuais que se complementam, em colaboração entre os indivíduos. O autor inclusive a compara com uma música de Mozart tocada no piano. Se tocada apenas por uma pessoa, a ação depende unicamente de seus processos cognitivos próprios, que estão sob seu poder de controle e decisão. No entanto, ao tocá-la em dueto, a ação depende da cooperação das ações entre ambos e somente será bem sucedida se as ações se complementarem colaborativamente. Nesse caso, os processos cognitivos serão compartilhados entre os participantes. O papel da interação social e da linguagem para a cognição

---

<sup>2</sup> “Storytelling is still a relevant activity for the person with dementia at all the different stages of the disease process for the simple reason that both the person with dementia and other family members have much of their identity invested in everyday stories, and they all continue to tell stories, even when the person with dementia has severe problems with animating the stories.”

<sup>3</sup> “All persons, but specially persons with Alzheimer’s disease, are dependent upon social interactions for the creation and maintenance of their personhood and sense of self.”

humana também foi pautado por Vygotski (2005), que apresenta a ideia de “uma relação dialética entre o linguístico e o cognitivo no tratamento da significação, plasmados no contexto enunciativo, isto é, instanciados em práticas sociais com linguagem” (MORATO, 2008, p. 159).

A colaboração da pessoa não acometida pela doença estabelece o que pode ser chamado de andaime<sup>4</sup> necessário a fim de facilitar a interação, seja estabelecendo significados conjuntos, relembrando experiências compartilhadas ou contribuindo com conhecimentos comuns a ambos. Esse andaime pode acontecer de três formas: i). pelas atividades de ativação de *frames*; ii). pelas contribuições e iii). pelos reparos. Todas essas ações podem favorecer a participação da pessoa com DA e também a compreensão de outros interlocutores (HYDÉN, 2018).

Em se tratando de narrativas que emergem interacionalmente, o andaime influi diretamente na sua continuidade quando as dificuldades ocasionadas pela doença se sobrepõem. Dessa forma, nas interações que acontecem dentro do contexto patológico abordado, o papel do andaime pode ser de significativa relevância, pois resulta em uma maior ou menor participação da pessoa acometida, ocupando uma posição central ou periférica no cenário cotidiano (CUSTODIO, 2019).

Devido à natureza das ações coordenadas que constituem o uso da linguagem, entendemos a narrativa que emerge nas interações cotidianas como constituída pelo trabalho de coconstrução entre falantes e ouvintes e apresentando uma estrutura mais flexível, delineada pelo contexto em que ocorre, diferente do modelo de Labov e Waletzky (1967), que exigia a presença de estruturas comuns às narrativas. Segundo Ochs e Capps (2001) as narrativas de experiências pessoais são histórias construídas dentro da interação que não apresentam um roteiro estabelecido, mas se desenrolam de acordo com a ação conjunta construída entre os participantes durante as mais variadas interações.

Por essa perspectiva, as autoras defendem a ideia de que as narrativas podem variar e que o que as caracteriza não seria uma estrutura fixa, mas cinco dimensões que podem ser observadas de uma forma mais fluida ou menos rígida, como se fosse em um *continuum* cujos polos

---

<sup>4</sup> Andaime (*scaffolding*) é um termo que foi cunhado a partir da teoria vygotskiana, por Wood, Bruner e Ross (1976) ao se referirem ao processo no qual um indivíduo mais experiente, no caso, o professor age dando um suporte interacional à criança a fim de superar suas dificuldades, agindo de forma autônoma.

são as possibilidades que os falantes podem alcançar nos elementos da história de acordo com o contexto interacional.

As dimensões que caracterizam as narrativas coconstruídas e que foram estabelecidas pelas autoras são:

- Narração (*tellership*): a história pode apresentar um narrador principal ou múltiplos conarradores.
- Historiabilidade (*tellability*): revela a importância da história ser contada em determinado contexto como sendo mais alta ou mais baixa em graus de relevância.
- Encaixe (*embeddedness*): demonstra a conexão da narrativa emergente com o discurso circundante, podendo surgir como destacável ou como encaixada à conversa.
- Linearidade (*linearity*): a história pode ser construída com progressão temporal e causal, ou apresentando uma ordem não-linear dos eventos narrados.
- Postura Moral (*moral stance*): a história pode demonstrar uma avaliação moral mais certa e constante ou mais incerta e fluída.

O quadro a seguir sintetiza as dimensões propostas pelas autoras:

QUADRO 1 – Dimensões da narrativa

Dimensões		Possibilidades
Narração	Um narrador ativo	→ Múltiplos co-narradores ativos
Historiabilidade	Alta	→ Baixa
Encaixe	Isolada	→ Encaixada
Linearidade	Ordem causal e temporal finalizada	→ Ordem causal e temporal aberta
Postura moral	Determinada, constante	→ Indeterminada, fluída

Fonte: Ochs e Capps (2001, p. 20).

A partir do modelo de Ochs e Capps (2001), consideramos as narrativas sobre histórias de vida de pessoas com DA, como *locus* de análise das estratégias de coconstrução de significados, da alternância de papéis e de colaboração, nas quais elas podem demonstrar o sentido sobre si mesmas e suas identidades e experiências em coconstrução com o interlocutor.

## 2 Referenciação

Diversas áreas de estudo da linguagem se debruçam sobre a relação estabelecida entre as palavras e o mundo e sobre como utilizamos a linguagem para referir à realidade do mundo. Estudos de cunho cognitivistas (ROSCH, 1975) pressupõem uma relação pré-estabelecida ao referenciarmos as palavras aos objetos do mundo e, assim sendo, as performances discursivas são medidas de acordo com seu grau de correspondência com o mundo (MONDADA; DUBOIS, 2003).

Por outro lado, o paradigma sociocognitivista aborda a linguagem, deslocando a concepção de uma referência pronta, de uma língua como um reflexo do mundo e considerando que os sentidos são negociados constantemente pelos sujeitos, mudando de acordo com suas percepções de mundo. Essa perspectiva fica clara nas palavras de Marcuschi (2001, p. 38) ao afirmar que “a referência não se dá apenas na relação linguagem-mundo”, ou seja, a realidade não é apreendida da mesma forma por todas as pessoas e assim, a língua manifesta-se como uma construção social e histórica, e a relação entre linguagem e mundo resulta em uma atividade cognitiva-interacional desempenhada pelos indivíduos (MIRA, 2019). Partilhando dessa concepção, Mondada e Dubois (2003, p. 273) consideram que:

As práticas linguísticas não são imputáveis a um sujeito cognitivo abstrato, racional, intencional e ideal, solitário face ao mundo, mas a uma construção de objetos cognitivos e discursivos na intersubjetividade das negociações, das modificações, das ratificações de concepções individuais e públicas do mundo.

Nesse jogo de construção de objetos negociados nas interações, podemos considerar que a realidade não está pronta, mas vai sendo ressignificada por meio das experiências que compartilhamos. As práticas de linguagem são desempenhadas por sujeitos que a todo instante reelaboram significados criativamente. Ao deslocar a ideia de uma referência única e pronta, a noção de referenciação compreende essa questão como “um fenômeno de natureza semântico-discursiva em que é possível observar a emergência de processos de significação que evidenciam as relações entre linguagem, cognição e interação”. (MIRA; CARNIN, 2017, p. 163.)

Por conseguinte, a referenciação possibilita ressignificar a realidade no âmbito discursivo, instaurando a noção de *objetos de discurso* dinâmicos cujo sentido é colaborativamente construído pelos interagentes durante o desenvolvimento da cena interacional. Ao negociar sentidos durante o processo textual-interativo da referenciação:

O sujeito, por ocasião da interação verbal, mobiliza um conjunto de estratégias de ordem sociocognitiva e opera sobre o material linguístico, procedendo a escolhas significativas para representar estados de coisas, com vistas à concretização de sua proposta de sentido (KOCH, 1999, 2002). [...] Desta maneira, “endereços” ou nódulos cognitivos já existentes podem ser, a qualquer momento, modificados ou expandidos, de modo que, durante o processo de compreensão, desdobra-se uma unidade de representação extremamente complexa, pelo acréscimo sucessivo e intermitente de novas categorizações e/ou avaliações acerca do referente. (KOCH, 2008, p. 203).

De acordo com Koch (2004), os falantes constroem os objetos de discurso em ações formulativas, metaformulativas e metadiscursivas que são desenvolvidas na interação. Portanto, os referentes não são estanques, mas constituem objetos de discurso que não preexistem naturalmente à atividade cognitiva e interativa dos falantes (APOTHÉLOZ; REICHLER-BÉGUELIN, 1995). Pelo contrário, é na e pela interação que objetos de discurso são modificados nas ações coordenadas em que os falantes se engajam para usar a linguagem em suas diversas configurações discursivas (CLARK, 1996; JUBRAN, 2006).

No escopo do presente trabalho, a atividade sociocognitiva da referenciação configura-se como o processo pelo qual a pessoa com DA e o interlocutor articulam conhecimentos mutuamente no propósito de favorecer a produção textual e interativa de narrativas. Considerando que, ao contar uma história, estamos produzindo também um texto oral, as narrativas desmistificam e estabelecem a coerência entre o passado, o presente e os fatos ainda não realizados (OCHS; CAPPS, 2001). A perspectiva textual-interativa possibilita compreender como a narrativa é construída na dinâmica interacional, mostrando de que forma os referentes são inseridos, ativados e negociados na tarefa de coconstrução (ou construção reflexiva) que marcam os elementos do ato de contar histórias (FLANNERY, 2010; MIRA, 2019).

### 3 Metodologia

A pesquisa proposta constitui-se como um estudo de caso, sendo uma investigação de cunho qualitativa-interpretativista, amparado pelo quadro teórico-metodológico da Linguística Textual e da Análise da Conversação (MARCUSCHI, 1998). A escolha dessas duas áreas é justificada pela intenção de desenvolver análises que contemplem as narrativas autobiográficas produzidas pela participante com DA por dois motivos a partir do textual-interativo, que “examina os princípios gerais de constituição do texto falado, mas dá um passo a mais na análise dos textos, ao considerar esses procedimentos no âmbito da construção do processo de interação entre falantes” (LEITE *et al.*, 2010, p. 52). A escolha dessa perspectiva procura evidenciar as estratégias de construção do texto falado no modelo proposto por Ochs e Capps (2001), no intuito de desenvolver análises que abarquem as estratégias referenciais e as estratégias de colaboração entre os interlocutores na interação oral, o partilhar de conhecimento entre ambos e o esforço e a disposição para produzir inferências.

Como instrumento de geração de dados, foi utilizada a entrevista aberta, sem roteiro preestabelecido, concebida como um evento de fala (MISHLER, 1986), dentro da qual o discurso é construído conjuntamente entre os participantes durante a interação face a face. Do ponto de vista interacional, as entrevistas se configuram sem uma rigidez estrutural, sendo similares a uma situação conversacional cotidiana em que os interagentes participam de forma simétrica da troca de turnos e nas dinâmicas de desenvolvimento dos tópicos. Os temas abordados nesses encontros são iniciados a partir de comentários de fatos do dia a dia ou de relatos da participante, que vive com DA, a respeito de suas atividades diárias, de viagens ou visitas aos familiares e de sua antiga rotina de trabalho como professora. É nesse contexto interacional que surgem as narrativas.

A geração de dados ocorreu por um período de cerca de 12 meses, em encontros mensais entre um pesquisador e a participante com DA. O tempo total de gravação em áudio e vídeo que compõe o *corpus* do presente estudo é de cerca de 18 horas. O critério de escolha dos dados, para esta análise, foi a recorrência de narrativas autobiográficas durante os encontros. Priorizamos esse tipo de realização discursiva, pois é uma das ações mais frequentes desempenhadas pela participante

nas interações em coconstrução com o interlocutor. Há também a recorrência de narrativas autobiográficas que são desencadeadas por tópicos introduzidos nas entrevistas abertas pela participante. O dado que será analisado no presente artigo foi gerado em abril de 2016, abrangendo em torno de 1 hora de gravação em vídeo. Como categorias de análise, mobilizaremos as dimensões da narrativa (OCHS; CAPPS, 2001) e a referenciação. Para propiciar uma melhor organização da análise, dividiremos o dado em três excertos.

O sistema de notação utilizado na transcrição dos dados tem como base as notações já utilizadas nos estudos do projeto NURC<sup>5</sup> (Norma Urbana Culta) por Marcuschi (1998) e adaptado por Custodio (2019). A identidade e o anonimato da participante foram preservados durante todo o processo de geração e transcrição de dados, conforme o Termo de Consentimento Livre e Esclarecido. A pesquisa teve a aprovação do Comitê de Ética em Pesquisa da Universidade do Vale do Rio dos Sinos (UNISINOS), projeto nº 15/191. Nas transcrições, a participante é identificada pelo nome fictício Joana. Os demais participantes das interações, o pesquisador e a cuidadora de Joana também receberam nomes fictícios.

### **3.1 A participante**

Os dados apresentados neste trabalho são provenientes de interações entre um pesquisador e Joana, que à época da realização da pesquisa tinha 70 anos. Diagnosticada com a Doença de Alzheimer há cerca de 5 anos, Joana encontra-se em uma fase moderada da doença e é ciente de sua condição. Uma característica de Joana que vale a pena ser destacada é o fato de que ela é bilíngue e possui um alto grau de letramento, tendo atuado profissionalmente ao longo de sua vida como professora universitária de língua inglesa. Em sua rotina, ela busca se envolver em diversos tipos de atividades, tais como exercícios físicos, conversas cotidianas, organização de livros, discos musicais, fotografias, recordações de viagens. Além dessas atividades, são frequentes as viagens e visitas à casa de familiares e amigos.

---

<sup>5</sup> O Projeto Norma Urbana Culta (NURC) teve início em 1969 e foi desenvolvido em cinco capitais brasileiras: Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre, tendo como objetivo analisar a linguagem oral de falantes de escolaridade de nível superior completo.

Atualmente, como sintomas visíveis da DA, Joana apresenta lapsos de memória recente, dificuldade de articulação fonológica no início de palavras, de acesso lexical, parafasias semânticas e lexicais e repetição de segmentos vocálicos, além da perda de visão, o que a impede de desempenhar suas atividades cotidianas sem auxílio. No entanto, apresenta preservadas a ordenação de estruturas sintáticas e a manipulação de regras pragmáticas no uso da linguagem.

Ela tem o apoio de seu círculo familiar mais próximo e de duas cuidadoras para desempenhar a maioria de suas atividades cotidianas. Apesar das dificuldades, Joana procura manter-se ativa fazendo exercícios físicos, organizando seus pertences pessoais, visitando amigos e familiares ou prestigiando atrações culturais sempre que possível. É bastante comum, nas situações de entrevistas, Joana iniciar as interações trazendo algum objeto importante para ela que queira mostrar ao entrevistador, como objetos e fotos que são recordações de viagem ou de família, e solicitar a leitura de textos diversos, devido a sua perda de visão. A participante apresenta-se sempre muito solícita a interagir e compartilhar suas experiências de vida.

#### 4 Análise de dados

Na interação que será analisada a seguir, participam Joana, Bete, uma de suas cuidadoras, e Fábio, um dos pesquisadores. Nesse segmento da interação, Bete utiliza uma expressão formulaica (*as cartas não mentem jamais*) que marca a transição das narrativas de Joana durante o encontro.

O tópico que inicia a interação é a festa de aniversário de Joana que havia acontecido recentemente. Em função de ser uma data marcante, na ocasião Joana completou 70 anos, é um acontecimento relevante que é narrado por também ter ocorrido após ao encontro anterior entre Joana e Fábio. Os detalhes da festa são relatados por Joana, que afirma ter vontade de ligar para que pessoas que estiveram na festa para conversar, agradecer o comparecimento e os presentes recebidos, embora ainda não tivesse tempo para realizar essa ação. No momento em que justifica o fato de não ter feito as ligações, Joana comenta que tem compromissos marcados e inicia uma narrativa sobre um momento ocorrido antes de iniciar o encontro com o pesquisador. O excerto 1 apresenta a transição da narrativa de um fato para outra que tem uma história autobiográfica, que tem como elemento de historiabilidade uma passagem da vida de Joana.

#### 4.1 Excerto 1 – A viagem a São Paulo

112	Joana	assim tu tem eu tenho umas ann: obrigações a a cumprir né
113		também vou de manhã no:... fazer:... academia então assim
114		tudo mar mar marcado agora
115	Fábio	[você tem uma agenda de compromissos
116	Joana	uma agenda mais ou menos hoje de tarde antes de tu chegar
117		pedi pra Bete "tu poderia sentar comigo ver um pouco de: o
118		lê algumas cartas da minha família?" e aí fo:i uma coisa
119		bem interessante porque primeiro eu já tava já tinha dito
120		pra Bete "vou ann: botar fora as cartas" aí depois que eu
121		já li algumas eu não vou porque eu vou dar pros meus
122		mãos irmãos ann: ver porque tem coisas da vida que eles
123		participaram e: a gente vai ann
124	Fábio	tá escrito lá nas cartas
125	Joana	tá escrito e a gente as cartas não vendem @@@
126	Fábio	é... @@@
127	Bete	não mentem
128	Joana	não mentem jamais @@@
129	Fábio	é... não mentem é...
130	Joana	entende? então assim... ann: aí eu me lembrei de: coisas
131		que aconteceu eu fui a operada uma vez em em São Paulo
132		e tá história @@@ do meu eu era pra ir pra São Paulo
133		pra um casamento de uma amiga que eu ia ser madrinha
134		também e aí eu fi aco aconteceu que na viagem tive:
135		ann: por eu tava num Fuca e: a gente viajou meu irmão
136		e eu ele é grande eu também @@@ nós no Fuca no ca
137		no tinha um motorista né então ele era de táxi
138		então nós dois tivemos que nos ver ali naquele espaço e
139	Fábio	o táxi era um Fusca?
140	Joana	é... e a e meu irmão e eu grandes né no carro... Fuca é
141		pequeno né?
142	Fábio	os dois atrás no banco...
143	Joana	é mas eu acho que não sei se meu irmão eu fiquei sempre
144		atrás mas de repente eu fiquei com dor e essa dor se
145		transfo se transformou numa: apendicite e então eu fui
146		pro: pra o casamento e fui operada e @@@ e aí minha

147	família foi pro Rio e eu fiquei na casa da minha amiga
148	que tinha casado e tava lá com uma prima minha que ficou
149	comigo depois da operação então todas essas coisas que eu
150	to te contando foram ao a conver a conversa da Bete
151	enquanto eu esperava o teu: a tua vinda aqui entende?

Na linha 116, Joana relata ao pesquisador o que ela fazia antes de sua chegada, demonstrando reconhecer a manutenção de regras pragmáticas essenciais à realização da interação. Esse relato é importante para que ele compreenda o teor da história que emerge na sequência da interação. A participante conta que Bete, sua cuidadora, estava lendo cartas antigas da família para ela, a seu pedido, visto que em decorrência da patologia, Joana está gradativamente perdendo a visão. Uma das coisas que Joana se ocupa em fazer com o auxílio de suas cuidadoras é verificar pertences pessoais a fim de ver se podem ser doados ou se devem ser guardados. Essa era a situação que suscitou a leitura das cartas.

Ao narrar a Fábio o que faziam previamente a sua chegada, Joana, nas linhas 117-8, reporta sua própria fala e já avalia o evento como tendo sido bem interessante (linha 119), visto que já havia decidido desfazer-se das cartas, conforme expressa sua fala reportada na linha 120, mas mudou de ideia após a leitura de Bete. A partir desse momento, Joana considera que as cartas são importantes e que devem ser entregues aos irmãos por contarem experiências vividas por eles (linhas 121-3). Ela relaciona esse tópico (as situações contadas nas cartas) ao clichê “as cartas não mentem jamais”, revelando um tom de brincadeira na interação.

Na linha 125, Joana demonstra dificuldade ao enunciar a palavra *mentem* substituindo-a por *vendem* (linha 125). Nesse momento da interação, a cuidadora Bete, que se mantinha em silêncio durante a conversa, oferece o andamento, corrigindo o enunciado para compor a expressão formulaica “as cartas não mentem”. Joana aceita a mudança e reforça o sentido da expressão, conforme pode ser observado na linha 128 (não mentem jamais). Da mesma forma, Fábio, a quem Joana se dirige no momento, retoma o sentido da expressão.

A expressão então assim (linha 130) conduz o interlocutor ao fechamento da brincadeira, isto é, a mudança de tópico e o encaixamento de outra narrativa. Em seguida, o enunciado aí eu já me lembrei de coisas que aconteceu (linhas 130-1) delimita o início da narrativa autobiográfica, uma das histórias das cartas.

Considerando o modelo de dimensões de Ochs e Capps (2001), o encaixe (*embeddedness*) pode caracterizar a narrativa construída interacionalmente como isolada do contexto discursivo ou encaixada, ou seja, ligada à sequência da interação. No caso da narrativa que Joana constrói a seguir, a história apresenta-se como encaixada ao contexto discursivo, vinculada ao tópico vigente e ao enquadre da interação, que estava organizado em torno das ações com Bete e do sentido da expressão formulaica (*as cartas não mentem jamais*), conforme mostra o segmento 113-128. Cabe destacar, que a expressão formulaica utilizada por Joana, atua como um recurso linguístico que introduz a narrativa que será contada a seguir, além de demonstrar a veracidade dos fatos, pois estavam contidos nas cartas. Ela narra os fatos em tom descontraído.

Nas linhas 131-2, Joana, em tom de riso, antecipa o desfecho da história, sumarizando o conteúdo do que será narrado e oferecendo indícios paralinguísticos a Fábio de que se trata de uma história engraçada. O propósito da viagem realizada a São Paulo é revelado por ela nas linhas 132-3, dizendo que iria ao casamento de uma amiga, da qual seria madrinha. Logo em seguida, Joana realiza o encaixe dos fatos narrados. A linearidade (*linearity*), uma das dimensões das narrativas propostas por Ochs e Capps (2001), abre a possibilidade para que as histórias não obedeçam unicamente a uma forma linear de sucessão de eventos. A narrativa de Joana, como é possível observar, configura-se em uma narrativa não-linear, traçando uma ordem de eventos que possivelmente a participante julgue como mais adequada à interação, estabelecendo uma atitude de colaboração com o interlocutor (CUSTODIO, 2019).

Dando continuidade à narrativa, na linha 135, Joana especifica o tipo de carro no qual ela e o irmão se deslocaram, um Fusca, muito conhecido por ser um carro pequeno. A fim de estabelecer o humor na narrativa e esclarecer por que o tipo de carro fora mencionado, Joana contrapõe à ideia de tamanho do carro categorizando a si mesma e ao irmão como grandes (linha 136) e repetindo o item lexical *Fuca*, enfatizando o paradoxo ao interlocutor.

Ao introduzir o referente *dor*, na linha 144, a participante imediatamente já utiliza da nominalização *essa dor* e o recategoriza como *apendicite* na linha 145, enfatizando que o que iniciou como uma simples dor passou a ser uma patologia grave que necessitava de intervenção cirúrgica e contribuindo, assim, para a progressão temática e desfecho da história. Logo em seguida, Joana utiliza o marcador

discursivo então, para sumarizar os acontecimentos principais da narrativa sobre a viagem e o paradoxo que cria o elemento de humor e então eu fui pro: pra o casamento e fui operada (linhas 145-6), o que é ratificado pelos risos posteriores. Na linha 149, novamente esse mesmo marcador é utilizado para indicar o término na narrativa e a retomada ao tópico inicial da conversa, as situações narradas nas cartas e o que ela estava fazendo antes da chegada do pesquisador.

As estratégias de referenciação utilizadas por Joana como a nominalização e a recategorização (linhas 144-5) consitituem um mecanismo que garante sua fluidez, ocasionando a retomada e a progressão tópica, além de proporcionar a continuidade da dinâmica interacional. É possível observar no excerto 1 que a dimensão da narração (*tellership*) é desenvolvida porque Joana ocupa a posição de narradora ativa, abrindo espaços para turnos de fala do interlocutor, mas colocando-se como narradora principal. Além disso, é importante destacar que, ao longo da interação, a participante transita entre o mundo da história e o aqui e agora interacional, apresentando os cenários, personagens, fazendo comentários e coordenando colaborativamente com interlocutor todos os movimentos discursivos realizados.

Em relação à dimensão do encaixe (*embeddedness*), a narrativa surge diretamente ligada ao tópico da interação, uma das histórias das cartas, ou seja, encaixada ao contexto discursivo, conforme já mencionado. Por fim, em se tratando da historiabilidade (*tellability*), que, de acordo com Ochs e Capps (2001) é a dimensão que revela o propósito de uma narrativa ser contada, a história da viagem justifica-se como uma experiência importante vivida por Joana, uma história significativa para si mesma e sobre sua família. Histórias como essa contada por Joana têm o que Linde (1993) denomina de “reportabilidade estendida” (*extended reportability*), porque podem ser recontadas várias vezes em diferentes contextos. Nesse excerto, a narrativa de Joana reflete um conteúdo da carta diretamente relacionado às histórias de sua vida. Na sequência, veremos que a historiabilidade dessa narrativa se revelará ainda maior com o desfecho da interação.

## 4.2 Excerto 2 – Histórias de família

152	Fábio	e Joana mas é as cartas são endereçadas pra você e pessoas
153		que mandaram pra você
154	Joana	é...
155	Fábio	ou as cartas são endereçadas pra outras pessoas
156	Joana	Não
157	Fábio	e você está com as cartas?
158	Joana	são da família assim da minha mãe pra pra pra os nossos
159		pra os filhos de mim pra minha mãe ann: pro meu pro pro
160		de mim pra minha família por exemplo eu tava em São Paulo
161		e: mandei carta pra pra minha mãe meu pai meus irmãos
162		depois tem uma outra carta que eu tô: no Rio ou em São
163		Paulo e no Rio porque eu fui numa viagem eu fui com
164		com meu irmão outra eu fui com a minha irmã outra eu fui
165		de avião então fu nessas cartas aparecem e aí eu comecei
166		a me lembrar de várias coisas e tem até cartas também de:
167		da da: de nós pra a cidade do interior que era a
168		casa onde meu avô morava
169	Fábio	uhum
170	Joana	ele era: médico lá então tinha: artes ali até coisas que
171		eu considereí assim bah o Carlos vai se lembrar Carlos é o
172		meu irmão né? vai se lembrar duma coisa que atrás da igrá
173		atrás do do hospital tinha uma: ...casinha de: ... de:
174		uma santinha assim
175	Bete	uma gruta
176	Joana	uma gruta que tinha uma santa e eu disse bah não me
177		lembrava disso então assim veio o: uma coisa de reve de
178		reviver aquele passado né
179	Fábio	que estavam lá nas cartas
180	Joana	é @@@ tão lá e eu achei interessante isso né...

Ao ser questionada por Fábio (linhas 152-3; 155), Joana explica qual era a origem e o destino das cartas e contextualizando o interlocutor, por meio do marcador metadiscursivo, por exemplo (linha 160), ela enumera situações vividas pela família que estão contidas nas cartas. De acordo com Koch (2004, p. 107) as exemplificações são inserções

que apresentam a “macrofunção cognitivo-interativa de facilitar a compreensão dos parceiros, pelo acréscimo de elementos necessários para esse fim”, uma estratégia textual-discursiva que atua na construção do sentido. Dessa forma, Joana explica o teor das histórias e os motivos que a levaram a mudar de ideia inicial de desfazer-se das cartas, como pode ser observado na linha 165 (eu comecei a me lembrar de várias coisas), que revela a importância das cartas e a vontade de mostrá-las aos irmãos.

Nas linhas 167-8, ao falar o nome da cidade onde o avô morava, Joana novamente utiliza de uma estratégia formulativa, introduzindo uma explicação que era a casa onde meu avô morava, a fim de prestar o esclarecimento necessário para o entendimento do interlocutor. Após introduzir o referente Carlos (linha 171), a participante interrompe momentaneamente a atividade discursiva, a fim confirmar o entendimento do interlocutor, conforme o turno das linhas 171-2 (Carlos é o meu irmão né?), que é ressaltado pelo uso do marcador discursivo né. Nesse caso, os dados mostram que Joana faz uso de uma estratégia metaformativa, por meio da qual “o locutor opera sobre os enunciados que produz, procedendo a reformulações, refletindo sobre a adequação dos termos empregados [...]” (KOCH, 2004, p. 122). Apesar de a DA ser uma patologia que acarreta o declínio cognitivo progressivo, com ênfase na memória, percebemos que, na relação linguagem – interação – cognição, a atividade metacognitiva se manifesta, enfatizando o caráter sociocognitivo da linguagem, ou seja, do reconhecimento metalinguístico das formas de uso da linguagem no processo de construção conjunta de sentidos que se desenrola na interação.

Em seguida, na linha 173, ao tentar dizer o que havia atrás do hospital onde seu avô trabalhava, Joana apresenta uma dificuldade de acesso lexical e utiliza de outros referentes como uma casinha e de uma santinha para criar o *frame* necessário para o entendimento do interlocutor. prontamente, Bete oferece apoio interacional introduzindo o referente gruta que é repetido por Joana na linha 176.

Nas linhas subsequentes (177-8; 180), Joana explicita que ouvir o conteúdo das cartas e relembra aquelas histórias a fizeram reviver aquele passado, fazendo uma avaliação positiva sobre o que as histórias lidas representam para ela e que ficará ainda mais explícito a seguir.

### 4.3 Excerto 3 – Preâmbulo

181	Fábio	e aí você vai decidir se você vai distribuir entre seus
182		Irmãos
183	Joana	sim isso eu vou: vou mas mostrar pra eles né e: é assim
184		como eu tô ann: quer dizer ta um momento na minha vida
185		interessante né a chegada dos meus 70 anos a o prenu
186		o prenu pre preâmbulo ante ante né momento antes que
187		eu tava pensando muito na minha vida... (SI) ((se emociona
188		e chora))eu pensei muito nos meus amigos ((fala com
189		voz engasgada)) as minhas ann: meus amigos minhas a minha
190		Família
191	Fábio	Uhum
192	Joana	as pessoas que a gente perdeu né... e eu tô muito: tocada
193		por muitas coisas entende assim ó ann: o que passou
194		na minha vida o que eu perdi que é essa:...a possibilidade
195		de ler e escrever que é a coisa mais triste pra mim e: ao
196		mesmo tempo ter esse acervo na casa que é espetacular
197		eu tenho discos maravilhosos os livros eu tô dando
198		porque é triste tu tá numa casa que tu tem livros e tu
199		não pode ler né
200	Fábio	“eu entendo Joana”
201	Joana	então... é: isso é: é uma coisa que toca né mas eu não
202		tô parando eu tô com a minha cabeça a milhão entende?
203		essa: essa possibilidade de ver as as a... as cartas da lê
204		da Bete me ler foi uma coisa bem bem interessante bem
205		engraçada assim também né? foi legal... e: então
206		isso tudo me traz de novo a vida que eu tinha entende

Após o questionamento de Fábio (linha 181), Joana realiza o desfecho dos tópicos mencionados anteriormente, citando a chegada dos seus 70 anos (linha 185) e rotulando-a como um preâmbulo no qual repensa sua vida. Emocionada, a participante enumera o que a levou repensar sua vida: os amigos e a família (linha 189-190), as pessoas que perdeu (linha 192) e, principalmente, a impossibilidade, em decorrência da doença, de ler e escrever (linha 194-5), avaliando como a coisa mais triste (linha 195) e contrapondo tal fato à existência de um acervo que construíra ao longo de sua vida e que agora não pode usufruir.

Cabe destacar que as práticas de leitura e escrita eram parte da identidade de Joana, não apenas como uma pessoa letrada, mas como profissional das Letras. Nesse momento da interação, Fábio oferece apoio interacional e emocional (linha 200) para a participante.

A seguir, na linha 201, Joana, com o dêitico *isso*, refere-se anaforicamente à impossibilidade de ler e recategoriza o fato como uma coisa que toca, e ressalta que, em oposição a esta perda, ela continua ativa: eu não estou parando eu tô com a minha cabeça a milhão (linhas 201-2). O referente coconstruído *cabeça a milhão* encapsula a ideia pré-concebida de quem tem uma doença neurodegenerativa perde suas funções mentais e não pode interagir. Além de encapsular essa situação, o referente exerce a função de uma anáfora indireta remetendo o interlocutor a tais elementos que não estão explícitos no discurso, mas que podem ser inferidos quando Joana ressalta sua condição positiva apesar de seu quadro patológico (CUSTODIO, 2019). Segundo Koch (2001, p. 76), as anáforas indiretas “caracterizam-se pelo fato de não existir no co-texto um antecedente explícito, mas um elemento de relação (por vezes uma estrutura complexa) que se pode denominar âncora [...] e que é decisivo para a interpretação [...]”. Ao negociar o referente com o interlocutor, Joana objetiva desconstituir a imagem de incapacidade devido à doença, apesar das perdas que teve.

Em seguida, nas linhas 203-5, Joana avalia novamente a possibilidade de ouvir as histórias das cartas como positiva, pois a fazem lembrar de situações que já havia esquecido e dar um novo sentido a elas, demonstrando a quinta dimensão da narrativa proposta por Ochs e Capps (2001), a postura moral (*moral stance*). Essa avaliação expressa pela participante exemplifica o argumento proposto por Hydén (2018) de que contar ou ouvir histórias sobre si contribuem, no contexto das doenças neurodegenerativas como a DA, a fim de construir uma imagem positiva de si mesmo e resgatar o sentido de suas vidas, considerando uma atitude colaborativa durante interação.

Conforme mencionado no início da análise, a historiabilidade (*tellability*) da narrativa se revela em sua real importância nesse momento da interação. A história contida na carta não trata apenas do passado da participante, mas de uma das histórias que revelam sua identidade, quem verdadeiramente é, para além da doença que enfrenta, representando suas experiências vividas com amigos e familiares, suas viagens. Essas memórias tornam-se mais explícitas quando Joana anuncia estar repensando sua vida

e, na linha 206, ao utilizar a expressão *isso tudo*, remete anaforicamente às histórias das cartas, às memórias e à possibilidade de ler e de escrever como algo que recorda a vida que tinha antes da patologia.

## **Considerações Finais**

O presente artigo teve como ponto de partida o pressuposto de que o uso da linguagem é, essencialmente, uma atividade conjunta em que os falantes se engajam em ações conjuntas para realizarem ações no mundo (CLARK, 1996). Uma dessas formas de agir pela linguagem no mundo é o ato de contar histórias: uma das ações em que mais nos engajamos cotidianamente. Em nossas análises, procuramos observar como o caráter colaborativo próprio da linguagem se manifesta, mais especificamente, por meio de uma narrativa autobiográfica produzida por uma pessoa que vive com a Doença de Alzheimer. A proposição de Clark (1996) frente à natureza de nossos dados permite realizar algumas considerações sobre a relação entre linguagem e interação no contexto de perdas cognitivas e linguísticas ocasionadas pela DA.

Primeiramente, as análises permitem observar que, apesar dos déficits linguísticos ocasionados pela DA, Joana age colaborativamente, coordenando suas ações ao contexto interacional e aos elementos que compõem a narrativa. A história autobiográfica de Joana demonstra a sua atividade sociointeracional percorrendo tanto o aqui e agora quanto o mundo da narrativa e orientando o interlocutor para tais movimentos.

As cartas lidas pela cuidadora atuam como um importante elemento de ancoragem que favorece a emergência da narrativa autobiográfica e reativa memórias que sinalizam para a exibição e ressignificação da identidade de Joana. As histórias sobre eventos passados são, para pessoas acometidas por doenças neurodegenerativas, mais importantes do que os próprios eventos em si (HYDÉN, 2018). Nas palavras de Joana, isso ressignifica reviver aquele passado. As cartas cumprem a função de um importante recurso semiótico por meio do qual experiências que não foram compartilhadas entre os interlocutores podem ser compartilhadas, reativadas na memória de Joana e ressignificadas. Podemos destacar que o alto grau de letramento de Joana, em razão de sua profissão e de sua inserção nas práticas de leitura e ensino de língua inglesa, constitui um fator que influencia em seu desempenho linguístico-cognitivo ao formular as narrativas e ao interagir com os interlocutores durante a interação.

A narrativa de Joana atende à característica fundamental de coconstrução, considerando uma concepção de narrativa que emerge na cena interacional, conforme o modelo proposto por Ochs e Capps (2001), caracterizado por dimensões mais fluídas, as quais foram possíveis de ser verificadas na sua performance narrativa. Tanto Bete quanto Fábio oferecem o andaimento necessário à Joana, atuando em momentos pontuais demonstrando participação, envolvimento, compartilhando de suas ideias e coconstruindo significados, principalmente quando alguma dificuldade surge, em função de sintomas da DA.

Considerando o papel da referenciação como estratégias de constituição da narrativa como texto oral, podemos observar que Joana retoma referentes, cria novos objetos de discurso conjuntamente, categoriza-os, rotula-os, dentre outros movimentos textuais-discursivos. As estratégias de referenciação utilizadas por ela e o papel colaborativo dos interlocutores são os elementos que constituem colaborativamente os mecanismos textuais e interacionais necessários para constituir e manter a narrativa no contexto interacional dos dados analisados, que se caracteriza por ser uma entrevista aberta semelhante a situações de conversas cotidianas.

Por fim, cabe-nos destacar o papel fundamental que contar histórias desempenha no que tange à questão do uso da linguagem na Doença de Alzheimer. Inserir as pessoas acometidas pela DA nas interações cotidianas, oferecer o apoio interacional necessário e dar espaço para que sejam ouvidas e contem suas histórias de vida contribui para que reconstruam o sentido de si mesmas. Claramente, é na interação com o outro que nos constituímos e ressignificamos experiências. Dessa forma, consideramos que é fundamental propiciar esse espaço de fala com pessoas que compartilhem de suas histórias ou que possam resgatá-las, assim como Bete fez, por meio de cartas ou de outros recursos semióticos.

A narrativa autobiográfica de Joana demonstrou a importância de recordar experiências de vida e de compartilhá-las com outras pessoas. A história fez com que a participante relembresse experiências vividas, repensasse sobre sua vida atual e, mais relevante ainda, se constituísse como uma pessoa que, apesar das dificuldades, não está se entregando à patologia, como é possível perceber em suas próprias palavras: eu tô com a minha cabeça a milhão entende?, mas que está se ressignificando na interação: isso tudo me traz de novo a vida que eu tinha.

## Agradecimentos

Katiuscia Custodio agradece à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado concedida.

## Declaração de Autoria

Este artigo foi desenvolvido por ambos os autores colaborativamente. Especificamente, a autora Katiuscia Custodio realizou a seleção e primeira versão da análise do dado com os principais pontos do referencial teórico. O autor Caio Mira concluiu a análise e escreveu parte do referencial teórico. Os dois autores realizaram a redação e a revisão final do texto.

## Referências

APOTHÉLOZ, D.; REICHLER-BÉGUELIN, M. J. Construction de la référence et stratégies de désignation. In: BERRENDONNER, A.; REICHLER-BEGUELIN, M. J. (org.). *Du syntagme nominal aux objects-de-discours*. Neuchâtel: Université de Neuchâtel Institute de Linguistique, 1995, p. 227-271.-

BASTOS, L. C.; BIAR, L. A. Análise de narrativa e práticas de entendimento da vida social. *D.E.L.T.A.*, São Paulo, n. 31 (especial), p. 97-126, 2015. DOI: <https://doi.org/10.1590/0102-445083363903760077>

CLARK, H. C. *Using Language*. New York: Cambridge University Press, 1996.

CUSTODIO, K. A. “Como é que vou dizer...”: a coconstrução de sentidos nas narrativas orais de uma pessoa com atrofia cortical posterior. 2019. 116f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2019.

DE FINNA, A. Narrative and Identities. In: DE FINNA, A.; GEORGAKOPOULOU, A. (org.). *The Handbook of Narrative Analysis*. West Sussex: Wiley Blackwell, 2015. p. 155-190. DOI: <https://doi.org/10.1017/CBO9781139051255.009>

DE FINNA, A.; GEORGAKOPOULOU, A. *Analyzing Narrative*. Discourse and Sociolinguistic Perspectives. New York; Cambridge: Cambridge University Press, 2012. DOI: <https://doi.org/10.1017/CBO9781139051255>

FLANNERY, M. De “grandes” a pequenas estórias: contribuições de uma nova perspectiva para a análise da narrativa. *Investigações*, Recife, v. 23, n. 2, p.117-142, 2010.

FLANNERY, M. *Uma introdução à análise linguística da narrativa oral: abordagens e modelos*. Campinas: Pontes Editores, 2015.

HYDÉN, L. C. Storytelling in Dementia: Collaboration and Commun Ground. In: HYDÉN, L. C.; ANTELIUS, E (org.). *Living with Demencia*. Palgrave: Londres, 2017. p. 116-134. DOI: [https://doi.org/10.1057/978-1-137-59375-7\\_7](https://doi.org/10.1057/978-1-137-59375-7_7)

HYDÉN, L. C. *Entangled Narratives: Collaborative Storytelling and the Re-Imagining of Dementia*. Oxford: Oxford University Press, 2018. DOI: <https://doi.org/10.1093/oso/9780199391578.001.0001>

JUBRAN, C. C. A. S. O tópico discursivo. In: JUBRAN, C. C. A. S.; KOCH, I.G.V. (org.). *Gramática do português culto falado no Brasil: a construção do texto falado*. Campinas, SP: Editora da Unicamp, 2006. v. 1. p. 89-132.

KIM, J. *Understanding Narrative Inquiry: The Crafting and Analysis of Stories as Research*. London: SAGE, 2016. DOI: <https://doi.org/10.4135/9781071802861>

KOCH, I. G. V. Expressões referenciais definidas e sua função textual. In: DUARTE, L. P. (org.). *Para sempre em mim: homenagem a Ângela Vaz Leão*. Belo Horizonte: CESPUC, 1999. p. 138-150.

KOCH, I. G. V. A referenciação como atividade cognitivo-discursiva e interacional. *Cadernos de Estudos Linguísticos*, Campinas, v. 41, p. 75-89, 2001. DOI: <https://doi.org/10.20396/cel.v41i0.8637002>

KOCH, I. G. V. A. *Desvendando os segredos do texto*. São Paulo: Cortez, 2002.

KOCH, I. G. V. A referenciação como construção sociocognitiva: o caso dos rótulos. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 16, n. 1, p. 201-213, 2008. DOI: <http://dx.doi.org/10.17851/2237-2083.16.1.201-213>.

KOCH, I. G. V. *Introdução à Linguística Textual: trajetória e grandes temas*. São Paulo: Martins Fontes, 2004.

LABOV, W.; WALETZKY, J. Narrative Analysis: Oral Versions of Personal Experience. In: HELM, J. (org.). *Essays on the Verbal and Visual Arts*. Seattle: University of Washington Press, 1967. p. 12-44.

LEITE, M. Q. *et al.* A análise da conversação no Grupo de Trabalho Linguística do Texto e Análise da Conversação da Associação Nacional de Pós-Graduação em Letras e Linguística. In: BENTES, A. C.; LEITE, M. Q. (org.). *Linguística de texto e análise da conversação: panorama das pesquisas no Brasil*. São Paulo: Cortez Editora, 2010. p. 49-87.

LINDE, C. *Life Stories: The Creation of Coherence*. New York: Oxford, 1993.

MARCUSCHI, L. A. O léxico: lista, rede ou cognição social?. In: NEGRI, L.; FOLTRAN, M. J.; OLIVEIRA, R. P. (org.). *Sentido e significação: em torno da obra de Rodolfo Ilari*. São Paulo: Editora Contexto, 2004. p. 263-284

MARCUSCHI, L. A. Anáfora indireta: o barco textual e suas âncoras. *Revista Letras*, Curitiba, n. 56, p. 217-258, 2001. DOI: <https://doi.org/10.5380/rel.v56i0.18415>

MARCUSCHI, L. A. *Análise da conversação*. São Paulo: Ática, 1998.

MIRA, C. COMO É QUE A GENTE DIZ? Uma análise das estratégias textual-interativas na narrativa de uma pessoa com Doença de Alzheimer. *Linguagem em (Dis)curso*, Tubarão, v. 19, p. 419-433, 2019.

MIRA, C.; CUSTODIO, K. A. Contribuições da noção de referenciação para análise da narrativa oral no contexto da atrofia cortical posterior. *Investigações*, Recife, v. 32, p. 1-23, 2019.

MIRA, C.; CARNIN, A. Histórias sobre o convívio com a Doença de Alzheimer: contribuições da noção de referenciação para a análise de narrativas no contexto de interações de um Grupo de Apoio. *Cadernos de Estudos Linguísticos*, Campinas, v. 59, p. 157-174, 2017. DOI: <https://doi.org/10.20396/cel.v59i1.8648426>

MISHLER, E. G. *Research Interviewing: Context and Narrative*. Cambridge: Harvard University Press, 1986.

MOITA LOPES, L. P. A performance narrativa do jogador Ronaldo como fenômeno sexual em um jornal carioca: multimodalidade, posicionamento e iconicidade. *Revista da Anpoll*, Belo Horizonte, v. 2, n. 27, p. 127-157, 2009. DOI: <https://doi.org/10.18309/anp.v2i27.146>

MONDADA, L.; DUBOIS, D. Construção dos objetos de discurso e categorização: uma abordagem dos processos de referenciação. In: CAVALCANTE, M. M.; RODRIGUES, B. B.; CIULLA, A. (org.). *Referenciação*. São Paulo: Contexto, 2003. p. 17-52.

MORATO, E. O caráter sociocognitivo da metaforicidade: contribuições do estudo do tratamento de expressões formulaicas por pessoas com afasia e com Doença de Alzheimer. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 16, n. 1, p. 157-177, 2008. DOI: <https://doi.org/10.17851/2237-2083.16.1.157-177>

MORATO, E. A noção de frame no contexto neurolinguístico: o que ela é capaz de explicar? *Cadernos de Letras da UFF*, Niterói, n. 41, p. 93-113, 2010.

MORATO, E. Das relações entre linguagem, cognição e interação – algumas implicações para o campo da saúde. *Linguagem em (Dis)curso*, Tubarão, v. 16, n. 3, p. 575-690, 2016.

NORRICK, N. R. Narrative contexts. In: NORRICK, N. R. (org.). *Conversational Narrative: Storytelling in Everyday Talk*. Philadelphia: John Benjamins Publishing Company, 2007. p. 105-134.

OCHS, E.; CAPPS, L. *Living Narrative: Creating Lives in Everyday Storytelling*. Cambridge: Harvard University Press, 2001.

ROSCH, E. Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology, General*, v. 104, n. 3, p. 192-233, 1975. DOI: <https://doi.org/10.1037/0096-3445.104.3.192>

SANTOS, E. N. *et al.* Educação em saúde na comunidade: dialogando sobre o envelhecimento e a Doença de Alzheimer. *Brazilian Journal of Surgery and Clinical Research – BJSCR*, Cianorte, PR, v. 27, n. 2, p. 32-36, 2019.

