v. 25, n. 1, e49797, 2025
Corpus Linguistics: Studies and Applications
ISSN: 