SHENK, D. There Was an Old Woman: Maintenance of Identity by People with Alzheimer's Dementia. In: DAVIS, B. H. (org.). *Alzheimer Talk, Text and Context: Enhancing Communication*. New York: Palgrave Macmillan, 2005. p. 3-17. DOI: [https://doi.org/10.1057/9780230502024\\_1](https://doi.org/10.1057/9780230502024_1)

VIEIRA, G. D. *et al.* A deposição de peptídeo beta-amilóide e as alterações vasculares presentes na doença de Alzheimer. *Journal of Health Biology Science*, Fortaleza, v. 2, n. 4, p. 218-223, 2014. DOI: <http://dx.doi.org/10.12662/2317-3076jhbs.v2i4.98.p218-223.2014>

VYGOTSKI, L. S. *Pensamento e linguagem*. São Paulo: Martins Fontes, 2005.

WOOD, D.; BRUNER, J.; ROSS, G. The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, [S.l.], n. 17, p. 89-100, 1976. DOI: <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

ZIDAN, M. *et al.* Alterações motoras e funcionais em diferentes estágios da Doença de Alzheimer. *Revista de Psiquiatria Clínica*, São Paulo, v. 39, n. 5, p. 161-165, 2012. DOI: <http://dx.doi.org/10.1590/S0101-60832012000500003>

## ANEXO

### Convenções de transcrição

Ocorrências	Sinais
Incompreensão de palavras ou segmentos	(SI)
Hipótese do que se ouviu	(hipótese)
Prolongamento de vogal e consoante	: (podendo aumentar de acordo com a duração)
Interrogação	?
Qualquer pausa	...
Pausa prolongada (medida em segundos)	(4.2)
Sobreposição de vozes	[apontando o local onde ocorre a superposição]
Citações literais, ou leituras de textos	“ ”
Risos	@@@
Tom mais baixo	°tom mais baixo°
Entonação enfática	MAIÚSCULA
Truncamento brusco	/
Silabação	- - -
Informação omitida por sigilo	XXX
Comentários do analista e designações gestuais	((minúscula))
Itálico	<i>palavras de língua estrangeira</i>
Indicação e continuidade de gestos significativos, com a descrição dos mesmos	<i>*início e fim do gesto*</i>

Fonte: Marcuschi (1998); Custodio (2019).





## **Apagamento do rótico em coda no Português Santomense (PST): uma análise sociolinguística**

### ***Rhotic deletion in coda in Santomean Portuguese (PST): an sociolinguistics analysis***

Nancy Mendes Torres Vieira

Universidade de São Paulo (USP), São Paulo, São Paulo / Brasil

nancy.vieira@usp.br

<http://orcid.org/0000-0002-7291-9759>

Amanda Macedo Balduino

Universidade de São Paulo (USP), São Paulo, São Paulo / Brasil

amanda.m\_b@hotmail.com

<http://orcid.org/0000-0002-1062-973X>

**Resumo:** O objetivo deste artigo é analisar o apagamento do rótico em coda no português de São Tomé (PST), uma variedade do português falada em São Tomé e Príncipe. Para análise desse processo, nos pautamos na Sociolinguística Variacionista e a análise quantitativa foi realizada nos softwares RStudio e Rbrul. Consideramos variáveis sociais (Sexo, Faixa Etária e Grau de Escolaridade) e linguísticas (Classe Gramatical, Tonicidade, Posição do Segmento e Contexto Fonológico Precedente) que poderiam favorecer ou desfavorecer o fenômeno. Examinando as 1523 ocorrências de rótico em coda, verificamos o apagamento do rótico em 56,53% dos dados. A variável mais significativa foi a Posição do Segmento com o maior índice de apagamento constatado quando o rótico estava em posição final (P.R. .71). A amostra foi, portanto, dividida em duas amostras menores analisadas separadamente: 712 ocorrências do rótico em coda medial e 811 em coda final. Na análise do fenômeno em coda medial, o índice de apagamento foi de 32,6% e os maiores índices de apagamento foram verificados entre os informantes do ensino fundamental (P.R. .62) e do sexo feminino (P.R. .59). Já em coda final, o índice de apagamentos foi de 77,5% e os falantes com ensino fundamental (P.R. .75) e mais jovens (P.R. .76) foram mais propensos ao apagamento, ademais, a classe verbal favoreceu o apagamento (P.R. .62). Os resultados são cojetados com trabalhos

prévios (BOUCHARD, 2017; BRANDÃO, 2018; BRANDÃO; DE PAULA, 2018; BRANDÃO *et al.*, 2017) e com resultados de estudos sobre variedades do PB e do PE.

**Palavras-chave:** português santomense; apagamento do rótico em coda; variáveis sociais e linguísticas; contato linguístico.

**Abstract:** The goal of this paper is to analyze rhotic deletion in coda in Santomean Portuguese (PST), a Portuguese variety spoken in Sao Tome and Principe. This process was analyzed based on Variationist Sociolinguistics. Quantitative analyses have been carried out by softwares RStudio and Rbrul. Thus, we consider social (sex, age and education level) and linguistic (grammatical class, syllable stress, position in morphological word and quality of preceding vowels) variables that might favor or not the phenomenon. By examining 1523 rhotic occurrences in coda, we verified r-deletion in 56,53% of the data. The most significant variable is Segmental Position with the highest rate of r-deletion in word-final syllables (P. R. 71). Then, the sample was divided into two samples separately analyzed: 712 occurrences in word-medial coda and 811 occurrences in word-final coda. In word-medial coda, 32,6% of rhotics are deleted. The highest rates of loss are associated to elementary school (P.R. .62) and women (P.R. .59). In word-final coda, 77.5% of rhotics are deleted, Speakers with elementary school (P.R. .75) and younger speakers (P.R. .76) are more likely to delete the rhotic in coda. Considering Grammatical Class, r-deletion is favored in verbs (P.R. .62). These outcomes are compared with previous studies (BOUCHARD, 2017; BRANDÃO, 2018; BRANDÃO; DE PAULA, 2018; BRANDÃO *et al.*, 2017) and with studies on PB and PE.

**Keywords:** Santomean Portuguese; deletion of rhotic in coda; social and linguistic variables; linguistic contact.

Recebido em 18 de novembro de 2020

Aceito em 18 de janeiro de 2021

## 1 Introdução

Neste artigo, examinamos o apagamento do rótico em coda silábica no português de São Tomé (PST). Mediante a Sociolinguística Variacionista (LABOV, 1972), visamos, assim, analisar o fenômeno de apagamento segmental em coda, amplamente reportado para variedades como o português brasileiro (PB) e o português europeu (PE), em uma variedade da língua portuguesa empregada em um contexto multilíngue. O presente estudo surgiu a partir do trabalho de Vieira e Balduino (2020), que discutiu o preenchimento da coda no português santomense (PST),

avaliando a possibilidade de apagamento nessa posição como resultado do contato dessa variedade do português com o santome, uma língua crioula desenvolvida e falada em São Tomé (ARAUJO; HAGEMEIJER, 2013; BANDEIRA, 2017; FERRAZ, 1979).

O estudo de Vieira e Balduino (2020) contabilizou os seguintes índices de apagamento: 61,53% (355/577) para o rótico, 24,21% (77/318) para a lateral e 5,27% (29/550) para a sibilante, concluindo, entre outras coisas, que uma possível convergência linguística entre o santome e o PST não é suficiente para explicar os diferentes percentuais de apagamento em coda na variedade de português falada em São Tomé (VIEIRA; BALDUINO, 2020, p. 29). Para as autoras, fatos relacionados ao contato são importantes para análise de fenômenos fonológicos no PST, porém devem ser cotejados em conjunto com aspectos estruturais das línguas, bem como com o espaço sociolinguístico onde a língua é falada, abarcando, portanto, fatores sociais relacionados ou não ao contato linguístico. O estudo de Vieira e Balduino (2020), embora ofereça uma descrição pioneira do processo de apagamento nas codas /r, l, S/, indicando ser este um fenômeno de implementação ampla no PST, não abarca a questão mediante modelagem estatística dos dados nem sob a perspectiva teórica da Sociolinguística Variacionista. Dessa forma, visando contribuir com a análise do apagamento da coda a partir de uma nova abordagem metodológica e focando no rótico como segmento-alvo, delimitamos como objetivos desse estudo: (i) analisar o apagamento do rótico no PST – segmento que apresentou o maior índice de apagamento no trabalho de Vieira e Balduino (2020) – a partir de uma amostra maior e à luz da Sociolinguística Variacionista, verificando os fatores sociais e linguísticos que influenciam esse fenômeno; e (ii) cotejar nossos resultados com trabalhos prévios que, utilizando a mesma metodologia, indicam ser o apagamento em coda um fenômeno recorrente no PST (BOUCHARD, 2017; BRANDÃO, 2018; BRANDÃO; DE PAULA, 2018; BRANDÃO *et al.*, 2017) e com dados de estudos sobre o mesmo processo em variedades do PB e do PE.

O artigo está organizado da seguinte forma: na seção 2, discutimos, brevemente, a consolidação do português como língua materna da população santomense; na seção 3, apresentamos o fenômeno examinado; na seção 4, definimos os procedimentos de exame adotados; na seção 5, expomos os resultados da análise dos dados e, por fim, tecemos as considerações finais, na seção 6.

## **2 Contexto linguístico de São Tomé**

A República de São Tomé e Príncipe (STP) é um país da Costa Oeste Africana localizado no Golfo da Guiné. Desde 1975, o português é a única língua oficial do arquipélago, entretanto, línguas autóctones como o santome (ISO 639-3: CRI), o angolar (ISO 639-3: AOA) e o lung'le (ISO 639-3: PRI), além do kabuverdianu (ISO 639-3: KEA), uma língua transplantada de Cabo Verde a STP, também são faladas pelos seus habitantes (AGOSTINHO, 2015; ARAUJO, 2020; BANDEIRA, 2017; HAGEMEIJER, 2018).

De acordo com Gonçalves (2010), Christofolletti (2013), Balduino (2018), entre outros autores, a convivência da língua portuguesa com as línguas autóctones, em STP, cria um contexto ecolinguístico complexo e desigual em que questões de standardização linguística interferem no emprego linguístico. A língua portuguesa, em geral, exerce efeitos glotocidas sobre as demais línguas do arquipélago, na medida em que é adquirida como língua materna (L1) em detrimento da aquisição das demais línguas faladas em STP.

Distintamente de outros países africanos em que o português configura uma das línguas oficiais, porém não é a língua majoritariamente falada pela população, como é o caso de Angola, Moçambique, Guiné-Bissau e Cabo Verde, em STP, o português é a língua materna de, cerca de, 98% da população de acordo com o último censo publicado (INE, 2012). A difusão do português enquanto língua materna deve-se, sobretudo, à convergência de fatores como a urbanização, democratização de ensino, dado que a alfabetização era e ainda é realizada em português, bem como às próprias dinâmicas sociolinguísticas do país, as quais estão circunscritas numa conjuntura multilíngue (ARAJO, 2020; GONÇALVES; HAGEMEIJER, 2015).

Conforme Araujo (2020), as elites locais desempenharam importante papel para a eleição do português como língua oficial bem como para sua difusão enquanto L1. Assim como nos demais territórios africanos que passavam por um processo de descolonização em relação a Portugal, o português foi eleito como língua oficial de STP não somente por corresponder a uma língua com tradição escrita, mas, também, como uma pretensa tentativa de diminuir conflitos étnicos locais. Para Araujo (2020), essa “neutralização política”, demarcada pela escolha do português como língua oficial, não pode deixar de ser avaliada em cotejo

com o fato de que esta já era empregada como língua de um grupo social, a elite santomense. Bouchard (2017) indica que a elite e seus descendentes, ainda que marginalizados pela elite portuguesa pré-independência, já tinham acesso à educação em Portugal e empregavam o português. Assim, no período pós-independência, o português foi preservado por tal grupo social, sendo promovido por instituições educacionais e midiáticas, ao passo que não foram criadas políticas linguísticas dedicadas à transmissão e preservação das línguas crioulas do arquipélago (ARAÚJO, 2020), fatos que contribuíram para sua marginalização.

Ainda que o português configure a língua mais usada e transmitida em STP, é possível observar divergências entre as variedades vernáculas de português do arquipélago e a norma de prestígio adotada por seus falantes e disseminada pela escolarização (BALDUINO, 2018; BOUCHARD, 2017). A variedade de português que goza de maior status social em STP, língua-alvo da população de STP, ainda é a variedade europeia (BALDUINO, 2018).

Em STP, circulam diferentes variedades vernáculas da língua portuguesa. Além de variedades urbanas, como o português santomense (PST) e o português principense (PP) (BALDUINO, 2019; SANTIAGO, 2019; SANTIAGO; AGOSTINHO, 2020), há registros de variedades como o português de Almoxarife (PA), empregado na costa da ilha de São Tomé, e do português dos Tongas (PT), falado na comunidade da roça Monte Café (BAXTER, 2018; FIGUEIREDO 2010; ROUGÉ, 1992). Ademais, outras variedades não registradas, como o português angular, falado na região dos Angolares, e o português falado pelos descendentes cabo-verdianos, também são observadas.

Essas variedades possuem estruturas e processos linguísticos singulares que as diferenciam da variedade europeia e as consolidam enquanto variedades da língua portuguesa. Tais processos linguísticos têm respaldo em fatores sociolinguísticos, entre os quais pode estar a influência das línguas autóctones. Isto posto, neste artigo, propomos uma análise do processo de apagamento do rótico em coda no PST, variedade vernácula da cidade de São Tomé. Para tanto, considerando que as dinâmicas sociais do país foram fundamentais para ascensão do português como L1 da população santomense, investigaremos possíveis variáveis linguísticas e sociais, que consolidam o espaço sociolinguístico no qual o PST está inserido, e podem estar correlacionadas com esse fenômeno.

### 3 Apagamento do rótico em coda silábica

Neste artigo, assume-se que a sílaba apresenta uma estrutura hierárquica, sendo organizada a partir de constituintes internos como: o *onset* (O) e a rima (R), subdividida em núcleo (N) e coda (C) (SELKIRK, 1982). Tendo como foco de análise a coda, observamos que neste constituinte são licenciadas consoantes como sibilantes /S/, lateral /l/, nasal /N/ e róticos no PST, (BALDUINO, 2019; BALDUINO; VIEIRA, 2020; VIEIRA; BALDUINO, 2020), assim como observado no português brasileiro (PB) e no europeu (PE) (CÂMARA JR., 1970; MATEUS; D'ANDRADE, 2000).

De acordo com diferentes trabalhos que analisam fenômenos cuja sílaba é o domínio, no português, a coda é propícia à variação e a apagamentos, os quais priorizam, especialmente no PB, a estrutura CV (ABAURRE; SÂNDALO, 2003; BISOL, 1999; BRANDÃO; MOTA; CUNHA, 2003; CALLOU; LEITE; MORAES, 1994; CALLOU; LEITE; MORAES, 2002; CÂMARA JR., 1970; MATEUS; RODRIGUES, 2003; QUEDNAU, 1993).

Em geral, o apagamento de codas pode ocorrer tanto em PB quanto em PE, sendo, entretanto, mais recorrente em PB (BRANDÃO; MOTA; CUNHA, 2003; CALLOU; SERRA, 2012; OLIVEIRA, 2018; RODRIGUES, 2012; SERRA; CALLOU, 2015), o que é fundamentado pelo fato de o PE ser uma modalidade de reforço consonântico, enquanto o PB tende a reforçar seu quadro vocálico (BRANDÃO; MOTA; CUNHA, 2003, p. 14).

O rótico apresenta não somente grande variabilidade, mas possui índices expressivos de apagamento em diversas variedades do PB (ALVES, 2015; BRANDÃO; MOTA; CUNHA, 2003; CALLOU; LEITE; MORAES, 1994; CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; MONGUILLHOT, 1997; OLIVEIRA, 2018; OUSHIRO; MENDES, 2014; SERRA; CALLOU, 2015), do PST (BOUCHARD, 2017; BRANDÃO, 2018; BRANDÃO; DE PAULA, 2018; BRANDÃO *et al.* 2017; VIEIRA; BALDUINO, 2020) e, ainda, em outras variedades africanas da língua portuguesa, como o português de Moçambique (BRANDÃO, 2018)

Agostinho (2016), Bouchard (2017), Brandão *et al.* (2017), Brandão e De Paula (2018) e Brandão (2018) chamam atenção para o fato de haver, no PST e no PP, alto grau de instabilidade na produção

dos róticos, sendo esse um segmento alvo de processos fonológicos como o apagamento que, por sua vez, é influenciado por variáveis linguísticas e sociais. De forma semelhante a variedades como o PB e o PE, o PST também apresenta diferentes realizações do rótico. Entretanto, distintamente de tais variedades, o rótico pode perder valor distintivo entre vogais, como indicado pelo par mínimo caro ['ka.rɔ] ~ ['ka.ʁɔ], e carro ['ka.rɔ] ~ ['ka.ʁɔ], em que o R-forte e o R-fraco alternam, inclusive, em diferentes produções de um mesmo falante. Nos demais contextos de ocorrência, essa alternância é também sustentada: como em *onset* no início de palavra (rato ['ra.tu] ~ ['ʁa.tu]), no segundo elemento de um *onset* complexo (prato ['pra.tu] ~ ['pʁa.tu]), e nas codas mediais e finais (carta ['kar.tɐ] ~ ['kaʁ.tɐ] e mar ['mar] ~ ['maʁ], respectivamente) (BALDUINO, em preparação).

Para Bouchard (2017), o uso do R-forte no PST, seja como C1 ou C2, seja em contexto intervocálico ou não, é influenciado pela variável faixa etária: o uso de [ʁ] pelas gerações mais jovens de São Tomé seria, assim, um indicativo de mudança linguística em curso, em que [r] está sendo substituído por [ʁ]. Essa hipótese é reforçada por Agostinho, Soares e Mendes (2020)<sup>1</sup> que discutem a possibilidade de fusão de quase-fonemas, isto é, de segmentos que apresentam relações fonológicas intermediárias, variando no grau de previsibilidade e contraste. Conforme o experimento conduzido pelos autores, há evidências que sugerem fusão e perda de contraste dos róticos no Português do Príncipe, variedade também falada em STP. Isso ocorreria em decorrência do seu status quase-fonêmico, assim como por sua baixa carga funcional no sistema e pelo contato linguístico com as línguas crioulas faladas na região (AGOSTINHO; SOARES; MENDES, 2020).

As propostas de Bouchard (2017) e de Agostinho, Soares e Mendes (2020) demonstram que o estabelecimento do fonema rótico no inventário fonológico do PST não é uma tarefa simples. Neste estudo, analisaremos o apagamento do rótico em coda concebendo a existência de

---

<sup>1</sup> “Merging of Quasi-Phonemes in Contact Situations: Evidence from Rhotics in Principense Portuguese”. Trabalho apresentado por Ana Livia Agostinho, Eduardo Soares, e Maiara Mendes no Annual Meeting on Phonology, em 2020 Santa Cruz: University of California, 2020.

duas realizações concorrentes no PST: [r] e [ʁ],<sup>2</sup> os quais serão referidos como rótico, posto que o estabelecimento de uma forma fonológica para tais fones corresponde a uma discussão que foge ao escopo desse trabalho. Não adotaremos, assim, a representação /R/ para róticos em coda, uma vez que esta pressupõe a neutralização de um fonema numa determinada posição, fato que não condiz ao comportamento do rótico no PST.<sup>3</sup>

#### 4 Metodologia

O estudo baseia-se nos pressupostos da Teoria da Variação e Mudança, também denominada Sociolinguística Variacionista, desenvolvida por Weinreich, Labov e Herzog (1968) e Labov (1972, 1994, 2001). A Sociolinguística é uma teoria linguística social que tem como objeto de estudo a língua, como é usada na vida cotidiana de uma comunidade, considerando os fatores sociais que se correlacionam com ela (LABOV, 1972).

A análise do apagamento do rótico no PST foi realizada com base em dados de um *corpus* coletado em São Tomé, o qual é constituído por entrevistas de fala espontânea gravadas em 2016<sup>4</sup> e 2019<sup>5</sup>. Durante a coleta dos dados, seguimos os seguintes procedimentos éticos, exigidos na realização de pesquisas com seres humanos: (i) os participantes da pesquisa eram maiores de 18 anos e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE), contendo informações gerais e esclarecimentos sobre a pesquisa, no qual concordaram, de

---

<sup>2</sup> É preciso ressaltar, no entanto, que os róticos, em coda, apresentam grande variabilidade no PST, sendo observáveis, ainda, realizações como [r], [h], [ʁ], ou mesmo uma produção vocalizada [w] ou lateralizada [ɬ].

<sup>3</sup> A neutralização é um processo natural pelo qual dois ou mais fonemas, opostos em um determinado contexto, deixam de apresentar oposição em outro ambiente. No PB, o rótico é neutralizado em coda, posição em que pode ser produzido de diferentes formas: [r], [ʁ], [h], [ɬ], [x], entre outros. Já em meio de palavra, nesta variedade, o r-forte e o r-fraco constituem fonemas distintos e geram oposição de pares mínimos: caro ['ka.ɾu] e carro ['ka.ɾu] ~ ['ka.ʁu]. No PST, como discutido ao longo da seção 3, isso não ocorre, visto que a oposição entre os róticos não é verificada: **caro** ['ka.ɾu] ~ ['ka.ʁu], e carro ['ka.ɾu] ~ ['ka.ʁu].

<sup>4</sup> Trabalho de campo de Ana Livia Agostinho e Amanda Macedo Balduino, não publicado.

<sup>5</sup> Trabalho de campo de Amanda Macedo Balduino, não publicado.

forma consciente, em participar do estudo; (ii) antes da gravação das entrevistas, os participantes da pesquisa foram informados de que seriam gravados e, durante as entrevistas, demonstraram estar cientes de tal fato, manifestando seu consentimento em participar do estudo; (iii) os nomes dos entrevistados foram omitidos na divulgação de dados do estudo.

Nesta análise, trabalhamos com doze entrevistas, cada uma com duração de, aproximadamente, 60 minutos, porém descartamos os 15 primeiros minutos, a fim de examinarmos uma produção mais próxima à fala vernacular dos informantes. Todos os informantes eram naturais da ilha de São Tomé, tendo o português como língua materna.

O *corpus* é composto por ocorrências de palavras com rótico em coda, sendo que, para cada entrevista, não foram consideradas mais do que seis realizações para uma mesma palavra, ou mais do que três realizações de um mesmo verbo no infinitivo. Esse método, por um lado, registra as possíveis realizações de um mesmo item lexical, já que um mesmo falante pode pronunciar uma mesma palavra de formas distintas. Por outro lado, ao restringirmos o número de ocorrências de cada item lexical, reduzimos a possibilidade de que o *corpus* fique enviesado, posto que tal método limita a influência de palavras de alta frequência que podem não representar o fenômeno de maneira efetiva na variedade em questão.<sup>6</sup>

Inicialmente foram contabilizadas 1564 ocorrências, então excluímos os casos de ressilabificação, em que o rótico passa a ocupar a posição de *onset* na sílaba seguinte, e esse número foi reduzido a 1523 ocorrências constituídas por 534 palavras.

Essas ocorrências foram submetidas a uma análise acústica, de modo a verificarmos o apagamento da coda nas palavras analisadas. Para tanto, utilizamos o *Praat*,<sup>7</sup> que possibilitou verificar a realização ou ausência do segmento na produção das palavras pelos informantes. O apagamento foi considerado quando não havia qualquer forma espectral

---

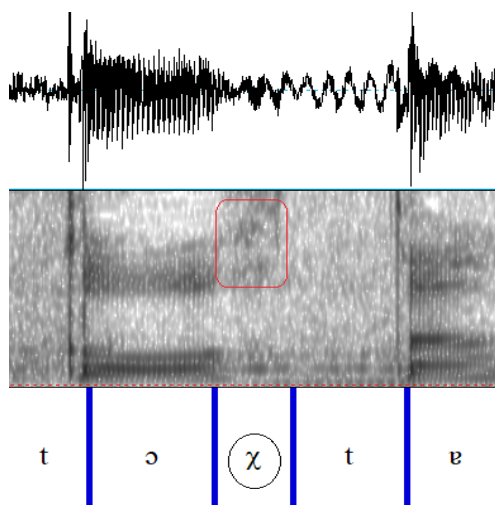
<sup>6</sup> Verbos no infinitivo, por exemplo, correspondem a itens em que a recorrência de apagamentos do rótico poderia ser numerosa, já que a estrutura dessa classe gramatical constitui contexto propício para o fenômeno. Daí a relevância de também serem observados itens não-verbais e limitarmos a ocorrência de verbos.

<sup>7</sup> *Praat* é o software utilizado para análise e síntese de fala, pelo qual é possível acessar informações acústicas do segmento, como a duração, o formato de onda, o espectrograma e, no caso deste estudo, sua ausência (BOERSMA; WEENINK, 2020).

que pudesse ser atribuída ao segmento alvo, cuja não realização já havia sido estabelecida, previamente, mediante uma análise de oitiva. Operamos, assim, com um exame espectral categórico, o qual não leva em consideração formas fonéticas gradientes (MENESES, 2012).

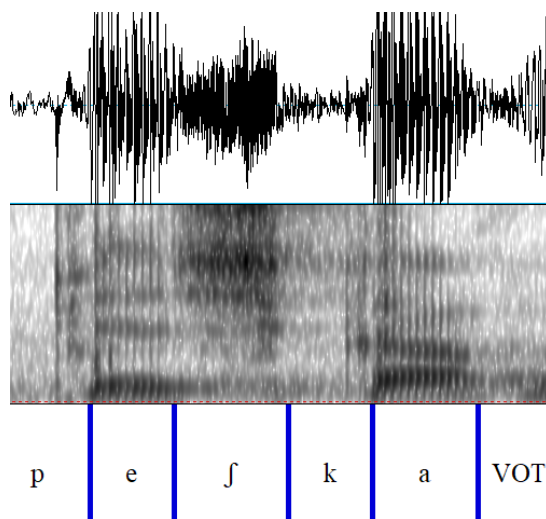
Nas Figuras 1 e 2, observamos exemplos de espectrogramas que ilustram a realização ou não de um segmento consonantal em coda. Enquanto na Figura 1 é possível verificarmos a realização da fricativa [χ], em coda medial, na Figura 2, o item lexical *pescar* é produzido sem o rótico, sendo a vogal [a] seguida pelo VOT da consoante oclusiva seguinte à palavra, porém não sendo identificado o rótico. A Figura 2 evidencia os casos nos quais o apagamento foi considerado para a análise.

FIGURA 1 – Espectrograma do item lexical *torta* com realização do rótico em coda medial



Fonte: elaboração própria

FIGURA 2 – Espectrograma do item lexical *pescar* sem realização do rótico em coda final



Fonte: elaboração própria

Feitos os exames espectrais que identificaram os casos de apagamento, partimos para a análise dos dados, nomeadamente, uma análise de regra variável<sup>8</sup> – em que a regra é o apagamento do rótico em posição de coda – e o tratamento quantitativo dos dados foi realizado no software RStudio.<sup>9</sup> A modelagem estatística dos dados foi feita pelo

<sup>8</sup> Análise de regra variável é uma forma de dar conta da variação estruturada, governada por regras, no uso da língua, que consiste num tipo de análise multivariada com o objetivo de separar, quantificar e testar a significância dos efeitos de fatores contextuais em uma variável linguística, aqui o rótico em coda que pode ou não ser apagado (GUY; ZILLES, 2007).

<sup>9</sup> O RStudio é uma interface funcional e mais amigável para o R que, por sua vez, consiste não apenas num software, mas também numa linguagem de programação voltada para a análise de dados, que pode ser utilizada para realizar computações estatísticas e gráficas, compilar *corpora*, produzir listas de frequências, entre outras diversas tarefas. O RStudio é o principal ambiente de desenvolvimento integrado para R que disponibiliza ferramentas adicionais diretamente na interface gráfica, como a visualização dos *scripts* abertos recentemente, o histórico de linhas de comando executadas, a lista de pacotes instalados, entre outras (OUSHIRO, 2014, p.134-136).

Rbrul,<sup>10</sup> um software que roda, a partir de um código, no RStudio. O modelo estatístico empregado foi o de regressão logística, no qual se verifica o efeito simultâneo de múltiplas variáveis independentes ou previsoras na aplicação da regra variável, calculando o chamado *peso relativo* para cada um dos fatores componentes das variáveis independentes.

O efeito dos fatores das variáveis independentes são valores probabilísticos que estão numa escala que vai de 0 a 1 em que 0 representa uma chance nula de algo ocorrer (0%), 1 representa certeza de que vai ocorrer (100%) e 0,5 é o ponto neutro, já que equivale a 50% de certeza de que algo aconteça. Dessa forma, um fator como “verbo” (da variável classe gramatical) com um *peso relativo* igual a 0,75 significaria que há uma chance de 75% de que a regra variável seja aplicada quando a palavra analisada é um verbo. Sintetizando, um peso relativo (P.R.) com um valor superior a 0,5 indica que o fator favorece a aplicação da regra, ao passo que um valor menor que 0,5 aponta que o fator a desfavorece. Já um valor igual a 0,5, por sua vez, indica que o fator não tem efeito na aplicação da regra.

O software também calcula o *input*, que representa o nível geral de uso de determinado valor da variável dependente, nesse caso, o apagamento de um segmento consonantal em coda. Tal valor deve se aproximar da frequência de aplicação da regra, calculada para a amostra total e quando isso não ocorre, é um indicativo de que a distribuição dos dados, através dos fatores analisados, não é equilibrada. Os P.R. são calculados em relação a esse nível geral. (GUY; ZILLES, 2007; OUSHIRO, 2017). Além disso, o software fornece o nível de significância do modelo, também chamado *valor-p* que é a probabilidade de se observar determinado resultado, por acaso, em caso de a hipótese nula ser verdadeira. O que nos remete a dois importantes conceitos da estatística inferencial: *hipótese nula* ( $H_0$ ) e *hipótese alternativa* ( $H_1$ ). Esta última é a hipótese que está sendo testada, como, por exemplo, a afirmação de

---

<sup>10</sup> O Rbrul é um software gratuito e idealizado por Daniel Ezra Johnson, software que realiza não apenas todas as funções que as versões Goldvarb e Varbrul realizam (regressão múltipla, tabulação cruzada, *step up/step down*), mas também outras funções como: rodar variáveis contínuas como variáveis independentes; variáveis contínuas como variáveis dependentes e, ainda, da conta de modelos mistos. Além disso, o Rbrul estabelece uma interface com as capacidades gráficas do R (GOMES, 2012). O programa e o manual de uso estão disponíveis na página <http://www.danielezrajohnson.com/rbrul.html>.

que há uma relação entre duas variáveis, enquanto a  $H_0$ , normalmente, é formulada como a negação da  $H_1$ , afirmando que não há relação entre as variáveis e que a distribuição dos dados observada resulta de uma flutuação aleatória e/ou erro de amostragem.

Se o valor  $p$  for muito baixo, rejeita-se a hipótese nula e a distribuição dos dados observada é considerada estatisticamente significativa, o que significa que a relação ou o efeito que está sendo testado são verdadeiros, já que, a probabilidade de ocorrerem por acaso é muito pequena, assim, quanto menor o valor de  $p$  mais significativo é o modelo. A comunidade científica costuma usar o limite de 0,05 (ou 5%) para considerar algo como muito pouco provável para acontecer ao acaso (GUY; ZILLES, 2007; OUSHIRO, 2017).

As variáveis sociais consideradas foram: Sexo, Faixa Etária e Grau de Escolaridade dos informantes. Trabalhamos com um mesmo número de homens e mulheres (seis homens e seis mulheres) a fim de verificar se há diferenças relacionadas ao Sexo, uma vez que estudos têm mostrado que a variação linguística se correlaciona com esta variável (CHESHIRE, 2004). Quanto à Faixa Etária, os informantes foram organizados em três níveis: o primeiro até 20 anos, o segundo entre 21 e 40 anos e o terceiro, acima de 40 anos. O Grau de Escolaridade é um indicador de *status* socioeconômico que, segundo estudos sociolinguísticos, está associado à utilização de traços diferentes da língua (LABOV, 1972) e está dividido da seguinte forma: ensino fundamental (da 4ª à 9ª classe), ensino médio (10ª à 12ª classe) e ensino superior (graduação e pós-graduação). A Tabela 1 mostra uma estratificação dos informantes conforme a faixa etária, o sexo e o grau de escolaridade.

QUADRO 1 – Estratificação dos informantes

Faixa Etária	Sexo		Grau de Escolaridade		
	Feminino	Masculino	Ensino Fundamental	Ensino Médio	Ensino Superior
18-20	2	2	0	4	0
21-40	3	2	2	1	2
40-52	1	2	1	0	2
Totais	6	6	3	5	4

Fonte: elaboração própria

No que se refere às variáveis linguísticas, analisamos: a Classe Gramatical da palavra (verbo e não-verbo), a Tonicidade da sílaba e a Posição do Segmento na palavra (final e medial), além do Contexto Fonológico Precedente quanto ao traço de Altura [-alto] (a, ε, o) e [+alto] (i, e, o, u) e os traços [-anterior] (a, o, u) e [+anterior] (i, e, ε).

Do total de 1523 ocorrências analisadas, 779 estão contidas em itens verbais e 744 em itens não verbais. Quanto à posição do segmento, trabalhamos com 811 ocorrências em coda final e 712 em coda medial. Considerando que estudos que analisam o apagamento do rótico no PB têm apresentando índices quase categóricos de apagamento desse segmento em coda final de itens verbais (CALLOU, 1987 apud OLIVEIRA, 2018, p. 26); CALLOU; LEITE; MORAES, 2002; CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; OLIVEIRA, 2018; OUSHIRO; MENDES, 2014), realizamos o cruzamento das variáveis: Posição do Segmento e Classe Gramatical da palavra, a fim de verificar a diferença entre os índices de apagamento em itens verbais e não-verbais em coda medial e final.

Após a análise inicial, a amostra, de 1523 ocorrências, foi dividida em duas amostras menores: 712 ocorrências do rótico em posição medial (coda interna) e 811 em posição final (coda externa). Feito isso, analisamos as duas amostras, separadamente, verificando a relevância das mesmas variáveis sociais e linguísticas, exceto a variável Posição do Segmento na palavra (final e medial), que deixa de ser pertinente – dado que na primeira amostra há apenas dados do rótico em coda medial e na segunda de coda final.

## 5 Análise e Resultados

Nesta seção, analisamos o apagamento do rótico, verificando as variáveis sociais e linguísticas que influenciam o fenômeno sem deixar de considerar questões relacionadas ao contato linguístico que podem estar relacionadas ao índice de apagamento. Em seguida, comparamos os resultados com estudos sobre variedades do PB e do PE.

Inicialmente, foram analisadas 1523 ocorrências, controlando oito variáveis, sendo: (i) três sociais: Sexo, Faixa Etária e Grau de Escolaridade dos informantes e (ii) cinco linguísticas: a Posição do Segmento, a Classe Gramatical da palavra, a Tonicidade da sílaba e o Contexto Fonológico Precedente que foi testado como duas variáveis: a primeira considerando o traço de altura [-alto] e [+alto] e a segunda, os traços [-anterior] e

[+anterior]. O apagamento foi constatado em 56,53% (861/1523) das ocorrências evidenciando a alta frequência percentual do fenômeno, no PST, e das oito variáveis testadas, foram selecionadas como relevantes, para a aplicação da regra de apagamento do rótico: Posição do Segmento, Escolaridade, Faixa Etária, Classe Gramatical e Sexo dos informantes conforme a Tabela 1. Quanto às variáveis Tonicidade, Contexto Fonológico Precedente quanto ao traço de altura [-alto] [+alto] e quanto aos traços [-anterior] [+anterior], foram excluídas – apontando que essas variáveis não estavam relacionadas, de modo significativo, com o apagamento do rótico.

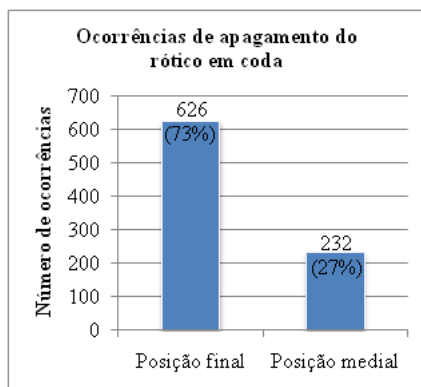
TABELA 1 – Variáveis atuantes no apagamento do rótico em coda no PST

Variável	Fatores	Apl./N	% R-Ø	P.R.
Posição	Final	629/811	77,5	.71
	Medial	232/712	32,6	.29
		861/1523		Range .42
Escolaridade	Fundamental	234/351	66,6	.69
	Médio	403/654	61,6	.40
	Superior	224/528	43,2	.39
		861/1523		Range .30
Faixa Etária	18-20 anos	351/531	66,1	.68
	21-40 anos	335/618	54,2	.42
	41-52 anos	175/374	46,7	.39
		861/1523		Range .29
Classe Gramatical	Verbos	575/779	73,8	.57
	Não-verbos	286/744	38,4	.43
		861/1523		Range .14
Sexo	Feminino	511/821	62,2	.55
	Masculino	350/702	49,5	.45
		861/1523		Range .10
Input: .59				$p < 0,0001$

Fonte: elaboração própria.

A Posição do Segmento foi a variável mais relevante para a aplicação da regra de apagamento com um *range*<sup>11</sup> de .42. O maior índice de apagamento (77,5%) é verificado quando o rótico está em posição final (P.R. .71). Do total de 861 apagamentos constatados, 626 (73%) ocorrem em fronteira de palavra (**açúcar** [a.'su.kɐ], **fazer** [fa.'ze], **maior** [maɪ.'ɔ]) e apenas 232 (27%) são constatados no interior dos vocábulos (**turma** ['tu.mɐ], **perto** ['pɛ.tu], **porque** [pu.'ke]), conforme o gráfico na Figura 3. Como apontado pelos estudos de Brandão *et al.* (2017), Brandão e De Paula (2018), Brandão (2018) e Bouchard (2017) a posição final é um contexto que favorece, consideravelmente, o apagamento do rótico em coda enquanto a posição medial o desfavorece (P.R. .29).

FIGURA 3 – Ocorrências de apagamento do rótico em coda medial e final



Fonte: elaboração própria

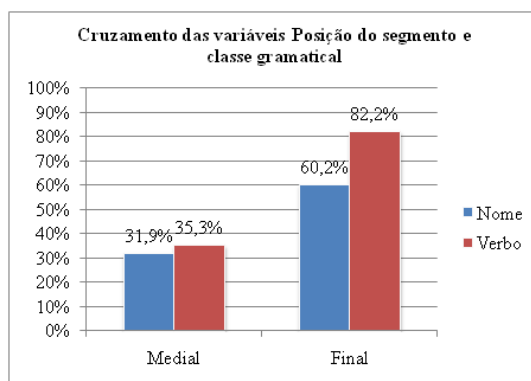
No que se refere à Escolaridade, o apagamento do rótico é favorecido por falantes com menor grau de escolaridade (P.R. .69) e desfavorecido por aqueles com ensino superior (P.R. .39) e ensino médio (P.R. .40), isto é, a probabilidade de um indivíduo mais escolarizado realizar o apagamento é bem menor do que a de um indivíduo menos escolarizado. Quanto à Faixa Etária, o maior índice de apagamento (66,1%) encontra-se entre os falantes mais jovens, com idade entre 18

<sup>11</sup> Diferença entre o menor e o maior peso relativo dos fatores de uma variável. Quanto maior o range mais significativa é a variável, uma vez que, os fatores favorecedores da aplicação da regra em análise estão mais próximos de 1 e aqueles que a desfavorecem estão mais próximos de 0.

e 20 (P.R. .68), e os menores índices estão entre os falantes de 21 a 40 anos (P.R. .42) e de 41 a 52 anos (P.R. .39), indicando que estes são mais conservadores.

A quarta variável selecionada é a Classe Gramatical da palavra que contém o rótico em coda, e o maior índice de apagamento (73,8%) ocorre nos verbos (P.R. .57), como em **comer** [ko.'me], em contraponto ao índice verificado nas palavras não-verbais (P.R. .43), tais como **turma** ['tu.mɐ]. Por fim, o sexo dos informantes é a última variável selecionada como relevante, e aponta que as mulheres favorecem o apagamento (P.R. .55), ao passo que os homens (P.R. .45) o desfavorecem. Levando em consideração que estudos sobre o mesmo fenômeno em variedades do PB têm apresentado índices quase categóricos de apagamento do rótico em coda final de itens verbais (CALLOU, 1987 apud OLIVEIRA, 2018, p. 26; CALLOU; LEITE; MORAES, 2002; CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; OLIVEIRA, 2018; OUSHIRO; MENDES, 2014), realizamos o cruzamento das variáveis: Posição do Segmento e Classe Gramatical da palavra, conforme o gráfico na Figura 4.

FIGURA 4 – Cruzamento das variáveis Posição do Segmento e Classe Gramatical da palavra



Fonte: elaboração própria

O gráfico na Figura 4 mostra que o maior índice de apagamento é constatado em coda final de itens verbais (82,2%), como no PB. Contudo, esse índice também é bem elevado em coda final de itens não verbais (60,2%) e o que mais se destaca é a diferença entre a frequência de aplicação da regra de apagamento nas posições final e medial.

Considerando tal diferença, a amostra, de 1523 ocorrências, foi dividida em duas amostras menores: 712 ocorrências do rótico em coda medial e 811 em coda final. Feito isso, analisamos as duas amostras separadamente. Nas duas análises foi averiguada a relevância das mesmas variáveis testadas na análise inicial: Grau de Escolaridade, Faixa Etária e Sexo dos informantes, Classe Gramatical da palavra, Tonicidade da sílaba e Contexto Fonológico Precedente quanto a (i) altura [-alto] [+alto] e (ii) anterioridade [-anterior] [+anterior]. O quadro 2 resume os resultados das duas análises, evidenciando a importância das variáveis sociais para o apagamento do rótico nos dois contextos, uma vez que, em ambos, duas das três variáveis sociais testadas foram selecionadas como relevantes, enquanto nenhuma das variáveis linguísticas se mostrou relevante no contexto de coda medial, e em contexto de coda final, apenas uma das variáveis linguísticas se revelou significativa.

QUADRO 2 – Resultados da análise de apagamento do rótico em coda no PST nas posições: medial e final

<b>Apagamento do rótico em coda medial</b>	<b>Apagamento do rótico em coda final</b>
Dados analisados: 712 ocorrências	Dados analisados: 811 ocorrências
Apagamento: 32,6%	Apagamento: 77,5%
Variáveis selecionadas:	Variáveis selecionadas:
<i>Escolaridade</i>	<i>Escolaridade</i>
<i>Sexo</i>	<i>Faixa Etária</i>
	<i>Classe Gramatical</i>

Fonte: elaboração própria.

Na análise do apagamento do rótico em coda medial, o índice de apagamento foi de 32,6% (232/712) e as variáveis selecionadas como relevantes à aplicação da regra de apagamento foram: Escolaridade e Sexo dos informantes. Os resultados para os fatores das variáveis selecionadas foram semelhantes aos da análise inicial (cf. TABELAS 1 e 2) e aos de Brandão *et al.* (2017). Os maiores índices de apagamento estão entre os informantes do ensino fundamental (P.R. .62) e do sexo feminino (P.R. .59) e os menores entre os falantes com ensino superior (P.R. .27) e do sexo masculino (P.R. .41).

TABELA 2 – Variáveis atuantes no apagamento do rótico em coda medial no PST

Variável	Fatores	Apl./N	% R-Ø	P.R.
Escolaridade	Fundamental	78/175	44,6	.62
	Médio	115/290	39,6	.61
	Superior	<u>39/247</u>	15,8	.27
		232/712		Range .35
Sexo	Feminino	143/373	38,3	.59
	Masculino	89/339	26,2	.41
		232/712		Range .18
<i>Input= 0.30</i>				<i>p &lt; 0,0001</i>

Fonte: elaboração própria.

Analisando o apagamento do rótico em coda final, contexto em que o fenômeno ocorre com muito mais frequência, sendo verificado em 77,5% (629/811) dos casos, o modelo estatístico selecionou as seguintes variáveis: Escolaridade, Faixa Etária e Classe Gramatical.

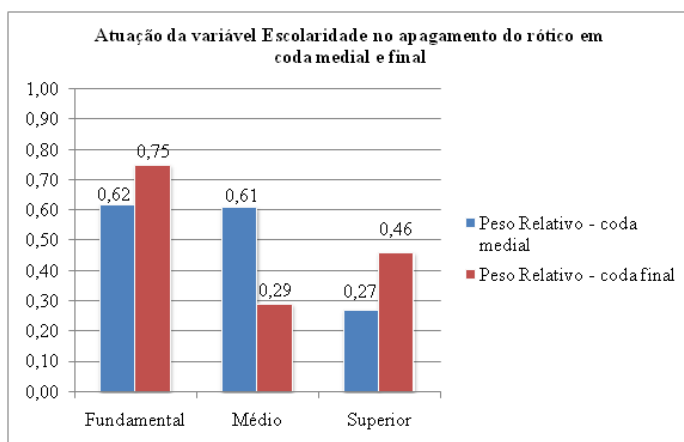
TABELA 3 – Variáveis atuantes no apagamento do rótico em coda final no PST

Variável	Fatores	Apl./N	% R-Ø	P.R.
Escolaridade	Fundamental	156/176	88,6	.75
	Médio	288/364	79,1	.29
	Superior	<u>185/271</u>	68,3	.46
		629/811		Range .46
Faixa Etária	18-20 anos	254/294	86,4	.76
	21-40 anos	236/329	71,7	.35
	41-52 anos	<u>139/188</u>	73,9	.37
		629/811		Range .41
Classe Gramatical	Verbos	526/640	82,2	.62
	Não-verbos	<u>103/171</u>	60,2	.38
		629/811		Range .24
<i>Input = 0.78</i>				<i>p &lt; 0,0001</i>

Fonte: elaboração própria.

A variável Escolaridade se mostrou relevante nos dois contextos analisados de forma semelhante, sendo a mais saliente para o apagamento em posição final, com um *Range* de .46. O gráfico na Figura 5 mostra uma comparação entre os pesos relativos dos fatores dessa variável (Fundamental, Médio e Superior) na análise do apagamento do rótico em coda medial e final. Falantes com ensino fundamental são os que mais apagam o rótico tanto em coda medial (P.R. .62) quanto em coda final (P.R. .75). Todavia, o menor índice de apagamento em coda medial é o dos falantes com ensino superior (P.R. .27), enquanto os falantes do ensino médio (P.R. .61) favorecem o fenômeno quase na mesma medida que os falantes do ensino fundamental. Já em coda final, apesar de os falantes com ensino superior (P.R. .46) desfavorecerem o apagamento, o menor índice é observado entre os falantes do ensino médio (P.R. .29). Contudo, de maneira geral esses resultados indicam, semelhantemente à análise inicial, que indivíduos mais escolarizados aplicam a regra de apagamento com menos frequência do que os menos escolarizados, o que também se verifica em Brandão *et al.* (2017), Brandão e De Paula (2018), Brandão (2018) e Bouchard (2017), e pode estar associado ao fato de que falantes com maior grau de escolaridade se aproximam mais da norma de prestígio: a variedade europeia, disseminada pela escola, que possui baixos índices de apagamento do rótico (BRANDÃO; MOTA; CUNHA, 2003; MATEUS; RODRIGUES, 2003).

FIGURA 5 – Atuação da variável Escolaridade no apagamento do rótico em coda medial e final



Fonte: elaboração própria.

Na Tabela 3, verifica-se que a Faixa Etária foi a segunda variável mais saliente, com um *Range* de .41, apontando, como na análise inicial, que falantes mais jovens, entre 18 e 20 anos (P.R. .76) são mais propensos ao apagamento, ao passo que, os mais velhos, entre 21 e 40 anos (P.R. .35) e entre 41 e 52 anos (P.R. .37), são mais conservadores, não sendo verificada uma grande diferença entre os índices dessas duas últimas faixas etárias.

A Classe Gramatical foi a única variável linguística selecionada como relevante para o apagamento do rótico em coda final e, como constatado, também em Brandão *et al.* (2017), Brandão e De Paula (2018), Brandão (2018) e Bouchard (2017), o maior índice de apagamento ocorre nos verbos (P.R. .62). Em itens verbais o fenômeno ocorre em 82,2% (526/640) dos casos.

No que diz respeito aos índices percentuais de apagamento em coda medial (32,6%) e final (77,5%) é interessante notar que o índice de apagamento em coda final verificado por Bouchard (2017) foi de 79,4%, bem próximo ao que verificamos nessa mesma posição. Já nos trabalhos de Brandão *et al.* (2017), Brandão e De Paula (2018) e Brandão (2018), baseados num *corpus* de 2009, os índices de apagamento medial e final foram, respectivamente, 4,4% e 44,7%, valores bem menores do que os constatados aqui a partir da análise de um *corpus* de 2016 e 2019, e por Bouchard (2017), que analisou um *corpus* coletado entre 2015 e 2017. Isso indica que, em um intervalo de menos de dez anos, houve uma mudança considerável na proporção de implementação do fenômeno, o que, em conjunto com a relevância da variável faixa etária na análise geral e em coda final, sugere uma mudança linguística em curso (cf. WEINREICH; LABOV; HERZOG, 1968).

Considerando esses resultados em comparação com o PB e o PE, destacamos que, como o PST, diversas variedades do PB também possuem índices elevados de apagamento do rótico, enquanto o PE possui índices menores<sup>12</sup> (ALVES, 2015; BRANDÃO; MOTA; CUNHA, 2003; CALLOU, 1987 *apud* OLIVEIRA, 2018, p. 25-26; CALLOU;

---

<sup>12</sup> Um trabalho que evidencia essa diferença é o de Brandão, Mota e Cunha (2003) que, realizando um estudo contrastivo entre o comportamento do rótico em final de vocábulo no PE e no PB, chega a resultados opostos para as variedades brasileira e europeia no que diz respeito aos índices de apagamento: 26% de apagamento e 74% de manutenção no PE e 78% de apagamento e 22% de manutenção no PB.

LEITE; MORAES, 1994; CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; MATEUS; RODRIGUES, 2003; MONGUILLHOT, 1997; OLIVEIRA, 2018; OUSHIRO; MENDES, 2014; SERRA; CALLOU, 2015).

Além disso, a diferença entre os índices de apagamento do rótico em coda medial e final, constatada no PST, também é verificada nas variedades no PB (ALVES, 2015; BRANDÃO; MOTA; CUNHA, 2003; GREGIS, 2001; MONARETTO, 2000; MONGUILLHOT, 1997; OUSHIRO; MENDES, 2014; PIMENTEL, 2003) e trabalhos que ultrapassam o domínio da sílaba apontam que esse seria um indício de que, ainda que o apagamento do rótico seja condicionado pela coda, sua produtividade envolveria domínios prosódicos maiores, como a fronteira de palavra prosódica (CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; OLIVEIRA, 2018). No PE, entretanto, segundo Mateus e Rodrigues (2003), o rótico é apagado com mais frequência em posição medial, quando antecede uma consoante oclusiva ou fricativa, já quando é seguido de vogal ou pausa raramente ocorre o apagamento.

Como no PST, em variedades do PB, os índices de apagamento do rótico em coda final são mais expressivos em verbos, do que em não-verbos, sendo, inclusive, constatados índices superiores a 80% como se verificou para o PST (CALLOU, 1987 *apud* OLIVEIRA, 2018, p. 26; CALLOU; LEITE; MORAES, 2002; CALLOU; SERRA, 2012; CALLOU; SERRA; CUNHA, 2015; OLIVEIRA, 2018; OUSHIRO; MENDES, 2014), todavia, o mesmo não acontece em variedades do PE. Brandão, Mota e Cunha (2003) constataram índices quase idênticos para verbos e nomes em coda final, e Mateus e Rodrigues (2003), analisando o apagamento em coda medial e final, constataram exatamente os mesmos índices de apagamento em verbos e não-verbos.

Quanto aos resultados relativos à atuação das variáveis sociais, apresentados neste artigo, encontramos similaridades em estudos sobre variedades do PB. Os maiores índices de apagamento, geralmente, estão entre: (i) os falantes mais jovens (BRANDÃO; MOTA; CUNHA, 2003; CALLOU, 1987 *apud* OLIVEIRA, 2018, p. 25; CALLOU; SERRA; CUNHA, 2015; MONARETTO, 2000; OUSHIRO; MENDES, 2014; PIMENTEL, 2003); (ii) menos escolarizados (ALVES, 2015; BRANDÃO; MOTA; CUNHA, 2003; MONARETTO, 2000; MONGUILLHOT, 1997); e (iii) do sexo feminino (CALLOU, 1987 *apud* OLIVEIRA, 2018, p. 25; GREGIS, 2001). Já no caso do PE, Mateus e Rodrigues (2003) afirmam

que o apagamento se deve mais a fatores linguísticos do que sociais. Em Brandão, Mota e Cunha (2003), apenas a variável social Faixa Etária se mostrou relevante e, como no PB e no PST, os informantes mais jovens apresentaram o maior índice de apagamento. Para as autoras, tanto em PB quanto em PE o apagamento do (R) é uma inovação, que, no PB, já está bastante disseminada (BRANDÃO; MOTTA; CUNHA, 2003).

Em suma, o apagamento do rótico em coda, como em **parto** ['par.tu] ~ ['pa.tu] e **chamar** [ʃa.'mar] ~ [ʃa.'ma], é influenciado por variáveis sociais e linguísticas, no entanto, esse fenômeno, apesar de ser produtivo no PE, PB e PST, é muito mais recorrente nas duas últimas, sugerindo uma tendência dessas duas variedades ao padrão silábico CV, haja vista que a elisão da coda – a realização de sílabas CVC como CV – é comum em muitas outras línguas (SELKIRK, 1982), o que pode ser relacionado ao fato de que a estrutura silábica CV é universal. Essa hipótese deve, no entanto, ser avaliada em conjunto com a análise de outros *templates* silábicos do PST, observando: (i) se o apagamento em coda é recorrente, também, tendo outros segmentos licenciados neste constituinte como alvo (BALDUINO; VIEIRA, 2020; VIEIRA; BALDUINO, 2020) em itens como **susto** ['suʃ.tu] e **palma** ['paɫ.mɐ], cujas codas correspondem a /S/ e /l/, e/ou (ii) se o fenômeno é implementado na dissolução de *onsets* complexos em palavras como **prato** ['pra.tu], em que o rótico ocupa a segunda posição do *cluster*. Ademais, fatores como a debilidade do rótico, especialmente, em fronteira de palavra, um domínio da prosódia, que parece caracterizar não só o PB, mas também o PST, também devem ser considerados.

## 6 Considerações Finais

Na análise inicial o apagamento do rótico no PST foi constatado em 56,53% (861/1523) das ocorrências analisadas, sendo que os maiores índices de apagamento foram verificados quando o rótico estava em posição final (P.R. .71), em verbos (P.R. .57) e entre indivíduos menos escolarizados (P.R. .69) e mais jovens (P.R. .68). Na análise do fenômeno em coda medial, o índice de apagamento foi de 32,6% (232/712) com os maiores índices de apagamento sendo verificados entre os informantes do ensino fundamental (P.R. .62) e do sexo feminino (P.R. .59). Na análise do apagamento em coda final, o índice de apagamento foi de 77,5% (629/811) e os falantes do ensino fundamental (P.R. .75) e mais jovens (P.R. .76)

foram os que mais apagaram o rótico. Quanto à Classe Gramatical, o maior índice de apagamento (82,2%) ocorreu nos verbos (P.R. .62).

A diferença entre os índices de apagamento verificados no presente estudo (32,6% em coda medial e 77,5% em coda final) e os índices (4,4% em coda medial e 44,7% em coda final) constatados nos trabalhos de Brandão *et al.* (2017), Brandão e De Paula (2018) e Brandão (2018), baseados num *corpus* de 2009, indica uma possível mudança linguística em curso (cf. WEINREICH; LABOV; HERZOG, 1968). Essa hipótese é reforçada pela relevância da variável Faixa Etária na análise geral e em coda final observada neste estudo, fato também apontado por Bouchard (2017), que trabalha com um *corpus* mais recente, coletado entre 2015 e 2017.

Comparando os resultados aqui obtidos para o PST com estudos sobre variedades do PB e do PE, concluímos que as variáveis sociais e linguísticas, geralmente, se comportam de forma semelhante no PST e no PB, com índices de apagamento expressivos. Já o PE apresenta índices de apagamento menores, e as variáveis linguísticas se comportam de forma diferente nas variedades do PE analisadas. Enquanto no PB e no PST o índice de apagamento é mais expressivo em coda final e em itens verbais, no PE não há diferenças entre os índices em itens verbais e não-verbais, e o fenômeno ocorre com mais frequência em posição medial, quando precede uma consoante oclusiva ou fricativa.

Assim, concluímos que o fenômeno, apesar de ser produtivo nas três variedades, é mais recorrente no PB e no PST, sugerindo uma preferência dessas duas variedades ao padrão silábico CV, comum em muitas outras línguas (SELKIRK, 1982), e que pode ser relacionado ao fato de a estrutura silábica CV ser universal – hipótese que precisa ser avaliada em conjunto com outros processos do PST que tenham a sílaba como domínio. Ademais, fatores, como a debilidade do rótico, especialmente, em fronteira de palavra, também devem ser considerados. Dessa forma, esse estudo, não somente corrobora os resultados de trabalhos prévios como os de Bouchard (2017), Brandão *et al.* (2017), Brandão e De Paula (2018) e Brandão (2018) quanto ao comportamento das variáveis sociais e linguísticas, como também aponta a necessidade de uma análise que discuta os padrões silábicos do PST à luz de outros fenômenos fonológicos, os quais podem não somente ser afetados pelo contato com o santome e outras línguas locais, como podem sugerir inovações em relação às estruturas silábicas previstas em outras variedades da língua portuguesa.

## **Agradecimento**

Agradecemos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento que permitiu a condução dessa pesquisa e elaboração deste artigo: processos 2015/25332-1 e 2017/26595-1.

## **Contribuições de cada autora**

Nancy Mendes Torres Vieira: Conceptualização; Curadoria de dados; Escrita – original; Metodologia; Análise formal; Escrita – análise e edição; Recursos.

Amanda Macedo Balduino: Conceptualização; Investigação; Coleta e curadoria de dados; Metodologia; Escrita – análise e edição; Recursos.

## **Referências**

ABAURRE, M. B. M.; SÂNDALO, M. F. S. Os róticos revisitados. In: HORA, D.; COLLISCHONN, G. (org.). *Teoria linguística: Fonologia e outros temas*. João Pessoa: Editora da Universidade Federal da Paraíba, 2003. p. 144-180.

AGOSTINHO, A. L. *Fonologia e Método Pedagógico do Lung'ie*. 2015. 446f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2015.

AGOSTINHO, A. L. Róticos em contexto intervocálico no Português da Ilha do Príncipe: fonologia e educação. In: ENCONTRO DA ABECS, IX., 2016, Brasília. *Anais [...]*. Brasília: Universidade de Brasília, 2016. p. 48-49. Disponível em: <https://encontroabecs.wordpress.com/cad-de-resumos>. Acesso em: 10 set. 2020

ALVES, M. A. Variação na produção/apagamento da vibrante pós-vocálica no falar florianopolitano. *Working Papers em Linguística*, Florianópolis, v. 16, n. 1, p. 20-35, 2015. DOI: <https://doi.org/10.5007/1984-8420.2015v16n1p20>.

ARAUJO, G. Há uma política linguística para o português em São Tomé e Príncipe?. In: SOUZA, S.; OLMO, F. C. (org.). *Línguas em português: a lusofonia numa visão crítica*. Porto: Universidade do Porto Press, 2020. p. 173-197.

ARAUJO, G.; HAGEMEIJER, T. *Dicionário Santome-Português/Português-Santome*. São Paulo: Hedra, 2013.

BALDUINO, A. M. *A nasalidade vocálica no português falado em São Tomé e Príncipe*. 2018. 296f. Dissertação (Mestrado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2018.

BALDUINO, A. M. Apagamento de /R/ e /S/ em coda no Português Principense. *Papia*, São Paulo, v. 29, n. 1, p. 25-39, 2019. Disponível em: [http://revistas.fflch.usp.br/papia/article/view/3366/pdf\\_1](http://revistas.fflch.usp.br/papia/article/view/3366/pdf_1). Acesso em: 23 jul. 2020.

BALDUINO, A. M. *Processos fonológicos no português de São Tomé e de Santo António do Príncipe*. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. (em preparação).

BALDUINO, A. M.; VIEIRA, N. M. T. Distribuição da lateral /l/ em coda no português santomense. *Estudos Linguísticos*, São Carlos, SP, v. 49, n. 2, p. 594-615, 2020. DOI: <https://doi.org/10.21165/el.v49i2.2490>.

BANDEIRA, M. *Reconstrução fonológica e lexical do protocrioulo do Golfo da Guiné*. 2017. 439f. Tese (Doutorado em Filologia e Língua Portuguesa) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2017.

BAXTER, A. O Português dos Tongas de São Tomé. In: OLIVEIRA, M. D.; ARAUJO, G. (org.). *O português na África Atlântica*. São Paulo: Humanitas/Fapesp, 2018. p. 297-324.

BISOL, L. A sílaba e seus constituintes. In: NEVES, M. H. M (org.). *Gramática do Português culto falado: novos estudos*. Campinas: Editora da Unicamp, 1999. p. 701-742.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer* (Version 5.3.82) Computer Program. 2020. Disponível em: <http://www.praat.org>. Acesso em: 20 jun. 2020.

BOUCHARD, M. E. *Linguistic Variation and Change in the Portuguese of São Tomé*. 2017. 389f. Dissertation (Doctoral) – Department of Linguistics, New York University, New York, 2017. DOI: <https://doi.org/10.5334/jpl.192>

BRANDÃO, S. F. Apagamento de R em coda externa em duas variedades africanas do português. *Diadorim*, Rio de Janeiro, v. 20, n. especial, p. 390-408, 2018. DOI: <https://doi.org/10.35520/diadorim.2018.v20n0a23283>.

BRANDÃO, S. F.; MOTA, M. A.; CUNHA, C. S. Um estudo contrastivo entre o português europeu e o português do Brasil: o –R final de vocábulo. In: BRANDÃO, S. F.; MOTA, M. A. (org.). *Análise contrastiva de variedades do português: primeiros estudos*. Rio de Janeiro: In-fólio, 2003. p. 163-180.

BRANDÃO, S. F.; DE PAULA, A. Róticos nas variedades santomense e moçambicana do Português. In: BRANDÃO, S. F. (org.). *Dois variedades africanas do Português: variáveis fonético-fonológicas e morfossintáticas*. São Paulo: Blucher, 2018. p. 93-118. DOI: <https://doi.org/10.5151/9788580393248>

BRANDÃO, S. F.; PESSANHA, D. B.; PONTES, S.; CORRÊA, M. Róticos na variedade urbana do Português de São Tomé. *Papia*, São Paulo, v. 27, n. 2, p. 293-315, 2017. Disponível em: <http://revistas.fflch.usp.br/papia/article/view/2762/pdf>. Acesso em: 20 jul. 2020.

CALLOU, D.; LEITE, Y.; MORAES, J. Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil. In: KOCH, I. G. V. (org.). *Gramática do português falado*. Campinas: Editora da Unicamp/FAPESP, 1994. p. 465-493.

CALLOU, D.; LEITE, Y.; MORAES, J. Processo(s) de enfraquecimento consonantal no português do Brasil. In: ABAURRE, M. B. M.; RODRIGUES, A. C. S. (org.). *Gramática do português falado VIII: novos estudos descritivos*. Campinas: UNICAMP/FAPESP, 2002. p. 537-555.

CALLOU, D.; SERRA, C. Variação do rótico e estrutura prosódica. *GELNE*, Natal, v. 14, n. especial, p. 41-58, 2012. Disponível em: <https://periodicos.ufrn.br/gelne/article/view/9363>. Acesso em: 20 jul. 2020.

CALLOU, D.; SERRA, C.; CUNHA, C. Mudança em curso no português brasileiro: o apagamento do R no dialeto nordestino. *Revista Abralín*, Curitiba, v. 14 n. 1, p. 195-219, 2015. DOI: <http://dx.doi.org/10.5380/rabl.v14i1.42491>.

CÂMARA Jr., J. M. *Estrutura da língua portuguesa*. Petrópolis: Vozes, 1970.

CHESHIRE, J. Sex and Gender in Variationist Research. In: TRUDGILL, P.; CHAMBERS, J. K.; SCHILLING-ESTES, N. (org.). *The Handbook of Language Variation and Change*. Oxford: Blackwell Publishing, 2004. p. 423-443. DOI: <https://doi.org/10.1111/b.9781405116923.2003.00024.x>

CHRISTOFOLETTI, A. *Ditongos no português de São Tomé e Príncipe*. 2013. 109f. Dissertação (Mestrado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.

FERRAZ, L. *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press, 1979.

FIGUEIREDO, C. *A concordância plural variável no sintagma nominal do português reestruturado da comunidade de almoxarife, São Tomé*. 2010. 792f. Tese (Doutorado em Linguística) – Faculdade de Ciências Sociais e Humanas, Universidade de Macau, Macau, 2010.

GOMES, C. A. Para além dos pacotes estatísticos Varbrul/Goldvarb e Rbrul: qual a concepção de gramática? *GELNE*, Natal, v. 14, n. especial, p. 259-272, 2012. Disponível em: <https://periodicos.ufrn.br/gelne/article/view/9363>. Acesso em: 20 jul. 2020.

GONÇALVES, R. *Propriedade de subcategorização verbal no português de S. Tomé*. 2010. 151f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2010.

GONÇALVES, R.; HAGEMEIJER, T. O português num contexto multilingue: o caso de São Tomé e Príncipe. *Revista Científica da Universidade Eduardo Mondlane*, Maputo, v. 1, n. 1, p. 84-103, 2015. Disponível em: <https://repositorio.ul.pt/bitstream/10451/31032/1/Goncalves%26Hagemeijer2015.pdf>. Acesso em: 25 jul. 2020.

GREGIS, H. *O apagamento da vibrante pós-vocálica em Porto Alegre*. 2001. 116f. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001.

GUY, G.; ZILLES, A. *Sociolinguística quantitativa*. São Paulo: Parábola Editorial, 2007.

HAGEMEIJER, T. From Creoles to Portuguese: Language Shift in São Tomé and Príncipe. In: LÓPEZ, L.; GONÇALVES, P.; AVELAR, J. (org.). *The Portuguese Language Continuum in Africa and Brazil*. Amsterdam; Philadelphia: John Benjamins, 2018. p. 169-184. DOI: <https://doi.org/10.1075/ihll.20.08hag>

INE - Instituto Nacional de Estatística de São Tomé e Príncipe. 2012. Disponível em: <https://www.ine.st/>. Acesso em: 10 jun. 2020.

LABOV, W. *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press, 1972.

LABOV, W. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell, 1994. v. 1.

LABOV, W. *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell, 2001. v. 2.

MATEUS, M. H.; D'ANDRADE, E. *The Phonology of Portuguese*. Oxford: Oxford Linguistics, 2000.

MATEUS, M. H. M.; RODRIGUES, C. A Vibrante em Coda em Português Europeu. In: HORA, D.; COLLISCHONN, G. (org.). *Teoria Linguística Fonologia e outros temas*. João Pessoa: Ed. Universitária, 2003. p.181-199.

MENESES, F. O. *As vogais desvozeadas no Português Brasileiro: investigação acústico-articulatória*. 2012. 123f. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2012.

MONARETTO, V. N. O. O apagamento da vibrante pós-vocálica nas capitais do sul do Brasil. *Letras de Hoje*. Porto Alegre, v. 35, n. 1, p. 275-284, 2000. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/14768/9834>. Acesso em: 20 jul. 2020.

MONGUILHOT, I. O. S. A vibrante em final de palavra na fala de Santa Catarina. In: ENCONTRO DO CÍRCULO DE ESTUDOS LINGÜÍSTICOS DO SUL, 2., 1997, Florianópolis. *Anais do II CELSUL*. Florianópolis: Universidade Federal de Santa Catarina, 1997. [1 CD-Rom]

OLIVEIRA, I. C. *Os róticos em coda silábica externa: o interior da região Sul no projeto AliB*. 2018. 133f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

OUSHIRO, L. Tratamento de dados com o R para análises sociolinguísticas. In: FREITAG, R. M. K. (org.). *Metodologia de coleta e manipulação de dados em Sociolinguística*. São Paulo: Editora Blucher, 2014. p. 133-177. DOI: <https://doi.org/10.5151/BlucherOA-MCMDS-10cap>

OUSHIRO, L. *Introdução à Estatística para Linguistas*, v.1.0.1. 2017. Disponível em: <https://rpubs.com/oushiro/iel>. Acesso em: 2 jul. 2020.

OUSHIRO, L.; MENDES, R. B. O apagamento de (-r) em coda nos limites da variação. *Veredas*, Juiz de Fora, v. 18, n. 2, p. 251-266, 2014. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/24963>. Acesso em: 2 jul. 2020.

PIMENTEL, R. M. *A variação linguística do fonema /r/ na posição pós-vocálica*. 2003. 105f. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

QUEDNAU, L. *A lateral pós-vocálica no português gaúcho: análise variacionista e representação não linear*. 1993. 110f. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1993.

RODRIGUES, M. C. Todas as codas são frágeis em português europeu? *Revista Linguística*, Rio de Janeiro, v. 8, n. 1, p. 138-149, 2012.

ROUGÉ, J. L. Les langues des Tonga. In: D'ANDRADE, E.; KIHM, A. (org.). *Actas do Colóquio sobre Crioulos de base lexical portuguesa*. Lisboa: Colibri, 1992. p. 171-76.

SANTIAGO, A. M. *As vogais do português do Príncipe*. Trabalho de conclusão de curso (Graduação em Letras) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2019.

SANTIAGO, A. M.; AGOSTINHO, A. L. Situação linguística do português em São Tomé e Príncipe. *A cor das Letras*, Feira de Santana, BA, v. 21, n. 1, p. 39-61, 2020. DOI: <https://doi.org/10.13102/cl.v21i1.4970>.

SELKIRK, E. The Syllable. In: HULST, H.; SMITH, N. (ed.). *The Structure of Phonological Representations*. Dordrecht: Foris, 1982. p. 337-383.

SERRA, C.; CALLOU, D. Prosodic Structure, Prominence and /r/-Deletion in Final Coda Position: Brazilian Portuguese and European Portuguese Contrasted. In: DOMINICIS, A. (org.). *pS-prominenceS: Prominences in Linguistics*. Viterbo: Disucom Press, 2015. p. 96-113.

VIEIRA, N. M. T.; BALDUINO, A. M. Apagamento de /R, S, l/ na coda no português de São Tomé: convergência linguística? *Papia*, São Paulo, v. 30, n. 1, p. 7-33, 2020. Disponível em: <http://www.revistas.fffch.usp.br/papia/article/view/3415/pdf>. Acesso em: 3 set. 2020.

WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical Foundations for a Theory of Linguistic Change. In: LEHMANN, W. P.; MALKIEL, Y. (org.). *Directions for Historical Linguistics*. Austin: University of Texas Press, 1968. p. 97-195.



## **Compilação, reciclagem e padronização de um Corpus Colaborativo de Linguística: percursos metodológicos**

### ***Compilation, recycling and standardization in a Collaborative Corpus of Linguistics: methodological approaches***

Guilherme Fromm

Universidade Federal de Uberlândia (UFU), Uberlândia, Minas Gerais / Brasil

guifromm@ufu.br

<http://orcid.org/0000-0001-5654-0135>

Márcio Issamu Yamamoto

Universidade Federal de Jataí (UFJ), Jataí, Goiás / Brasil

marcioiy@ufg.br

<http://orcid.org/0000-0001-9792-8187>

**Resumo:** O objetivo deste texto é relatar uma experiência de trabalho com um *corpus* colaborativo, desenvolvida nos últimos 10 anos (2010-2020) com alunos de diversas turmas de graduação, pós-graduação e alunos de Iniciação Científica na Universidade Federal de Uberlândia (já parcialmente descrita por Fromm, 2013, e Fromm e Yamamoto, 2013). O trabalho parte da metodologia de elaboração do *corpus* (incluindo seu histórico) como um passo para um tipo de pesquisa específica (terminográfica), passa por uma análise para indicar e solucionar problemas de compilação e termina na sua adequação e padronização para reuso. O resultado é um *corpus* robusto, bem balanceado, bilíngue (inglês/português) e que poderá ser usado em inúmeras outras pesquisas na área de Linguística.

**Palavras-chave:** Linguística; Linguística de *Corpus*; *Corpus* colaborativo; Árvore de Domínio; Terminografia.

**Abstract:** This text describes a work experience with a collaborative corpus, developed during the last 10 years (2010-2020) with students in several undergraduate and graduate classes and students working on scientific undergraduate research projects at Universidade Federal de Uberlândia (already partially described by Fromm, 2013, and Fromm and Yamamoto, 2013). The work starts from the corpus elaboration methodology

(including its history) as a step towards a specific type of research (terminographic), goes through an analysis to point out and solve compilation problems and ends in its adequacy and standardization for reuse. The result is a robust, well-balanced, bilingual (English/Portuguese) corpus that can be used in numerous other studies in the area of Linguistics.

**Keywords:** Linguistics; Corpus Linguistics; Collaborative corpus; Domain Tree; Terminography.

Recebido em 19 de outubro de 2020

Aceito em 24 de fevereiro de 2021

## 1 Apresentação

As experiências com a Linguística de *Corpus* (doravante LC) fazem parte de todo o percurso na pós-graduação dos autores e dirigem nossas pesquisas há mais de 15 anos. Pretendemos, com este artigo, apresentar dois trabalhos acadêmicos sequenciais compartilhados pelos pesquisadores na elaboração e na reelaboração de um mesmo *corpus* num período de dez anos, dando maior ênfase aos processos metodológicos aplicados aos *corpora* compilados na pesquisa: (1) projeto de elaboração de *corpus* acadêmico bilíngue na área de Linguística e (2) projeto de reelaboração e construção de um vocabulário bilíngue de Linguística (YAMAMOTO, 2018).

Partimos do pressuposto de que uma pesquisa de caráter empírico, descritivo, estatístico e probabilístico que envolva *corpora* deva considerar que o pesquisador vá trabalhar com a metodologia da LC.<sup>1</sup> Há duas formas básicas para o trabalho com *corpus*: compilar o próprio *corpus* (oral ou escrito) ou usar *corpora* já prontos (normalmente disponíveis na internet). No primeiro caso, os *corpora*, elaborados sob medida para pesquisas, são geralmente pequenos, devido ao trabalho envolvido e à restrição de tempo; no segundo caso, nem sempre o conteúdo desses *corpora* é totalmente aproveitável em pesquisas propostas.

Os *corpora* bilíngues a serem aqui descritos foram compilados a várias mãos. Eles começaram como um mero exercício em sala de aula, com *corpora* descartáveis, evoluindo, em projetos posteriores,

---

<sup>1</sup> Reconhecemos que nem toda pesquisa que trabalha com um *corpus* (= conjunto de textos para análise) tem esse caráter descrito nem exige a utilização da LC.

para tornarem-se dois *corpora*, que podem ser reaproveitados por outros pesquisadores.

Para o desenvolvimento do artigo, além desta apresentação, pretendemos mostrar o histórico que nos levou ao trabalho com esses projetos (1 e 2), alguns pressupostos teóricos adotados na elaboração de *corpus*, a descrição da metodologia de trabalho em si, o percurso, do começo ao fim, entre o *corpus* inicial (do primeiro projeto) e o *corpus* final (do segundo projeto) e nossas considerações finais sobre todo o projeto.

## 2 Histórico

Nosso projeto partiu de uma ideia de reaproveitamento, em sala de aula, de *corpora* que os alunos elaboravam para a disciplina de graduação “Língua Inglesa: Estudos Descritivos e Linguística de Corpus 1” (de 60 horas, no curso de Inglês e Literaturas de Língua Inglesa), ministrada por um dos autores. A disciplina tinha dois objetivos de trabalho práticos: compilação e análise de *corpora* (com o consequente aprendizado de como trabalhar com ferramentas de análise lexical) e o uso deles para construção de vocabulários especializados por meio de um trabalho terminográfico a partir de um *software* desenvolvido por um dos autores (FROMM, 2007).<sup>2</sup> Em determinado momento, no primeiro ano em que a disciplina foi ministrada (2010), surgiu uma dúvida metodológica: deveríamos desenvolver novos *corpora* para cada turma, os chamados *corpora* descartáveis (VARANTOLA, 2002), ou poderíamos trabalhar num projeto maior que reaproveitasse esses *corpora*? Desdobrando um trabalho secundário desenvolvido na tese de um dos autores (FROMM, 2007), optamos por trabalhar com a compilação de um *corpus* que pudesse dialogar com cada aluno, por se tratar de sua própria área de atuação: Linguística.

A partir de um protótipo de Árvore de Domínio (taxonomia) da área de Linguística (FROMM, 2007), começamos os trabalhos com os alunos: cada grupo, composto por dois estudantes, deveria trabalhar com uma subárea da Linguística (teórica ou aplicada) de sua escolha, de forma

---

<sup>2</sup> VoTec. O resultado dos trabalhos dos alunos pode ser acessado em: <http://treino.votec.ileel.ufu.br>, área de Linguística. O vocabulário de Logística, por exemplo, foi uma das tentativas de trabalho colaborativo dos alunos que não prosperou (houve desinteresse pelo tema; percebemos, naquele momento, que o assunto a ser analisado é um passo metodológico importante quando da elaboração de *corpora* por parte dos alunos).

bílingue (português e inglês), compilando *corpora* comparáveis (mesmo assunto) de 500 mil palavras daquela subárea e na língua de trabalho escolhida (a dupla trabalhava uma subárea, como *Etimologia*, e cada aluno se dedicava a uma língua). Como resultado geral, obtivemos, para cada dupla, um *subcorpus* de aproximadamente um milhão de palavras que poderia ser usado para vários tipos de pesquisas linguísticas.

### 3 Pressupostos teóricos

Apresentaremos, nos próximos subtópicos, alguns conceitos fundamentais sobre o que é um *corpus*, como ele se torna colaborativo e o que é um trabalho terminográfico (para o qual os *corpora* foram compilados).

#### 3.1 Tamanho, balanceamento e representatividade

O **tamanho**, o **balanceamento** e a **representatividade** são conceitos e princípios inerentes à LC que nortearam a elaboração dos *corpora* (e posterior exploração dos mesmos para trabalhos terminográficos).

Em primeiro lugar, quanto ao **tamanho** de um *corpus*, Berber Sardinha (2004, p. 26-27) propõe três tipos de abordagens: (a) impressionística, (b) histórica e (c) estatística, e alguns parâmetros para o dimensionamento do tamanho do *corpus*: (1) ser representativo da comunidade de fala pesquisada; (2) atender o objetivo da pesquisa. Na (a) abordagem impressionística, o autor coloca como salvaguarda, já que não há maneira de saber o tamanho que representaria uma amostra ideal do recorte que está sendo feito, tornar seu tamanho o maior possível. A (b) abordagem histórica compara a frequência de palavras de acordo com a época da publicação das obras e descreve como essa frequência pode mudar com o tempo. Por fim, a (c) abordagem estatística baseia-se em dados matemáticos para definir a adequação do tamanho de um *corpus* e sua representatividade. Em conclusão, a questão de categorizar tamanho de *corpora* atualmente, tendo em vista os mega *corpora* (como os da BYU<sup>3</sup>), é um tanto complicada dada a velocidade com a qual a tecnologia de coleta (e análise) se desenvolve.

---

<sup>3</sup> Disponíveis em: <https://www.english-corpora.org>. Acesso em: 18 maio 2020.

Em segundo lugar, há o **balanceamento**: “processo pelo qual se garante que dois *corpora* sejam construídos de maneira similar quanto à origem, gênero, extensão, período de produção dos textos...” (TAGNIN; BEVILACQUA, 2013, p. 215), relacionado ao número de palavras e textos constituintes dos *corpora*.

Por fim, a **representatividade** está ligada ao princípio de que o *corpus* cumpre uma função representativa da língua (ou variedade linguística) e de uma comunidade representada. Logo, deve responder às perguntas: representativo de quê (amostragem) e para quem, considerados dados como frequência, número de palavras e aspecto probabilístico da linguagem (BERBER SARDINHA, 2004, p. 19-25).

### 3.2 Tipologia de *corpus*

A tipologia de *corpus*, em LC, busca descrever o desenho de um *corpus* e seu propósito de uso. A tipologia dos *corpora* desta pesquisa é baseada em Berber Sardinha (2004) e em Teixeira (2008), detalhada na seção 4.3.

Citamos algumas características apontadas por Berber Sardinha (2004, p. 20-21) relacionadas ao perfil dos *corpora* desta pesquisa. O autor inicia a descrição de tipologia de *corpus* com o critério escrito ou falado, sendo que o escrito pode ser impresso ou não. O segundo critério, o tempo, pode ser subdividido em: sincrônico, diacrônico, contemporâneo ou histórico.<sup>4</sup> O próximo critério, a seleção, se subdivide entre amostragem (estático, em oposição ao dinâmico), monitor (subdividido em dinâmico ou orgânico), e equilibrado.<sup>5</sup> O quarto critério, conteúdo, trata do texto

---

<sup>4</sup> O sincrônico abarca um período de tempo, enquanto o diacrônico abarca vários; o contemporâneo descreve o *corpus* de um tempo presente e o histórico o de um período de tempo passado.

<sup>5</sup> O *corpus* de amostragem envolve textos ou fragmentos de várias tipologias e gêneros textuais, como uma amostra finita da linguagem, em oposição ao *corpus* monitor, que descreve o estado atual da língua; o *corpus* dinâmico ou orgânico é flexível, pode ser aumentado ou diminuído diferentemente do estático (monitor); e o equilibrado é composto por textos em quantidades ou qualidades (tipologia, por exemplo) semelhantes. Neste projeto, o critério dinâmico ou orgânico descreve bem nossos *corpora*, no sentido de que podem ser aumentados com a adição de novas subáreas da Linguística; já critério equilíbrio é baseado no número de *tokens* que compõe os *corpora*. Por nossos *corpora* serem destinados a uma pesquisa terminográfica na área de Linguística, compilamos,

regional ou dialetal, especializado e multilíngue, o que nesta pesquisa significa *corpora* especializados da área de Linguística e multilíngues por abranger a Língua Portuguesa (doravante LP) e a Língua Inglesa (doravante LI).

Quanto à autoria, pode ser de aprendiz ou de língua nativa, isto é, conter textos de falantes não nativos ou nativos. O critério disposição interna trata se os textos do *corpus* são comparáveis (ex.: original e traduzido) ou se é alinhado (com linhas de tradução abaixo do texto original). A finalidade do *corpus* pode ser de (a) estudo, de (b) referência, de (c) treinamento ou teste. O objetivo do *corpus* de (a) estudo é descrever a língua; o de (b) referência contrasta com outros *corpora*; e o de (c) treinamento ou teste serve como subsídio para teste de ferramentas lexicais. Em nossos projetos, a finalidade do *corpus* é de estudo, já que os dados serão extraídos para definição terminológica.

Em comparação com a tipologia de Berber Sardinha (2004), a tipologia de Teixeira (2008) difere quanto ao item nível de codificação, o qual abrange metadados como cabeçalho e etiquetagem. Em ambos os projetos descritos neste artigo, a etiquetagem não se fez necessária, pois a ferramenta de palavras-chave do WST cumpriu a tarefa de isolar os termos, objeto das pesquisas terminográficas. Quanto aos cabeçalhos, eles foram inseridos manualmente e neles constam o título do artigo, a data de coleta e o *site* de origem.

### 3.3 O que é um *corpus* colaborativo?

Acreditamos que, em primeiro lugar, devamos situar o que consideramos um *corpus* colaborativo. É importante especificar, pois a ideia pode advir de algo generalista: Gardner, Krowne e Xiong (2010), por exemplo, consideram a Wikipedia como um grande *corpus* colaborativo.<sup>6</sup>

A proposta de compilação de *corpora* também pode abranger vários níveis de colaboração, como explicitada por Mello (2014), que

---

especificamente, textos especializados nessa área. Assim sendo, nosso foco foi a compilação de textos do gênero acadêmico, subdivididos em artigos científicos, resenhas e material instrucional, como manuais, aulas, livros, apostilas e resenhas.

<sup>6</sup> Em nenhum momento, no texto referido, os autores citam a Wikipedia como ideia de um *corpus* com a mesma conotação dada pela Linguística de *Corpus*. A proposta aqui é mais genérica, trabalhando com a definição de *corpus* como conjunto de textos que, no caso, são construídos colaborativamente.

trabalha com *corpora* orais (espontâneos de fala) no projeto C-ORAL-BRASIL, envolvendo a descrição de textos das gravações, anotações e a inserção de metadados sociolinguísticos, por exemplo; todo o trabalho é pensado coletivamente e deve ser gerido por um ou mais grupos de estudo, em nível nacional ou internacional. Nessa empreitada de compilação, há vários estágios a serem desenvolvidos pelos grupos: planejamento da arquitetura do *corpus* (com uma grande preocupação em relação às questões de tamanho e balanceamento), gravações (e todos os aparatos técnicos necessários), transcrições (obedecendo a critérios rigorosíssimos dentro de um projeto maior, no qual se insere), revisões das transcrições (inclusive com análises estatísticas), segmentação e alinhamento dos formatos de saída do *corpus*.

Nossa proposta, aqui, é muito mais restrita e voltada para a sala de aula. Bowker (2003) desenvolveu projetos semelhantes com alunos de tradução.<sup>7</sup> Embora não tivéssemos conhecimento de seu texto na época que começamos a trabalhar com nosso *corpus* colaborativo, percebemos que suas preocupações foram semelhantes às nossas: devemos usar um *corpus* já existente ou montamos um *corpus* específico para nossos experimentos? Os alunos têm fácil acesso aos computadores para compilar *corpora*? Um trabalho coletivo não poderia gerar um *corpus* maior e mais representativo? Tudo o que está disponível na internet pode virar *corpus* de estudo? Nossos *corpora* precisam de etiquetagem? Todas essas questões levam a uma pergunta básica: qual a metodologia a ser adotada para o trabalho colaborativo? Discutiremos, mais adiante, quais foram as nossas propostas metodológicas.

Diante do exposto, definimos *corpus* colaborativo como um *corpus* compilado por vários pesquisadores, em momentos diferentes, construído especificamente para atender à construção de um vocabulário bilíngue de Linguística. Como resultado, obtivemos um *corpus* mais representativo das subáreas da Linguística, sem necessidade de etiquetagem, já que o foco da pesquisa foram os termos da área.

---

<sup>7</sup> Diferente de nós, que trabalhamos com *corpora* comparáveis (textos de línguas diferentes, pertencentes à mesma área ou subárea de conhecimento, porém não se constituem como traduções), a autora trabalhou com seus alunos *corpora* paralelos (no par francês-inglês, tradução do mesmo texto de uma língua à outra). O tamanho adotado para cada *corpus* também foi diferente: enquanto partimos do pressuposto de um *corpus* com, pelo menos, 500 mil palavras em cada subárea, a autora trabalhou com *corpora* entre 20 e 50 mil palavras.

### 3.4 Trabalho terminográfico

Existem várias teorias que podem ser usadas para trabalhos terminográficos. Durante todo o projeto relatado aqui, pautamo-nos por duas delas. A primeira diz respeito à tipologia para obras de consulta ao léxico, sistematizada por Barbosa (2001), que divide a elaboração de obras lexicográficas e terminográficas entre (a) dicionário, (b) vocabulário e (c) glossário. De acordo com a autora (2001, p. 36), o (a) dicionário busca compilar e registrar as palavras de frequência regular situadas no campo de várias normas; o (b) vocabulário registra vocábulos delimitados por uma situação comunicativa, podendo estar relacionados a usos em espaços geográficos e grupos sociais diferentes e pertencentes ao mesmo universo de discurso; e o (c) glossário registra palavras provenientes de um texto ou discurso específico relacionadas ao mesmo tempo, espaço geográfico e contextos de uso da palavra. O vocabulário pode se subdividir em: (a) vocabulário fundamental e (b) vocabulário técnico-científico e especializados. O vocabulário fundamental abarca os vocábulos frequentes e regularmente distribuídos de uma comunidade ou de um- grupo social específico. Em nosso contexto, o vocabulário técnico-científico e especializados, que é um recorte de língua (de grupos profissionais), foi o escolhido para o trabalho com os alunos. Este vocabulário trabalha com uma área de especialidade, sua unidade é o termo (com significado restrito e alta frequência), o verbete pode apresentar mais de uma acepção (mas apenas dentro da área escolhida) e a perspectiva desse vocabulário é sincrônica e sinfásica, ou seja, relativo ao estilo de língua, tais quais familiar, formal ou literário.

A segunda, Teoria Comunicativa da Terminologia, de Cabré (2000), elabora o conceito de “termo”. Entre as várias inovações propostas em sua teoria em relação à Teoria Geral da Terminologia de Wüster (publicada originalmente em 1931), de caráter prescritivo e que a precede, temos: (1) caráter descritivo que perpassa todo um projeto; (2) uso de *corpora* na coleta dos termos; (3) concepção que o termo é, basicamente um substantivo (embora outras palavras lexicais, como adjetivos e verbos, também possam aparecer como características de uma determinada área); (4) um termo **está** termo dentro de determinada área (geralmente com significado monossêmico), mas pode, ao mesmo tempo, pertencer à língua geral, com significados polissêmicos.

### 3.5 A importância da árvore de domínio

A construção da árvore de domínio ou conceitual<sup>8</sup> é um passo metodológico da Terminologia, como parte do percurso semasiológico,<sup>9</sup> que permite o levantamento e a delimitação conceitual dos termos a serem pesquisados e definidos numa obra terminográfica. Segundo Aubert (2001), a árvore de domínio, com suas áreas e subáreas definidas, contribui para evitar o que o autor denomina riscos de “ruído” e de silêncio, isto é, a presença de termos não relevantes a uma subárea ou a ausência de termos essenciais a uma área ou subárea específica. A árvore pode servir tanto ao pesquisador quanto aos leitores da obra desenvolvida, proporcionando-lhes uma visão da abrangência do tema e do trabalho desenvolvido.

A árvore de domínio contribui para a sistematização dos termos mais frequentes (procedimento este possível graças à LC) e para a categorização por subáreas numa pesquisa (ALVES *et al.*, 2010; BARROS, 2004; FROMM, 2015, 2018). Ao adotarmos a LC como abordagem e metodologia para análise dos dados linguísticos, essa sistematização se dará a partir da lista de palavras-chave (candidatos a termos que podem ser classificados, posteriormente, nos subdomínios da árvore de proposta). Em seguida, por exemplo, analisando os cotextos e contextos em linhas de concordância, podemos identificar os semas ou traços semânticos que servirão à elaboração de um verbete.

### 3.6 Definição terminográfica baseada em *corpus*

Existem várias maneiras de construir uma definição, seja para fins lexicográficos, seja para fins terminográficos (como bem explicado no trabalho de CARDOSO, 2017). De modo geral, e para nosso projeto,

---

<sup>8</sup> A árvore de domínio é um esquema representativo de como se dividem as principais subáreas de uma grande área de conhecimento (como a árvore do CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico –, por exemplo). Pode ser em formato de diagrama ou disposta em formato topicalizado. Ela pode ser levantada através de pesquisa com especialistas na área e/ou (como propomos) através da análise de *corpus* especializado.

<sup>9</sup> Percurso semasiológico, no contexto da Terminologia e da Terminografia, é quando partimos do termo para chegarmos ao seu conceito e definição, contrário ao percurso onomasiológico, que parte do conceito para chegar ao termo. Definir o que é Biolinguística, por exemplo, faz parte do percurso semasiológico.

em específico, usa-se o modelo de definição GPDE (Gênero Próximo, Diferenças Específicas) (FROMM; YAMAMOTO, 2020), também conhecido como Aristotélico ou por compreensão. O que Fromm (2007) propôs, no entanto, é que a definição, cujo modelo foi aplicado no *software* que elaborou para sua tese, fosse elaborada única e exclusivamente por meio de uma Análise Componencial (ILARI, 2003) de semas e palavras advindos de excertos coletados no *corpus* de pesquisa; ou seja, a definição terminográfica, no nosso projeto, deve ser baseada em *corpus*. Para os excertos, procuramos levantar aqueles que contivessem mais contextos definitórios e explicativos. Segundo Aubert (2001, p. 69), o contexto definitório descreve “o conjunto completo dos traços conceptuais distintivos do termo” e o contexto explicativo descreve alguns “traços conceptuais pertinentes específicos do termo”, geralmente relacionados à “materialidade, finalidade, funcionamento e similares”. Em relação à validação dos termos por parte de especialistas, preferimos usar a abordagem proposta por Tagnin (2012): não necessariamente precisamos da validação de especialistas para termos extraídos a partir de *corpora* sólidos e bem montados e que possam ser considerados altamente confiáveis. Respaldados em Tagnin (2012, p.169) consideramos que o maior especialista de uma ciência que possa validar verbetes sobre a área é o *corpus* que montamos sobre ela, não os profissionais que dela fazem parte, pois é difícil algum pesquisador/usuário ter noção de sua área como um todo.

### 3.7 Reciclagem de *corpus*

Na LC, o termo **reciclar**<sup>10</sup> pode ser usado com mais de um significado. Ele pode ser utilizado, por exemplo, para descrever a análise da padronização lexical, nas linhas de concordância, aplicada ao ensino e aprendizagem de língua estrangeira e de segunda língua (PÉREZ-PAREDES; SÁNCHEZ-TORNEL; CALERO, 2012; SINCLAIR, 2003). Sinclair (2003, p. XIV-XIX) subdivide esse procedimento de análise em sete outros passos: início, interpretação, consolidação, relato, reciclagem, resultado e repetição. Nesse contexto, o resultado, segundo o autor, significa elaborar uma lista de hipóteses finais sobre as informações provenientes das linhas de concordâncias e associá-las ao relatório final do nó da análise inicial.

---

<sup>10</sup> *Recycle* ou *reuse* em inglês.

Outro conceito existente na LC aplicado à reciclagem de *corpora* consiste na reutilização de *corpora* para atender objetivos específicos, distintos daqueles de seu projeto original. Por exemplo, os passos metodológicos que fazem parte do processo de reciclagem em contexto de tradução com *corpora* paralelos são: a coleta de material, a extração dos dados reutilizáveis e o refinamento e a aplicação dos dados extraídos (TIEDEMANN, 2003). A coleta de material trata do levantamento dos *corpora* paralelos utilizados para análise das possíveis traduções. Esses *corpora* etiquetados e alinhados podem ser aplicados, como recurso computacional, na lexicografia e na tradução automática.

O objetivo da reciclagem do *corpus* de Linguística (do primeiro projeto) que propomos é, primeiramente, não perder o produto resultante de um trabalho prévio. Em segundo lugar, reduzir tempo de uma nova pesquisa, caso *corpora* prévios possam ser reutilizados para análises linguísticas de um novo projeto ou para dar continuidade a um projeto preexistente, como foi nosso caso.

Quando os *corpora* compilados são grandes e provenientes de um período longo de coleta (abrangendo anos), é necessário considerar algumas variáveis: (1) **disponibilidade** atual do *corpus*; (2) **limpeza** a ser feita; e (3) **tipologia**, quando se trata da fidelidade ao tipo ou gênero textual.

A **disponibilidade** de um *corpus* trata da questão da disponibilidade, na Internet, dos arquivos que o compõem; isto é, devido à volatilidade da rede mundial, alguns textos podem ou não continuar disponíveis por muito tempo. Caberá aos pesquisadores se os textos já coletados servirão ou não aos objetivos de suas pesquisas e, então, definirem se vão mantê-los ou excluí-los de seus *corpora*. Essa indisponibilidade na Internet se deve a fatores como: a mudança do endereço na web, a retirada do arquivo da Internet ou a alocação do arquivo em outro *site* ou base de dados.

A **limpeza** é relacionada à remoção de dados não relevantes ao conteúdo linguístico que se objetiva analisar em um *corpus* (elementos pré e pós-textuais, referências bibliográficas, dados de quadros e tabelas que não estejam inseridos como figuras, elementos textuais em língua estrangeira etc.). Utilizamos os comandos de *Control F* e *Substituir* para localizar e deletar esses dados de forma manual.

A padronização e o balanceamento da **tipologia** ou o **gênero dos textos** são também fatores importantes ao se considerar a composição de *corpora*: o desenho dos *corpora* deve prever quais tipos e quantos

textos (ou *tokens*,<sup>11</sup> isto é, o número de itens ou ocorrências de um vocábulo que comporão o *corpus*) farão parte desses *corpora*, sob pena de, se mal planejados, apresentar resultados de análise que podem ser desacreditados.

Por fim, a partir de um *corpus* inicial, após ser limpo e equilibrado, é possível obter um novo *corpus* alterado a fim de atender aos propósitos do projeto de pesquisa. Em nosso caso, o *corpus* inicial serviu à construção do *corpus* final (reciclado) e à construção do vocabulário bilíngue de Linguística (YAMAMOTO, 2018).

### 3.8 Padronização de *corpora*

A ideia de padronização se refere ao ato de equilibrar o tamanho, o gênero e o subgênero textual e as demais características dos *corpora*, que devem ser previstas no desenho dos mesmos. O tamanho de cada *corpus* (de cada projeto) e dos sub*corpora* que o compõem, no nosso caso, deveu-se a uma pesquisa prévia (FROMM, 2007).<sup>12</sup> Também foram feitas pesquisas (notadamente com o buscador Google) para descobrir, a partir da área pesquisada (Linguística), a prevalência de gêneros (textos acadêmicos e instrucionais) e subgêneros (teses, dissertações, artigos científicos, resenhas e manuais).

## 4 Metodologia

Destacamos, nos próximos subitens, os passos seguidos na elaboração do *corpus* colaborativo de Linguística (primeiro projeto) e sua posterior reciclagem (segundo projeto). Para os *corpora* aqui descritos, adotamos as abordagens impressionística, mencionada por Berber

---

<sup>11</sup> Na frase “O maracujá e o abacate são frutas” há 7 *tokens*.

<sup>12</sup> Em sua tese (FROMM, 2007), um dos autores notou, em suas conclusões, que para a elaboração de trabalhos terminográficos baseados em *corpora*, um *corpus* de 30 mil palavras para cada subárea da computação (*corpus* escolhido para testar o *software* desenvolvido para a tese) não era suficiente para a elaboração de definições terminológicas. A partir dessa constatação, tomou-se como ponto de partida, para a compilação dos *corpora* de Linguística aqui descritos, um tamanho cerca de 16 vezes superior para cada subárea, de modo que os requisitos de elaboração de um vocabulário criado única e exclusivamente a partir de *corpora* (especialmente no tocante à elaboração das definições) pudessem ser contemplados.

Sardinha (2004), e histórica, baseada na experiência da compilação de *corpus* para um trabalho terminográfico anterior (FROMM, 2007).

#### 4.1 A elaboração da árvore de domínio

Existem várias maneiras de elaborar a taxonomia de um domínio da ciência. Para os nossos *corpora*, foi pensada uma pergunta básica para a escolha das subáreas que compõem a árvore: qual seu objeto de estudo? Por exemplo, a Morfologia analisa as classes de palavras, os morfemas, as unidades mínimas de sentido etc. Uma pergunta básica como essa norteou todo o design dos nossos *corpora*. Não entraram aqui, por exemplo, teorias associadas às diferentes áreas da Linguística<sup>13</sup> ou metodologias<sup>14</sup> usadas em pesquisas na área. Linguística de *Corpus* e a Linguística Computacional, por exemplo, são muito mais abordagens e metodologias do que áreas específicas.

Voltando à pergunta mencionada no parágrafo anterior (qual o objeto de estudo?), a LC, ao nosso ver, não tem **um** objeto específico de pesquisa, já que pode ser usada por quase todas as áreas elencadas na árvore, por isso a consideramos uma abordagem (computacional) e metodologia (muito bem delimitada) de pesquisa.

Inicialmente a árvore foi concebida por Fromm (2007) e a metodologia adotada para a elaboração foi a consulta a livros da área de Linguística. O autor decidiu pela divisão da grande área entre Linguística Teórica e Linguística Aplicada, inserindo nelas subáreas: 25 subordinadas à Linguística Teórica e 6 à Linguística Aplicada.

Nas pesquisas terminográficas tradicionais, a elaboração da árvore de domínio, com ajuda de especialistas da área a ser estudada, é essencial para que os termos sejam classificados. Nesta pesquisa, essa árvore possui certa plasticidade, pois pode ser alterada à medida que o *corpus* é compilado. Na Figura 1, observamos a árvore prototípica que deu início ao projeto (FROMM, 2007, p. 39).

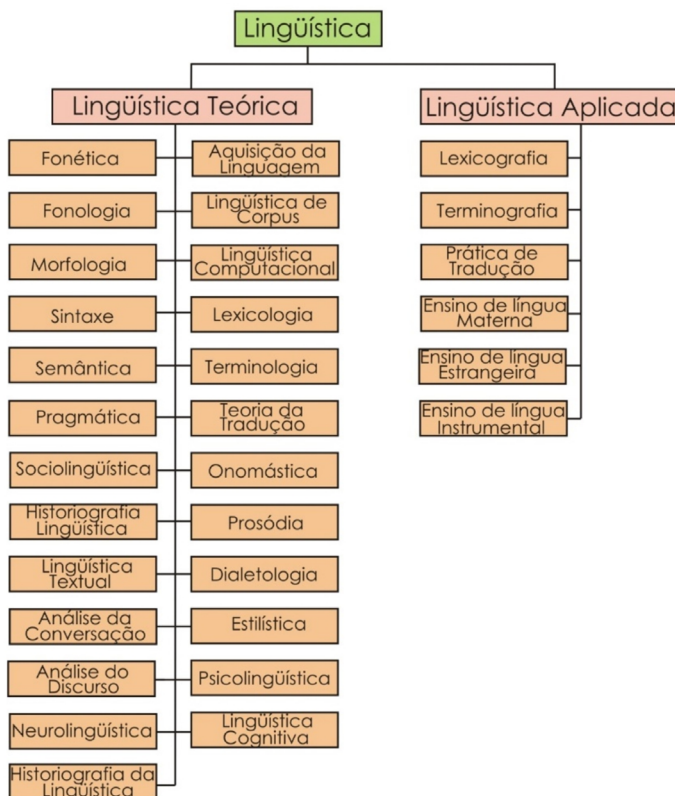
---

<sup>13</sup> Note-se que uma mesma teoria (como o Gerativismo ou a Linguística Sistêmico-Funcional) pode ser encontrada como base teórica em várias subáreas da nossa árvore; ou, ainda, uma subárea pode ser analisada por diferentes teorias linguísticas.

<sup>14</sup> Essa é a percepção atual dos autores, diferente do que é apresentado na figura 1, quando os questionamentos sobre teoria, metodologia e objeto de estudo ainda estavam sendo discutidos.

FIGURA 1 – Primeira versão da Árvore de Domínio da Linguística

## *Árvore do Campo da Lingüística*



Fonte: Fromm (2007, p. 39).

Com a compilação dos *corpora* (LP e LI) e o desenvolvimento da pesquisa nas diversas turmas que cursaram a disciplina, o formato da árvore foi mudando. Um passo adotado quanto à escolha de qual área a dupla de alunos trabalharia pode ser verificado na Figura 2. Conforme as áreas eram pesquisadas, as cores nos retângulos correspondentes a cada subárea mudavam:

- a) vermelho para as subáreas consideradas completas (com, no mínimo, 500 mil *tokens*), impedindo que novos grupos escolhessem aquelas subáreas;

- b) amarelo para as subáreas que ainda não estavam completas, pois ainda faltavam *tokens* em uma ou ambas as línguas<sup>15</sup> – no caso, uma nova dupla trabalharia para completar o que estava faltando (os textos já coletados eram repassados ao novo grupo);
- c) verde para as áreas ainda não trabalhadas e que poderiam ser livremente escolhidas pelos alunos.

FIGURA 2 – Versão intermediária da Árvore de Domínio da Linguística



Fonte: Elaborada pelos autores

<sup>15</sup> O desenrolar da compilação do *corpus* nos mostrou que determinadas subáreas possuíam grande quantidade de textos em língua inglesa, mas poucos textos em língua portuguesa, indicando um desbalanço no estado da arte de língua para língua. No caso específico da Ecolinguística, foram necessários 4 anos para que a quantidade mínima fosse atingida em língua portuguesa.

A plasticidade da árvore, já citada anteriormente, pode ser notada entre as três figuras aqui apresentadas. Nota-se, por exemplo, que **Pesquisa Narrativa** aparece somente na Figura 2; na época, incluída por sugestão de uma especialista. Posteriormente, decidimos que certas áreas, como Pesquisa Narrativa, seriam consideradas metodologias/abordagens e não áreas com objetos de pesquisa bem delimitados, sendo retiradas das árvores (e seus respectivos *subcorpora* do *corpus* de LA de cada língua).

A elaboração da árvore a partir das informações advindas do *corpus* nem sempre fica clara. Por exemplo, quando da condução dos estudos de mestrado de um dos autores (YAMAMOTO, 2015), observamos que, apesar de semelhantes em relação à abordagem diacrônica da língua, as subáreas da Etimologia, Filologia e Linguística Histórica haviam se consolidado como disciplinas em épocas diferentes (já constando da FIGURA 2 também), e eram distintas quanto ao objeto de estudo. Essas subáreas foram fundidas ou desmembradas, entrando e saindo da árvore (e dos *corpora*) várias vezes.

A plasticidade na construção da árvore de domínio da Linguística fez parte de uma correspondência biunívoca entre a montagem da taxonomia da área e o *corpus* dessa área. A partir da árvore inicial (FIGURA 1) e passando pelas diversas fases de sua elaboração (como na FIGURA 2), conforme o *corpus* era elaborado, os pesquisadores percebiam, nas fontes consultadas, que novas subáreas da Linguística (como a Biolinguística e a Linguística Diacrônica, já presentes na FIGURA 3) iam surgindo com o tempo (num período de 10 anos). Conforme o *corpus* demonstrava a existência de novas subáreas, as mesmas iam sendo incorporadas à árvore e novos levantamentos foram feitos para completar os 500 mil *tokens* em cada uma delas. Essa metodologia de elaboração de árvore e compilação de *corpus*, acreditamos, poderia ser considerada uma complementação à definição de pesquisa direcionada pelo *corpus* (*corpus driven*) proposta por Viana e Tagnin (2015, p. 323): “[...] estudo que se desenvolve conforme dados apresentados pelo *corpus*, sem pressuposições teóricas”.

FIGURA 3 – Versão atual da Árvore de Domínio da Linguística



Fonte: Elaborada pelos autores.

Algumas decisões sobre como nomear áreas e subáreas na árvore, assim como a profundidade de análise, foram tomadas como um consenso entre os autores deste texto. Um exemplo é a questão de quantos níveis deveríamos descer na estrutura. Algumas áreas, como a Onomástica, por exemplo, ainda podem conter subníveis (Toponímia e Antroponímia, no caso). Devido à complexidade do trabalho e à indisponibilidade de tempo, decidiu-se por uma árvore de apenas 3 níveis.

No segundo nível da árvore, podemos notar uma renomeação de Linguística Teórica para Linguística e, finalmente, para Linguística Descritiva (LD). Para fazermos nossa escolha, consideramos o caráter descritivo da Linguística como ciência, e não o caráter prescritivo da Gramática. No intuito de corroborar nossa escolha, fizemos a busca das definições nos *corpora* de estudo em LP e encontramos a definição de LD em contraste com a LA, conforme Albuquerque (2017, p. 227):


[...] as duas principais subáreas da linguística de acordo com a necessidade de pesquisa e contribuições científicas são a linguística descritiva e a linguística aplicada. A primeira por sua clara importância documental e analítica das línguas do mundo. A segunda por suas contribuições significativas no campo educacional, que procuram criar uma ponte entre teoria e prática, entre universidade e comunidade.

Temos aí uma citação que compara as duas Linguísticas e as contrasta, de forma que os argumentos são claros e objetivos, atendendo nossa proposta. Com exceção da terminologia adotada na Figura 2 (que pode ser considerada um erro, pois a subárea de segundo nível e a área principal, no primeiro nível, possuíam a mesma designação), a preferência pela nomeação Linguística Descritiva pode ser considerada como uma decisão política dos pesquisadores.

É importante notar, no entanto, que a árvore proposta na Figura 3<sup>16</sup> (derivada da pesquisa de doutorado de um dos autores) não é definitiva;<sup>17</sup> nenhuma árvore de domínio pode ser considerada finalizada, pois subáreas aparecem (e, por que não, desaparecem), conforme o desenvolvimento científico. E, já que esses projetos tratam da compilação de *corpora* a partir de uma correspondência biunívoca,<sup>18</sup> temos que reconhecer as limitações desses projetos: nem todas as possíveis subáreas da Linguística podem estar, em termos de quantidade de textos, igualmente representadas na internet. Um exemplo dessa questão foi a área de **Etimologia** (assim

<sup>16</sup> Há outras versões da árvore, como as propostas em Fromm e Yamamoto (2013) e Fromm (2018).

<sup>17</sup> Especialmente porque, conforme mencionado, não continuamos a pesquisa para um possível quarto nível.

<sup>18</sup> Dados levantados a partir do *corpus* sugerem a (re)elaboração dos subcampos da árvore  os subcampos da árvore (pesquisados pelos autores e sugeridos por colegas pesquisadores) indicam os *subcorpora* a serem levantados.

como a de Ecolinguística, já citada), cujo *corpus* demorou vários anos para ser compilado durante os projetos; nossa interpretação do fato é que, como novas subáreas aparecem (e, com elas, paulatinamente, novos textos vão sendo escritos), outras podem estar com uma baixa produção textual, ao menos no momento de compilação do *corpus*;<sup>19</sup> logo, a compilação de seus respectivos *corpora* precisaria passar por uma metodologia de coleta que não simplesmente o *download* de textos disponíveis na rede (como, por exemplo, escanear textos impressos, passá-los por um OCR<sup>20</sup> e disponibilizá-los eletronicamente).

#### 4.2 Os gêneros adotados

Desde o início, o gênero adotado para a compilação das diversas fases dos *corpora* foi o acadêmico, subdividido pelos subgêneros: tese, dissertação e artigo científico. Posteriormente, na pesquisa de doutorado de um dos autores, por sugestão de um dos membros de sua banca de mestrado (YAMAMOTO, 2015), foram incluídos materiais instrucionais (manuais).<sup>21</sup>

#### 4.3 O *corpus* colaborativo

A tipologia (de acordo com Teixeira (2008)) do *corpus* colaborativo de Linguística (primeiro projeto), compilado pelos alunos, é descrita no Quadro 1.

---

<sup>19</sup> O *corpus* desta pesquisa abarca produções compiladas entre os anos de 2010 e 2020, cujos textos podem ter sido publicados antes de 2010 até 2018, sendo, portanto, um *corpus* contemporâneo.

<sup>20</sup> Em inglês, *Optical Character Recognition*, que, em português, significa Reconhecimento óptico de caracteres é uma tecnologia que permite a conversão de imagens, textos em PDF, documentos escaneados, entre outros, em textos legíveis e editáveis.

<sup>21</sup> Manuais, aqui, são livros gerais de Linguística voltados aos alunos, pois eles contêm traços semânticos mais fáceis de serem identificados (= apresentam mais contextos definitórios) e usados em pesquisas terminológicas ou terminográficas.

QUADRO 1 – Tipologia do *corpus* colaborativo do primeiro projeto

<b>Língua</b>	Bílingue (inglês e português)
<b>Modo</b>	Escrito (textos acadêmicos: artigos científicos, dissertações e teses)
<b>Data de publicação</b>	Sincrônico (levantamento realizado entre 2010 e 2018), fechado <sup>22</sup>
<b>Seleção</b>	Estático
<b>Conteúdo</b>	Especializado (Linguística)
<b>Autoria</b>	Falantes nativos/não nativos (inglês e português), individual/coletivo
<b>Disposição Interna</b>	Comparável
<b>Uso na pesquisa</b>	Estudo (análise terminológica/terminográfica)
<b>Tamanho</b>	Grande (mais de 10 milhões de palavras)
<b>Nível de Codificação</b>	Com cabeçalhos, <sup>23</sup> sem etiquetas <sup>24</sup>

Fonte: Elaborado pelos autores.

Em relação ao tamanho do *corpus* desse primeiro projeto, podemos verificar os dados na Tabela 1.

<sup>22</sup> Mesmo com algumas subáreas incompletas, foi fechado por causa do encerramento do trabalho com os alunos neste *corpus* de Linguística.

<sup>23</sup> Informações de coleta delimitadas por um sinal de menor e maior, que permite ignorá-las nas buscas feitas nas ferramentas de análise (ver FIGURA 4).

<sup>24</sup> Nunca houve, por parte dos pesquisadores, preocupação em etiquetar as diferentes versões do corpus. Como o objetivo da compilação era o trabalho terminográfico, as ferramentas básicas (lista de palavras, palavras-chave e concordanciador) dos programas de análise lexical (conforme o nome já sugere, os mesmos foram desenvolvidos para trabalhos a partir do léxico) não necessitavam de etiquetagem para o trabalho de elaboração de um vocabulário na área de Linguística. Embora a teoria de Cabré indique um destaque para substantivos como candidatos a termos, é muito mais fácil, para os pesquisadores, usarem a ferramenta palavras-chave do que etiquetar o corpus (e ter que revisar todas as etiquetas, além do problema de achar um bom etiquetador em língua portuguesa).

TABELA 1 – Tamanho do *corpus* colaborativo

	Português	Inglês
Número de textos	1.873	1.911
<i>Tokens</i>	26.856.704	29.369.816
<i>Types</i> <sup>25</sup>	353.700	429.384
<i>Type/token ratio (TTR)</i> <sup>26</sup>	1,32%	1,46%
Número Total de <i>Tokens</i>	56.226.520	

Fonte: Elaborada pelos autores.

#### 4.4 Armazenamento de *corpora*

A organização e o arquivamento de *corpora* em uma máquina são passos muito importantes, dos quais depende o acesso eficaz aos arquivos do projeto, isto é, o *corpus*, as listas de palavras, as listas de palavras-chave e de outros documentos necessários à produção das obras terminográficas.

##### 4.4.1 *Corpus* colaborativo

Podemos descrever a estruturação de compilação do projeto inicial através da Figura 4. Não houve, no começo do projeto, uma preocupação em salvar os arquivos com algum título, tipo de arquivo ou codificação específicos; os alunos apenas salvavam os artigos que baixavam em formato .txt com seus nomes completos (retângulo vermelho). Foram inseridos, a partir do segundo ano, pequenos cabeçalhos (no começo dos textos), nos quais constavam data e endereço de coleta (retângulo verde).

<sup>25</sup> *Types* são o total de formas, ou o número de vocábulos presentes em um dado *corpus*, ou seja, na frase “O maracujá e o abacate são frutas” há 6 *types*, já que o artigo *o* repete duas vezes.

<sup>26</sup> *Type/token ratio (TTR)* é o resultado proveniente da divisão das formas (*types*) pelos itens (*tokens*) de um *corpus* expressa em porcentagem. Esta razão indica a riqueza lexical de um *corpus*, de forma que quanto maior for esta razão, maior será o número de palavras diferentes contidas naquele *corpus*.

Para montar a estrutura de compilação, simplesmente reproduzimos a estrutura da árvore de domínio da Linguística (quadro azul).<sup>27</sup>

FIGURA 4 – Estruturação inicial do projeto



Fonte: Elaborada pelos autores.

#### 4.4.2 Corpus reciclado

Para a execução do segundo projeto, organizamos uma pasta em um disco rígido onde foram arquivados os *corpora* de Linguística, em LP e LI, e essa pasta foi compartilhada entre os pesquisadores por meio do OneDrive, ferramenta usada para a disponibilização de dados nas nuvens. Na Figura 5, podemos ver as pastas destes *corpora*.

<sup>27</sup> Obviamente a estruturação das pastas foi sendo alterada de acordo com as revisões da árvore de domínio. Nesta figura, podemos perceber que a estrutura está mais próxima da árvore mostrada na Figura 2.

FIGURA 5 – Pasta para arquivamento dos *corpora* de Linguística

OneDrive > Normalizado G e M - versão de ajuste			
Nome		Data de modificação	Tipo
Arvore de domínio		05/12/2019 17:08	Pasta de arquivos
Concord		02/03/2020 01:45	Pasta de arquivos
Ing_Corpus de Linguística		22/01/2020 16:08	Pasta de arquivos
KWL		14/11/2019 10:53	Pasta de arquivos
Metodologias		13/12/2019 11:38	Pasta de arquivos
Port_Corpus de Linguística		13/12/2019 11:36	Pasta de arquivos

Fonte: Elaborada pelos autores.

A Figura 5 mostra que os *corpora* de Linguística foram arquivados numa pasta RAIZ, nomeada **Normalizado G e M** (iniciais dos pesquisadores). Dentro dessa pasta, há as subdivisões nas duas línguas do projeto, ou seja, em LP, Port\_Corpus de Linguística, e em LI, Ing\_Corpus de Linguística. Ao abrir essas pastas, encontramos três subpastas: as duas subáreas da Linguística e a pasta com o *corpus* de manuais de Linguística. Veja a Figura 6.

FIGURA 6 – Subpastas: LD, LA e manuais de Linguística

OneDrive > Normalizado G e M - versão de ajuste > Port_Corpus de Linguística			
Nome		Data de modificação	Tipo
Linguística Aplicada		20/04/2019 08:49	Pasta de arquivos
Linguística Descritiva		22/10/2019 23:24	Pasta de arquivos
Manuais L PT		03/03/2020 04:48	Pasta de arquivos



























Fonte: Elaborada pelos autores.

Na Figura 6, encontramos a pasta de Linguística Aplicada, com suas 18 subáreas, a pasta de Linguística Descritiva, com suas 29 subáreas e o *corpus* com os 13 manuais de Linguística em LP. O mesmo acontece para a LI e cinco manuais de Linguística, que podem ser vistos na Figura 7.



FIGURA 8 – *Subcorpus* de Biolinguística: nomeação dos arquivos

Normalizado G e M - versão de ajuste &gt; Port\_Corpus de Linguística &gt; Linguística Descritiva &gt; Biolinguística

Nome	Data de modificação	Tipo	Tamanho
 L_BL_PT_A_TXT1.txt	05/03/2020 11:48	Documento de Te...	88 KB
 L_BL_PT_A_TXT2.txt	05/03/2020 11:48	Documento de Te...	113 KB
 L_BL_PT_A_TXT3.txt	05/03/2020 11:48	Documento de Te...	81 KB
 L_BL_PT_A_TXT4.txt	05/03/2020 11:48	Documento de Te...	120 KB
 L_BL_PT_A_TXT5.txt	05/03/2020 11:48	Documento de Te...	148 KB
 L_BL_PT_A_TXT6.txt	05/03/2020 11:48	Documento de Te...	50 KB
 L_BL_PT_A_TXT7.txt	05/03/2020 11:48	Documento de Te...	178 KB
 L_BL_PT_A_TXT8.txt	12/03/2020 15:25	Documento de Te...	109 KB
 L_BL_PT_A_TXT11.txt	05/03/2020 11:48	Documento de Te...	60 KB
 L_BL_PT_A_TXT13.txt	05/03/2020 11:48	Documento de Te...	96 KB
 L_BL_PT_A_TXT14.txt	05/03/2020 11:48	Documento de Te...	80 KB
 L_BL_PT_A_TXT16.txt	05/03/2020 11:48	Documento de Te...	30 KB
 L_BL_PT_A_TXT23.txt	12/03/2020 15:41	Documento de Te...	85 KB
 L_BL_PT_A_TXT24.txt	12/03/2020 15:46	Documento de Te...	89 KB
 L_BL_PT_A_TXT25.txt	05/03/2020 11:48	Documento de Te...	1.046 KB
 L_BL_PT_A_TXT26.txt	05/03/2020 11:48	Documento de Te...	54 KB
 L_BL_PT_A_TXT28.txt	05/03/2020 11:48	Documento de Te...	104 KB
 L_BL_PT_A_TXT29.txt	05/03/2020 11:48	Documento de Te...	110 KB
 L_BL_PT_D_TXT15.txt	05/03/2020 11:48	Documento de Te...	336 KB
 L_BL_PT_D_TXT30.txt	12/03/2020 16:01	Documento de Te...	165 KB
 L_BL_PT_MIA_TXT18.txt	05/03/2020 11:48	Documento de Te...	10 KB
 L_BL_PT_T_TXT12.txt	05/03/2020 11:48	Documento de Te...	1.045 KB
 L_BL_PT_T_TXT17.txt	05/03/2020 11:48	Documento de Te...	656 KB
 L_BL_PT_T_TXT20.txt	05/03/2020 11:48	Documento de Te...	411 KB
 L_BL_PT_T_TXT21.txt	05/03/2020 11:48	Documento de Te...	414 KB
 L_BL_PT_T_TXT31.txt	05/03/2020 12:17	Documento de Te...	1.048 KB

Fonte: Elaborada pelos autores.

O primeiro arquivo traz a codificação **L\_BL\_PT\_A\_TXT1**. A letra **L** significa que o arquivo pertence à subárea da Linguística Descritiva. Essa codificação foi criada antes que o binômio Linguística Descritiva fosse adotado pelos pesquisadores. Caso esse mesmo procedimento fosse adotado hoje, seria mais adequado usar as letras **LD**, que melhor representariam primeira subdivisão da árvore, ou mesmo um sistema de renomeação automática como a proposta pela plataforma do ToGatherUp de Oliveira (2019). Em contraste à essa codificação, há a abreviação **LA** para a subárea da Linguística Aplicada. Em seguida, há a abreviação **BL**, adotada para a

subárea da Biolinguística, existente no terceiro nível da árvore de domínio da Linguística. Depois, apresentamos a língua do *subcorpus* – neste caso, o português –, codificado por **PT**. No caso do inglês, a codificação foi **EN**, como mencionado previamente. Depois da língua, expomos o gênero ou subgênero, de quatro formas diferentes: **A**, **D**, **MIA** e **T**. A letra **A** foi usada para **artigos científicos**; **D** para **dissertação** de mestrado; **MIA** para **Material instrucional**; e **T** para **teses** de doutorado. Finalmente, há o número do texto, a partir do número 1, apesar de, em geral, não seguir a sequência correta, já que arquivos foram removidos para que houvesse o balanceamento dos *corpora*. Não consideramos necessária a indicação do ano de publicação dos textos, pois não era um trabalho de perspectiva histórica ou diacrônica, mas caso esta perspectiva fosse adotada, o ano de publicação teria sido necessário. O mesmo procedimento pode se aplicar ao quesito de autoria das obras, caso seja relevante, como foi no segundo projeto: o dado de autoria facilitou a organização do *corpus* de manuais em sua primeira fase, o que não persistiu quando do carregamento dos textos na plataforma do ToGatherUp.

Esses passos metodológicos foram adotados para a organização e o arquivamento dos *corpora* de Linguística, cujo dimensionamento final totalizou 49,89 milhões de palavras. O próximo questionamento foi a forma de arquivar com segurança os *corpora*. Apesar de haver opções on-line, como o Google Drive, da empresa Google, e o OneDrive, da Microsoft, podendo ser gratuitas ou pagas, optamos pela plataforma ToGatherUp (OLIVEIRA, 2019).<sup>28</sup>

Essa plataforma permite que o trabalho terminográfico seja mais prático e simplificado, devido a algumas funções automáticas, tais como, organização, codificação/nomeação dos arquivos, arquivamento de textos em diretórios (personalizado para os projetos), visualização do dimensionamento do *corpus* (quantidade de textos e palavras) e a inserção de cabeçalho e de metadados nos textos. Outras vantagens são a segurança dos dados, a sustentabilidade (economia de *hardware*, economia de espaço), a inovação e a flexibilidade. Há outras funções e vantagens que não trataremos aqui devido ao foco deste artigo, mas que podem ser consultadas de acordo com a referência bibliográfica. Como a escolha dessa plataforma foi posterior ao processo de reciclagem

---

<sup>28</sup> Plataforma de acesso restrito a pesquisadores, disponível em: [http://web-dev.ileel.ufu.br/togatherup/projetos/computacao/mod\\_login.php](http://web-dev.ileel.ufu.br/togatherup/projetos/computacao/mod_login.php). Acesso: 20 maio 2020.

dos *corpora* desta pesquisa, já havíamos desenvolvido nossa própria codificação para os arquivos. Para que o leitor possa ter uma visão dessa plataforma, trazemos as Figuras 9 e 10 e, em seguida, explicaremos seu funcionamento.

FIGURA 9 – Imagem parcial da plataforma ToGatherUp

ID	Arquivo	Descrição	Disciplina	Data	Título
212	PT-LG-LD-CI-AC-IN-10Aug2019-212.txt	Linguística Descritiva	Dialetologia	10Aug2019	O LÉXICO DA REGIÃO NORTE DO BRASIL
211	PT-LG-LD-CI-AC-IN-10Aug2019-211.txt	Linguística Descritiva	Dialetologia	10Aug2019	O DIALETO CAIPIRÁ
210	PT-LG-LD-CI-AC-IN-10Aug2019-210.txt	Linguística Descritiva	Dialetologia	10Aug2019	Linguagem e memória no envelhecimento: um estudo neurolinguístico
209	PT-LG-LD-CI-AC-IN-10Aug2019-209.txt	Linguística Descritiva	Dialetologia	10Aug2019	Língua e identidade portuguesa
208	PT-LG-LD-CI-AC-IN-10Aug2019-208.txt	Linguística Descritiva	Dialetologia	10Aug2019	INVESTIGAÇÕES GEOSOCIOLINGÜÍSTICAS: CONSIDERAÇÕES PARA UMA DESCRIÇÃO DOS FENÔMENOS DA VARIAÇÃO
207	PT-LG-LD-CI-AC-IN-10Aug2019-207.txt	Linguística Descritiva	Dialetologia	10Aug2019	Geolinguística pluridimensional: desafios metodológicos
206	PT-LG-LD-CI-AC-IN-10Aug2019-206.txt	Linguística Descritiva	Dialetologia	10Aug2019	Empréstimos linguísticos na visão do gramático Eduardo Carlos Pereira: um enfoque na perspectiva da História das Ideias Linguísticas

Fonte: Selecionada pelos autores.

FIGURA 10 – Imagem parcial da plataforma ToGatherUp

O LÉXICO DA REGIÃO NORTE DO BRASIL	<a href="http://www.sbpnet.org.br/livro/61ra/simposios/si...">http://www.sbpnet.org.br/livro/61ra/simposios/si...</a>	8695	00:05:00
O DIALETO CAIPIRÁ	<a href="http://www.letras.ufscar.br/linguagem/edicao21/p...">http://www.letras.ufscar.br/linguagem/edicao21/p...</a>	263475	00:05:00
Linguagem e memória no envelhecimento: um estudo neurolinguístico	<a href="http://www.revistainvestigacoes.com.br/volume-25-N...">http://www.revistainvestigacoes.com.br/volume-25-N...</a>	29330	00:05:00
Língua e identidade portuguesa	<a href="http://www.revistainvestigacoes.com.br/volume-22-N...">http://www.revistainvestigacoes.com.br/volume-22-N...</a>	25176	00:05:00
INVESTIGAÇÕES GEOSOCIOLINGÜÍSTICAS: CONSIDERAÇÕES PARA UMA DESCRIÇÃO DOS FENÔMENOS DA VARIAÇÃO	<a href="http://e-revista.unioeste.br/index.php/linguaelet...">http://e-revista.unioeste.br/index.php/linguaelet...</a>	35401	00:05:00
Geolinguística pluridimensional: desafios metodológicos	<a href="http://celtsul.org.br/Encontros/08/geolinguistica_p...">http://celtsul.org.br/Encontros/08/geolinguistica_p...</a>	11848	00:05:00
Empréstimos linguísticos na visão do gramático Eduardo Carlos Pereira: um enfoque na perspectiva da História das Ideias Linguísticas	<a href="http://www.revistainvestigacoes.com.br/volume-25-N...">http://www.revistainvestigacoes.com.br/volume-25-N...</a>	29039	00:05:00

Fonte: Selecionada pelos autores.

Objetivando a concisão, descreveremos somente os dados relativos ao registro e arquivamento dos textos. A primeira coluna, à esquerda, de 212-206, contém os números (parciais) dos arquivos carregados no ToGatherUp. Caso não haja a exclusão de nenhum deles, os números são correspondentes à quantidade de textos carregados na plataforma. A segunda coluna mostra a codificação dos arquivos .txt com dados da língua, do nível na árvore de domínio, a data de carregamento etc.; na terceira e na quarta colunas, há a classificação dos textos em Linguística Descritiva ou Aplicada (nível 2), e na subárea do próximo nível da taxonomia, neste caso, a Dialectologia; a quinta coluna registra a data de carregamento do arquivo: 10Aug2019, correspondendo ao dia, mês e ano; a sexta coluna registra o título do texto; a sétima, o endereço da Internet onde o texto se encontra disponível; a oitava coluna registra a quantidade de palavras-ocorrência (*tokens*) do texto (leitura automatizada pela plataforma); e a última coluna registra o tempo necessário utilizado para a busca e o processamento do texto, por parte do pesquisador, antes do carregamento na plataforma.<sup>29</sup>

#### 4.5 A reciclagem

A reciclagem do *corpus* de Linguística do primeiro projeto significa que o pesquisador tem em mãos, como foi no nosso caso, subcorpora de **dimensionamento** maior ou menor que o tamanho predeterminado como objetivo da pesquisa no segundo projeto. A partir desse *corpus* existente, o pesquisador o recicla de forma a aumentar ou reduzir os subcorpora, até que atinja o dimensionamento adequado ao seu projeto de pesquisa terminográfico em questão, neste caso, de 500 mil *tokens*.

No que diz respeito à **limpeza**, objetivamos, em nosso trabalho, explorar os contextos definitórios e explicativos dos termos, provenientes dos textos acadêmicos na área de Linguística (portanto, vários dados apresentados nos textos não eram relevantes para nossa pesquisa). No tocante à **tipologia** ou ao **gênero dos textos**, observamos que, na análise da composição dos gêneros textuais, havia aqueles que não atendiam ao

---

<sup>29</sup> Uma das características da plataforma é calcular o tempo gasto nas várias fases de compilação, pré-análise e nomeação dos arquivos. Consideramos importante o pesquisador apresentar, em seu trabalho, uma quantificação de tempo gasto numa compilação de *corpus*, de forma a deixar clara ao leitor a quantificação do tempo para a execução de projetos dessa natureza.

gênero acadêmico, o que os desqualificava para comporem os *corpora*, como, por exemplo, páginas genéricas da Internet, páginas de *blogs* e textos descritivos que pecavam na cientificidade da informação.<sup>30</sup> Apresentamos, na sequência, dados mais detalhados da reciclagem dos *corpora* do primeiro projeto.

#### 4.5.1 Redimensionamento de *corpus*

Em nossa pesquisa terminográfica, o dimensionamento estabelecido como padrão ou ideal foi de 500 mil *tokens* para cada subárea da Linguística; o planejado, no entanto, não se concretizou de maneira uniforme quanto à quantidade de *tokens* levantados pelos alunos: algumas duplas não conseguiram chegar ao número mínimo de *tokens*, já outras se entusiasmaram na coleta e extrapolaram o valor. Essa questão (além do próprio redimensionamento após a limpeza) explica o surgimento de *subcorpora* pequenos ou grandes. Classificamos os *subcorpora* compilados no projeto inicial em quatro tipos diferentes, abaixo relacionados, e mostramos como trabalhamos com cada tipo, no sentido de adequá-los ao *corpus* reciclado de Linguística do novo projeto.

- (1) Inexistente: fizemos a compilação dos *tokens* a partir do zero.<sup>31</sup>
- (2) Pequeno (menor que 500 mil *tokens*): compilamos o *subcorpus* até que atingisse o dimensionamento de 500 mil *tokens*.
- (3) Aproximado: foi mantido como estava ou redimensionado, se após a limpeza (de dados não relevantes) contivesse menos que 500 mil *tokens*.
- (4) Grande (maior que 500 mil *tokens*): após a limpeza, fizemos a leitura no WordSmith Tools 7.0 (WST; SCOTT, 2015) e usamos a ferramenta PLOT para eliminar os textos com menor densidade terminológica.<sup>32</sup>

---

<sup>30</sup> Como especialistas na própria área em estudo, tivemos condições de fazer esse tipo de análise. Para o levantamento de *corpora* em outras áreas, por exemplo, teríamos que consultar especialistas nas referidas áreas para uma certificação das fontes de coleta mais relevantes.

<sup>31</sup> Isso significa que uma determinada subárea da Linguística estava prevista na árvore do primeiro projeto, mas não foi feita a compilação para ela ou se trata de uma nova subárea, inserida na árvore de domínio do segundo projeto.

<sup>32</sup> Densidade terminológica (BARROS, 2004) é usada no sentido de identificar textos nos quais havia uma maior ou menor ocorrência de palavras-chave de determinada subárea em específico. Usamos a ferramenta PLOT do WST para identificar os textos

#### 4.5.2 Limpeza do *corpus*

Tendo em vista os objetivos de identificar termos e seus contextos, na limpeza do *corpus* do segundo projeto, eliminamos dados textuais e metadados (como marcação XML adicionada aos textos) que não atendiam ao nosso objetivo: extrair os traços conceituais úteis à construção das definições terminológicas e enciclopédicas do VoBLing<sup>33</sup> a partir dos contextos definitórios e dos contextos explicativos dos termos.

Em relação aos dados textuais eliminados dos artigos, temos: o nome da revista, o volume, a edição, o nome de autores, os metadados dos autores, as notas de rodapé, a bibliografia e os anexos. Quanto às dissertações, teses e trabalhos de conclusão de curso, eliminamos as informações pré-textuais e pós-textuais, tais como: os dados institucionais, os dados do autor, as abreviaturas, o sumário, os agradecimentos, a bibliografia ou as referências bibliográficas, os anexos e os apêndices. Outros dados eliminados foram as citações em língua estrangeira, as notas de rodapé, quadros e tabelas. Esses dados foram eliminados (manualmente), pois interfeririam na contagem de *tokens*, de forma que o dimensionamento dos *corpora* ficaria comprometido e o balanceamento dos *corpora* seria prejudicado.

#### 4.6 O *corpus* reciclado

O dimensionamento dos *subcorpora* de LD e LA foi de aproximadamente 47,7 milhões de *tokens*, provenientes dos gêneros científicos, sem incluir os manuais, pois eles compuseram um *corpus* à parte. Vale ressaltar que o *subcorpus* de Linguística Matemática (LM) foi menor que o das outras subáreas; para esta subárea, na LI, foi possível compilar os 500 mil *tokens*, o que não ocorreu da mesma forma na LP, limitada a 220 mil *tokens*. Logo, os *corpora* de LM totalizaram 440 mil *tokens*, em LP e LI, e não 1 milhão de *tokens*, padrão estabelecido

---

nos quais os 10 termos mais frequentes daquela subárea apresentavam menor frequência. Em seguida, fomos eliminando esses textos até que o número de *tokens* almejado fosse atingido (500 mil).

<sup>33</sup> VoBLing, projeto de vocabulário bilíngue de Linguística, português-inglês. O mesmo se valer de uma nova versão da plataforma já usada (VoTec), com adição de vários campos novos, como nota (que contém uma definição enciclopédica do termo), etimologia, som etc. Disponível em: <http://vobling.votec.ileel.ufu.br/>

como ideal. Já o dimensionamento dos *corpora* de manuais foi de 2,2 milhões de *tokens* nas duas línguas, já balanceados. Nosso balanceamento não englobou gêneros textuais diversos, senão o acadêmico. Logo, o balanceamento quanto aos gêneros textuais não se fez necessário e não fez parte do desenho dos projetos à exceção do gênero manuais de Linguística. O número de *tokens* no corpus de manuais foi quantificado, pois fez parte do desenho posterior, já do projeto final a fim de que obtivéssemos um maior número de contextos definitórios e explicativos para a extração de traços semânticos úteis à construção das definições dos termos.

A tipologia do *corpus* reciclado é demonstrada no Quadro 2.

QUADRO 2 – Tipologia do *corpus* reciclado

<b>Língua</b>	Bilíngue (inglês e português)
<b>Modo</b>	Escrito (textos acadêmicos: artigos científicos, dissertações e teses; textos instrucionais: manuais)
<b>Data de publicação</b>	Sincrônico (readequação e novos levantamentos realizados entre 2018 e 2020), fechado
<b>Seleção</b>	Amostragem, estático
<b>Conteúdo</b>	Especializado (Linguística)
<b>Autoria</b>	Falantes nativos/não nativos (inglês e português), individual/coletivo
<b>Disposição interna</b>	Comparável
<b>Uso na pesquisa</b>	Estudo (análise terminológica/terminográfica)
<b>Tamanho</b>	Grande (mais de 10 milhões de palavras)
<b>Nível de Codificação</b>	Com cabeçalhos, sem etiquetas

Fonte: Elaborado pelos autores.

Em relação ao tamanho do *corpus* reciclado, podemos verificar os dados na Tabela 2:

TABELA 2 – Tamanho do *corpus* reciclado

	Português	Inglês
Número de textos	1665	1911
<i>Tokens</i>	24.597.756	25.294.698
<i>Types</i>	356.989	356.911
<i>Type/token ratio (TTR)</i>	1,45%	1,41%
TOTAL <i>tokens</i>	49.892.454	

Fonte: Elaborada pelos autores.

Percebemos, comparando as Tabelas 1 e 2, que houve uma diminuição na quantidade de *tokens* entre os *corpora* do primeiro e do segundo projetos, embora o *corpus* reciclado do segundo projeto contenha mais subáreas na árvore de domínio (e, consequentemente, no *corpus*) e contenha textos de uma tipologia diversa do primeiro projeto (os manuais). Também percebemos que houve um melhor equacionamento nas TTRs no *corpus* do segundo projeto entre LP e LI, garantindo um melhor balanceamento. Acreditamos que a padronização levou a esse fenômeno.

## 5 Do *corpus* inicial ao *corpus* final

Nesta seção, compartilhamos algumas reflexões provenientes do percurso da reciclagem dos *corpora*, desde a coleta manual até ao arquivamento. Trataremos de aspectos metodológicos da coleta e do processamento dos *corpora* que optamos por adotar.

Como ponto de partida, é importante refletir sobre os gêneros textuais e sua constituição para um projeto terminográfico. Isso porque, na pesquisa de mestrado de um dos autores, observamos que trabalhar somente com artigos, dissertações e teses não foi suficiente para obtermos contextos definitórios e explicativos com traços semânticos suficientes para a construção de definições terminológicas.

Em segundo lugar, quando se trata da busca do gênero acadêmico na Internet, há arquivos como dissertações e teses de acesso público, cujos direitos autorais são cedidos para fins educacionais ou para o progresso da ciência, contudo encontram-se bloqueados em formato PDF. Nesse

caso, foi necessário usar programas (*software*) ou *sites* da Internet<sup>34</sup> que permitiam que os arquivos fossem salvos em formato txt.

Ainda tratando-se do gênero textual e sua disponibilidade, é importante considerar se esse gênero está disponível em formato digital. Do contrário, será necessário partir de um formato físico impresso até chegar no formato digitalizado. No segundo projeto, foi necessário escanear manuais de Linguística, converter os textos e corrigi-los. O escaneamento ainda é um procedimento moroso, que demanda uma quantidade considerável de tempo. Em seguida, como os manuais foram salvos em formato de imagem, foi necessário decodificá-lo para o formato de textos, usando programas específicos<sup>35</sup> e, em geral, pagos. Após a decodificação para textos, restou a correção textual, já que a maioria dos programas de OCR não decodifica a acentuação e diacríticos da língua portuguesa corretamente. Em nossa pesquisa, utilizamos o Microsoft Word para correção textual e salvamos os arquivos em formato .txt para o processamento pelo WST.<sup>36</sup>

Finalmente, há que considerar a codificação do arquivo txt, pois o WST requer que todos os arquivos sejam da mesma codificação para o correto processamento textual. Em um trabalho colaborativo, pode ocorrer que os membros salvem os arquivos em formato .txt, porém com codificações diferentes como UTF-8, UTF-16 ou ANSI, ocasionando a incompatibilidade de leitura. Para solucionar esse problema, utilizamos o utilitário *Text Converter* do WST para padronização dos arquivos em formato UTF-16 LE (segundo recomendação do manual do programa).

---

<sup>34</sup> Exemplo de *sites* para manuseio de arquivos PDF: <https://freemypdf.com/>, <https://www.ilovepdf.com/>.

<sup>35</sup> Neste projeto, usamos o OCR OmniPage (<https://www.kofax.com/Products/omnipage/>), a versão paga do Adobe Acrobat (<https://www.adobe.com/br/>) e o ABBYY Screenshot Reader (<https://www.abbyy.com/pt-br/>).

<sup>36</sup> Na fase de coleta do *corpus* inicial, no primeiro projeto, tanto o WST quanto o AntConc foram usados pelos alunos para verificar o tamanho dos sub*corpora* levantados e, na sequência dos trabalhos, para elaborar um projeto terminográfico. A partir do segundo projeto, de reciclagem do *corpus*, apenas o WST foi usado, devido à robustez computacional e à possibilidade de salvar os resultados.

## 6 Considerações finais

Dez anos de trabalhos contínuos na elaboração e na reelaboração de *corpora* na área de Linguística nos trouxeram uma rica experiência. Do começo, quando éramos novatos na universidade e a LC era coisa para *nerds*, até o presente, quando encontramos mais colegas trabalhando com a LC e os alunos escolhendo nossas disciplinas de graduação e pós-graduação por causa da LC, o panorama foi se alterando.

Nesse período, formamos (os autores e colegas da universidade) toda uma geração de alunos que consegue pensar em trabalhos de pós-graduação usando a abordagem e a metodologia da Linguística de *Corpus*. Uma geração que já começa a pensar em não mais apenas consumir *software* de análise lexical, mas também desenvolver *software* específicos para suas pesquisas e que trabalha, ao mesmo tempo, com descrição e aplicação (em sala de aula).

A elaboração de uma árvore de domínio da Linguística, baseada em *corpus*, nos trouxe um conhecimento geral da área que, até hoje, poucos colegas têm; o conhecimento de retrabalhar continuamente a compilação das fontes textuais de uma ciência também nos proporcionou autoridade para comentar sobre a taxonomia dessa ciência.

Para análises linguísticas, notamos que nosso *corpus* colaborativo (que se constituiu como estático),<sup>37</sup> assim como *corpora* advindos de qualquer área da ciência, deveria ter um caráter monitor: esse *corpus* colaborativo pode ter sua compilação (e, conseqüentemente, sua árvore de domínio) sempre reavaliada e ampliada. Novas funcionalidades, como as etiquetas (começando com a morfossintática) e a disponibilização desse *corpus* em plataforma eletrônica (com as ferramentas básicas da LC) para pesquisas online, devem ser agregadas para que esse *corpus* seja sempre útil para atuais e futuros pesquisadores.

Por fim, gostaríamos de agradecer, imensamente, às dezenas de alunos de graduação, pós-graduação e iniciação científica que nos ajudaram a tornar o *corpus* do primeiro projeto uma experiência didática e de pesquisa muito importante para aprimorar nossa formação como professores e pesquisadores.

---

<sup>37</sup> Estático porque seu desenho não pressupunha um projeto contínuo. Cada vez que o *corpus* foi trabalhado, o mesmo era reaberto e encerrado ao final do período, sem expectativa de nova atualização.

### **Contribuição dos autores**

O primeiro autor (FROMM) foi o responsável pela descrição do *corpus* original, o segundo autor (YAMAMOTO) foi o responsável pela descrição do segundo *corpus*. O restante do texto foi escrito e revisto pelos dois autores.

### **Referências**

ALBUQUERQUE, D. B. de. Múltiplos olhares em Linguística e Linguística Aplicada, 2016. *Ecolinguística*: Revista Brasileira de Ecologia e Linguagem (ECO-REBEL), Brasília, v. 3, n. 1, p. 227-237, 2017. Disponível em: <https://periodicos.unb.br/index.php/erbel/article/view/26263/23021> Acesso em: 26 mar. 2020.

ALVES, I. M. *et al.* (org.). *Estudos lexicais em diferentes perspectivas*. São Paulo: FFLCH/USP, 2010.

AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilíngue*. 2. ed. São Paulo: FFLCH/CITRAT, 2001.

BARBOSA, M. A. Dicionário, vocabulário, glossário: concepções. In: ALVES, I. M. (org.). *A constituição da normalização terminológica no Brasil*. 2. ed. São Paulo: FFLCH/CITRAT, 2001. p. 23-45.

BARROS, L. A. *Curso básico de terminologia*. São Paulo: EDUSP, 2004.

BERBER SARDINHA, T. *Linguística de Corpus*. São Paulo: Manole, 2004.

BOWKER, L. Towards a Collaborative Approach to Corpus Building in the Translation Classroom. In: BAER, J. B.; KOBAYASHI, G. S. (ed.). *Beyond the Ivory Tower: Rethinking Translation Pedagogy*. Amsterdam; Philadelphia: John Benjamins, 2003. p. 193-210.

CABRÉ, M. T. Hacia una teoría comunicativa de la terminología: aspectos metodológicos. In: CABRÉ, M. T. *La Terminología*: representación y comunicación. Barcelona: IULA, 2000. p. 129-150.

CARDOSO, S. A. F. *TermosTeo*: a elaboração de vocabulários monolíngues de termos da Teologia em um estudo conduzido por *corpus*. 2017. 340f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, Uberlândia, 2017.

FROMM, G. Vocabulário de linguística: treinamento em terminografia bilíngue, uso de corpora e ambiente de gestão terminológica. In: ISQUERDO, A. N.; DAL CORNO, G. O. M. (org.). *As ciências do léxico: lexicologia, lexicografia, terminologia*. Campo Grande: Editora UFMS, 2018. p. 309-328.

FROMM, G. Vocabulário de Linguística: treinamento em Terminografia Bilíngue, uso de corpora e ambiente de gestão terminológica. In: ENCONTRO INTERMEDIÁRIO DO GT DE LEXICOLOGIA, LEXICOGRAFIA E TERMINOLOGIA DA ANPOLL, 10., 2015, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: ANPOLL, 2015. p. 1-5.

FROMM, G. A questão da taxonomia num *corpus* colaborativo para construção de um vocabulário na área de linguística. In: SIMPÓSIO INTERNACIONAL DE LETRAS E LINGÜÍSTICA - SILEL, 2013, Uberlândia. *Anais [...]*. Uberlândia: EDUFU, 2013. p. 1-9.

FROMM, G. Ensino de terminologia: trabalhando com site e banco de dados. *Debate Terminológico*, Porto Alegre, v. 6, p. 2-22, 2010.

FROMM, G. VoTEC: a construção de vocabulários eletrônicos para aprendizes de tradução. 2007. 215f. Tese (Doutorado em Estudos Linguísticos e Literários em inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007. Disponível em: [https://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855/publico/TESE\\_GUILHERME\\_FROMM.pdf](https://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855/publico/TESE_GUILHERME_FROMM.pdf). Acesso em: 30 mar. 2020.

FROMM, G.; YAMAMOTO, M. I. Terminologia, Terminografia, Tradução e Linguística de *Corpus*: a criação de um vocabulário bilíngue sobre Linguística. In: TAGNIN, S.; BEVILACQUA, C. (org.). *Corpora na Terminologia*. São Paulo: Hub Editorial, 2013. p. 129-152.

FROMM, G.; YAMAMOTO, M. I. A microestrutura em verbetes da área da Linguística. *Revista Estudos da Linguagem*, Belo Horizonte, v. 28, n. 1, p. 205-234, 2020. DOI: <http://dx.doi.org/10.17851/2237-2083.28.1.205-234>. Disponível em: <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/15255>. Acesso em: 23 mar. 2020.

GARDNER, J.; KROWNE, A.; XIONG, L. Automatic Invocation Linking for Collaborative Web-Based Corpora. In: CHBEIR, R.; BADR, Y.; ABRAHAM, A. (org.). *Emergent Web Intelligence: Advanced Semantic Technologies*. Londres: Springer-Verlag, 2010. p. 23-45.

ILARI, R. *Introdução ao estudo do léxico: brincando com as palavras*. 2. ed. São Paulo: Contexto, 2003.

MELLO, H. Methodological Issues for Spontaneous Speech *Corpora* Compilation: The Case of C-ORAL-BRASIL. In: RASO, T.; MELLO, H. (ed.). *Spoken Corpora and Linguistic Studies*. Amsterdam: John Benjamins Publishing, 2014. p. 27-68. DOI: <https://doi.org/10.1075/scl.61.01mel>

OLIVEIRA, F. P. de. *ToGatherUp*: um protótipo de ferramenta para a construção de *corpora*. 2019. 219f. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, 2019. DOI: <http://dx.doi.org/10.14393/ufu.di.2019.679>.

PÉREZ-PAREDES, P. L.; SÁNCHEZ-TORNEL, M.; CALERO, J. A. M. Learners' Search Patterns During *Corpus*-Based Focus-on-Form Activities. *International Journal of Corpus Linguistics*, [S.l.], v. 17, n. 4, p. 482-515, 2012. DOI: <https://doi.org/10.1075/ijcl.17.4.02par>

SCOTT, M. *WordSmith Tools*. Version 7. Liverpool: Lexical Analysis Software, 2015.

SINCLAIR, J. *Reading Concordances: An Introduction*. London: Longman, 2003.

TAGNIN, S.; BEVILACQUA, C. *Corpora na Terminologia*. São Paulo: Hub Editorial, 2013.

TAGNIN, S. E. O. *Corpus*-Driven Terminology in Brazil. In: POUPET, A. L. B.; XATARA, C. (org.). *Cahiers de Lexicologie – Dynamique de la Recherche en Lexicologie, Lexicographie et Terminologie au Brésil*. Paris: Classiques Garnier, 2012. p. 169-182.

TEIXEIRA, E. D. *A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual*. 2008. 439f. Tese (Doutorado, Programa de Pós-Graduação em Estudos Linguísticos e Literários em Inglês) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2008.

TIEDEMANN, J. *Recycling Translations: Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing*. Uppsala: Acta Universitatis Upsaliensis, 2003.

VARANTOLA, K. Disposable *Corpora* as Intelligent Tools in Translation. *Cadernos de Tradução*, Florianópolis, v. 1, n. 9, p. 171-189, 2002. Disponível em: <https://periodicos.ufsc.br/index.php/traducao/article/view/5985/5689>. Acesso em: 15 maio 2020.

VIANA, V.; TAGNIN, S. E. O. (org.). *Corpora na Tradução*. São Paulo: Hub Editorial, 2015.

YAMAMOTO, M. I. *Linguística histórica e Linguística de corpus: caminhos que se cruzam para desvelar a história da linguagem: um vocabulário bilíngue português-inglês*. 2015. 118f. Dissertação (Mestrado em Linguística Letras e Artes) – Instituto de Letras e Linguística, Universidade Federal de Uberlândia, 2015. Disponível em <https://repositorio.ufu.br/handle/123456789/15483>. Acesso em: 15 mar. 2020.

YAMAMOTO, M. I. Vocabulário bilíngue português/inglês de linguística geral. *Revista Philologus*, v. 24, p. 272-297, 2018. disponível em <http://www.filologia.org.br/rph/ano24/70supl/023.pdf>. Acesso em: 15 mar. 2020.



**Flavors of the progressive in the New Romania:  
the perfective progressive periphrasis in  
Brazilian Portuguese and Argentinian Spanish**

***Sabores do progressivo na România Nova: a perífrase perfectiva  
progressivo no português brasileiro e no espanhol argentino***

Romina Trebisacce

Universidad Autónoma de Entre Ríos (UADER), Entre Ríos / Argentina

Universidad de Buenos Aires (UBA), Buenos Aires / Argentina

rtrebisacce@gmail.com

<http://orcid.org/0000-0003-3587-3234>

Victoria Ferrero

Universidad de Buenos Aires (UBA), Buenos Aires / Argentina

victoria\_ferrero@hotmail.com

<http://orcid.org/0000-0001-6210-5113>

Renato Miguel Basso

Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo / Brasil

rmbasso@gmail.com

<http://orcid.org/0000-0003-2580-0365>

**Abstract:** In this paper, we analyze the perfective progressive periphrasis (PPP) in Brazilian Portuguese (BrP) and Argentinian Spanish (AS) in a comparative way. Based on different linguistic tests, we make two statements regarding the PPP in comparison with the imperfective progressive periphrasis (IPP). Firstly, we claim that the PPP has a progressive and perfective meaning. Secondly, we claim that the PPP allows iterative readings when combined with telic events (i.e., achievements in BrP and AS and accomplishments just in AS). We propose a syntactic and semantic analysis which accounts for these observations in a compositional way: while the gerund form expresses a progressive meaning (present in both periphrases), the auxiliary on the PPP expresses a perfective meaning which allows the iterative readings observed in this periphrasis.

**Keywords:** verbal aspect; actionality; periphrasis; semantics; syntax.

**Resumo:** Neste artigo, analisamos a perífrase perfectiva progressiva (PPP) no português brasileiro (PB) e no espanhol argentino (EA), de modo comparativo. Baseado em testes linguísticos, fazemos duas afirmações sobre a PPP em comparação com a perífrase imperfectiva progressiva. Em primeiro lugar, afirmamos que a PPP tem significado progressivo e perfectivo. Em segundo lugar, afirmamos que a PPP permite leituras iterativas quando combinada com eventos télicos (*achievements* no PB e no EA e *accomplishments* somente no EA). Propomos uma análise sintático-semântica que dá conta dessas observações de um modo composicional: ao passo que a forma do gerúndio expressa o significado progressivo (presente em ambas as perífrases), o auxiliar na PPP expressa um significado perfectivo que permite a leitura iterativa observada nessa perífrase.

**Palavras-chave:** aspecto verbal; acionalidade; perífrase; semântica; sintaxe.

Submitted on September 22th, 2020

Accepted on January 20th, 2021

## Introduction

Differently from English, Portuguese and Spanish have a particular verbal periphrasis, namely a *perfective* progressive periphrasis (PPP), illustrated in examples (1), in contrast with the much more common *imperfective* progressive periphrasis (IPP), in (2):<sup>1</sup>

- (1) a. João esteve correndo. – Portuguese  
b. Juan estuvo corriendo. – Spanish
- (2) a. João estava correndo. – Portuguese  
b. Juan estaba corriendo. – Spanish
- (3) John was running.

The main difference between these two periphrases is the auxiliary verb which can be perfective (‘esteve’, ‘estuvo’) or imperfective (‘estava’, ‘estaba’). Note that the English translation for both cases is the same (cf. (3)).

---

<sup>1</sup> In this paper, the first examples (i.e., a.) will always be from Brazilian Portuguese and the second examples (i.e., b.) will be from Argentinian Spanish.

The aim of this paper is to present a contrastive semantic analysis for the PPP, comparing Brazilian Portuguese (BrP) and Argentinian Spanish (AS) data. In order to do that, we focus on two concerns. In the first place, we analyze the aspectual meaning of this periphrasis. Since it is composed of imperfective morphology (realized in the gerund form) and perfective morphology (realized in the auxiliary), one of the main questions of this paper is whether this periphrasis conveys perfective or imperfective meaning. After applying several tests, we conclude that it is not clear the aspectual value of the PPP: while some tests show its perfective meaning, in other ones the meaning expressed seems to be imperfective.

In the second place, and considering the morphological similarity between the PPP and the IPP, we study the meanings these periphrases have when combined with Aktionsart classes (activities, achievements, accomplishments, states, and semelfactives). The meanings resulting from these combinations (in particular, from the combination of the PPP and the IPP with telic predicates) allow us to make some observations about their semantic similarities and differences: while they both (IPP and PPP) behave equally in expressing durative and homogeneous events, just the PPP allows iterative readings.

Regarding these observations, we propose a compositional analysis of the PPP, in which the gerund form expresses progressive meaning and gives rise to durative and homogeneous events and the perfective morphology in the auxiliary expresses a perfective meaning that allows for iterative readings. This model not only can explain why the PPP displays perfective and imperfective meaning, but it can also account for the similarities and differences observed between the PPP and the IPP. In our proposal, durative and homogeneous meaning present in both periphrases is due to progressive meaning in gerund form, while iterative meaning, allowed just in the PPP, is due to the perfective meaning present in the auxiliary's perfective morphology.

This paper is organized as follows. In the first section, we investigate the aspectual status of the PPP – is it perfective or imperfective? Our analysis is based on four main linguistic tests, namely: (i) interrupting an event in the PPP; (ii) closed temporal intervals and the PPP; (iii) temporal progression; and (iv) the culmination of the event. In the second section, we show the interpretations that result from combining the PPP with the different aspectual classes. In both sections, we deal with BrP and

AS data.<sup>2</sup> In the third section, we discuss the similarities and differences between the IPP and the PPP in the two languages investigated, and in the fourth section we summarize the different and similar interpretations of the PPP in BrP and AS. Finally, in the fifth section we propose a semantic analysis for both periphrases, focusing on the PPP. In the Conclusion, we discuss some open questions and our results.

## 1 What does it take to be perfective?

As we mention in the Introduction, the PPP contains an auxiliary verb in the perfective form and a main verb in the progressive form (i.e., the gerundive form), normally associated with the imperfective aspect. Given this configuration, it could be asked whether the PPP is perfective or imperfective. In other words, what is the (grammatical) aspect resulting from combining a perfective auxiliary with a progressive main verb?<sup>3</sup>

In order to answer this question, we need to determine what should be understood by “perfective” and “imperfective” aspects. A common assumption in the literature is the one based on Klein’s (1994) proposal, according to which the perfective aspect involves the time of the event being included in the topic time, and the imperfective aspect involves the topic time being included in the event time. One way of formally capturing these ideas can be found in Bohnemeyer (2014), which presents the formulas below, in which *P* is a variable for event predicates,  $\tau(e)$  represents the event time (i.e., the duration of the event),

---

<sup>2</sup> As for the data presented in this paper, it is important to make some comments. All the examples presented in this paper, as well as the readings we discuss, represent our own intuition. However, since in some cases the readings we point out are not easy to get, we ask different linguists and non-linguists for several semantic intuitions.

<sup>3</sup> Squartini (1998) also investigates the aspectual behaviour of the PPP, given the coexistence of two apparently incompatible aspectual values (perfective and imperfective) in the periphrasis. However, he concludes that the periphrasis is perfective. The progressive information actually affects the actionality of the event, imposing restrictions related to durativity. Rather than an interaction between perfectivity and imperfectivity, the periphrasis would show perfective aspect and a durative actionality value. In this paper, even though we examine the aspectual value shown by the PPP, we also present a systematic comparison between the behaviour of the PPP and the IPP when combined with the aspectual classes. Thus, we explore the different interpretations each one gives rise to in BrP as well as in AS.

$t_T$  represents topic time, and  $g$  represents the variable assignment function parameter with respect to a model  $M$ :

$$(4) \quad [[PF]]^{M,g} = \lambda P \exists e [\tau(e) \subseteq t_T \wedge P(e)]$$

$$(5) \quad [[IMPF]]^{M,g} = \lambda P \exists e [t_T \subseteq \tau(e) \wedge P(e)]$$

The intuition behind (4) is that a perfective event does not evolve past a certain topic time, and the imperfective, as stated in (5), does just the opposite, because an imperfective (ideally) evolves beyond a certain topic time. The examples in (6) and (7) illustrate these points, respectively. While in (6) the event of John painting the picture is included in the topic time (it is presented as closed at a certain topic time), in (7) the event continues beyond the topic time.

- (6) a. João pintou o quadro.  
 b. Juan pintó el cuadro.  
 John paint<sub>.3.Person.Perf</sub> the picture.

- (7) a. João estava pintando/pintava o quadro.  
 b. Juan estaba pintando/pintaba el cuadro.  
 John be<sub>.3.Person.Imperf</sub> painting/ paint<sub>.3.Person.Imperf</sub> the picture.

However, to show that the PPP is perfective or imperfective, it is important to go beyond these definitions and investigate their consequences. Therefore, aligned with the definitions in (4) and (5), we use several tests, already employed in the relevant literature (DOWTY, 1979; SQUARTINI, 1998), in order to analyze the perfective meaning.

- (i) interruption of the event – combined with “when” clauses, imperfectives result in an interpretation in which the event introduced by the “when” clause occurs within the temporal interval of the imperfective event, but that is not the case with perfective events (BONOMI, 1997);
- (ii) closed interval adverbials – perfectives can naturally combine with closed time intervals, such as ‘o dia todo’/ ‘todo el día’ (*the whole day*), but not imperfectives (Squartini, 1998);
- (iii) temporal progression/succession of events – usually, concatenated perfective events are interpreted in the order of their appearance, whereas imperfective events do not usually impose any temporal ordering (SQUARTINI, 1998);
- (iv) the implied culmination of the event– when an accomplishment is in the perfective, it is implied that the event has culminated, whereas when we take an accomplishment in the imperfective this is not implied (DOWTY, 1979).

In sections 1.1 to 1.4 we will explore these tests using data from BrP and AS, contrasting PPP and IPP, which is by default imperfective. In section 1.5 we present an interim summary of our conclusions.

### 1.1 Interrupting an event in the PPP

As we can see in (8), only imperfective forms, synthetic or periphrastic, give rise to situations where the event described by the predicate can be interrupted by the event introduced by the “when” clause (i.e., Maria arrived while João/Juan was swimming). In contrast, when the eventuality is expressed by a perfective form (9), this is not possible.

- (8) a. João nadava/estava nadando quando Maria chegou.  
 b. Juan nadaba/estaba nadando cuando María llegó.  
 John swim<sub>.3.Person.Imperf</sub>/be<sub>.3.Person.Imperf</sub> swimming when Mary arrived.
- (9) a. ?João nadou quando Maria chegou.  
 b. ?Juan nadó cuando María llegó.  
 John swim<sub>.3.Person.Perf</sub> when Mary arrived.

The sentence in (9) is hard to accept. However, if we accept it, we can claim that it can only mean that the event of John swimming immediately succeeds (or precedes) the event of Mary arriving. That is to say, it can never mean that the arriving of Mary occurs within (the duration of) John’s swimming event.

If we consider the meaning of perfective and imperfective that we have presented above, data (8-9) easily follow. Given that in the perfective meaning the time of the event is included in a topic time and the event is presented as concluded (or bounded), the eventuality is not available for being interrupted.

If PPP had an imperfective meaning we would expect the same behaviour we have observed in imperfective examples (8) and in (10), below. However, this is not the case, as the sentences in (11) show:

- (10) a. Ontem, João estava pintando o quadro quando Pedro chegou.  
 b. Ayer, Juan estaba pintando el cuadro cuando Pedro llegó.  
 Yesterday, John be<sub>.3.Person.Imperf</sub> painting the picture when Pedro arrived.

- (11) a. ?Ontem, João esteve pintando o quadro quando Pedro chegou.  
 b. ?Ayer, Juan estuvo pintando el cuadro cuando Pedro llegó.  
 Yesterday, John be<sub>.3.Person.Perf</sub> painting the picture when Pedro arrived

Sentences in (11) are really odd, and the only interpretation available for cases like (11) – if there is any – is the one in which the event of painting a picture immediately precedes (or succeeds) the event of Pedro arriving – both in BrP and in AS. In other words, according to this test the PPP and perfective forms behave similarly.

## 1.2 Closed temporal intervals and the PPP

The PPP can combine with durational adverbials, whose function is to delimit the time in which an event takes/took place. As it is shown in the examples below, the periphrasis occurs naturally with “durante x tempo” and “durante x tiempo” (12), with “o dia todo” and “todo el día” (13), or with “até x” and “hasta x” (14).

- (12) a. Ontem, João esteve trabalhando *durante seis horas*.  
 b. Ayer, Juan estuvo trabajando *durante seis horas*.  
 Yesterday, John be<sub>.3.Person.Perf</sub> working for six hours.
- (13) a. Ontem, João esteve trabalhando *o dia todo*.  
 b. Ayer, Juan estuvo trabajando *todo el día*.  
 Yesterday, John be<sub>.3.Person.Perf</sub> working the day whole (the whole day).
- (14) a. Ontem, João esteve trabalhando *até às sete*.  
 b. Ayer, Juan estuvo trabajando *hasta las siete*.  
 Yesterday, John be<sub>.3.Person.Perf</sub> working until seven.

Delimiting adverbials introduce a closed temporal interval and, as Squartini (1998) argues, the compatibility with these durational phrases is also a test for identifying perfectivity in a verbal form. As we can see in the next examples, establishing a delimited temporal interval where the event takes place is consistent with the perfective value (15). However, it is inconsistent with the imperfective, which rejects these durational adverbials (16) unless some other information is provided, such as a larger context in which the imperfective event can be anchored.

- (15) a. João correu durante duas horas.  
 b. Juan corrió durante dos horas.  
 John run<sub>.3.Person.Perf</sub> for two hours.
- (16) a. #João corria/estava correndo durante duas horas.  
 b. #Juan corría/estaba corriendo durante dos horas.<sup>4</sup>  
 John run<sub>.3.Person.Imperf</sub>/be<sub>.3.Person.Imperf</sub> for two hours.

As exhibited, the PPP can be combined with delimited adverbials (17), while the IPP cannot (18). Due to that contrast of acceptability, we can affirm that, as the IPP is imperfective, the PPP shows a perfective value also with respect to this test.

- (17) a. João esteve trabalhando durante seis horas.  
 b. Juan estuvo trabajando durante seis horas.  
 John be<sub>.3.Person.Perf</sub> working for six hours.
- (18) a. #João estava trabalhando durante seis horas.  
 b. #Juan estaba trabajando durante seis horas.  
 John be<sub>.3.Person.Imperf</sub> working for six hours.

### 1.3 Temporal progression

It is a well-known fact that the perfective moves the flux of a narrative forward, whereas the imperfective is responsible for description and background (cf. KAMP; ROHER, 1983). This difference appears in contexts in which there is a temporally ordered sequence of events (SQUARTINI, 1998), as exemplified in (19):

- (19) a. João conversou com a Maria, jantou e pensou sobre o seu dia seguinte.  
 b. Juan conversó con María, cenó y pensó sobre su próximo día.  
 John talk<sub>.3.Person.Perf</sub> to Mary, have<sub>.3.Person.Perf</sub> dinner, and think<sub>.3.Person.Perf</sub> about the next day.  
 ‘John talked to Maria, had dinner and thought about his next day’

<sup>4</sup> The symbol “#” indicates that the sentence is pragmatically anomalous. This means that (16) and (18), below, are not acceptable “out of the blue”, they need a bigger context, such as “Last year, ...”, which is not necessary for the perfective forms.

The events in (19), expressed by the perfective forms *conversou/conversó*, *jantou/cenó* and *pensou/pensó*, are interpreted as succeeding one another in time, whereas the imperfective forms of the same verbs in (20) denote events which are taken to be ongoing simultaneously. That is to say, events in imperfective are not required to be concatenated.

- (20) a. João estava conversando com a Maria, jantando e pensando no seu dia seguinte.  
 b. Juan estaba conversando con María, cenando y pensando sobre su próximo día.  
 John be<sub>.3.Person.Imperf</sub> talking to Mary, be<sub>.3.Person.Imperf</sub> having dinner, and be<sub>.3.Person.Imperf</sub> thinking about the next day.  
 ‘João was talking to Maria, having dinner and thinking about his next day’

Following the same reasoning of the other tests, if the PPP is perfective, we expect it to behave as the perfectives in (19). However, this is not the case. Sentence (21) is not natural, and if acceptable it does not convey any order among the events:

- (21) a. ? João esteve chegando em casa, (esteve) conversando com Maria, (esteve) jantando e (esteve) pensando no seu dia seguinte.  
 b. ? Juan estuvo llegando a casa, (estuvo) conversando con María, (estuvo) cenando y (estuvo) pensando en su próximo día.  
 John be<sub>.3.Person.Perf</sub> arriving home, be<sub>.3.Person.Perf</sub> talking to Mary, be<sub>.3.Person.Perf</sub> having dinner, and be<sub>.3.Person.Perf</sub> thinking about the next day.

#### 1.4 The culmination of the event

We can find another difference between perfective and imperfective when these forms are combined with accomplishments.<sup>5</sup> When the predicate is in the perfective form, as in (22), the implication<sup>6</sup>

<sup>5</sup> In section 2.1 we investigate in more detail the combination of telic events with imperfective meaning.

<sup>6</sup> We prefer to use a more neutral form such as “implication” and “implies” because according to some (ALTSHULER, 2013; BASSO, 2007; PIRES DE OLIVEIRA; BASSO, 2010, a.o.) the interpretation that the event is finished which result from combining a telic event with (past) perfective aspect is not an entailment. Since the nature of this interpretation (if it is an implicature or an entailment) is not particularly important here, we will remain agnostic. What is important for us is whether the PPP and the perfective behave similarly.

is that the event has culminated, that is, the picture has been finished. In turn, when someone states (23) it is not implied that the event has culminated. That is to say, whereas (22) implies (24), (23) does not.

- (22) a. O João pintou o quadro.  
 b. Juan pintó el cuadro.  
 John paint<sub>.3.Person.Perf</sub> the picture.
- (23) a. O João pintava/estava pintando o quadro.  
 b. Juan pintaba/estaba pintando el cuadro.  
 John paint<sub>.3.Person.Imperf</sub> / be<sub>.3.Person.Imperf</sub> painting the picture.
- (24) a. O quadro está pintado/terminado.  
 b. El cuadro está pintado/terminado.  
 The picture is painted/finished.

In this sense, if PPP had perfective meaning, we would expect it to behave as (22), that is, to imply the culmination of the event. However, this is not the case.

When someone utters (25), the culmination of the event is not implied. In other words, it could be the case (and it is, in fact, the main interpretation that we have) that John had been painting the picture for a while and had stopped doing it without finishing it.

- (25) a. O João esteve pintando o quadro.  
 b. Juan estuvo pintando el cuadro.  
 John be<sub>.3.Person.Perf</sub> painting the picture

This point can be further stressed with the use of, for instance, “mas não terminou”/“pero no lo terminó” (“but did not finish it”) combined with the examples (22), (25), and also (26), with the IPP:

- (22') a. # O João pintou o quadro, mas não terminou.  
 b. # Juan pintó el cuadro, pero no lo terminó.  
 John paint<sub>.3.Person.Perf</sub> the picture, but did not finish it.
- (25') a. O João esteve pintando o quadro, mas não terminou.  
 b. Juan estuvo pintando el cuadro, pero no terminó.  
 John be<sub>.3.Person.Perf</sub> painting the picture, but did not finish it

- (26) a. O João estava pintando o quadro, mas não terminou.  
 b. Juan estaba pintando el cuadro, pero no ha terminado.  
 John be<sub>3, Person, Imperf</sub> painting the picture, but did no finished it

This point is somewhat related to what we saw in 1.1, but now the interruption has to do with reaching the *telos* of a telic event. And in this case, the contrast between (22') and (25') shows that the PPP and the perfective do not behave similarly since the *telos* of an accomplishment is not implied to be reached in the case of (25'), which is similar to (26). Let us summarize what we have found so far.

### 1.5 Interim summary

To summarize, on the one hand, the PPP behaves like the perfective in two respects: (i) an event in the PPP cannot be interrupted by another event, and (ii) the PPP does allow the combination with a durational adverbial introducing a closed temporal interval. On the other hand, it shows a non-perfective behavior regarding two senses: (i) the perfective, with coordinated events, does not convey any order among the events in contexts of temporally ordered sequences; however, the PPP does not move the narrative forward, and (ii) while the perfective implies the culmination of the event (that the *telos* was reached), the PPP does not.

This conclusion is the same for BrP and AS data, but it is not conclusive regarding the aspectual meaning of the PPP, since it sometimes seems to be perfective and sometimes imperfective. In the next section, we will explore the interpretation which results from the combination of the PPP with events from the different Vendlerian classes. This analysis will help us understand the behavior of the PPP, especially with telic events (an important feature for the test in 1.4), and also its differences with respect to the IPP.

## 2 PPP and actionality

In this section we analyze the behavior of the PPP when combined with different kinds of events. By studying the combination between PPP, IPP and events from different aspectual classes we will be able to reach some partial conclusions about the meaning of these periphrases.

We follow Vendler's famous classification of events, according to which there are four classes which differ in actional values, characterized by features such as telicity (i.e., whether a predicate has a culmination point), duration (i.e., whether a predicate holds in time) and dynamicity (i.e., whether a predicate involves changes, in contrast to a state). These classes are: states, activities, accomplishments and achievements. The literature also recognizes a fifth class, the so called semelfactives, which are atelic and non-durative events.<sup>7</sup> The table 1, adapted from Smith (1991, p. 20), shows these classes according to some relevant semantic features.

TABLE 1 – Types of events

	Telic (culmination point)	durative (holds in time)	Dynamic (involves change)
activities	–	+	+
states	— <sup>8</sup>	+	–
semelfactives	–	–	+
achievements	+	–	+
accomplishments	+	+	+

Source: Authors

In section 2.1, we will analyze telic predicates (namely, achievements and accomplishments) and in section 2.2, atelic predicates (namely, activities, states, and semelfactives).

## 2.1 PPP with telic predicates

As we have said before, telic predicates present a culmination point, that is to say, among other characteristics, they cannot go on

<sup>7</sup> It is widely agreed that there is also a class of “degree achievement”, which are events that involve a scalar development, such as “dry”, “grow”, “fatten”. In this paper we will not consider degree achievements, but the conclusions we present can be extended to this class of events as well.

<sup>8</sup> It is worth noting that, although the [Telic] feature is defined as negative for states in the table above, the author mentions that it is an irrelevant specification for eventualities [-dynamic].

indefinitely. Even though achievements and accomplishments both share the telic feature, they constitute different classes because of another semantic feature: their duration. Accomplishments are durative, whereas achievements have no temporal structure (i.e., they have no duration).

Given this difference, literature concerning IPP has claimed that these predicates behave differently when combined with this periphrasis: while accomplishments in IPP have a progressive reading according to which their *telos* is not reached (cf. (27)), achievements have a “preparatory phase” reading (cf. (28)). The same kinds of interpretation are obtained in BrP and AS:

- (27) a. O João estava correndo um quilômetro (mas não conseguiu terminar).  
 b. Juan estaba corriendo un kilómetro (pero finalmente no llegó).  
 John be<sub>.3.Person.Imperf</sub> running a mile (but he didn't do it).
- (28) a. O João estava chegando ao trabalho.  
 b. Juan estaba llegando al trabajo.  
 John be<sub>.3.Person.Imperf</sub> arriving at work.

Sentences such as (27), as expected, have a progressive reading, and in these cases, we cannot claim that the *telos* was reached, as it is clear by the information presented in the adjunct. As for data such as (28), these sentences do not describe the event denoted by the predicate but all kinds of previous events connected with the event described by the predicate (for instance, John could be about to cross the entrance, or about to park his car).

Regarding the PPP, accomplishments and achievements also behave in a different way when combined with this periphrasis. However, these predicates in PPP obtain different readings and interestingly these readings are unlike the ones we have observed when combined with IPP. This is another piece of evidence that shows that PPP and non-periphrastic perfectives have differences in meaning, despite some similarities.

Let us begin with accomplishments. As well as with IPP, accomplishments in PPP seem to denote situations in which the *telos* is not reached. For instance, if someone says (29) or (30), we cannot conclude that João/Juan has actually reached the pharmacy or finished preparing lunch:

- (29) a. João esteve correndo até a farmácia. —/—> João correu até a farmácia.  
 b. Juan estuvo corriendo hasta la farmacia. —/—> Juan corrió hasta la farmacia.  
 John be<sub>.3.Person.Perf</sub> running to the drugstore.
- (30) a. João esteve preparando o almoço. —/—> João preparou o almoço.  
 b. Juan estuvo preparando el almuerzo. —/—> Juan preparó el almuerzo.  
 John be<sub>.3.Person.Perf</sub> preparing lunch.

In addition, they accept continuations that emphasize the non-culmination of the event (such as *but he/she didn't P*).

- (31) a. João esteve correndo até a farmácia (mas não chegou).  
 b. Juan estuvo corriendo hasta la farmacia (pero no llegó).  
 John be<sub>.3.Person.Perf</sub> running to the drugstore (but he did not get there).
- (32) a. João esteve preparando o almoço (mas não terminou).  
 b. Juan estuvo preparando el almuerzo (pero no terminó).  
 John be<sub>.3.Person.Perf</sub> preparing lunch, but he did not finish it.

Finally, PPP does not allow temporal phrases such as “in x time”. Since “in x time” presupposes that the *telos* has been reached (BASSO, 2007; BERGAMINI-PEREZ; BASSO, 2016) and since accomplishments in PPP denote situations where the *telos* is not necessarily reached, combining these predicates with such temporal phrases is not allowed:

- (33) a. ?João esteve correndo até a farmácia em dez minutos.  
 b. ?Juan estuvo corriendo hasta la farmacia en diez minutos.  
 John be<sub>.3.Person.Perf</sub> running to the drugstore in ten minutes.
- (34) a. ?João esteve preparando o almoço em dez minutos.  
 b. ?Juan estuvo preparando el almuerzo en diez minutos.  
 John be<sub>.3.Person.Perf</sub> preparing lunch in ten minutes.

It is important to mention that both sentences are acceptable in BrP (but not in AS) if one considers not an episodic event but a capacity that João used to have. For instance, for some time in his life, João was able to run to the drugstore in 10 minutes, but he is no longer capable of doing that (this interpretation is made more salient if one uses the adverb ‘já’). That is, sentences (33) and (34) are acceptable just in case these

events take place more than once in the past. However, regarding their episodic meaning, neither (33) nor (34) are well construed.

Data such as (29-34) show that accomplishments in PPP have a similar meaning with what we find in accomplishments in IPP: in both cases the *telos* of the event is not guaranteed to be reached.<sup>9</sup> Therefore, we can conclude that PPP and IPP have the effect of implying that the *telos* of telic events is not reached. This is particularly interesting for the PPP, since it is usually assumed in the literature that telic events in the perfective entail that the *telos* is reached.<sup>10</sup> This is the reason why the perfective is odd when combined with the phrase “and he finally finished it” that seems to be redundant.

- (35) a. ?João pintou um retrato em vinte minutos e finalmente o terminou.  
 b. ?Juan pintó un retrato en veinte minutos y finalmente lo terminó.  
 John paint<sub>.3.Person.Perf</sub> a portrait in twenty minutes and he finally finished it.

In contrast, the oddness of (36) is due to the combination of the temporal phrase “in twenty minutes” and the fact that PPP implies the non-culmination of the event, as observed in examples (33-34).

- (36) a. ?João esteve pintando um retrato em vinte minutos e finalmente o terminou.  
 b. ?Juan estuvo pintando un retrato en veinte minutos y finalmente lo terminó.  
 John be<sub>.3.Person.Perf</sub> painting a portrait in twenty minutes and he finally finished it.

<sup>9</sup> Interestingly, there seems to be some differences between the meaning of accomplishments in PPP and in IPP. In AS a sentence such as “La semana pasada, Juan estuvo armando un rompecabezas” (‘Last week, John be<sub>.3PersonPerf</sub> making a puzzle’) can be interpreted as different events of making (and finishing) a puzzle (preferably, the same puzzle). In AS, this reading is not available with IPP “La semana pasada, Juan estaba armando un rompecabezas” (‘Last week, John be<sub>.3PersonImperf</sub> making a puzzle’). As for BrP neither PPP nor IPP – “Na semana passada, João esteve montando um quebra-cabeça” (‘Last week, John be<sub>.3PersonPerf</sub> making a puzzle’) and “Na semana passada, João estava montando um quebra-cabeça” (‘Last week, John be<sub>.3PersonImperf</sub> making a puzzle’) – give rise to the mentioned reading. This difference will be revised in more detail in sections 3 and 4. On the other hand, with the same sentence, both the PPP and the IPP in AS as well as in BrP can trigger a reading where there is a non-continuous event of making the same puzzle. We can have an event of P fragmented in different scenes, and they are all temporally located in the interval of one week.

<sup>10</sup> Basso (2007) is an exception and calls the configuration in which a perfective telic event has not reached its *telos* a case of “detelicization”. We will turn to this topic later.

At this point, it is worth noting that this effect of not reaching the *telos* does not seem to be similar to the ones produced in other contexts, such as the temporal phrase “for x time” when combined with perfective events, as in (37). It has been said that the “for x time” phrase has also the effect of detelicizing the perfective telic event it combines with. Since this adverbial only measures the duration of the event, its presence can pragmatically lead to the conclusion that the event has not reached its *telos*. But given that this is an implicature (cf. BASSO, 2007; PIRES DE OLIVEIRA; BASSO, 2010), the continuation with “and he finally finished it” is not semantically prohibited. In this respect, both the perfective and the PPP behave similarly:

- (37) a. João pintou um retrato por vinte minutos e finalmente o terminou.  
 b. Juan pintó un retrato por veinte minutos y finalmente lo terminó.  
 John paint<sub>.3.Person.Perf</sub> a portrait and he finally finished
- (38) a. João esteve pintando um retrato por vinte minutos e finalmente o terminou.  
 b. Juan estuvo pintando un retrato por veinte minutos y finalmente lo terminó.  
 John be<sub>.3.Person.Perf</sub> painting a portrait for twenty minutes and he finally finished it.

Hence, based on the data presented above, we can conclude that accomplishments in PPP and IPP give rise to progressive, durative, non-telic readings, given that they allow for adjuncts that emphasizes the non-culmination of the events.

Let us consider now achievements, which are telic non-durative events (that is, they have no internal temporal structure). As mentioned before, achievements in IPP not only denote situations where the *telos* is not reached but also describe the preparatory phase of the event denoted by the predicate (i.e., all the activities directly related to the event of winning or waking up).

- (39) a. João estava ganhando a corrida.  
 b. Juan estaba ganando la carrera.  
 John be<sub>.3.Person.Imperf</sub> winning the race.
- (40) a. João estava acordando.  
 b. Juan estaba despertándose.  
 John be<sub>.3.Person.Imperf</sub> waking up.

As observed by Rothstein (2004), the reason why achievements in IPP denote a preparatory phase is because they describe events without internal structure. According to her account, given that the progressive operator<sup>11</sup> combines with predicates containing stages (or internal structure), when it is combined with achievements the result is the preparatory phase reading. Interestingly, achievements in PPP differ from achievements in IPP in that they can describe two kinds of situations depending on lexical information of the verb: i, one in which there is a repetition of events of P (iterative reading), as in (41); ii. another one in which previous events anteceding P are described (preparatory phase reading), as in (42).

- (41) a. João esteve acordando por toda a manhã.<sup>12</sup>  
 b. Juan estuvo despertándose toda la mañana.  
 John be<sub>.3.Person.Perf</sub> waking up (the whole morning).
- (42) a. João esteve ganhando a corrida.  
 b. Juan estuvo ganando la carrera.  
 John be<sub>.3.Person.Perf</sub> winning the race.

The main difference between situations described in (41) and (42) is the fact that only in (41) we can conclude that the event of P took place (multiple times). In (42), in contrast, the event does not need to reach the *telos* (i.e., John could have lost the race after all).

<sup>11</sup> Rothstein (2004) follows Landman (1992) in claiming that the progressive operator requires predicates which have stages. More specifically, he claims that an assertion of the form *x is VP-ing* is true iff there is an event *e* going on which is a stage of an event *e'*, where *e'* is in the denotation of the VP. An event *e* is a stage of event *e'* if it develops into *e'*; in this case *e'* is a continuation of *e*. Therefore, for a sentence in the progressive to be true it must be the case that its predicate has events *e* which are stages developing into *e'*. For instance, in order to *John was painting a portrait* to be true we expect *paint a portrait* to have stages, that is, we expect events *e* of *paint a portrait* developing into the event *e'* of *paint a portrait*.

<sup>12</sup> Iterative meaning in sentence (41) can be difficult to get in BrP. However, this meaning is accessible if we imagine a context where João was sleeping but he was interrupted in several moments by different situations. In that case, João has been waking up the whole morning in different moments.

- (43) a. João esteve acordando por toda a manhã <sup>??</sup>(mas não acordou).  
 b. Juan estuvo despertándose toda la mañana <sup>??</sup>(pero no se despertó).  
 John be<sub>.3.Person.Perf</sub> waking up the whole morning <sup>??</sup>(but he didn't wake up).
- (44) a. João esteve ganhando a corrida, mas no fim ele não ganhou.  
 b. Juan estuvo ganando la carrera, pero finalmente no la ganó.  
 John be<sub>.3.Person.Perf</sub> winning the race, but finally he didn't won.

We also can detect this in the entailments in (45) and (46).

- (45) a. João esteve acordando por toda a manhã. —> ele acordou  
 b. Juan estuvo despertándose toda la mañana —> se despertó.  
 John be<sub>.3.Person.Perf</sub> waking up the whole morning —> he woke up.
- (46) a. João esteve ganhando a corrida por 20 minutos —/—> ele ganhou a corrida  
 b. Juan estuvo ganando la carrera por 20 minutos —/—> la ganó.  
 John be<sub>.3.Person.Perf</sub> winning the race for 20 minutes —/—> he won the race.

Given the data above, we make two observations. On the one hand, IPP as well as PPP force achievements to describe durative eventualities (either they describe a preparatory phase or an iterative event). On the other hand, examples (41-46) show that PPP differs from IPP in giving rise to an iterative meaning when it combines with some kinds of achievements.<sup>13</sup> These observations suggest two things. Firstly, that the progressive meaning triggering the durative reading could be expressed by the gerund form, which is present in both periphrases. Secondly, that the difference between the IPP and the PPP (i.e., the fact that only the PPP allows for the preparatory phase reading and for the iterative reading) is due to the aspectual value of the auxiliary form. In this sense, while the duration in both cases would be given by “-ndo” (“-ing”) form, the possibility of expressing the iterative reading would depend on the perfective auxiliary ‘esteve/estuvo’.

<sup>13</sup> As we have already noticed, the PPP gives rise to iterative readings when combined with certain kind of predicates; particularly, achievement denoting a reversible change of state. In a durative situation, this property allows the iteration of the non-durative eventuality, producing the interpretation of a repetition of events. Predicates as “fechar a porta” / “cerrar la puerta” (close the door), “assustar-se” / “asustarse” (get frightened) are also examples that show this property and follow the same behavior.

To sum up, even if achievements in PPP and in IPP are alike in describing durative situations, they behave differently because of the readings they give rise to. Achievements in PPP can describe situations in which the event denoted by the predicate takes place more than just once. Since this interpretation is available only when they are combined with PPP (and not with IPP) it could be the case that this reading is possible because of the meaning of the periphrasis.

## 2.2 PPP with atelic predicates

Atelic events do not have a culmination point (i.e., a *telos*). One distinctive characteristic of these predicates is that they are homogeneous, that is, they are composed of parts that are considered identical. Among atelic events, we can find activities, semelfactives and states, being the main difference between them the fact that only the last ones are non-dynamic.

The literature on atelic events in English has shown that activities and semelfactives accept the IPP, while states reject it.<sup>14</sup>

- (47) John is running.
- (48) John is knocking the door.
- (49) a. \*John is being tall.  
b. \*John is hating apples.  
c. \*John is knowing French.

In BrP and in AS, states behave in a different way. The IPP can combine with “transient” states (50) (similar to stage level predicates) but they are odd with permanent states (51) (similar to individual level predicates).<sup>15</sup>

- (50) a. João estava sendo bom.  
b. Juan estaba siendo bueno.  
John be<sub>.3.Person.Imperf</sub> being good.

<sup>14</sup> As pointed by and anonymous reviewer, to whom we thank, in some cases IPP with states are acceptable in English. Cf. Guimarães (2017).

<sup>15</sup> Basso and Ilari (2004) provides a deep investigation of progressive states in BrP.

- (51) a. \*João estava sendo alto.  
 b. \*Juan estaba siendo alto.  
 John be<sub>.3.Person.Imperf</sub> being tall.

When the IPP combines with a non-permanent state, the meaning is almost like the one of activities: it denotes a situation where the predicate takes place at intervals of time (and not at every instant of time). In these cases, non-permanent states behave as non-agentive activities.

In cases where the IPP is acceptable, namely, (50), the actionality of the predicate does not suffer any kind of change: the sentence describes a durative and atelic situation which holds in time. Non-permanent states (behaving as activities) and states in PPP (cf., (52) and (53)) also keep their basic aspectual meaning, given that they already are durative and atelic.

- (52) a. João esteve sendo bom.  
 b. Juan estuvo siendo bueno.  
 John be<sub>.3.Person.Perf</sub> being good.
- (53) a. \*João esteve sendo alto.  
 b. \*Juan estuvo siendo alto.  
 John be<sub>.3.Person.Perf</sub> being tall.

As for semelfactives, even though they are inherently instantaneous (i.e., non-durative) and atelic predicates (SMITH, 1991; ROTHSTEIN, 2004), when they combine with certain operators (such as the progressive one), they can describe atelic and durative situations in which there is repetition. As we have seen before, when PPP combines with a telic predicate, the *telos* does not seem to be reached. On the other hand, when they combine with non-durative predicates, they give rise to durative situations by iterating the predicate or by focusing on a preparatory phase. Therefore, since activities, semelfactives and states can describe durative and atelic eventualities, they are expected to conserve their aspectual meaning. Hence, they accept durative temporal phrases (*for x time*) and reject temporal phrases which demand a *telos* (*in x time*).

- (54) a. ?João esteve correndo em vinte minutos. (activity)  
 b. ?Juan estuvo corriendo en veinte minutos.  
 John be<sub>.3.Person.Perf</sub> running in twenty minutes.
- (55) a. João esteve correndo durante/por vinte minutos. (activity)  
 b. Juan estuvo corriendo durante veinte minutos.  
 John be<sub>.3.Person.Perf</sub> running for twenty minutes.
- (56) a. ?João esteve tossindo em 20 minutos. (semelfactive)  
 b. ?Juan estuvo tosiendo en 20 minutos.  
 John be<sub>.3.Person.Perf</sub> coughing in twenty minutes.
- (57) a. João esteve tossindo durante/por vinte minutos. (semelfactive)  
 b. Juan estuvo tosiendo durante/por veinte minutos.  
 John be<sub>.3.Person.Perf</sub> coughing for twenty minutes.
- (58) a. ?João esteve sendo bom em vinte minutos. (non-permanent state)  
 b. ?Juan estuvo siendo bueno en veinte minutos.  
 John be<sub>.3.Person.Perf</sub> being good in twenty minutes.
- (59) a. João esteve sendo bom durante/por vinte minutos. (non-permanent state)  
 b. Juan estuvo siendo bueno durante veinte minutos.  
 John be<sub>.3.Person.Perf</sub> being good for twenty minutes.

Furthermore, atelic predicates in PPP allow the following inferences, since they denote situations where the *telos* is not reached.

- (60) a. João esteve conversando com um médico → João conversou com um médico.  
 b. Juan estuvo conversando con un doctor → Juan conversó con un doctor.  
 John be<sub>.3.Person.Perf</sub> talking with a doctor → John talked with a doctor.
- (61) a. João esteve tossindo → João tossiu.  
 b. Juan estuvo tosiendo → Juan tosió.  
 John be<sub>.3.Person.Perf</sub> coughing → John coughed.
- (62) a. João esteve sendo bom → João foi bom.  
 b. Juan estuvo siendo bueno → Juan fue bueno.  
 John be<sub>.3.Person.Perf</sub> being good → John was good.

To sum up, we can sketch the following conclusions. PPP seems to be sensitive to the three aspectual meanings presented in the introduction, that is, dynamicity, duration and telicity. First of all, as well as IPP, this periphrasis requires predicates with some degree of dynamicity. That is the reason why it cannot combine with permanent states. Secondly, it combines with durative predicates. If the predicate it combines with is a non-durative one (achievements or semelfactives, for instance), PPP forces an interpretation in which the situation described is durative, via repetition/iteration or a (durative) preparatory phase. Finally, when PPP combines with telic events, it does not imply that the *telos* is reached, and in this respect it differs from the perfective.

It seems that the PPP and the IPP are very similar, but it is important to note that they are not identical. In the next section, we explore in more detail the differences as well as the similarities between these two periphrases.

### 3 PPP versus IPP: similarities and differences

The aim of this section is to make a systematic comparison of both periphrases in order to find out the similarities but particularly the differences in meaning they present.

As it has been shown in previous sections, PPP and IPP behave alike in certain contexts. First of all, in both cases when the periphrases are combined with telic predicates, the culmination of the event is not implied. That is to say, the interpretation that the *telos* was reached only arises with the non-periphrastic perfective. Both the PPP and the IPP give rise to a reading where the event could be culminated, but the opposite situation is implied. With respect to the examples below, only in (63) we could affirm that the room is arranged, meanwhile in (64), (65) and (66) the interpretation is that it is not arranged, even though this possibility is also allowed.

- (63) a. João arrumou o quarto ontem.  
 b. Juan ordenó la habitación ayer.  
 John arrange<sub>3.Person.Perf</sub> the room yesterday.

- (64) a. João arrumava o quarto ontem.<sup>16</sup>  
 b. Juan ordenaba la habitación ayer.  
 John arrange<sub>.3.Person.Imperf</sub> the room yesterday.
- (65) a. João estava arrumando o quarto ontem.  
 b. Juan estaba ordenando la habitación ayer.  
 John be<sub>.3.Person.Imperf</sub> arranging the room yesterday.
- (66) a. João esteve arrumando o quarto ontem.  
 b. Juan estuvo ordenando la habitación ayer.  
 John be<sub>.3.Person.Perf</sub> arranging the room yesterday.

That is the reason why PPP and IPP accept phrases that emphasize that the event has not actually culminated (“but s/he didn’t P”) and they reject temporal adjuncts that imply the culmination of the event (“in x time”). However, even though PPP has the same effect in implying the non-culmination of the event it combines with, it seems to conserve its perfective meaning since the event (with or without its *telos*) is presented as closed, i.e., as no longer being the case, no longer happening or occurring, as the test in 1.1 with “when” clauses shows.

Second, both periphrases do not move the flux of narrative forward. Hence, when a narrative presents events expressed by these periphrases, they can be overlapped, and they do not have a necessary relation of succession.

Third, the PPP as well as the IPP denote durative situations, and when they are combined with non-durative events (achievements or semelfactives), these periphrases trigger a durative reading, either a “preparatory phase” one or an iterative one. Even though they both describe durative situations, the readings available for each periphrasis are not the same.

<sup>16</sup> It is worth noting that the non-periphrastic imperfective in BrP can also have a modal value when combined with a punctual temporal adjunct, besides the imperfective interpretation (restricted to formal written contexts). In contrast to the IPP, which has a pure progressive meaning, a synthetic imperfective form as “João tomava um café agora” (“John have<sub>.3.Person.Imperf</sub> a coffee now”) can also mean “John would have a coffee now”. Thus, the sentence with the non-periphrastic imperfective form would have the same interpretation as “João tomaria um café agora”. However, AS does not show such a contrast; the differences between the IPP and the synthetic imperfective form are mainly stylistic.

In what concerns the differences between these periphrases, we also already have shown some. The main superficial difference is related to the form of the auxiliary in each construction: while in the IPP the verb “*estar*” has an imperfective form, in the PPP it shows a perfective morphology. This aspectual contrast present in the auxiliary brings up a differentiated behavior that, we claim, is due to the fact that perfective aspect in the PPP presents the event as closed, whereas the imperfective aspect in IPP does not.

Firstly, unlike the IPP, an event in PPP cannot be interrupted by another event, as we can see in the contrast (67-68).

- (67) a. ? No exato momento em que vi Maria, ela esteve alcançando o pico da montanha.  
 b. ? En el exacto momento en que vi a María, ella estuvo alcanzando la cima de la montaña.

In the exact moment I saw Mary, she be<sub>.3.Person.Perf</sub> reaching the top of the mountain.

- (68) a. No exato momento em que vi Maria, ela estava alcançando o pico da montanha.  
 b. En el exacto momento en que vi a María, ella estaba alcanzando la cima de la montaña.

In the exact moment I saw Mary, she be<sub>.3.Person.Imperf</sub> reaching the top of the mountain.

Second, PPP allows for the combination with delimiting durational adverbs, which establish a closed temporal interval: either “*durante x tempo*” / “*durante x tiempo*”, that has been presented before, or other kinds of adverbs that describe a delimited period of time.

- (69) a. #João estava pintando o quadro durante três horas.  
 b. #Juan estaba pintando el cuadro durante tres horas.  
 John be<sub>.3.Person.Imperf</sub> painting the picture for three hours.

- (70) a. João esteve pintando o quadro durante três horas.  
 b. Juan estuvo pintando el cuadro durante tres horas.  
 John be<sub>.3.Person.Perf</sub> painting the picture for three hours.

Third, the iterative meaning that arises with achievements in PPP is due to the same fact: since in the PPP the event is presented as closed,

but also as having duration, the only possibility is for the (individual, singular, or punctual) event to be repeated.

- (71) a. João esteve acordando por toda a manhã. —> ele acordou  
 b. Juan estuvo despertándose toda la mañana —> se despertó.  
 John be<sub>.3, Person, Perf</sub> waking up the whole morning —> he woke up.

Similarly, in AS, accomplishments in PPP can give rise to a reading where it expressed a situation of performing and finishing the same event (over and over) for a period of time.

- (72) a. Juan estuvo armando un rompecabezas la semana pasada.  
 John be<sub>.3, Person, Perf</sub> doing a puzzle last week.

However, this is not allowed with the IPP. (73) can only have two kinds of meanings: (i) it can be the background of a main perfective predicate ('He was making a puzzle when...'); (ii) it can mean that the event of making (but not finishing) a puzzle holds in different moments of the period denoted by the temporal phrase. These two meanings are also present in BrP; thus, with respect to the IPP, they behave similarly.

- (73) a. Juan estaba armando un rompecabezas la semana pasada.  
 John be<sub>.3, Person, Imperf</sub> making a puzzle last week.

We will come back to this difference between AS and BrP in the next section.

#### 4 PPP and IPP in Argentina and in Brazil

In this section we focus on the differences found in the two languages regarding the interpretation of these periphrases. As it has been shown above, differences between BrP and AS are restricted to the PPP. More concretely, as we will see below, they concern the meaning obtained when the PPP combines with telic events.

Let us begin with the interpretation of accomplishments in PPP.

- (74) a. ?João esteve correndo até a farmácia em dez minutos.  
 b. ?Juan estuvo corriendo hasta la farmacia en diez minutos.  
 John be<sub>.3, Person, Perf</sub> running to the drugstore in ten minutes.

In BrP, a sentence like (74) became acceptable if taken to express a past ability that João had, in other words, if the event of “John running to the store in 10 minutes” is interpreted as a capacity that John used to have for some time in his life. This non-episodic interpretation is not allowed in AS. In fact, the only way to obtain this non-episodic interpretation in AS is with imperfective morphology on the main verb (i.e., without the periphrasis and the gerund):

- (75) a. Juan corría hasta la farmacia en diez minutos.  
 John run<sub>.3.Person.Imperf</sub> to the drugstore in ten minutes

Another difference between BrP and AS regarding accomplishments in progressive periphrases has to do with the possibility of expressing a non-episodic event of doing/holding and finishing the eventuality expressed by the predicate. As observed above, the main reading of sentences such as (76) is the one in which it is not implied that the *telos* of the event is reached. That is, the one in which the event of John making a puzzle is not finished.

- (76) a. João esteve montando um quebra-cabeça na semana passada.  
 b. Juan estuvo armando un rompecabezas la semana pasada.  
 John be<sub>.3.Person.Perf</sub> doing a puzzle last week.

However, as observed in section 3, in AS there is another interpretation for a sentence such as (76.b) according to which John has actually made (and finished) the puzzle. For this interpretation to be obtained it is necessary for the event of making a puzzle to be more than one. That is to say, a sentence such as (76.b) can also mean that John made a puzzle over and over in a bounded period of time (i.e., last week). This interpretation is allowed in AS, but does not arise in BrP.

- (77) a. João estava montando um quebra-cabeça na semana passada.  
 b. Juan estaba armando un rompecabezas la semana pasada.  
 John be<sub>.3.Person.Perf</sub> making a puzzle last week.

In AS, (77.b.) could only denote two kinds of situations. Firstly, it can be a background; in this case, the IPP works as a frame for a main perfective predicate introduced by a “when clause” (*Last week, John was making a puzzle, when he realized his dog was missing*). Secondly,

it can mean a unique non-continuous event of making a puzzle (the same puzzle) that holds in different moments of the interval denoted by the temporal phrase; in this case, we have an event of P fragmented in different scenes, they all temporally located in the interval of one week. Thus, in AS, (75.b.) does not give rise to the reading of more than one event of making (and finishing) the puzzle. In this respect, BrP (77.a.) shows the same behavior.

To sum up, the periphrases differ in BrP and AS in two respects, regarding their meaning when combined with accomplishments. Firstly, in BrP a sentence with an accomplishment in PPP and a temporal adjunct that demands the *telos* (as “em x tempo” / “en x tiempo”) gives rise to an (even if hard to grasp) acceptable (past) habitual interpretation; in AS, however, this interpretation is not possible. Secondly, in AS the PPP can trigger the interpretation of more than one event of P, which does not arise with the IPP. In BrP, iterative interpretations with accomplishments are not allowed with the PPP.

With all this data about the PPP in both languages, we propose a semantic analysis below.

## 5 A formal analysis

In this section, we sketch a semantic analysis for the two periphrases, which we claim accounts for the interpretations that we presented above. In order to summarize the discussion presented so far, we repeat here the main readings our analysis should account for. Firstly, we should explain the fact that the imperfective periphrasis as well as the perfective periphrasis give rise to durative and “detelicized” events. As observed in section 2.1, when they are combined with punctual predicates, they give rise to preparatory phase or iterative readings (both of them durative events) and, when they are combined with accomplishments, they give rise to events whose *telos* is not reached. That is to say, our analysis should be able to propose a mechanism, available for the two periphrases, that could derive such readings. Secondly, we should explain the fact that only the perfective progressive periphrasis can give rise to iterative readings with achievements and some accomplishments, as observed in sections 2.1 and 3. That is, the analysis we sketch here should present a mechanism available in the perfective periphrasis – but not in the imperfective one – capable of deriving these iterative readings.

Considering these remarks, we propose a compositional analysis of the progressive periphrases according to which (i) the gerund form, realized in both periphrases, expresses an operator which is the responsible for “detelicizing” and making the event durative – roughly, the result will be a continuous or progressive interpretation, and (ii) the perfective operator in the auxiliary of the PPP is able to make the eventuality iterative. In the following we explain this analysis in more detail.

We claim that the gerund form expresses a semantic operator PROG (78) having the following denotation, as proposed by Rothstein (2004, p. 46-47):

$$(78) \quad \neg \text{NDO}(\text{VP}) \rightarrow \lambda e. \text{PROG}(e, \lambda e'. \text{VP}(e') \wedge \text{Ag}(e') = x).$$

$$(79) \quad |\text{PROG}(e, P)|_{w, g} = 1, \text{ iff } \exists e' \exists w' : \langle e', w' \rangle \in \text{CON}(g(e), w) \text{ and } |P|_{w', g(e')} = 1 \text{ where } \text{CON}(g(e), w) \text{ is the continuation branch of } g(e) \text{ in } w.$$

In a simplified way, what (79) says is that, for the progressive to be true, there must be a stage  $e'$  of  $e$  in a world  $w'$  that is part of the event  $e$  in a world  $w$ ; i.e., a progressive event is one which can develop in a bigger event of the same type.

Therefore, when the predicate combines with PROG (expressed by the gerund form) in both cases, PPP and IPP, PROG has two results: (i) it makes the event denoted by the predicate durative because it says that the event must contain stages  $e'$  of  $e$ , and (ii) it makes the event “detelicized” given that  $e'$  must be true in  $w'$  being  $e$  (the denotation of the predicate) evaluated in a world  $w$  which is not necessarily the same as  $w'$ .<sup>17</sup> In short, PROG results in events that have stages and its true-conditions focus on one stage holding in a world which is not necessarily the word where the denotation of  $P$  is evaluated.

<sup>17</sup> It is worth noting that the evaluation of  $e$  (the denotation of the predicate) in a different world from the world at which the stage  $e'$  is evaluated is what has allowed literature on the subject to avoid the imperfective paradox, noticed since Dowty (1979). If the stage  $e'$  were evaluated at the same word where the denotation of the predicate is, as proposed initially by Bennett and Partee (1978), the meaning of the progressive would be wrong for telic predicates, since it would have to be the case that the denotation of the predicate is true. That is to say, for *John reading the book* to be true, it would also be true *John read the book* (the denotation of  $P$ ), but the truth of the denotation of a telic predicate implies that the *telos* is reached, something missing in the meaning of the progressive, since there is no necessity for the *telos* to be reached.

In consequence, being PROG the meaning expressed by the gerund, lets us explain why both periphrases have the same effect on telic predicates: if it takes an accomplishment, it describes a durative event whose *telos* is not (necessarily) reached; if it takes an achievement, it forces a durative temporal structure, via repetition or a preparatory phase, and does not guarantee that the *telos* is reached.

Regarding atelic events, semelfactives and activities do not behave equally. Since the former are instantaneous events, they behave as achievements in that the progressive operator forces a durative temporal structure, in this case only via repetition. As for activities, given that they are homogeneous and durative events, they already have stages. Therefore, their temporal structure does not suffer any change.

Let us now consider the meaning of the perfective auxiliary in order to derive the iterative readings. As presented in section 1, the perfective in the auxiliary is an operator PERF that says that the event has to be included in the topic time.

$$(80) \quad [[PF]]^{Mg} = \lambda P \exists e[\tau(e) \subseteq t_T \wedge P(e)]$$

We claim that the instruction for these cases is the following: for a predicate in the perfective to be true, it must be the case that the event expressed by the predicate is included (i.e., presented as closed) *at least* once in the topic time. That is what allows explaining the two meanings available in AS in sentences such as (81): it can mean that the ongoing event of making a puzzle is included in the topic time (i.e., presented as closed but not necessarily finished) one time or more than one. The possibility of being included more than once is what explains AS iterative interpretation and BrP semelfactive interpretation.

- (81) a. João esteve montando um quebra-cabeça (algumas vezes) na semana passada.  
 b. Juan estuvo armando un rompecabezas (una y otra vez) la semana pasada.  
 John be<sub>.3.Person.Perf</sub> making a puzzle (over and over) last week.

As for the interpretation of achievements with the perfective operator, we claim that since the time span created by the PROG operator is bigger than the instantaneous event, the event time is included more than once in the topic time, giving rise to the iterative interpretation (when the “preparatory phase” interpretation is not available).

- (82) a. João esteve fechando a porta por toda a manhã.  
 b. Juan estuvo cerrando la puerta toda la mañana.  
 John be<sub>.3.Person.Perf</sub> closing the door (the whole morning).

In contrast, given that the imperfective meaning states that the topic time must be included in the event time (83), it is not possible for sentences like (84-85) to denote an iterative event. In these cases, the ongoing event denoted by the predicate must evolve beyond a certain topic time.

- (83)  $[[\text{IMPF}]]^{M,g} = \lambda P \exists e [t_T \subseteq \tau(e) \wedge P(e)]$
- (84) a. João estava montando um quebra-cabeça.  
 b. Juan estaba armando un rompecabezas.  
 John be<sub>.3.Person.Imperf</sub> making a puzzle.
- (85) a. João estava fechando a porta.  
 b. Juan estaba cerrando la puerta.  
 John be<sub>.3.Person.Imperf</sub> closing the door.

To sum up, while the homogeneous and durative structure is due to the progressive operator placed in the gerund form, the iterative reading available for the perfective periphrasis (but not for the imperfective one) depends on the meaning of PERF, which states that the event time should be included *at least* one time in the topic time ( $\tau(e) \subseteq t_T$ ).

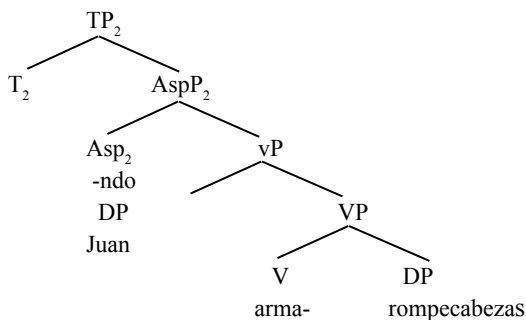
Concerning the syntax of these periphrases, we claim that the gerund form (as well as all non-finite forms) is a cluster of phrases that includes a defective TP, which is semantically null<sup>18</sup> but syntactically necessary in order to be bound to the TP<sub>1</sub> above it;<sup>19</sup> an AspP, which is

<sup>18</sup> Being semantically null is expressed by an identity function that takes a function and returns the same function.

<sup>19</sup> The presence of TP in constructions such as the ones studied in this paper could be discussed since this form does not seem to present information of that kind. We claim that its presence is needed because there are some instances of gerunds expressing a different time than the one expressed in the main predicate (i.e., “Estudando todas as noites, João pôde passar no exame”/“Estudiando todas las noches, Juan pudo aprobar el examen” (‘Studying all nights, John could pass the exam’)). We claim that when TP<sub>2</sub> is bound by TP<sub>1</sub> we have constructions where the gerund and the main predicate constitute the same predication, for instance, in cases of periphrases.

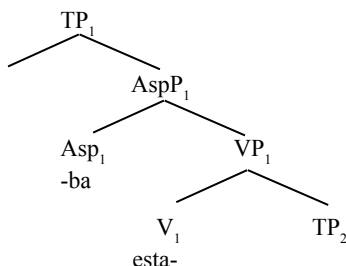
responsible for the imperfectivity (in the case of gerund); and, finally, the vP-VP domain where the lexical and descriptive content is. The progressive operator we presented in this paper is located in the head of AspP. Therefore, “-ndo” (“-ing”) in ASP is a function that takes a  $\langle s, t \rangle$  and returns the sets of events of P that are ongoing event. That is, it specifies the domain of events.

- (86) Gerund form for “Juan armando rompecabeza”/“João montado quebra-cabeças”  
John making puzzles



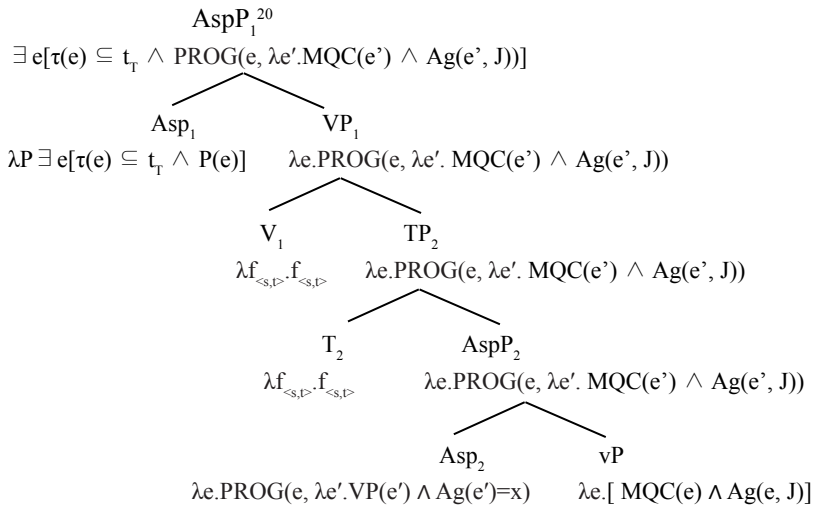
The auxiliary, in turn, is represented syntactically as a VP whose complement is a defective TP (it always takes a verbal phrase) and without a vP above it (there is no need to insert a subject). This prevents the auxiliary from having an argument structure. On the contrary, the auxiliary VP does have full TP and AspP, and it is in AspP that we found the denotation of perfective and imperfective presented in this paper. Semantically, the auxiliary VP takes a  $\langle s, t \rangle$  and returns the same type if the verb is a copulative one. That is to say, if the auxiliary has no lexical content (i.e., if it is a copula) the head of VP is an identity function.

- (87) Auxiliary form



Therefore, the semantic derivation of a sentence such as (84) would be as follows:

- (88) a. João esteve montando um quebra-cabeça. (“montar um quebra-cabeça” = MQC)  
 b. Juan estuvo armando un rompecabezas.  
 John was making a puzzle.



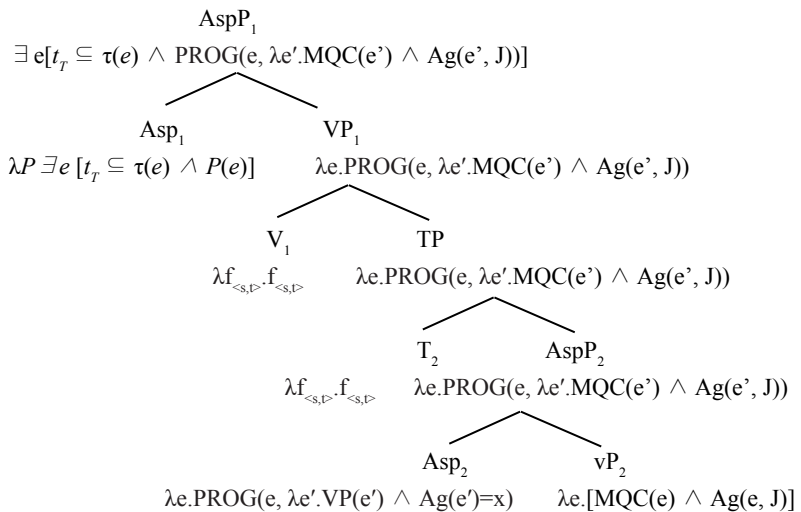
As we can see in (88),  $\text{Asp}_2$  (i.e., the progressive operator expressed by *–ndo*) takes the *vP* (containing the *VP*), which is the event of making a puzzle, and returns the set of ongoing events of making a puzzle. Since  $\text{T}_2$  and  $\text{V}_1$  are semantically null, they express an identity function. Then,  $\text{Asp}_1$  (i.e., the perfective operator expressed by the auxiliary) takes the set of ongoing events of making a puzzle, expressed by the  $\text{VP}_1$ , and returns that the event time of making a puzzle should be included at least once in the topic time. We can see this derivation below:

<sup>20</sup> It is worth mentioning that some parts of the representations for the semantic calculus are intentionally omitted, for the sake of simplicity.

- (89) a.  $[[vP]] = \lambda e.[MQC(e) \wedge Ag(e, J)]$   
 b.  $[[Asp_2]] = \lambda e.PROG(e, \lambda e'.VP(e') \wedge Ag(e')=x)$   
 c.  $[[AspP_2]] = \lambda e.PROG(e, \lambda e'.VP(e') \wedge Ag(e')=x)(\lambda e.[MQC(e) \wedge Ag(e, J)])$   
 $= \lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$   
 (...)
   
 d.  $[[VP_1]] = \lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$   
 e.  $[[Asp_1]]^{M.g} = \lambda P \exists e[\tau(e) \subseteq t_T \wedge P(e)]$   
 f.  $[[AspP_1]] = \lambda P \exists e[\tau(e) \subseteq t_T \wedge P(e)] (\lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))])$   
 $= \exists e[\tau(e) \subseteq t_T \wedge PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$

As for the imperfective progressive periphrasis, the derivation is exactly like the one presented in (88), except for denotation of  $Asp_1$ , which takes the set of ongoing events of making a puzzle, expressed by the auxiliary in  $VP_1$ , and returns the meaning that states that the topic time should be included in the event time of making a puzzle.

(90)



- (91) a.  $[[vP]] = \lambda e.[MQC(e) \wedge Ag(e, J)]$   
 b.  $[[Asp_2]] = \lambda e.PROG(e, \lambda e'.VP(e') \wedge Ag(e')=x)$   
 c.  $[[AspP_2]] = \lambda e.PROG(e, \lambda e'.VP(e') \wedge Ag(e')=x)(\lambda e.[MQC(e) \wedge Ag(e, J)])$   
 $= \lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$   
 (...)
   
 d.  $[[VP_1]] = \lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$   
 e.  $[[Asp_1]]^{M,g} = \lambda P \exists e[t_r \subseteq \tau(e) \wedge P(e)]$   
 f.  $[[AspP_1]] = \lambda P \exists e[t_r \subseteq \tau(e) \wedge P(e)] (\lambda e.[PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))])$   
 $= \exists e[t_r \subseteq \tau(e) \wedge PROG(e, \lambda e'.MQC(e') \wedge Ag(e', J))]$

To sum up, while the progressive meaning in *-ndo* allows us to explain the similar behavior of both periphrases (i.e., it makes events durative and homogeneous), the alternation between perfective and imperfective meaning in the auxiliary is what explains the differences between IPP and PPP regarding iterative readings.

## 6 Concluding remarks

In this paper, we have analyzed the perfective progressive periphrasis (PPP) and the imperfective progressive periphrasis (IPP) in Argentinian Spanish (AS) and in Brazilian Portuguese (BrP), emphasizing in the differences they present when combined with the Vendlerian aspectual classes; in particular, with telic events. We have proposed a compositional model that can account for the similarities and differences they show: while the gerund form (present in both periphrases) expresses a progressive meaning that gives rise to durative and homogeneous events, the auxiliary in PPP expresses a perfective meaning that allows for iterative readings.

In order to describe the behavior of the PPP in BrP and in AS, and its differences with respect to the IPP, we firstly discussed the aspectual value of the periphrases, in section 1. We showed that considerations over the perfectiveness or imperfectiveness of the periphrasis are inconclusive, since it behaves as perfective with respect to some tests and as imperfective with respect to other tests. Secondly, in section 2, we studied the interpretations arisen when they combine with the aspectual classes. We showed that while progressive periphrases do not change the temporal structure of atelic events, the combination of these periphrases with telic events gives rise to durative progressive (i.e., non-telic) readings. We also

observed that the main difference between the IPP and the PPP is that the latter can give rise to iterative readings, not allowed with the IPP. That is, with telic predicates, both the IPP and the PPP trigger an interpretation in which the *telos* is not reached, but the particular readings they give rise are different. On one hand, with achievements they both result in a durative event, but with a different nature: while the IPP triggers only a preparatory phase reading, the PPP can also trigger an iterative reading. The same mechanism is available for accomplishments: while both trigger detelicized events interpretations, only the PPP can also give rise to an iterative reading, as data from AS shows. The particular differences between AS and BrP are studied in section 4.

Considering these observations, we have proposed a formal model that allows us to account for the fact that, while both the IPP and the PPP give rise to durative progressive events, only the PPP triggers iterative interpretations, which are not allowed with IPP. As we pointed out in the compositional analysis, the gerund form “-ndo”, realized in both periphrases, is responsible for the durative progressive meaning, meanwhile the different interpretations (e.g., the possibility of iteration with the PPP) are due to the aspectual value of the auxiliary. In particular, we proposed that the gerund expresses a progressive operator PROG (following ROTHSTEIN, 2004), which states a durative an ongoing event. On the other hand, we claimed that the perfective value of the auxiliary is an operator PERF that includes the time of the event in the topic time at least once. Thus, this mechanism allowed us to account for the progressive readings of the PPP, since the event is presented as closed but not necessarily completed. At the same time, we could account for the iterative readings of the PPP, impossible with the IPP, by claiming that, following the instruction of the PERF operator, the event time must be included in the topic time once or more than once.

### **Authorship statement**

Romina Trebisacce and Victoria Ferrero were responsible for the Argentinian Spanish data and Renato Basso, for the Brazilian Portuguese data. The syntactic analysis was developed by Romina Trebisacce and Victoria Ferrero, and the semantic counterpart was developed by Renato Basso. Together they organized and structured the text, and came up with the final analysis. The whole process was collaborative.

## References

ALTSHULER, D. G. There Is No Neutral Aspect. *Semantics and Linguistic Theory*, Ithaca, NY, v. 23, p. 40-62, 2013. DOI: <https://doi.org/10.3765/salt.v23i0.2681>

BASSO, R. M. *Telicidade e detelicização: semântica e pragmática do domínio tempo-aspectual*. 2007. 288f. Dissertation (Master in Linguistics) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2007.

BASSO, R. M.; ILARI, R. Estativos e suas características. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 4, n. 1, p. 15-26, 2004. DOI: <https://doi.org/10.1590/S1984-63982004000100003>

BENNETT, M.; PARTEE, B. H. *Toward the Logic of Tense and Aspect in English*. Bloomington: Indiana University Linguistics Club, 1978.

BERGAMINI-PEREZ, J. F.; BASSO, R. M. Adjuntos temporais e measure phrases: uma proposta semântica. *Cadernos de Estudos Lingüísticos*, Campinas, v. 58, n. 2, p. 345-367, 2016. DOI: <https://doi.org/10.20396/cel.v58i2.8647159>

BOHNEMEYER, J. Aspect vs. Relative Tense: The Case Reopened. *Natural Language & Linguistic Theory*, [S.l.], v. 32, n. 3, p. 917-954, 2014. DOI: <https://doi.org/10.1007/s11049-013-9210-z>

BONOMI, A. Aspect, Quantification and When-Clauses in Italian. *Linguistics and Philosophy*, [S.l.], v. 20, n. 5, p. 469-514, 1997. DOI: <https://doi.org/10.1023/A:1005388230492>

DOWTY, D. *Word Meaning and Montague Grammar*. Dordrecht: Kluwer Academic Publishers, 1979. DOI: <https://doi.org/10.1007/978-94-009-9473-7>

GUIMARÃES, P. A. L. *Verbos de estado e morfologia de progressivo: um estudo comparativo entre o português do Brasil e o inglês dos Estados Unidos da América*. 2017. 202f. Dissertation (Master in Linguistics) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2017.

KAMP, H.; ROHRER, C. Tense in Texts. In: BÄUERLE, R.; SCHWARZE, C.; VON STECHOW, A. (ed.). *Meaning, Use and Interpretation of Language*. Berlin; New York: de Gruyter, 1983. p. 250-269.

KLEIN, W. *Time in Language*. London: Routledge, 1994.

LANDMAN, F. The Progressive. *Natural Language Semantics*, [S.l.], v. 1, n. 1, p. 1-32, 1992. DOI: <https://doi.org/10.1007/BF02342615>

PIRES DE OLIVEIRA, R.; BASSO, R. M. Sobre a semântica e a pragmática do perfectivo. *Revista Letras*, Curitiba, v. 81, p. 123-140, 2010. DOI: <https://doi.org/10.5380/rel.v81i0.17327>

ROTHSTEIN, S. *Structuring Events*. Oxford: Blackwell, 2004. DOI: <https://doi.org/10.1002/9780470759127>

SMITH, C. S. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers, 1991.

SQUARTINI, M. *Verbal Periphrases in Romance: Aspect, Actionality, and Grammaticalization*. Berlin: Mouton de Gruyter, 1998. DOI: <https://doi.org/10.1515/9783110805291>.





## Modelos de procesamiento sintáctico y sus implicaciones para el estudio del lenguaje

### *Syntactic processing models and their implications for the study of language*

Noelia Ayelén Stetie

Universidad de Buenos Aires (UBA), Consejo Nacional de Investigaciones Científicas  
y Técnicas (CONICET), Buenos Aires / Argentina

nstetie@filo.uba.ar

<http://orcid.org/0000-0001-7602-6942>

**Resumen:** Comprender cómo se procesa el lenguaje ha sido interés central de la psicolingüística hace varias décadas ya que entender los procesos involucrados aporta al estudio del lenguaje, pero, también, al estudio y comprensión de la mente. En este artículo se presenta una revisión de los modelos de procesamiento sintáctico y sus aportes a la comprensión acerca del funcionamiento del lenguaje y de la cognición. En primer lugar, se retoman algunas de las discusiones clásicas acerca del procesamiento del lenguaje y los procesos cognitivos en general. En segunda instancia, se realiza una revisión de distintos modelos de procesamiento sintáctico propuestos en las últimas décadas, sus características y supuestos acerca de la facultad del lenguaje. Por último, se comparan las diferentes propuestas en función de los debates clásicos presentados y su posición acerca de la universalidad de ciertos procesos cognitivos.

**Palabras-clave:** procesos cognitivos; comprensión del lenguaje; procesamiento sintáctico; adjunción.

**Abstract:** Understanding how language is processed has been a theme of an utmost interest for psycholinguistics for several decades, since fully acknowledging the processes involved contributes not only to the study of language, but also to both the study and understanding of the mind. This article presents a review of syntactic parsing models and their contributions to the comprehension of the functioning of language and cognition. First of all, some of the classical discussions about language processing and cognitive processes in general are analyzed. Secondly, different models of syntactic parsing proposed in the last decades as well as their characteristics and assumptions about the language faculty are reviewed. Finally, the different proposals are compared

in terms of the classical debates presented and their position on the universality of certain cognitive processes.

**Keywords:** cognitive processes; language comprehension; syntactic parsing; attachment.

Recebido em 11 de fevereiro de 2021

Aceito em 29 de março de 2021

## 1 Introducción

A lo largo de los años, se han propuesto varios modelos de procesamiento del lenguaje que explican muy bien algunos factores, pero que presentan explicaciones *ad hoc* para otros. El objetivo de este artículo es presentar un estado del arte sobre los modelos de procesamiento sintáctico para dar cuenta de qué mecanismos intervienen en el procesamiento del lenguaje y cómo lo hacen. Sin embargo, no es posible realizar dicha tarea sin enmarcarla en una discusión más general sobre el lenguaje, los procesos cognitivos y su arquitectura funcional. Debido a la metodología experimental de la psicolingüística y a los desafíos que esta conlleva –la contrastación empírica siempre es, por necesidad, fragmentaria–, las investigaciones suelen presentar recortes muy específicos de problemas complejos y esto lleva a que, en muchos casos, se pierdan de vista los debates más generales que están detrás.

Para poder reflexionar sobre los compromisos teóricos y epistemológicos de los distintos modelos, es importante enmarcar la discusión sobre los resultados de las investigaciones experimentales y los modelos de procesamiento sintáctico en un debate más amplio sobre el lenguaje y su arquitectura funcional. El presente trabajo se dividirá en cuatro apartados. Primero se expondrán algunos ejes que han acompañado la discusión sobre el lenguaje y la mente/cerebro en las últimas décadas: procesamiento serial vs. en paralelo, autonomía vs. interacción, flujo informativo abajo-arriba<sup>1</sup> vs. arriba-abajo,<sup>2</sup>

---

<sup>1</sup> *Bottom-up*. En este artículo, siempre que sea posible, se utilizarán términos en español. Sin embargo, debido a que la mayoría de la literatura en psicolingüística está en inglés, se brindarán aclaraciones de los términos en inglés.

<sup>2</sup> *Top-down*.

universalidad vs. variabilidad. En segundo lugar, se presentará un estado del arte sobre los modelos de procesamiento sintáctico. Luego se realizará una comparación entre qué es entendido como un procesamiento óptimo para cada uno de los modelos de procesamiento sintáctico presentados. Por último, se concluirá el artículo con una reflexión sobre lo expuesto y una presentación de nuevos desafíos.

## 2 Algunos debates sobre la facultad del lenguaje

La arquitectura del lenguaje proporciona el marco dentro del cual tiene lugar el procesamiento del lenguaje y, por consiguiente, es responsable de todas las restricciones fundamentales de la conducta de esta facultad. En este sentido, se destacan dos grandes perspectivas que se podrían llamar sintactocéntricas y no sintactocéntricas (JACKENDOFF, 2009). Las primeras consideran a la sintaxis como único componente generativo; mientras que las segundas sostienen que cada componente o interfaz puede generar estructuras y tiene sus propias reglas de combinación.

Por un lado, una de las propuestas clásicas sobre la arquitectura del lenguaje es la del Programa Minimalista (CHOMSKY, 1995, 2015), que asume una organización modular de la mente y del lenguaje (FODOR, 1983) y un procesamiento serial de base derivacional. Coloca la capacidad generativa del lenguaje en el componente sintáctico y considera a la fonología y a la semántica como componentes interpretativos, periféricos. Es decir que las reglas combinatorias de las sintaxis son las únicas que especifican la estructura, lo que vuelve a la sintaxis el único componente generativo y, también, el mediador entre los demás.

Por otro lado, propuestas como la de Jackendoff (2003, 2010) conciben una arquitectura en paralelo.<sup>3</sup> Este enfoque considera la

---

<sup>3</sup> El término *en paralelo* muchas veces se utiliza con distintas acepciones. Por un lado, una arquitectura en paralelo es aquella que propone que diferentes componentes o módulos procesan información de distintas fuentes en paralelo y no de forma serial. Por otro lado, procesamiento en paralelo puede referirse también a que, frente a dos interpretaciones posibles en un determinado punto, el sistema sigue procesando ambas y llega al final con varias alternativas de interpretación para una misma oración. Si bien de una arquitectura en paralelo se espera que haya procesamiento en paralelo de distintas fuentes de información, no necesariamente se espera que se computen en paralelo diferentes representaciones para la misma oración, ya que los diversos

estructura lingüística como el producto de una serie de capacidades generativas paralelas –pero que interactúan– vinculadas a la fonología, la sintaxis y la semántica. No solo la sintaxis es independiente y con capacidad generativa, sino que también lo son el componente semántico y el fonológico. A su vez, Jackendoff (2010) propone que la relación entre los distintos componentes está mediada por un conjunto de componentes de interfaz, que establecen enlaces óptimos entre las diversas estructuras y sus partes, lo que corre a la sintaxis del lugar de mediadora obligada.

La arquitectura del lenguaje no solo influye en el debate sobre la serialidad o el paralelismo, también lo hace en la discusión acerca de la interacción versus la autonomía. En términos generales, un modelo serial propone módulos autónomos con, frecuentemente, supremacía de la sintaxis; mientras que uno en paralelo propone interacción entre los distintos tipos de información y niveles de representación que se están procesando. Pero también se podría preguntar si las diferentes etapas en el procesamiento son temporalmente discretas –seriales– o si se superponen. En una propuesta en la que hay superposición de etapas, también llamado procesamiento en cascada,<sup>4</sup> se permite que la información fluya de un nivel al siguiente antes de que haya completado su procesamiento (MCCLELLAND, 1979). Si las etapas se solapan, podría haber filtraciones entre estas, lo que explicaría cierta interacción moderada en el procesamiento del lenguaje (BOLAND; CUTLER, 1996).

Otro debate complejo en las ciencias cognitivas y en el estudio del lenguaje en particular es la dirección del flujo de información: procesamiento abajo-arriba o arriba-abajo (EZQUERRO, 1995). Las teorías del procesamiento guiado por los datos, o teorías abajo-arriba, sostienen que el procesamiento se arraiga en la realidad empírica y está guiado por los datos sensoriales, sin influencia de procesos de alto orden. Por otro lado, las teorías arriba-abajo consideran que es posible la influencia de los conocimientos y esquemas de una persona en la percepción. Estas sostienen que en el procesamiento interviene información de alto orden que causa que se espere la presentación de un estímulo en particular. A estas teorías también se las conoce con el

---

componentes podrían ir imponiendo restricciones sobre una única representación. Como así también se podrían proponer modelos con un procesamiento serial, pero que computen en simultáneo distintas interpretaciones de la oración.

<sup>4</sup> *Cascade processing*.

nombre de procesamiento guiado conceptualmente, ya que consideran que las funciones mentales superiores, las experiencias, conocimientos, motivaciones y antecedentes culturales influyen en la percepción del mundo –y de la información lingüística.

Aunque persisten algunas propuestas excesivamente estancas, actualmente en las ciencias cognitivas hay evidencia suficiente de que es necesario pensar en propuestas que conciban algunos procesos con un flujo de información abajo-arriba, pero que acepten que también pueden ser mediados por expectativas, es decir, que conviven con una dirección de la información arriba-abajo. Entender en qué dirección va el flujo de la información sirve para pensar los límites entre percepción y cognición y si efectivamente es posible trazarlos y hay percepción no mediada por la cognición. Si se asume un procesamiento estrictamente abajo-arriba, se estaría considerando que es posible una percepción “pura” y universal sin mediación de la cognición.

Esto se creía que sucedía en algunos procesos de primer orden, como los vinculados a la visión y la audición. La ilusión Müller-Lyer (BERMOND; VAN HEERDEN, 1996; NIJHAWAN, 1991; REDDING; HAWLEY, 1993) fue utilizada durante mucho tiempo como evidencia de un estricto procesamiento abajo-arriba y del encapsulamiento y la universalidad de algunos procesos modulares (FODOR, 1983) ya que el efecto de esta ilusión en muchas situaciones va en contra de los conocimientos y la propia experiencia de los sujetos (EVERETT, 2013). Sin embargo, una investigación (SEGALL; CAMPBELL; HERSKOVITS, 1963) de esta y otras ilusiones perceptivas en sociedades que representaban variedad de categorías culturales encontró que la percepción de la ilusión de Müller-Lyer varía hasta el punto de no existir en algunas culturas. Otra ilusión sensorial muy conocida es el efecto McGurk (MACDONALD; MCGURK, 1978; MCGURK; MACDONALD, 1976; TIIPPANA, 2014), que demuestra un cambio categórico en la percepción auditiva inducido por un discurso visual incongruente. McGurk y MacDonald (1976) grabaron una voz articulando una consonante y la emparejaron con una cara articulando otra consonante. Aunque la señal acústica del habla se reconocía bien por sí sola, se oía como una tercera consonante tras el emparejamiento con el habla visual incongruente. Estas ilusiones sensoriales evidencian la interacción temprana en niveles básicos del procesamiento, lo que invita a considerar distintas direcciones e interacciones en cuanto al flujo de la información.

Por último, caracterizar el procesamiento del lenguaje trae a colación las discusiones sobre la facultad del lenguaje y las distinciones entre total universalidad y determinismo relativista. ¿Cuál es el límite y alcance de la universalidad? Aunque en la actualidad pareciera no haber controversias sobre el hecho de que se adquieren idiosincráticamente algunos aspectos del lenguaje —e.g. vocabulario, acentos—, existe también cierto consenso sobre el hecho de que algunos principios estructurales del lenguaje son universales e innatos (CHOMSKY, 1986; PINKER, 1995). De ser así, existe la posibilidad de que algunos de los procedimientos utilizados para procesar el lenguaje también sean universales e innatos.

La discusión sobre la universalidad y la variabilidad no solo se traslada al procesamiento del lenguaje, sino que abarca diferentes lecturas o interpretaciones, vinculadas a las lenguas, las personas y las situaciones. En relación con el primer punto, la pregunta es si existe un mecanismo que sigue reglas universales comunes a todas las lenguas, si lo universal es este mecanismo en sí y no hay un conjunto de reglas específico o si efectivamente hay un conjunto de principios que se parametrizan para cada lengua o que cambian según el tipo de estructura a analizar. En cuanto al segundo eje, el interrogante gira en torno a explicar el comportamiento a nivel individual: independientemente de la lengua, ¿todas las personas procesan el lenguaje de forma universal o hay diferencias individuales sustanciales en el procesamiento sintáctico? Y si hay diferencias individuales, ¿a qué se deben y cómo se explican? Por último, también se puede considerar la universalidad en relación con las situaciones: ¿siempre se procesa el lenguaje de forma universal o hay distintos requerimientos de la tarea que modulan el procesamiento sintáctico?

En la siguiente sección se realizará un recorrido por las principales propuestas teóricas sobre el procesamiento sintáctico para entender qué papel le otorga cada una a la universalidad y cómo responden a los distintos interrogantes mencionados.

### **3 Modelos de procesamiento sintáctico**

Luego de repasar algunas discusiones clásicas acerca del procesamiento del lenguaje y del funcionamiento de la mente/cerebro en general, se presentará la discusión sobre algunos modelos de procesamiento en el nivel oracional. Se entiende al procesamiento

sintáctico –o *parsing*– como los procesos mentales que segmentan y analizan secuencias de palabras en unidades constituyentes (TRAXLER; GERNSBACHER, 2011, p. 147). Es decir, procesos que se encargan de pasar la estructura lineal de la oración a una estructura jerárquica.<sup>5</sup> A su vez, algunas propuestas toman el concepto de analizador sintáctico –o *parser*– para referirse al dispositivo que lleva a cabo dichos procesos.

Desde la psicolingüística se han postulado diferentes propuestas sobre qué conjunto específico de mecanismos se utiliza al procesar lenguaje y cómo se relacionan las diferentes instancias entre sí. Para comprender una oración se utiliza información sintáctica, semántica, fonológica y pragmática, pero la pregunta es cuándo y cómo. ¿Son instancias separadas durante el procesamiento? ¿Se le da más importancia a un tipo de información sobre otra? Es necesario proponer un conjunto de procesos que dé cuenta de las características generales que se le están adjudicando al lenguaje –e.g. componentes y subcomponentes, información transmitida entre ellos, relaciones temporales entre sus actividades– y de los resultados obtenidos en estudios experimentales. Además, es necesario que se postulen estrategias que no sean simplemente clasificaciones arbitrarias, sino que se deriven de, y por lo tanto revelen, los principios organizativos y las características fundamentales del lenguaje.

### 3.1 Primeros pasos: modularidad y generativismo

Teniendo como marcos la modularidad de la mente –y del lenguaje– propuesta por Fodor (1983) y el generativismo –sintactocéntrico– de Chomsky (1957), Lyn Frazier y Janet Fodor (1978) propusieron uno de los primeros modelos de procesamiento sintáctico que no solo explicaba satisfactoriamente los datos experimentales encontrados hasta el momento, sino que además lo hacía a partir de estrategias de procesamiento que estaban en consonancia con una determinada arquitectura de la facultad del lenguaje. La Máquina de

---

<sup>5</sup> Detrás de esta afirmación opera el supuesto de que en la mente/cerebro se manipulan estructuras jerárquicas y no estructuras lineales. Aunque la mayoría de los modelos de procesamiento sintáctico revisados en este trabajo operan con este supuesto, todavía no está cerrado el debate. Véase la discusión entre Ding, Melloni, Tian y Poeppel (2017) y Franck y Christiansen (2018) para una revisión.

salchichas<sup>6</sup> (FRAZIER; FODOR, 1978) o el Modelo de vía muerta<sup>7</sup> (FRAZIER; RAYNER, 1982), como es conocido actualmente, sostiene un procesamiento serial en etapas.

La primera instancia está a cargo del Empaquetador de frases preliminar,<sup>8</sup> la máquina de salchichas propiamente dicha. Su tarea consiste en asignar estructura a grupos de palabras adyacentes en la cadena léxica y transmitirlos como paquetes de frases separados a la segunda etapa, a cargo del Supervisor de estructura oracional,<sup>9</sup> encargado de conectar las distintas frases y de realizar un seguimiento de las dependencias entre los elementos que están ampliamente separados en la oración. La distinción en dos dispositivos distintos –el Empaquetador y el Supervisor– no depende de qué tipo de operaciones puede realizar cada uno, sino solo de con cuánto material pueden operar (FRAZIER; FODOR, 1978). En otras palabras, es el supuesto del límite de la memoria de trabajo lo que motiva la división del *parser* sintáctico en dos instancias: cuanto más estructurado es el material que almacena, menor es la demanda de almacenamiento (BADDELEY, 2007; MILLER, 1956).

¿Cómo hace entonces el segundo mecanismo para adjuntar frases entre sí? Frazier y Fodor (1978) sostienen que el *parser* se guía por un comportamiento universal para establecer los lazos de adjunción: el principio de adjunción mínima,<sup>10</sup> que estipula que al enfrentarse con una estructura con más de una posibilidad de adjunción, se seleccionará la estructura que menos nodos sintácticos implique.<sup>11</sup> Por ejemplo, en (1) la ambigüedad de adjunción reside en el sintagma “con el paraguas”: una interpretación posible de la oración es adjuntarlo al verbo –“golpeó con el paraguas”– y otra al sintagma nominal –“el hombre con el paraguas”–. Según el principio de adjunción mínima, en esta estructura la opción preferida sería adjuntar al sintagma verbal, ya que es lo que menos nodos genera, priorizando así la economía estructural.

---

<sup>6</sup> *Sausage machine model.*

<sup>7</sup> *Garden path model.*

<sup>8</sup> *Preliminary Phrase Packager* o PPP.

<sup>9</sup> *Sentence Structure Supervisor* o SSS.

<sup>10</sup> *Minimal attachment.*

<sup>11</sup> Como ya se mencionó, la mayoría de los modelos de procesamiento sintáctico que se presentarán operan bajo el supuesto de que el *parser* trabaja con estructuras jerárquicas, no secuenciales, como los árboles sintácticos.

1. La niña golpeó al hombre con el paraguas.

Independientemente de qué tipo de constituyente se vaya a unir, o cuáles sean las alternativas de adjunción, el modelo propone que siempre se va a favorecer la adjunción más simple, la que implique postular menos nodos sintácticos. La explicación se relaciona con las propias características del sistema: las presiones del tiempo y las limitaciones de la memoria de trabajo. La estrategia de adjunción mínima generará la menor demanda posible en la memoria; incluso si resulta ser incorrecta, probar la adjunción mínima primero asegura que el reanálisis,<sup>12</sup> de ser necesario, también se haga de forma eficiente, agregando nodos adicionales al marcador de frase (FRAZIER; FODOR, 1978).

Sin embargo, la adjunción mínima no siempre es la respuesta; cuando ambas interpretaciones de la oración generen la misma cantidad de nodos, el *parser* realizará el análisis cumpliendo con el principio de cierre tardío,<sup>13</sup> los nuevos ítems se agregarán a la cláusula que esté siendo procesada en ese momento. Frazier y Fodor (1978) argumentan que existe evidencia considerable –en inglés– de que el *parser* prefiere mantener las frases abiertas el mayor tiempo posible, en lugar de cerrarlas prontamente. Según este principio, en (2) el procesador sintáctico adjunta a la cláusula que se está procesando en ese momento y la interpretación preferida es que “el cerrajero estaba en el local”.

2. El policía disparó al empleado del cerrajero que estaba en el local.

A modo de resumen, las primeras etapas del modelo se conciben de forma modular: un sistema automático, inconsciente, rápido, obligatorio, específico de dominio, encapsulado y hasta innato (FODOR, 1983). Ese sería el módulo, o submódulo, más precisamente, sintáctico. El modelo considera que la estructura de la frase se construye siguiendo exclusivamente información sintáctica y que la información extrasintáctica –semántica, prosódica, pragmática– se integra en estadios posteriores. En caso de que esto lleve a incongruencias en la estructura propuesta, el *parser* se verá obligado a reanalizar. De esta forma, el *parser* controla el manejo de los recursos activando solamente

<sup>12</sup> Para profundizar sobre la noción de reanálisis se sugiere Fodor y Ferreira (2013).

<sup>13</sup> *Late closure*. Este es una reformulación del principio de asociación a la derecha –*right association*– propuesto por Kimball (1973).

aquella información que sea necesaria en el momento preciso, y es extremadamente rápido ya que no se detiene a tomar una decisión sobre cuál es la mejor interpretación posible de un estímulo, sino que analiza y segmenta siguiendo unos principios fijos.

### 3.2 ¿Principios universales?

El modelo anterior se asume universal en dos direcciones: universal en el sentido de que en todas las lenguas se aplicaría el mismo procesamiento sintáctico, siguiendo los dos principios mencionados –adjunción mínima y cierre tardío–; pero el término universal también refiere a que se seguirían los mismos principios para el procesamiento sintáctico de todas las diferentes estructuras dentro de una misma lengua.

Cuetos y Mitchell (1988) decidieron poner a prueba la primera hipótesis y realizaron una tarea experimental con hispanohablantes y angloparlantes. Les hicieron leer oraciones como (2) y luego responder mediante un cuestionario qué interpretación preferían de la oración. Los hablantes de inglés respondieron en consonancia con lo propuesto por Frazier y Fodor (1978), pero los hispanohablantes preferían adjuntar la cláusula relativa al primer sintagma nominal (N1), fenómeno conocido como cierre temprano<sup>14</sup> o adjunción alta.<sup>15</sup>

Este estudio recibió algunas críticas vinculadas con la metodología utilizada; es muy probable que haya diferencias en los resultados entre medidas *online*<sup>16</sup> y *offline*.<sup>17</sup> Las medidas *online* aportan evidencia sobre el procesamiento sintáctico mientras este se va realizando y proporcionan información sobre las primeras preferencias de adjunción. Las medidas *offline*, en cambio, brindan datos sobre cuál es la interpretación final de la oración, lo cual no implica que haya sido ni la primera ni la más sencilla. Si se asume que la información no sintáctica interactúa en una etapa posterior, luego de haber establecido la estructura, para brindar la interpretación final de la oración, bien podría ser que en un primer proceso, rápido, automático e inconsciente, el *parser* haya optado por una construcción jerárquica determinada y que luego se haya visto obligado

---

<sup>14</sup> *Early closure*.

<sup>15</sup> *High attachment*.

<sup>16</sup> En línea, simultáneas al procesamiento de la oración.

<sup>17</sup> Fuera de línea, posteriores al procesamiento de la oración, más susceptibles de ser interpretativas y no automáticas.

a modificarla por información de índole no sintáctica. Considerando esto, Mitchell y Cuetos (1991) realizaron un segundo estudio con medidas *online* –medición de tiempos de reacción en una tarea de lectura autoadministrada.<sup>18</sup> Allí encontraron la misma preferencia por la adjunción alta o cierre temprano en los hablantes de español; la opción preferida seguía siendo la adjunción al N1 –“el empleado que estaba en el local” en (2).

Varios estudios posteriores en español encontraron las mismas preferencias de adjunción (CARREIRAS; CLIFTON, 1993; DUSSIAS, 2001; FERNÁNDEZ, 2002, 2003; IGOA; CARREIRAS; MESSEGUER, 1998). Esto generó investigación en otras lenguas y los resultados fueron en la misma línea: no todos se comportan como los angloparlantes.<sup>19</sup> Varias lenguas parecen presentar una preferencia por la adjunción alta o cierre temprano: holandés (BRYSSBAERT; MITCHELL, 1996; DESMET; BRYSSBAERT; BAECKE, 2002; MITCHELL; BRYSSBAERT, 1998); alemán (HEMFORTH; KONIECZNY; SCHEEPERS, 2000); croata (LOVRIC; FODOR, 2000); francés (ZAGAR; PYNTE; RATIVEAU, 1997); ruso (FEDOROVA; YANOVICH, 2004); polaco (NOWAK, 2000); griego (PAPADOPOULOU; CLAHSSEN, 2003); gallego (FRAGA; GARCIA-ORZA; ACUÑA, 2005); afrikaans (MITCHELL *et al.*, 2000); y japonés (JUN; KOIKE, 2008; KAMIDE; MITCHELL, 1997). Mientras que otras lenguas presentan, al igual que el inglés (FERNÁNDEZ, 2002, 2003; FRAZIER; CLIFTON, 1996; FRAZIER; FODOR, 1978), una preferencia por la adjunción baja o cierre tardío: rumano, noruego, sueco (EHRlich; FERNÁNDEZ; FODOR; STENSHOEL; VINERANU, 1999); euskera o vasco (GUTIERREZ-ZIARDEGI; CARREIRAS; LAKA, 2004); y árabe (ABDELGHANY; FODOR, 1999; QUINN; ABDELGHANY; FODOR, 2000).

Estos resultados cuestionan, en principio, un modelo como el de vía muerta, que asumía las mismas estrategias de análisis y segmentación para todas las estructuras de las lenguas y para todas las lenguas, y llevan a nuevas propuestas como el Modelo construal (FRAZIER; CLIFTON, 1996). En este cambia el lugar que ocupa la universalidad: consideran

<sup>18</sup> *Self-paced reading*.

<sup>19</sup> Llama la atención como mucho de lo que se cree saber sobre la mente/cerebro y el procesamiento del lenguaje solo fue testeado en poblaciones específicas (HENRICH; HEINE; NORENZAYAN, 2010).

universalidad a través de las lenguas, pero variabilidad entre las diferentes estructuras.

Este modelo propone que en las lenguas hay dos tipos de estructuras o relaciones que se procesan de forma diferente: las primarias y las secundarias (FRAZIER; CLIFTON, 1996). En las primarias – que se establecen de forma obligatoria entre dos constituyentes– el procesamiento sintáctico se rige por los principios de adjunción mínima y cierre tardío. La diferencia radica en las relaciones secundarias, cuando la relación entre dos constituyentes no es obligatoria, sino optativa. En ese caso, el procesamiento sintáctico se realiza siguiendo el principio construal, que sostiene que los constituyentes que mantienen relaciones secundarias se adjuntarán al dominio temático actualmente en proceso, es decir, al último constituyente que hubiese recibido un papel temático (FRAZIER; CLIFTON, 1997).

3. El periodista entrevistó a la hija del coronel que tuvo un accidente.

En una oración como (3), el dominio temático en procesamiento está constituido por todo el sintagma nominal completo “la hija del coronel”, que a su vez incluye dos potenciales sujetos de la cláusula relativa, “hija” y “coronel”. A diferencia de otras preposiciones, “de” no funciona como asignador temático y, por lo tanto, no asigna rol temático al segundo sintagma (N2). Frente a estos casos, el procesador sintáctico se guiará por principios interpretativos de base pragmática (FRAZIER; CLIFTON, 1997). Uno de estos es el de referencialidad, que sostiene que los núcleos son referenciales, es decir, que introducen entidades –como participantes– en el modelo del discurso. Los modificadores restrictivos, como las cláusulas relativas, buscarán preferentemente lugares de adjunción que sean referenciales. Así, si se dispone de dos posibles opciones de adjunción, se elegirá la primera.

Esto es lo que sucedería en español y en varias de las lenguas documentadas, sin embargo no explica las preferencias de adjunción del inglés. Según los autores del modelo, estos principios, de índole pragmática, se impedirían si entraran en conflicto con otra información, como la máxima de claridad de Grice (1975), según la cual los hablantes para ser cooperativos deben emplear, siempre que resulte posible, expresiones no ambiguas. Como en inglés existe el genitivo sajón –’s–, para referir a la hija del coronel, se usaría otra forma –“the colonel’s

*daughter*”–. Si una lengua posee estructuras alternativas para expresar posesión, es posible que entonces tenga preferencias por una adjunción baja o cierre temprano. Sin embargo, esta explicación no da cuenta de lo que sucede en otras lenguas como el holandés, el croata y el afrikaans, que también poseen el genitivo sajón, pero tienen preferencias de adjunción hacia el N1 (CARREIRAS; MESEGUER, 1999; FERNÁNDEZ, 2003), ni tampoco puede explicar qué sucede con otras lenguas con adjunción baja, pero que no cuentan con el genitivo sajón.

Tanto este modelo como el de vía muerta son seriales y modulares, están inscriptos en una tradición sintactocéntrica y priorizan la información sintáctica, dándole más peso y considerándola accesible desde una primera instancia. La diferencia recae en que el Modelo construal considera que, únicamente en el establecimiento de adjunciones de estructuras secundarias, algunos principios pragmáticos pueden influir tempranamente. Además proponen procesos que se dan de abajo hacia arriba y no consideran la potencial influencia de expectativas e información de alto orden durante el procesamiento sintáctico. Esta última característica, el análisis sintáctico ciego a otra información, no ha estado exenta de críticas ya que “es como si la sintaxis tuviera que cargar dos maletas cerradas con llave, las cuales entrega en un punto de verificación a los componentes que tienen las llaves adecuadas” (JACKENDOFF, 2010, p. 184). Para los modelos seriales, la información extrasintáctica recién se integra en una última etapa, lo que supone, por un lado, economía de procesamiento, pero, por otro lado, implica que el análisis sea más costoso cuando la zona crítica es más tardía o más compleja sintácticamente. Si se comete un error, por ejemplo en la adjunción de dos constituyentes, este se arrastraría hasta el final, cuando se podría haber evitado en un primer lugar si se hubiera considerado la interacción con información no sintáctica.<sup>20</sup>

### 3.3 ¿Universalidad y parametrización?

Otra propuesta que busca dar cuenta de las diferencias en las preferencias de adjunción encontradas en distintas lenguas es la Hipótesis de recencia y proximidad del predicado<sup>21</sup> (GIBSON *et al.*,

---

<sup>20</sup> Los defensores de estos modelos consideran, sin embargo, que estos errores serían la excepción y no la norma.

<sup>21</sup> *Recency and predicate proximity*.

1996). Los autores parten de los hallazgos de Cuetos y Mitchell (1988), pero proponen otra explicación: consideran que efectivamente existe un principio universal, al que llamarán recencia (GIBSON *et al.*, 1996), pero sostienen que este es a su vez modulado por otros factores.

Los autores consideran que las preferencias iniciales del *parser* son guiadas por los pesos relativos de dos factores con fuerzas opuestas: el principio de recencia y la proximidad del predicado. Para verificar esto, proponen la lectura de oraciones con tres posibles lugares de adjunción de una cláusula relativa. Su hipótesis es que si efectivamente solo operara el principio de recencia, al forzar la adjunción a los distintos sintagmas nominales, los tiempos de lectura irán aumentando a mayor distancia de la cláusula relativa. Es decir, al manipular el número de los tres sintagmas nominales, obtendrían que los sujetos leen más rápido (4), luego (5) y por último (6).

4. las lámparas cerca de las pinturas de la casa que fue dañada en la inundación
5. las lámparas cerca de la pintura de las casas que fue dañada en la inundación
6. la lámpara cerca de las pinturas de las casas que fue dañada en la inundación

En una tarea de lectura autoadministrada<sup>22</sup> encontraron que los tiempos eran menores en las oraciones que forzaban una adjunción baja –al tercer sintagma nominal, *la casa* en (4)–. No obstante, la segunda estructura en generar tiempos más cortos fue la (6), en donde la relativa adjunta al N1, es decir, el más lejano (GIBSON *et al.*, 1996). Por un lado, esto sugiere que las preferencias de adjunción del español no son las mismas cuando hay dos antecedentes nominales que cuando hay tres. Por otro lado, es interesante que el orden de preferencia de adjunción no procede de forma monotónica desde el lugar más reciente al más lejano –o viceversa–. Esto implicaría que el principio de recencia no es el único factor que está operando en las preferencias de adjunción, ya que si así

---

<sup>22</sup> La tarea fue con ventana móvil no acumulativa (GIBSON *et al.*, 1996). Esta situación es poco ecológica, ya que sobrecarga la memoria de trabajo al no habilitar la relectura, es posible que el costo en la memoria de trabajo produzca que el factor de recencia tenga más peso.

lo fuera la segunda estructura más fácil de leer –con tiempos de lectura menores– debería ser la que adjunta al N2 –*la pintura* (5)– y no la que adjunta al N1 –*la lámpara* (6).

La Hipótesis de recencia y proximidad del predicado también es conocida con el nombre de Modelo de dos factores. El modelo propone que hay dos factores modulando las preferencias de adjunción: uno prioriza la adjunción baja, es decir al sintagma más reciente –recencia–, y el otro favorece la adjunción alta, al lugar más alejado –proximidad del predicado (GIBSON *et al.*, 1996). Ambos están motivados por efectos de memoria de trabajo e integración estructural. El principio universal de recencia requiere que se agreguen nuevos elementos a las estructuras construidas más recientemente, según las limitaciones de la memoria de trabajo. Por otro lado, debido a que –casi– todas las oraciones incluyen un predicado en su núcleo, los autores plantean la hipótesis de que la estructura central del predicado –el predicado y sus argumentos– es considerada en mayor medida como lugar posible de adjunción por el *parser*. Si los recursos son escasos, los sitios de adjunción asociados con un predicado estarán más disponibles que otros, porque se deben mantener activos estos sitios para seguir adjuntado argumentos y poder comprender la frase (GIBSON *et al.*, 1996). En definitiva, la propuesta de Gibson *et al.* (1996) es que el *parser* se guía por recencia y proximidad del predicado, dos factores que compiten entre sí; el primero es universal, mientras que el otro presenta variación en distintas lenguas, generando así una preferencia por la adjunción alta en las lenguas que tiene más peso.<sup>23</sup>

Una explicación posible de la variación entre lenguas está vinculada al orden de palabras (GIBSON *et al.*, 1996; GIBSON; PEARLMUTTER, 1998; ROTHMAN, 2010). Cuanto mayor sea la distancia promedio entre un verbo y sus argumentos, más fuertemente se necesita activar el predicado inicialmente en esa lengua para permitir adjunciones de mayor distancia. Cuanto más activado esté el predicado, más se preferirá la adjunción a este en una ambigüedad, y mayor será el costo asociado con infringir la proximidad del predicado y adjuntar siguiendo el principio de recencia. Esto explicaría la preferencia por adjunción baja en lenguas con un orden de palabras SVO rígido –como

---

<sup>23</sup> Como el principio de recencia es el resultado de los requisitos generales de la memoria de trabajo, Gibson *et al.* (1996) consideran que la variación entre lenguas no podría localizarse allí, sino en el otro factor.

el inglés, en donde ganaría el factor de recencia—, y la preferencia de adjunción alta para lenguas que tienen o permiten otro orden de palabras como VSO, en el que el verbo se separa de sus argumentos —el español, en donde el factor preponderante sería la proximidad del predicado—.

Si la variación está vinculada al orden de palabras, surge entonces otro interrogante: cuál es la relación entre la gramática de una lengua y el *parser*. Una respuesta posible es que la configuración del *parser* está relacionada con el contacto de los sujetos con la lengua, es decir que se establece mediante un proceso computacional en el que los parámetros gramaticales establecidos durante la adquisición se utilizan para personalizar el *parser* (MITCHELL; CUETOS, 1991). Otra explicación de la variación está vinculada con la parametrización: ya sea que las preferencias de procesamiento estén vinculadas a parámetros de la facultad del lenguaje o que el *parser* se parametrize, independientemente de la gramática, en relación a los pesos relativos que otorga al factor proximidad del predicado (GIBSON *et al.*, 1996).

### 3.4 ¿Mecanismo universal y preferencias particulares?

A partir de los estudios en distintas lenguas, varias propuestas se vieron obligadas a reformular la aplicación de la noción de universalidad a los modelos de procesamiento sintáctico: el Modelo construal (FRAZIER; CLIFTON, 1996) propone un procesamiento sintáctico universal para todas las lenguas, pero no para todas las estructuras; la Hipótesis de recencia y proximidad del predicado (GIBSON *et al.*, 1996), en cambio, considera un factor universal en el procesamiento sintáctico y otro parametrizable que explicaría las diferencias de adjunción halladas en distintas lenguas. Además, los autores que desencadenaron estas investigaciones presentaron otra explicación alternativa a la que llamaron Hipótesis de ajuste lingüístico<sup>24</sup> (CUETOS; MITCHELL; CORLEY, 1996; MITCHELL; CUETOS, 1991; MITCHELL *et al.* 1995). Este modelo pone la universalidad en el mecanismo de procesamiento sintáctico, pero sostiene que este se adapta al funcionamiento de cada lengua, y hasta al de cada individuo. Proponen una estrategia de aprendizaje sensible a la frecuencia: ante una estructura ambigua el *parser* optará inicialmente por la resolución que haya demostrado ser la apropiada en esa lengua

---

<sup>24</sup> *Linguistic Tuning Hypothesis*.

más frecuentemente en el pasado (MITCHELL; CUETOS, 1991). Si el primer análisis de la oración se realiza en base a la frecuencia, surge el interrogante acerca de cuál específicamente: ¿frecuencia de combinaciones de ítems léxicos concretos o de combinaciones estructurales determinadas –también llamadas frecuencia de grano fino<sup>25</sup> y grano grueso<sup>26</sup> respectivamente–?

Inicialmente, la Hipótesis de ajuste lingüístico sostenía que el *parser* mantiene registros estadísticos y basa el análisis inicial de la oración en frecuencias de grano grueso (MITCHELL; CUETOS, 1991), es decir, un nivel puramente estructural –sintáctico. Por ejemplo, estructuras como N + adjetivo + cláusula relativa –“el muchacho simpático que volvió de viaje”– son comunes en español, pero no en inglés. La estrategia sintáctica en español que permite “saltar al modificador” y adjuntar la cláusula relativa al sustantivo podría estar extendiéndose a estructuras similares como las de cláusulas relativas con doble antecedente nominal y así explicar las diferencias entre lenguas (CUETOS; MITCHELL, 1988).

En este sentido, Brysbaert y Mitchell (1996) afirman que la Hipótesis de ajuste lingüístico es una variante del Modelo de vía muerta, ya que, según ambos, el *parser* calcula un único análisis inicial basado en consideraciones estructurales. La diferencia radica en que en el Modelo de vía muerta las decisiones están determinadas por principios universales, mientras que, para el segundo, las decisiones estructurales se basan en registros estadísticos de la forma en que la ambigüedad se resuelve con mayor frecuencia en una lengua en particular.

Los autores realizaron estudios de corpus en español (CUETOS *et al.*, 1996) y en inglés (MITCHELL *et al.*, 1995) en los que contabilizaron la frecuencia de estructuras gramaticales con adjunción alta y baja y encontraron una correspondencia con los resultados obtenidos en los estudios previos (CUETOS; MITCHELL, 1988). Sin embargo, la correlación entre las preferencias de adjunción de los sujetos y los estudios de corpus no se replicó en otras investigaciones realizadas en inglés (GIBSON; SCHÜTZE, 1999) ni en holandés (BRYLSBAERT; MITCHELL, 1996; MITCHELL; BRYLSBAERT, 1998). Esto implicaría que versiones del modelo que consideren la frecuencia de grano grueso no pueden dar cuenta de los resultados obtenidos en tareas experimentales.

---

<sup>25</sup> *Fine-grained.*

<sup>26</sup> *Coarse-grained.*

No obstante, hay otras propuestas que consideran la importancia de la frecuencia de grano fino o de grano mixto en el establecimiento inicial de la estructura oracional, entre ellas, los Modelos de satisfacción de restricciones<sup>27</sup> (MACDONALD; PEARLMUTTER; SEIDENBERG, 1994; MACDONALD; SEIDENBERG, 2006; MCRAE; MATSUKI, 2013; MCRAE; SPIVEY-KNOWLTON; TANENHAUS, 1998; TANENHAUS; SPIVEY-KNOWLTON; EBERHARD; SEDIVY, 1995; TRUESWELL; TANENHAUS; GARNSEY, 1994). Estos proponen un procesamiento en paralelo que incluye información sintáctica, semántica y pragmática, e incluso prosódica, en simultáneo, con una fuerte perspectiva integracionista. Consideran que a medida que avanza el procesamiento de una frase u oración, se activan parcialmente unidades o conjuntos de unidades correspondientes a cada tipo de información codificada en las representaciones léxicas. La activación está modulada por restricciones, como la frecuencia o el contexto del discurso, y son la fuerza y la consistencia de las restricciones las que determinan qué interpretación finalmente dominará. Además, la fuerza relativa de varias restricciones da como resultado grados variables de preferencia por una u otra estructura alternativa.

A diferencia de otros modelos de procesamiento sintáctico preocupados por la administración de los recursos cognitivos, argumento que utilizan para justificar un procesamiento serial, este tipo de modelos se basan en la lógica de asignación de recursos y explican así cómo es posible un procesamiento en paralelo en vez de uno serial (LEVY, 2008). Consideran que el procesamiento de las oraciones se desarrolla en paralelo en dos sentidos: por un lado se activan todas las fuentes de información disponible que aplican restricciones sobre las diferentes alternativas posibles de interpretación de la oración, y, por otro lado, esas alternativas compiten entre sí y la decisión de cuál es la adecuada depende de cuál reciba mayor activación.

En otras palabras, los Modelos de satisfacción de restricciones sostienen una competencia simultánea de candidatos, que reciben activación de diversas fuentes de información, sin primacía en la sintaxis; algunas versiones hasta consideran a la representación de la estructura oracional como un epifenómeno del procesamiento léxico (MACDONALD *et al.*, 1994). A estas propuestas también se las

---

<sup>27</sup> *Constraint Satisfaction Models.*

denomina Modelos lexicalistas, porque sostienen que la representación léxica de una palabra incluye no solo información sobre su ortografía, pronunciación y significado(s), sino también sus funciones gramaticales, los tipos de estructuras sintácticas en las que participa y sus frecuencias (MACDONALD; SEIDENBERG, 2006).

Sin embargo, no todas las propuestas priorizan restricciones léxicas. Algunas versiones más recientes conocidas como Modelos basados en el uso<sup>28</sup> o en la experiencia<sup>29</sup> o Enfoque producción-distribución-comprensión<sup>30</sup> (GENNARI; MACDONALD, 2009; HSIAO; MACDONALD, 2016; MACDONALD; THORNTON, 2009; WELLS *et al.* 2009) consideran que en los procesos de comprensión de oraciones también intervienen las diferencias individuales basadas en la experiencia y otorgan un papel importante al aprendizaje estadístico. Plantean que las restricciones en el sistema de producción del lenguaje, como la accesibilidad y la preferencia por estructuras más cortas (STALLINGS; MACDONALD; O'SEAGHDHA, 1998), promueven ciertas estructuras y pares léxicos sobre otros. Estas presiones de producción, a lo largo del tiempo y a través del uso, crean patrones de distribución en el lenguaje que los sujetos perciben y se convierten en las restricciones probabilísticas que guían el proceso de comprensión en un sistema basado en restricciones (MACDONALD; THORNTON, 2009).

A este tipo de enfoques también se los conoce como Modelos probabilísticos,<sup>31</sup> estadísticos,<sup>32</sup> o basados en expectativas<sup>33</sup> (LEVY, 2008), propuestas que de la mano de la lingüística computacional han proliferado en las últimas dos décadas. Mientras que los modelos previos consideran que se procesan categorías sintácticas abstractas siguiendo determinadas reglas para formar estructuras jerárquicas, estas propuestas sostienen que el procesamiento del lenguaje depende del seguimiento de los pesos estadísticos de las combinaciones frecuentes en el nivel léxico y estructural (FRANCK; CHRISTIANSEN, 2018). Estos modelos describen distribuciones de probabilidad sobre los datos lingüísticos, es

---

<sup>28</sup> *Usage-based accounts.*

<sup>29</sup> *Experience-based theories.*

<sup>30</sup> *Production-distribution-comprehension (PDC) account.*

<sup>31</sup> *Probabilistic models.*

<sup>32</sup> *Statistical models.*

<sup>33</sup> *Expectation-based theories.*

decir, asignan probabilidades condicionales a representaciones lingüísticas –e.g. palabras, N-gramas, frases– basadas en la frecuencia a nivel léxico y estructural y en la información contextual y generan expectativas sobre las próximas palabras (MACDONALD; THORNTON, 2009). Estos cálculos de probabilidades se interpretan comúnmente como un reflejo de alguna forma de predicción, expectativa o anticipación graduada en la comprensión del lenguaje (HUETTIG, 2015; KUPERBERG; JAEGER, 2015; ZUNINO, 2019). La dificultad de procesamiento no está vinculada a los recursos cognitivos o a la cantidad de nodos sintácticos de una estructura, sino a cómo se relaciona el estímulo lingüístico con las predicciones que establece el *parser* (LEVY, 2008).

Los modelos presentados en este apartado sostienen que el *parser* registra frecuencias y computa pesos estadísticos, lo cual implicaría activación o preferencia gradual frente a distintas posibilidades de adjunción, a diferencia de las propuestas anteriores que, al considerar la aplicación de reglas –universales o no–, proponen decisiones o preferencias categóricas.

### 3.5 Diferencias individuales

Las diferentes preferencias de adjunción en distintas lenguas –a N1 o N2 en cláusulas de relativo con doble antecedente nominal– se han puesto en duda en los últimos años. Algunas investigaciones señalan que dichas preferencias no son tan fuertes como cabría esperar, ya que en tareas *offline* de cuestionarios la preferencia de adjunción, hacia uno u otro sintagma, ronda el 60% y con gran variación entre sujetos y entre ítems (FERNÁNDEZ, 2002; SWETS *et al.* 2007). Por ejemplo, 55% N1 en español reportado en Igoa, Carreiras y Meseguer (1998); 60% N1 en holandés reportado en Desmet *et al.* (2002); 55% N1 en italiano reportado en De Vincenzi y Job (1993). Además algunos estudios encontraron evidencia contradictoria: lenguas que inicialmente se creía que tenían una preferencia por adjunción alta, luego mostraron preferencia por adjunción baja y viceversa. Tal es el caso del portugués brasileño (baja en MIYAMOTO, 1999; alta en MAIA; MAIA, 2005; RIBEIRO, 2005), el italiano (DE VINCENZI; JOB, 1993) y el búlgaro (SEKERINA; FERNÁNDEZ; PETROVA, 2004).

Esto podría sugerir que las preferencias de adjunción son graduales, lo cual estaría en línea con los modelos de procesamiento sintáctico no reglado presentados en el apartado anterior. Además, si

las preferencias no están fuertemente sesgadas, se podrían cambiar fácilmente variando la presencia o ausencia de otras restricciones lingüísticas. Por ejemplo, las manipulaciones en los estímulos utilizados, ya sea en el sintagma nominal complejo (GILBOY *et al.* 1995), en el verbo de la cláusula matriz (ROHDE; LEVY; KEHLER, 2011), o en la cláusula relativa (GRILLO; COSTA, 2014; LOVRIĆ; FODOR, 2000; QUINN *et al.*, 2000), dan como resultado una variación en las preferencias de adjunción. También se ha reportado que la variación entre los participantes –en términos de experiencia del lenguaje o capacidad de lectura– afecta sus preferencias (DUSSIAS, 2001; FERNÁNDEZ, 2003; FRENCK-MESTRE; PYNTE, 1997).

La variabilidad encontrada en las preferencias de adjunción en distintas lenguas podría deberse, entre otros factores, a una gran variabilidad entre sujetos. Al comienzo se planteó un interrogante en torno a la universalidad: ¿todas las personas procesan el lenguaje de la misma forma o hay diferencias individuales independientes de la lengua? En este sentido, hay poco consenso acerca de si existen diferencias individuales significativas en el procesamiento sintáctico y, de ser así, qué las explica.<sup>34</sup> De los diferentes modelos revisados se desprenden distintas hipótesis sobre las diferencias individuales y su relación con el procesamiento sintáctico. Modelos como el de ajuste lingüístico (MITCHELL; CUETOS, 1991), los de satisfacción de restricciones (MACDONALD; SEIDENBERG, 2006) y los basados en el uso (MACDONALD; THORNTON, 2009) consideran que los sujetos basan sus preferencias de adjunción en frecuencias –de grano fino, grueso o mixto– registradas en el *parser* y en los ítems léxicos, esto implica que las preferencias de adjunción estarían basadas en la experiencia previa que hayan tenido los sujetos. Por otro lado, varios modelos otorgan un rol importante a las restricciones de la memoria sobre el procesamiento del lenguaje. Por ejemplo, el Modelo de vía muerta (FRAZIER;

---

<sup>34</sup> En los últimos años, tal vez frente a la dificultad de encontrar respuestas únicas y consensuadas, ha habido un crecimiento de estudios de diferencias individuales. Aunque comprender cómo y por qué un individuo determinado se comporta de forma diferente a otro aporta también a la comprensión global de los procesos cognitivos y a la búsqueda de los mecanismos comunes a toda la especie (VOGEL; AWH, 2008), estos estudios tienen consecuencias epistemológicas y teóricas distintas que aquellos que buscan entender el funcionamiento de los rasgos comunes, ya que, en la mayoría de los casos, estas diferencias individuales no son sistematizables.

RAYNER, 1982) sostiene que el principio de adjunción mínima es una consecuencia de las limitaciones de la memoria de trabajo y asume que el sistema calcula una única interpretación de la oración porque si no sería demasiado costoso para la memoria.<sup>35</sup>

Si efectivamente existen diferencias individuales en el procesamiento sintáctico, estas están vinculadas o se refieren a cuestiones fundamentales más amplias sobre la arquitectura de la mente y el sistema de procesamiento del lenguaje. Una discusión central es hasta qué punto los sistemas cognitivos son modulares en lugar de dirigidos por sistemas de dominio general (FODOR, 1983). Al entender si la variabilidad en la capacidad de los sistemas generales de dominio –como la memoria de trabajo y las funciones ejecutivas– está asociada con el procesamiento sintáctico, se pueden comprender mejor aspectos vinculados a la arquitectura general de la mente, como hasta qué punto el lenguaje es modular y en qué medida recluta sistemas de dominio general. En esta línea, algunas teorías (CAPLAN; WATERS, 1999; WATERS; CAPLAN, 2003) proponen que el procesamiento del lenguaje se divide en etapas iniciales –automáticas– y posteriores –interpretativas– con solo la última sujeta a diferencias individuales en memoria de trabajo y otras habilidades cognitivas; es decir, caracterizan de forma diferente al procesamiento *online* frente al *offline*, y proponen que solo este último sería sensible a las diferencias cognitivas entre los individuos.

En la literatura psicolingüística se han propuesto explicaciones –y predictores– de las diferencias individuales en el procesamiento del lenguaje de dos tipos: vinculadas a las demandas cognitivas y centradas en factores específicos del lenguaje. Dentro de las primeras, se encuentran explicaciones vinculadas a la memoria de trabajo (GIBSON, 2000; SWETS *et al.*, 2007), al control inhibitorio (NOVICK; TRUESWELL; THOMPSON-SCHILL, 2010; VAN DYKE; JOHNS; KUKONA, 2014); y a la velocidad de procesamiento (CAPLAN *et al.*, 2011; SALTHOUSE, 1996). En el segundo grupo, algunos factores de interés han sido la experiencia y el conocimiento del lenguaje (MACDONALD; CHRISTIANSEN, 2002; SCHWERING; MACDONALD, 2020), la práctica y frecuencia específica con el tipo de estímulos a evaluar (PAYNE

---

<sup>35</sup> Esto ha sido problematizado y hay trabajos que evidencian que se puede mantener activa más de una representación y que las representaciones erróneas de la oración no siempre decaen luego del reanálisis (SLATTERY *et al.* 2013).

*et al.*, 2014); y la habilidad fonológica (ACHESON; MACDONALD, 2011).<sup>36</sup>

Swets *et al.* (2007) realizaron un estudio en el que midieron las preferencias de adjunción de los participantes y el *span* de la memoria de trabajo para determinar sus efectos en el procesamiento sintáctico. Encontraron que los lectores con *spans* de memoria de trabajo bajo eran menos propensos a utilizar estrategias de recencia o cierre tardío para desambiguar que los lectores con *spans* altos.<sup>37</sup> Es decir, los hablantes con poca memoria de trabajo prefieren adjuntar las cláusulas relativas al N1, contradiciendo la estrategia de cierre tardío, que asume que la recencia es una estrategia para maximizar los recursos de procesamiento.

Una posible explicación se basa en la Hipótesis de la prosodia implícita (FODOR, 2002), según la cual, a medida que se lee una oración, la forma en que una voz interna la divide en unidades prosódicas puede usarse como información discriminativa cuando la información sintáctica y semántica deja la entrada estructural ambigua. Fodor (2002) propuso esta teoría como una explicación de las diferencias interlingüísticas en las adjunciones de cláusulas relativas: la ruptura prosódica antes de la relativa podría interpretarse como marca de una discontinuidad estructural en el árbol sintáctico y así inducir preferencias de adjunción al N1, ya que se construiría un árbol en el que todo el sintagma nominal complejo es modificado por la cláusula relativa en lugar de solo N2.

Swets *et al.* (2007) hicieron un segundo experimento para evaluar si la segmentación prosódica estaba mediando las diferencias individuales. Para eso, presentaron los estímulos divididos en tres partes.<sup>38</sup> Efectivamente, el formato de presentación de las oraciones resultó en una mayor tendencia a adjuntar a N1, tanto en inglés como en holandés (SWETS *et al.*, 2007). Los autores propusieron dos explicaciones posibles de la relación entre la capacidad de la memoria de trabajo y las preferencias de adjunción. La primera es la ya mencionada en relación a la prosodia implícita; la segunda se vincula con el Modelo de la máquina de salchichas (FRAZIER; FODOR, 1978), ya que podría ser que las personas

---

<sup>36</sup> Para un recuento más exhaustivo de las diferencias individuales se sugiere revisar James *et al.* (2018).

<sup>37</sup> Los mismos resultados fueron reportados en un estudio posterior (JAMES *et al.*, 2018).

<sup>38</sup> La línea indica la segmentación elegida para la presentación de los estímulos: La doncella de la princesa / que se rascó en público / estaba terriblemente avergonzada.

con diferentes *spans* de memoria, en la primera etapa del procesamiento –el Empaquetador de frases preliminar–, empaqueten las frases en *chunks* de distinto tamaño, ya que este número estaría limitado por la capacidad de la memoria de trabajo. Esto operaría de forma similar a las rupturas prosódicas e influiría en las decisiones finales de adjunción.

Estos hallazgos sugieren que las diferencias individuales en relación con la memoria de trabajo podrían estar asociadas con diferentes estrategias de procesamiento del lenguaje, lo que a su vez podría verse como evidencia en contra de principios o reglas universales en el procesamiento sintáctico. No obstante, deberían estudiarse en profundidad las diferencias encontradas, particularmente en procesamiento *online*, ya que Swets *et al.* (2007) y James *et al.* (2018) evaluaron procesamiento *offline*.

### 3.6 Diferencias situacionales

La tercera pregunta planteada en torno a la universalidad es si el lenguaje se procesa siempre de forma universal; además de los ya mencionados efectos individuales, también hay efectos situacionales que influyen en el procesamiento sintáctico. En las últimas dos décadas se han realizado investigaciones que sugieren que en determinadas situaciones el *parser* realiza un procesamiento plano o superficial,<sup>39</sup> subespecificado<sup>40</sup> o de interpretación mínima suficiente<sup>41</sup> (CHRISTIANSON *et al.*, 2006; FERREIRA, 2003; FERREIRA; CHRISTIANSON; HOLLINGWORTH, 2001; KLIN *et al.*, 2006; STEWART; HOLLER; KIDD, 2007; SWETS *et al.*, 2008).

Los modelos ya revisados suponen que el objetivo general del sistema de procesamiento del lenguaje es ofrecer una representación precisa y detallada de la entrada lingüística y que dicho procesamiento se realiza siempre de la misma manera y en profundidad. Es importante señalar que la metodología experimental propia de la psicolingüística podría estar forzando a los sujetos a trabajar más duro de lo habitual para llegar a una interpretación única y definitiva, ya sea porque a los participantes se les suelen hacer preguntas de comprensión que revelan la

---

<sup>39</sup> *Shallow language processing.*

<sup>40</sup> *Underspecified language processing.*

<sup>41</sup> *Good-enough processing.*

naturaleza de la interpretación, o simplemente debido a las características de demanda de la situación experimental (FERREIRA; BAILEY; FERRARO, 2002). En circunstancias cotidianas de comunicación, puede ser que las ambigüedades de adjunción queden sin resolver. Generalmente las tareas que se deben realizar en función del estímulo lingüístico son relativamente mínimas—e.g. confirmar, asentir, ejecutar acciones motoras simples— y rara vez se requiere demostrar la precisión o la naturaleza detallada de la comprensión de un enunciado. Además, los enunciados en diálogos se producen rápidamente uno tras otro, por lo que es posible que el sistema no tenga tiempo para considerar todas las fuentes de información relevantes y calcular una estructura específica y detallada para cada uno. En cambio, el *parser* podría basarse en una interpretación mínima suficiente que se perfecciona si es necesario.

Una evidencia de que la comprensión de oraciones puede ser superficial o estar mediada por un procesamiento plano son las ilusiones semánticas (véase SANFORD; STURT, 2002 para una revisión). Barton y Sanford (1993, p. 482) presentaron la siguiente situación a un grupo de participantes: “Un avión cae en la frontera entre México y Estados Unidos. Las autoridades siguen discutiendo dónde enterrar a los sobrevivientes. ¿Dónde creés que deberían hacerlo?”. Sorprendentemente, la mitad respondió “enterrarlos donde quieran sus familiares” y no se dio cuenta de que los sobrevivientes eran personas vivas que no deberían ser enterradas. Es decir, la interpretación resultante era inconsistente con la entrada lingüística.

Este fenómeno es explicado por las propuestas de procesamiento predictivo y de influencias arriba-abajo en el procesamiento del lenguaje que consideran que, en varias ocasiones, las personas realizan un procesamiento heurístico, también llamado plano o superficial, en el que influyen las expectativas en la comprensión del lenguaje. Una de las propuestas más prominentes es la del Enfoque de interpretación mínima suficiente (FERREIRA *et al.*, 2002; FERREIRA; PATSON, 2007). Este propone que con el procesamiento abajo-arriba, incremental y lento, compite una heurística arriba-abajo, rápida y frugal, que funciona en interacciones y sirve para respaldar la comunicación acelerada. El sistema cognitivo, en muchas situaciones, se basaría en un pequeño conjunto de heurísticas, en lugar de en algoritmos de composición, para construir el significado de las oraciones. Dichas heurísticas pueden ser tanto

semánticas, basándose en el conocimiento del mundo y en los marcos semánticos, como estructurales, basándose en la frecuencia de aparición de ciertas estructuras<sup>42</sup> (FERREIRA *et al.*, 2002).

Estas propuestas no niegan que el sistema de procesamiento del lenguaje utilice algoritmos sintácticos, sino que habilitan una interacción temprana con la información semántica. Consideran que dicho procesamiento profundo convive con un procesamiento superficial o de interpretación mínima suficiente y que este tipo de procesamiento basado en heurísticas opera generalmente más rápido (KARIMI; FERREIRA, 2015; TOWNSEND; BEVER, 2001). La interpretación heurística podría entonces seleccionarse porque está disponible más rápidamente o porque el sistema tiene alguna razón para preferirla —e.g. se ajusta al conocimiento del mundo. Esto explicaría por qué las personas tienden a interpretar mal oraciones inverosímiles e interpretar como correctas oraciones agramaticales (FERREIRA; PATSON, 2007).

#### 4 El mejor *parser*

Las propuestas debatidas consideran diferentes explicaciones acerca de cómo se procesa el lenguaje, pero siempre con el énfasis en la optimidad de procesamiento, lo cual denota la multiplicidad de interpretaciones acerca de a qué consideran óptimo. La discusión gira en torno a varios aspectos que se retomarán a continuación.

En primer lugar, el debate se vincula con la relación entre almacenamiento y procesamiento. Si el almacenamiento de los ítems léxicos incluye información gramatical, semántica y de frecuencia de cada ítem, el procesamiento se libera. Mientras que si se postula economía de almacenamiento, se genera más sobrecarga en el procesamiento. En el primer punto, están los Modelos lexicalistas (MACDONALD; SEIDENBERG, 2006) al proponer representaciones léxicas ricas en las que los ítems léxicos cargan con información de distinta índole; mientras que otros modelos no consideran que se almacene información sobre la frecuencia y para explicar las preferencias de adjunción utilizan un conjunto de principios o reglas derivacionales. A su vez, si en los ítems

---

<sup>42</sup> En el caso de lenguas SVO, dicha heurística implicaría asumir que el primer sintagma nominal es el agente de la acción y el siguiente es la entidad afectada por la acción (TOWNSEND; BEVER, 2001).

léxicos está registrada la frecuencia, la discusión es cuál específicamente: combinaciones a nivel léxico –N-gramas– o frecuencias estructurales.

En relación con el procesamiento, tampoco hay consenso sobre qué se procesa: reglas que computan categorías sintácticas abstractas –principios– o pesos estadísticos y combinaciones frecuentes en el nivel léxico y estructural. Algunos modelos proponen que el *parser* opera mediante principios de análisis sintácticos fijos –universales o parametrizables– para construir estructuras jerárquicas; otro grupo sostiene que se realiza un procesamiento plano o superficial basado en heurísticas o marcos interpretativos –semánticos o sintácticos– para construir la estructura de la oración; y otros consideran que los procesos de comprensión sopesan la probabilidad de interpretaciones alternativas basadas en el aprendizaje a partir de experiencias anteriores y computan pesos estadísticos, eliminando la estructura jerárquica como base de organización sintáctica. Para este último grupo, la universalidad en el procesamiento del lenguaje está justamente en la computación, en el proceso (MACDONALD; SEIDENBERG, 2006; MACDONALD; THORNTON, 2009).

En tercer lugar, surge el interrogante acerca de si es óptimo un procesamiento en paralelo de distintos tipos de información. Se han revisado modelos que proponen un procesamiento serial y abajo-arriba en el que el *parser* maneja un tipo de información a la vez, generalmente comenzando con la información sintáctica –Modelo de vía muerta (FRAZIER; FODOR, 1978; FRAZIER; RAYNER, 1982)–, aunque también se han presentado propuestas en las que restricciones prosódicas –Hipótesis de la prosodia implícita (FODOR, 2002)– o pragmáticas –Modelo construal (FRAZIER; CLIFTON, 1996)– intervienen tempranamente en la construcción de la estructura oracional. Por otro lado, hay propuestas que consideran que todas las fuentes de información se procesan en paralelo y aplican restricciones sobre una única representación oracional –algunas variantes de los Modelos de satisfacción de restricciones, como el Modelo competición-integración<sup>43</sup> de McRae *et al.* (1998)–, o generan varias representaciones de la oración que compiten entre sí –Modelos de satisfacción de restricciones (MACDONALD; SEIDENBERG, 2006)– o de las cuales solo se persigue

---

<sup>43</sup> *Competition-integration model.*

una a la vez –Modelo de carrera irrestricta<sup>44</sup> de Van Gompel, Pickering y Traxler (2000). Estas propuestas descansan en el hecho de que se ha encontrado evidencia temprana de la interacción semántica (MCRAE *et al.*, 1998; TRUESWELL *et al.*, 1994), mientras que los modelos seriales argumentan que eso se explica porque el procesamiento sintáctico ocurrió previamente de forma automática y rápida.

Una interpretación posible de esta discusión es que los modelos modulares se centran en el procesamiento temprano, durante el cual no habría interacción, mientras que los modelos interactivos examinarían el procesamiento tardío, momento en el que sí se daría dicha interacción (FRIEDERICI, 2011). En este sentido, surgen otras preguntas acerca del decurso temporal de los eventos: aunque se conciba un procesamiento serial, ¿las distintas etapas son necesariamente discretas o podrían solaparse? ¿Podría ser que se esté confundiendo por etapas seriales a procesos encapsulados en paralelo que simplemente tardan diferentes tiempos en realizarse?

Estrechamente vinculado está el debate acerca de cómo se integra la información de diferentes fuentes y a qué componentes se les otorga capacidad generativa. Los modelos seriales consideran cierta autonomía de la sintaxis: es la única consultada para el establecimiento de la estructura oracional y la integración con las demás fuentes de información se produce en estadios posteriores. No obstante, la información sintáctica no es la única capaz de asistir en la construcción de la representación de la oración, tal como lo evidencian las tareas realizadas bajo el Enfoque de interpretación mínima suficiente o procesamiento superficial (FERREIRA; PATSON, 2007). Si hay evidencia de que la semántica o la prosodia intervienen tempranamente, no es posible aceptar un modelo estrictamente sintactocéntrico. Es necesario encontrar el vínculo entre las diferentes propuestas teóricas, las características que se le adjudican al lenguaje y a la mente y el decurso temporal del procesamiento del lenguaje.

Otro punto de la discusión gira en torno a los recursos, más precisamente la disputa entre las teorías de limitación de los recursos<sup>45</sup> versus las teorías de asignación de los recursos<sup>46</sup> (LEVY, 2008). Las primeras proponen que algunas estructuras sintácticas requieren de un

---

<sup>44</sup> *Unrestricted race model.*

<sup>45</sup> *Resource-requirement o resource-limitation theories.*

<sup>46</sup> *Resource-allocation theories.*

determinado recurso cognitivo, no específicamente lingüístico –e.g. memoria–, más que otras y que ese recurso es escaso; esto da lugar a una mayor dificultad de procesamiento. Estas teorías también consideran un determinado enfoque en la resolución de ambigüedades: el *parser* solo puede perseguir una alternativa a la vez y, al enfrentarse a una ambigüedad, elige la que minimiza los recursos consumidos. En esto se basan los modelos que proponen un procesamiento serial (FRAZIER; CLIFTON, 1996; FRAZIER; FODOR, 1978; FRAZIER; RAYNER, 1982; GIBSON *et al.*, 1996; MITCHELL; CUETOS, 1991). Por otro lado, las teorías de asignación de los recursos consideran que el *parser* asigna diferentes cantidades de recursos a distintas interpretaciones que está analizando en paralelo y la dificultad de procesamiento surge cuando esos recursos resultan ser asignados de manera ineficiente; la dificultad de integración de una nueva palabra corresponde a la cantidad de reasignación necesaria para reflejar el efecto de la palabra en la clasificación de preferencias (LEVY, 2008). En esta línea están los Modelos de satisfacción de restricciones (MACDONALD; SEIDENBERG, 2006) al proponer la activación de distintas fuentes de información en paralelo y la competencia entre distintas representaciones de la oración. Si en términos evolutivos, un sistema óptimo implica hacerlo de la manera más veloz posible, ¿puede ser que eso implique ser redundante con el almacenamiento o tener varios sistemas procesando en paralelo?

Esta postura suele ser criticada por contradecir el principio de economía cognitiva: el sistema cognitivo siempre favorecería aquellos procesos que minimicen recursos y esfuerzos al procesar información (COLMAN, 2015). Algunos investigadores (FERREIRA; PATSON, 2007; GIGERENZER; TODD; THE ABC GROUP, 1999) han señalado que un sistema con recursos limitados y que debe tomar decisiones rápidamente funciona mejor si se basa en un pequeño conjunto de heurísticas rápidas y frugales en lugar de intentar ejecutar algoritmos que consultan cada pieza de información potencialmente relevante. En esta línea, el Modelo de vía muerta propone una heurística simple que gobierna todas las decisiones de análisis: el principio de adjunción mínima (o de todo mínimo<sup>47</sup> en Fodor y Inoue (1998)), que toma el supuesto básico de que el *parser* es un dispositivo de mínimo esfuerzo. Así el *parser* puede construir rápidamente una representación que pueda

---

<sup>47</sup> *Minimal everything principle.*

soportar algún tipo de interpretación, aunque no sea la mejor. Desde este punto de vista, entonces, las representaciones de interpretación mínima suficiente surgen porque el sistema hace uso de un conjunto de heurísticas que le permiten realizar la menor cantidad de trabajo necesario para llegar a un significado para la oración.

Una pregunta interesante en relación con los debates presentados es si en todos los casos las opciones son mutuamente excluyentes; la evidencia señalada parecería indicar que ninguna de las vías por sí sola es suficiente para explicar todos los procesos observados durante el procesamiento del lenguaje. En este sentido, parece importante pensar en propuestas que integren las distintas evidencias y resultados reportados, como recientemente lo han hecho los modelos que consideran varias rutas o vías<sup>48</sup> en el procesamiento del lenguaje (HAGOORT; BAGGIO; WILLEMS, 2009; JACKENDOFF, 2010; KARIMI; FERREIRA, 2015; KUPERBERG, 2007; VAN HERTEN; CHWILLA; KOLK, 2006).

Una de estas propuestas, el Modelo de equilibrio cognitivo<sup>49</sup> (KARIMI; FERREIRA, 2015), considera que coexisten los dos tipos de procesamiento –heurístico y algorítmico– y cuando el sistema de procesamiento del lenguaje encuentra las primeras palabras en una oración, entra en un estado de desequilibrio cognitivo (PIAGET, 1985), debido a que la oración contiene información no procesada que plantea incertidumbre y, por lo tanto, debe procesarse e integrarse con los esquemas existentes para recuperar el equilibrio cognitivo. Consideran que el procesamiento guiado por heurísticas y el procesamiento algorítmico comienzan al mismo momento, solo que debido a que el primero se basa en la implementación de reglas simples o marcos semánticos, información arriba-abajo, generalmente producirá una representación antes que la ruta algorítmica, que deriva el significado de una manera abajo-arriba, organizando y combinando el estímulo lingüístico incremental con reglas lingüísticas sucesivas y bien definidas. De esta forma, cuando la salida del procesamiento heurístico está disponible, la ruta algorítmica aún no ha terminado y, por lo tanto, la salida intermedia de la ruta heurística influye en el procesamiento algorítmico en curso. Esto da como resultado la formación de representaciones de interpretación mínima suficiente: cuando la salida del procesamiento heurístico está disponible, el sistema

---

<sup>48</sup> *Multistream models of language processing.*

<sup>49</sup> *Online cognitive equilibrium.*

alcanza un estado de equilibrio y, por lo tanto, prefiere permanecer en ese estado, lo que genera que no se asignen más recursos para el procesamiento algorítmico.

## 5 Reflexiones finales

Al comienzo del trabajo se propusieron tres preguntas en torno a la universalidad en el procesamiento del lenguaje. En relación con la primera, si el *parser* sigue reglas universales, se repasaron algunos modelos que consideran que el *parser* está esencialmente preconfigurado y sigue estrategias universales comunes a todas las lenguas; otros que consideran que algunas estrategias son universales y otras varían entre las distintas lenguas, debido a una posible parametrización; otros que sostienen que el *parser* funciona de forma universal en todas las lenguas, pero que la variación se presenta en las diferentes estructuras dentro de cada lengua; y otros que proponen que el *parser* registra las frecuencias de uso de determinadas estructuras y que las preferencias de adjunción dependen de dichas frecuencias, en donde lo universal está en la existencia de ese mecanismo que realiza las computaciones estadísticas. Indefectiblemente, esto lleva a pensar que si la universalidad tal vez esté vinculada con la granularidad, entonces esta se encuentre a nivel de grano grueso, mientras que a nivel de grano mixto y fino haya variación entre las distintas lenguas.

La segunda pregunta planteada es si, en el procesamiento del lenguaje, hay diferencias individuales independientes de la lengua. Se revisaron algunas críticas hechas a los estudios de ambigüedades de adjunción con cláusulas relativas y se señaló el rol que las diferencias individuales en la memoria de trabajo podrían tener en el procesamiento sintáctico.

En relación con la última pregunta, si el procesamiento del lenguaje se realiza siempre de la misma forma o varía según la tarea requerida, se revisó evidencia que sugiere que no siempre se realiza un procesamiento algorítmico ni se consulta toda la información potencialmente relevante, sino que en determinadas ocasiones se realiza un procesamiento de interpretación mínima suficiente, basado en heurísticas o marcos interpretativos generales.

Las últimas décadas han sido testigos de importantes avances en la comprensión acerca de cómo se procesa el lenguaje. Está claro que el procesamiento sintáctico es altamente incremental y que varias fuentes de

información no sintácticas tienen efectos rápidos sobre la resolución de ambigüedades. Las propuestas mencionadas en este trabajo representan las hipótesis más consensuadas acerca del procesamiento del lenguaje, sin embargo, ninguna puede explicar por completo la evidencia empírica existente. En este sentido, se planteó la importancia de pensar propuestas que integren toda la evidencia encontrada –a veces aparentemente contradictoria–, de incluir poblaciones más amplias y representativas en los estudios y de pensar situaciones experimentales más ecológicas, a fin de seguir avanzando en el entendimiento acerca de cómo se procesa el lenguaje.

### Agradecimientos

El presente artículo es fruto del trabajo de investigación en el marco del Doctorado en Lingüística en la Universidad de Buenos Aires. Agradezco a la Dra. Gabriela Zunino por su acompañamiento y dirección y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina, por financiar dicha investigación.

### Referencias

ABDELGHANY, H.; FODOR, J. D. Low Attachment of Relative Clauses in Arabic. In: ARCHITECTURES AND MECHANISMS OF LANGUAGE PROCESSING, 5., 1999, Edinburgh, UK. *Trabajo presentado*. Edinburgh, UK: Cambridge University, 1999.

ACHESON, D. J.; MACDONALD, M. C. The Rhymes that the Reader Perused Confused the Meaning: Phonological Effects During On-line Sentence Comprehension. *Journal of Memory and Language*, Cambridge, v. 65, n. 2, p. 193-207, 2011. DOI: <https://doi.org/10.1016/j.jml.2011.04.006>

BADDELEY, A. *Working Memory, Thought, and Action*. Oxford: OuP, 2007. DOI: <https://doi.org/10.1093/acprof:oso/9780198528012.001.0001>

BARTON, S. B.; SANFORD, A. J. A Case Study of Anomaly Detection: Shallow Semantic Processing and Cohesion Establishment. *Memory & Cognition*, New York, v. 21, n. 4, p. 477-487, 1993. DOI: <https://doi.org/10.3758/BF03197179>

BERMOND, B.; VAN HEERDEN, J. The Muller-Lyer Illusion Explained and Its Theoretical Importance Reconsidered. *Biology and Philosophy*, Amsterdam, v. 11, n. 3, p. 321-338, 1996. DOI: <https://doi.org/10.1007/BF00128785>

BOLAND, J. E.; CUTLER, A. Interaction with Autonomy: Multiple Output Models and the Inadequacy of the Great Divide. *Cognition*, Amsterdam, v. 58, n. 3, p. 309-320, 1996. DOI: [https://doi.org/10.1016/0010-0277\(95\)00684-2](https://doi.org/10.1016/0010-0277(95)00684-2)

BRYLSBAERT, M.; MITCHELL, D. C. Modifier Attachment in Dutch: Deciding between Garden-Path, Construal and Statistical Tuning Accounts of Parsing. In: WORKSHOP ON COMPUTATIONAL MODELS OF HUMAN SYNTACTIC PROCESSING, 1996, Wassenaar. *Trabajo presentado*. Wassenaar: Netherlands Institute for Advanced Studies, 1996.

CAPLAN, D.; DEDE, G.; WATERS, G.; MICHAUD, J.; TRIPODIS, Y. Effects of Age, Speed of Processing, and Working Memory on Comprehension of Sentences with Relative Clauses. *Psychology and Aging*, Cambridge, v. 26, n. 2, p. 439-450, 2011. DOI: <https://doi.org/10.1037/a0021837>

CAPLAN, D.; WATERS, G. Issues Regarding General and Domain-Specific Resources. *Behavioral and Brain Sciences*, Cambridge, v. 22, n. 1, p. 114-122, 1999. DOI: <https://doi.org/10.1017/S0140525X99441780>

CARREIRAS, M.; CLIFTON, C. Relative Clause Interpretation Preferences in Spanish and English. *Language and Speech*, New York, v. 36, n. 4, p. 353-372, 1993. DOI: <https://doi.org/10.1177/002383099303600401>

CARREIRAS, M.; MESEGUER, E. Procesamiento de oraciones ambiguas. In: VEGA RODRÍGUEZ, M. D.; CUETOS VEGA, F. (ed.). *Psicolingüística del español*. Madrid: Trotta, 1999. p. 163-203.

CHOMSKY, N. *Syntactic Structures*. The Hague: Mouton, 1957. DOI: <https://doi.org/10.1515/9783112316009>

CHOMSKY, N. *Knowledge of Language: Its Nature, Origin, and Use*. Westport: Greenwood Publishing Group, 1986.

CHOMSKY, N. *The Minimalist Program*. Cambridge: Massachusetts Institute of Technology, 1995.

CHOMSKY, N. *The Minimalist Program: 20th Anniversary Edition*. Cambridge: Massachusetts Institute of Technology, 2015.

CHRISTIANSON, K.; WILLIAMS, C. C.; ZACKS, R. T.; FERREIRA, F. Younger and Older Adults “Good-Enough” Interpretations of Garden-Path Sentences. *Discourse Processes*, Londres, v. 42, n. 2, p. 205-238, 2006. DOI: [https://doi.org/10.1207/s15326950dp4202\\_6](https://doi.org/10.1207/s15326950dp4202_6)

COLMAN, A. M. *A Dictionary of Psychology*. 4th ed. Oxford: Oxford University Press, 2015.

CUETOS, F.; MITCHELL, D. C. Cross-Linguistic Differences in Parsing: Restrictions on the Use of the Late Closure Strategy in Spanish. *Cognition*, Amsterdam, v. 30, n. 1, p. 73-105, 1988. DOI: [https://doi.org/10.1016/0010-0277\(88\)90004-2](https://doi.org/10.1016/0010-0277(88)90004-2)

CUETOS, F.; MITCHELL, D. C.; CORLEY, M. M. Parsing in Different Languages. In: CARREIRAS, M.; SEBASTIÁN-GALLÉS, N.; GARCÍA-ALBEA, J. (ed.). *Language Processing in Spanish*. New York: Psychology Press, 1996. p. 145-187.

DE VINCENZI, M.; JOB, R. Some Observations on the Universality of the Late-Closure Strategy. *Journal of Psycholinguistic Research*, New York, v. 22, n. 2, p. 189-206, 1993. DOI: <https://doi.org/10.1007/BF01067830>

DESMET, T.; BRYLSBAERT, M.; BAECKE, C. D. The Correspondence Between Sentence Production and Corpus Frequencies in Modifier Attachment. *The Quarterly Journal of Experimental Psychology: Section A*, New York, v. 55, n. 3, p. 879-896, 2002. DOI: <https://doi.org/10.1080/02724980143000604>

DING, N.; MELLONI, L.; TIAN, X.; POEPEL, D. Rule-Based and Word-Level Statistics-Based Processing of Language: Insights from Neuroscience. *Language, Cognition and Neuroscience*, New York, v. 32, n. 5, p. 570-575, 2017. DOI: <https://doi.org/10.1080/23273798.2016.1215477>

DUSSIAS, P. E. Sentence Parsing in Fluent Spanish-English Bilinguals. In: NICOL, J. (ed.). *One Mind, Two Languages: Bilingual Language Processing*. New York: Wiley-Blackwell, 2001. p. 159-176.

EHRlich, K.; FERNÁNDEZ, E. M.; FODOR, J. D.; STENSHOEL, E.; VINEREANU, M. Low Attachment of Relative Clauses New Data from Swedish. In: ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING, 12., 1999, New York. *Trabajo presentado*. New York, NY: CUNY Graduate Center, 1999.

EVERETT, C. *Linguistic Relativity: Evidence Across Languages and Cognitive Domains*. Berlin: De Gruyter Mouton, 2013. DOI: <https://doi.org/10.1515/9783110308143>

EZQUERRO, J. Teorías de la arquitectura de lo mental. In: BRONCANO, F. (ed.). *La mente humana*. Madrid: Trotta, 1995. p. 97-150.

FEDOROVA, O.; YANOVICH, I. Relative Clause Attachment in Russian: The Role of Constituent Length. In: ARCHITECTURES AND MECHANISMS OF LANGUAGE PROCESSING, 10., 2004, Aix de Provence. *Trabajo presentado*. Aix de Provence: 2004.

FERNÁNDEZ, E. M. Relative Clause Attachment in Bilinguals and Monolinguals. In: HEREDIA, R.; ALTARRIBA, J. (ed.). *Advances in Psychology: Bilingual Sentence Processing*. Amsterdam: North-Holland, 2002. p. 187-215. DOI: [https://doi.org/10.1016/S0166-4115\(02\)80011-5](https://doi.org/10.1016/S0166-4115(02)80011-5)

FERNÁNDEZ, E. M. *Bilingual Sentence Processing: Relative Clause Attachment in English and Spanish*. New York: John Benjamins Publishing, 2003. DOI: <https://doi.org/10.1075/lald.29>

FERREIRA, F. The Misinterpretation of Noncanonical Sentences. *Cognitive Psychology*, Cambridge, v. 47, n. 2, p. 164-203, 2003. DOI: [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7)

FERREIRA, F.; BAILEY, K. G.; FERRARO, V. Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, New York, v. 11, n. 1, p. 11-15, 2002. DOI: <https://doi.org/10.1111/1467-8721.00158>

FERREIRA, F.; CHRISTIANSON, K.; HOLLINGWORTH, A. Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis. *Journal of Psycholinguistic Research*, New York, v. 30, n.1, p. 3-20, 2001. DOI: <https://doi.org/10.1023/A:1005290706460>

FERREIRA, F.; PATSON, N. D. The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, Londres, v. 1, n. 1/2, p. 71-83, 2007. DOI: <https://doi.org/10.1111/j.1749-818X.2007.00007.x>

FODOR, J. A. *The Modularity of Mind*. Cambridge: MIT Press, 1983. DOI: <https://doi.org/10.7551/mitpress/4737.001.0001>

FODOR, J. D. Prosodic Disambiguation in Silent Reading. *North East Linguistics Society*, New York, v. 32, p. 113-132, 2002.

FODOR, J. D.; FERREIRA, F. (ed.). *Reanalysis in Sentence Processing*. Berlin: Springer Science & Business Media, 2013.

FODOR, J. D.; INOUE, A. Attach Anyway. In: FODOR, J.; FERREIRA, F. (ed.). *Reanalysis in Sentence Processing*. Dordrecht: Springer, 1998. p. 101-141. DOI: [https://doi.org/10.1007/978-94-015-9070-9\\_4](https://doi.org/10.1007/978-94-015-9070-9_4)

FRAGA, I.; GARCIA-ORZA, J.; ACUÑA, J. C. Relative Clause Disambiguation in Galician: New Evidence of High Attachment in Romance Languages. *Psicológica*, Valencia, v. 26, n. 2, p. 243-260, 2005.

FRANK, S. L.; CHRISTIANSEN, M. H. Hierarchical and Sequential Processing of Language. *Language, Cognition and Neuroscience*, New York, v. 33, n. 9, p. 1213-1218, 2018. DOI: <https://doi.org/10.1080/23273798.2018.1424347>

FRAZIER, L.; CLIFTON, C. *Construal*. Cambridge: MIT Press, 1996.

FRAZIER, L.; CLIFTON, C. Construal: Overview, Motivation, and Some New Evidence. *Journal of Psycholinguistic Research*, New York, v. 26, n. 3, p. 277-295, 1997. DOI: <https://doi.org/10.1023/A:1025024524133>

FRAZIER, L.; FODOR, J. D. The Sausage Machine: A New Two-Stage Parsing Model. *Cognition*, Amsterdam, v. 6, p. 291-325, 1978. DOI: [https://doi.org/10.1016/0010-0277\(78\)90002-1](https://doi.org/10.1016/0010-0277(78)90002-1)

FRAZIER, L.; RAYNER, K. Making and Correcting Errors During Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology*, Cambridge, v. 14, p. 178-210, 1982. DOI: [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)

FRENCK-MESTRE, C.; PYNTE, J. Syntactic Ambiguity Resolution While Reading in Second and Native Languages. *The Quarterly Journal of Experimental Psychology*, New York, v. 50, n. 1, p. 119-148, 1997.

FRIEDERICI, A. D. The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, New York, v. 91, n. 4, p. 1357-1392, 2011. DOI: <https://doi.org/10.1152/physrev.00006.2011>

GENNARI, S. P.; MACDONALD, M. C. Linking Production and Comprehension Processes: The Case of Relative Clauses. *Cognition*, Amsterdam, v. 111, n. 1, p. 1-23, 2009. DOI: <https://doi.org/10.1016/j.cognition.2008.12.006>

GIBSON, E. The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In: MARANTZ, A.; MIYASHITA, Y.; O'NEIL, W. (ed.). *Image, Language, Brain*. Cambridge: MIT Press, 2000. p. 95-126.

GIBSON, E.; PEARLMUTTER, N. Constraints on Sentence Comprehension. *Trends in Cognitive Science*, Amsterdam, v. 2, n. 7, p. 262-268, 1998. DOI: [https://doi.org/10.1016/S1364-6613\(98\)01187-5](https://doi.org/10.1016/S1364-6613(98)01187-5)

GIBSON, E.; PEARLMUTTER, N.; CANSECO-GONZALEZ, E.; HICKOK, G. Recency Preference in the Human Sentence Processing Mechanism. *Cognition*, Amsterdam, v. 59, n. 1, p. 23-59, 1996. DOI: [https://doi.org/10.1016/0010-0277\(95\)00687-7](https://doi.org/10.1016/0010-0277(95)00687-7)

GIBSON, E.; SCHÜTZE, C. T. Disambiguation Preferences in Noun Phrase Conjunction do not Mirror Corpus Frequency. *Journal of Memory and Language*, Cambridge, v. 40, n. 2, p. 263-279, 1999. DOI: <https://doi.org/10.1006/jmla.1998.2612>

GIGERENZER, G.; TODD, P. M.; THE ABC GROUP. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press, 1999.

GILBOY, E.; SOPENA, J. M.; CLIFTON, C.; FRAZIER, L. Argument Structure and Preferences in the Processing of Spanish and English Complex NPs. *Cognition*, Amsterdam, v. 54, p. 131-167, 1995. DOI: [https://doi.org/10.1016/0010-0277\(94\)00636-Y](https://doi.org/10.1016/0010-0277(94)00636-Y)

GRICE, H. P. Logic and Conversation. In: COLE, P; MORGAN, J. L. (ed.). *Syntax and Semantics*. Speech Acts. New York: Academic Press, 1975. p. 41-58. DOI: [https://doi.org/10.1163/9789004368811\\_003](https://doi.org/10.1163/9789004368811_003)

GRILLO, N.; COSTA, J. A Novel Argument for the Universality of Parsing Principles. *Cognition*, Amsterdam, v. 133, n. 1, p. 156-187, 2014. DOI: <https://doi.org/10.1016/j.cognition.2014.05.019>

GUTIERREZ-ZIARDEGI, E.; CARREIRAS, M.; LAKA, I. Bilingual Sentence Processing: Relative Clause Attachment in Basque and Spanish. In: ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING, 17, 2004, Maryland. *Trabajo presentado*. College Park, MD: University of Maryland, 2004.

HAGOORT, P.; BAGGIO, G.; WILLEMS, R. M. Semantic Unification. In: GAZZANIGA, M. (ed.). *The Cognitive Neurosciences*. 4th ed. Cambridge: MIT Press, 2009. p. 819-836.

HEMFORTH, B.; KONIECZNY, L.; SCHEEPERS, C. Syntactic Attachment and Anaphor Resolution: The Two Sides of Relative Clause Attachment. In: CROCKER, M.W.; PICKERING, M.; CLIFTON, C. JR. (ed.). *Architectures and Mechanisms for Language Processing*. Cambridge: Cambridge University Press, 2000. p. 259-281. DOI: <https://doi.org/10.1017/CBO9780511527210.012>

HENRICH, J.; HEINE, S. J.; NORENZAYAN, A. Most People Are Not WEIRD. *Nature*, Londres, v. 466, n. 7302, p. 29, 2010. DOI: <https://doi.org/10.1038/466029a>

HSIAO, Y.; MACDONALD, M. C. Production Predicts Comprehension: Animacy Effects in Mandarin Relative Clause Processing. *Journal of Memory and Language*, Cambridge, v. 89, p. 87-109, 2016. DOI: <https://doi.org/10.1016/j.jml.2015.11.006>

HUETTIG, F. Four Central Questions about Prediction in Language Processing. *Brain Research*, Amsterdam, v. 1626, p. 118-135, 2015. DOI: <https://doi.org/10.1016/j.brainres.2015.02.014>

IGOA, J. M.; CARREIRAS, M.; MESEGUER, E. A Study on Late Closure in Spanish: Principle-Grounded vs. Frequency-Based Accounts of Attachment Preferences. *The Quarterly Journal of Experimental Psychology Section A*, Londres, v. 51, n. 3, p. 561-592, 1998. DOI: <https://doi.org/10.1080/713755775>

JACKENDOFF, R. Précis of Foundations of Language: Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*, Cambridge, v. 26, n. 6, p. 651-707, 2003. DOI: <https://doi.org/10.1017/S0140525X03000153>

JACKENDOFF, R. The Parallel Architecture and Its Place in Cognitive Science. In: HEINE, B.; NARROG, H. (ed.). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2009. p. 646-778. DOI: <https://doi.org/10.1093/oxfordhb/9780199544004.013.0023>

JACKENDOFF, R. *Fundamentos del lenguaje: mente, significado, gramática y evolución*. México D.F.: Fondo de Cultura Económica, 2010.

JAMES, A. N.; FRAUNDORF, S. H.; LEE, E. K.; WATSON, D. G. Individual Differences in Syntactic Processing: Is there Evidence for Reader-Text Interactions? *Journal of Memory and Language*, Cambridge, v. 102, p. 155-181, 2018. DOI: <https://doi.org/10.1016/j.jml.2018.05.006>

JUN, S. A.; KOIKE, C. Default Prosody and Relative Clause Attachment in Japanese. *Japanese-Korean Linguistics*, Stanford, v. 13, p. 41-53, 2008.

KAMIDE, Y.; D. C. MITCHELL. Relative Clause Attachment: Non-Determinism in Japanese Parsing. *Journal of Psycholinguistic Research*, New York, v. 26, p. 247-254, 1997. DOI: <https://doi.org/10.1023/A:1025017817290>

KARIMI, H.; FERREIRA, F. Good-Enough Linguistic Representations and Online Cognitive Equilibrium in Language Processing. *Quarterly Journal of Experimental Psychology*, Londres, v. 69, n. 5, p. 1013-1040, 2015. DOI: <https://doi.org/10.1080/17470218.2015.1053951>

KIMBALL, J. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, Amsterdam, v. 2, p. 15-47, 1973. DOI: [https://doi.org/10.1016/0010-0277\(72\)90028-5](https://doi.org/10.1016/0010-0277(72)90028-5)

KLIN, C. M.; GUZMÁN, A. E.; WEINGARTNER, K. M.; RALANO, A. S. When Anaphor Resolution Fails: Partial Encoding of Anaphoric Inferences. *Journal of Memory and Language*, Cambridge, v. 54, n. 1, p. 131-143, 2006. DOI: <https://doi.org/10.1016/j.jml.2005.09.001>

KUPERBERG, G. R. Neural Mechanisms of Language Comprehension: Challenges to Syntax. *Brain Research*, Amsterdam, v. 1146, p. 23-49, 2007. DOI: <https://doi.org/10.1016/j.brainres.2006.12.063>

KUPERBERG, G. R.; JAEGER, T. F. What do We Mean by Prediction in Language Comprehension? *Language, Cognition and Neuroscience*, New York, v. 31, n. 1, p. 32-59, 2015. DOI: <https://doi.org/10.1080/23273798.2015.1102299>

LEVY, R. Expectation-Based Syntactic Comprehension. *Cognition*, Amsterdam, v. 106, n. 3, p. 1126-1177, 2008. DOI: <https://doi.org/10.1016/j.cognition.2007.05.006>

LOVRIC, N.; FODOR, J. D. Relative Clause Attachment in Sentence Parsing. In: ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING, 13, 2000, La Jolla. *Trabajo presentado*. San Diego, CA, 2000.

MACDONALD, J.; MCGURK, H. Visual Influences on Speech Perception Processes. *Perception & Psychophysics*, New York, v. 24, n. 3, p. 253-257, 1978. DOI: <https://doi.org/10.3758/BF03206096>

MACDONALD, M. C.; CHRISTIANSEN, M. H. Reassessing Working Memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, New York, v. 109, n. 1, p. 35-54, 2002. DOI: <https://doi.org/10.1037/0033-295X.109.1.35>

MACDONALD, M. C.; PEARLMUTTER, N. J.; SEIDENBERG, M. S. Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review*, New York, v. 101, n. 4, p. 676-703, 1994. DOI: <https://doi.org/10.1037/0033-295X.101.4.676>

MACDONALD, M. C.; SEIDENBERG, M. S. Constraint Satisfaction Accounts of Lexical and Sentence Comprehension. In: TRAXLER, M. J.; GERNSBACHER, M.A. (ed.). *Handbook of Psycholinguistics*. Amsterdam: Elsevier/Academic Press, 2006. p. 581-611. DOI: <https://doi.org/10.1016/B978-012369374-7/50016-X>

MACDONALD, M. C.; THORNTON, R. When Language Comprehension Reflects Production Constraints: Resolving Ambiguities with the Help of Past Experience. *Memory & Cognition*, New York, v. 37, n. 8, p. 1177-1186, 2009. DOI: <https://doi.org/10.3758/MC.37.8.1177>

MAIA, M.; MAIA, J. A compreensão de orações relativas por falantes monolíngües e bilíngües de Português e de Inglês. In: MAIA, M.; FINGER, I. (org.). *Processamento da Linguagem*. Pelotas: Educat, 2005. p. 163-178.

MCCLELLAND, J. L. On The Time Relations of Mental Processes: An Examination of Systems of Processes in Cascade. *Psychological Review*, New York, v. 86, n. 4, p. 287-330, 1979. DOI: <https://doi.org/10.1037/0033-295X.86.4.287>

MCGURK, H.; MACDONALD, J. Hearing Lips and Seeing Voices. *Nature*, Londres, v. 264, n. 5588, p. 746-748, 1976. DOI: <https://doi.org/10.1038/264746a0>

MCRAE, K.; MATSUKI, K. Constraint-Based Models of Sentence Processing. In: VAN GOMPEL, R. (ed.). *Sentence Processing*. Sussex: Psychology Press, 2013. p. 51-77.

MCRAE, K.; SPIVEY-KNOWLTON, M. J.; TANENHAUS, M. K. Modeling the Influence of Thematic Fit (and Other Constraints) in On-Line Sentence Comprehension. *Journal of Memory and Language*, Cambridge, v. 38, n. 3, p. 283-312, 1998. DOI: <https://doi.org/10.1006/jmla.1997.2543>

MILLER, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, New York, v. 63, n. 2, p. 81-97, 1956. DOI: <https://doi.org/10.1037/h0043158>

MITCHELL, D. C.; BRYSSBAERT, M. Challenges to Recent Theories of Crosslinguistic Variation in Parsing: Evidence from Dutch. In: HILLERT, D. (ed.). *Sentence Processing: A Crosslinguistic Perspective*. Leiden: Brill, 1998. p. 313-335. DOI: [https://doi.org/10.1163/9780585492230\\_018](https://doi.org/10.1163/9780585492230_018)

MITCHELL, D. C.; BRYSSBAERT, M.; GRONDELAERS, S.; SWANEPOEL, P. Modifier Attachment in Dutch: Testing Aspects of Construal Theory. In: KENNEDY, A.; RADACH, R.; HELLER, D.; PYNTE, R. (ed.). *Reading as a Perceptual Process*. Amsterdam: North-Holland, 2000. p. 493-516. DOI: <https://doi.org/10.1016/B978-008043642-5/50023-1>

MITCHELL, D. C.; CUETOS, F. The Origins of Parsing Strategies. *Current Issues in Natural Language Processing*, [S.l.], p. 1-12, 1991. Disponible en: [https://www.researchgate.net/publication/239666294\\_The\\_origins\\_of\\_parsing\\_strategies](https://www.researchgate.net/publication/239666294_The_origins_of_parsing_strategies). Accesado en: 10 jan. 2021.

MITCHELL, D. C.; CUETOS, F.; CORLEY, M. M.; BRYSSBAERT, M. Exposure-Based Models of Human Parsing: Evidence for the Use of Coarse-Grained (Nonlexical) Statistical Records. *Journal of Psycholinguistic Research*, New York, v. 24, n. 6, p. 469-488, 1995. DOI: <https://doi.org/10.1007/BF02143162>

MIYAMOTO, E.T. *Relative Clause Processing in Brazilian Portuguese and Japanese*. 1999. 125f. Thesis (Doctor of Philosophy) – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, 1999.

NIJHAWAN, R. Three-Dimensional Müller-Lyer Illusion. *Perception & Psychophysics*, New York, v. 49, n. 4, p. 333-341, 1991. DOI: <https://doi.org/10.3758/BF03205989>

NOVICK, J. M.; TRUESWELL, J. C.; THOMPSON-SCHILL, S. L. Broca's Area and Language Processing: Evidence for the Cognitive Control Connection. *Language and Linguistics Compass*, Londres, v. 4, n. 10, p. 906-924, 2010. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00244.x>

NOWAK, A. On Relative Clause Attachment in Polish: Evidence for Late Closure and Against Case Matching. In: ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING, 13, 2000, La Jolla. *Trabajo presentado*. San Diego, CA, 2000.

PAPADOPOULOU, D.; CLAHSSEN, H. Parsing Strategies in L1 and L2 Sentence Processing: A Study of Relative Clause Attachment in Greek. *Studies in Second Language Acquisition*, Cambridge, UK, v. 25, n. 4, p. 501-528, 2003. DOI: <https://doi.org/10.1017/S0272263103000214>

PAYNE, B. R.; GRISON, S.; GAO, X.; CHRISTIANSON, K.; MORROW, D. G.; STINE-MORROW, E. A. Aging and Individual Differences in Binding During Sentence Understanding: Evidence from Temporary and Global Syntactic Attachment Ambiguities. *Cognition*, Amsterdam, v. 130, n. 2, p. 157-173, 2014. DOI: <https://doi.org/10.1016/j.cognition.2013.10.005>

PIAGET, J. *The Equilibrium of Cognitive Structures*. Chicago: University of Chicago Press, 1985.

PINKER, S. *El instinto del lenguaje*. Madrid: Alianza, 1995.

QUINN, D.; ABDELGHANY, H.; FODOR, J. D. More Evidence of Implicit Prosody in Reading: French and Arabic Relative Clauses. *In: ANNUAL CUNY CONFERENCE ON HUMAN SENTENCE PROCESSING*, 13, 2000, La Jolla. *Trabalho apresentado*. San Diego, CA, 2000.

REDDING, G. M.; HAWLEY, E. Length Illusion in Fractional Müller-Lyer Stimuli: An Object-Perception Approach. *Perception*, New York, v. 22, n. 7, p. 819-828, 1993. DOI: <https://doi.org/10.1068/p220819>

RIBEIRO, A. J. Late Closure em Parsing no Português do Brasil. *In: MAIA, M.; FINGER, I. (org). Processamento da linguagem*. Pelotas: Educat, 2005. p. 51-70.

ROHDE, H.; LEVY, R.; KEHLER, A. Anticipating Explanations in Relative Clause Processing. *Cognition*, Amsterdam, v. 118, n. 3, p. 339-358, 2011. DOI: <https://doi.org/10.1016/j.cognition.2010.10.016>

ROTHMAN, J. On the Typological Economy of Syntactic Transfer: Word Order and Relative Clause High/Low Attachment Preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Language Teaching*, Berlin, v. 48, n.2-3, p. 245-273, 2010. DOI: <https://doi.org/10.1515/iral.2010.011>

SALTHOUSE, T. A. The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, New York, v. 103, n. 3, p. 403-428, 1996. DOI: <https://doi.org/10.1037/0033-295X.103.3.403>

SANFORD, A. J.; STURT, P. Depth of Processing in Language Comprehension: Not Noticing the Evidence. *Trends in Cognitive Sciences*, Amsterdam, v. 6, n. 9, p. 382-386, 2002. DOI: [https://doi.org/10.1016/S1364-6613\(02\)01958-7](https://doi.org/10.1016/S1364-6613(02)01958-7)

SCHWERING, S. C.; MACDONALD, M. C. Verbal Working Memory as Emergent from Language Comprehension and Production. *Frontiers in Human Neuroscience*, Lausanne, v. 14, p. 1-19, 2020. DOI: <https://doi.org/10.3389/fnhum.2020.00068>

SEGALL, M. H.; CAMPBELL, D. T.; HERSKOVITS, M. J. Cultural Differences in the Perception of Geometric Illusions. *Science*, Washington, v. 139, n. 3556, p. 769-771, 1963. DOI: <https://doi.org/10.1126/science.139.3556.769>

SEKERINA, I. A.; FERNÁNDEZ, E. M.; PETROVA, K. A. Relative Clause Attachment in Bulgarian. *In: ANNUAL WORKSHOP ON FORMAL APPROACHES TO SLAVIC LINGUISTICS*, 12th, 2004, Ottawa. *Proceedings* [...]. Ottawa: University of Ottawa, 2004. p. 375-394.

SLATTERY, T. J.; STURT, P.; CHRISTIANSON, K.; YOSHIDA, M.; FERREIRA, F. Lingering Misinterpretations of Garden Path Sentences Arise from Competing Syntactic Representations. *Journal of Memory and Language*, Cambridge, v. 69, n. 2, p. 104-120, 2013. DOI: <https://doi.org/10.1016/j.jml.2013.04.001>

STALLINGS, L. M.; MACDONALD, M. C.; O'SEAGHDHA, P. G. Phrasal Ordering Constraints in Sentence Production: Phrase Length and Verb Disposition in Heavy-NP Shift. *Journal of Memory and Language*, Cambridge, v. 39, n. 3, p. 392-417, 1998. DOI: <https://doi.org/10.1006/jmla.1998.2586>

STEWART, A. J.; HOLLER, J.; KIDD, E. Shallow Processing of Ambiguous Pronouns: Evidence for Delay. *Quarterly Journal of Experimental Psychology*, New York, v. 60, n. 12, p. 1680-1696, 2007. DOI: <https://doi.org/10.1080/17470210601160807>

SWETS, B.; DESMET, T.; CLIFTON, C.; FERREIRA, F. Underspecification of Syntactic Ambiguities: Evidence from Self-Paced Reading. *Memory & Cognition*, New York, v. 36, n. 1, p. 201-216, 2008. DOI: <https://doi.org/10.3758/MC.36.1.201>

SWETS, B.; DESMET, T.; HAMBRICK, D. Z.; FERREIRA, F. The Role of Working Memory in Syntactic Ambiguity Resolution: A Psychometric Approach. *Journal of Experimental Psychology: General*, New York, v. 136, n. 1, p. 64-81, 2007. DOI: 10.1037/0096-3445.136.1.64

TANENHAUS, M. K.; SPIVEY-KNOWLTON, M. J.; EBERHARD, K.; SEDIVY, J. C. Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, Washington, v. 268, p. 1632-1634, 1995. DOI: <https://doi.org/10.1126/science.7777863>

TIIPPANA, K. What is the McGurk Effect? *Frontiers in Psychology*, Lausanne, v. 5, n. 725, p. 1-3, 2014. DOI: <https://doi.org/10.3389/fpsyg.2014.00725>

TOWNSEND, D.; BEVER, T. G. *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press, 2001. DOI: <https://doi.org/10.7551/mitpress/6184.001.0001>

TRAXLER, M.; GERNSBACHER, M. A. (ed.). *Handbook of Psycholinguistics*. London: Elsevier, 2011.

TRUESWELL, J. C.; TANENHAUS, M. K.; GARNSEY, S. M. Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, Cambridge, v. 33, n. 3, p. 285-318, 1994. DOI: <https://doi.org/10.1006/jmla.1994.1014>

VAN DYKE, J. A.; JOHNS, C. L.; KUKONA, A. Low Working Memory Capacity is Only Spuriously Related to Poor Reading Comprehension. *Cognition*, Amsterdam, v. 131, n. 3, p. 373-403, 2014. DOI: <https://doi.org/10.1016/j.cognition.2014.01.007>

VAN GOMPEL, R. P.; PICKERING, M. J.; TRAXLER, M. J. Unrestricted Race: A New Model of Syntactic Ambiguity Resolution. In: KENNEDY, A.; RADACH, R.; HELLER, D.; PYNTE, R. (ed.). *Reading as a Perceptual Process*. Amsterdam: North-Holland, 2000. p. 621-648. DOI: <https://doi.org/10.1016/B978-008043642-5/50029-2>

VAN HERTEN, M.; CHWILLA, D. J.; KOLK, H. H. When Heuristics Clash with Parsing Routines: ERP Evidence for Conflict Monitoring in Sentence Perception. *Journal of Cognitive Neuroscience*, Cambridge, v. 18, n. 7, p. 1181-1197, 2006. DOI: <https://doi.org/10.1162/jocn.2006.18.7.1181>

VOGEL, E. K.; AWH, E. How to Exploit Diversity for Scientific Gain: Using Individual Differences to Constrain Cognitive Theory. *Current Directions in Psychological Science*, New York, v. 17, n. 2, p. 171-176, 2008. DOI: <https://doi.org/10.1111/j.1467-8721.2008.00569.x>

WATERS, G. S.; CAPLAN, D. The Reliability and Stability of Verbal Working Memory Measures. *Behavior Research Methods, Instruments, & Computers*, Dordrecht, v. 35, n. 4, p. 550-564, 2003. DOI: <https://doi.org/10.3758/BF03195534>

WELLS, J. B.; CHRISTIANSEN, M. H.; RACE, D. S.; ACHESON, D. J.; MACDONALD, M. C. Experience and Sentence Processing: Statistical Learning and Relative Clause Comprehension. *Cognitive Psychology*, Cambridge, v. 58, n. 2, p. 250-271, 2009. DOI: <https://doi.org/10.1016/j.cogpsych.2008.08.002>

ZAGAR, D.; PYNTE, J.; RATIVEAU, S. Evidence for Early Closure Attachment on First Pass Reading Times in French. *The Quarterly Journal of Experimental Psychology Section A*, New York, v. 50, n. 2, p. 421-438, 1997.

ZUNINO, G. M. Procesamiento de lenguaje: ¿de qué hablamos cuando hablamos de predicción? *Quintú Quimün, Revista de lingüística*, General Roca, v. 3, p. 1-38, 2019.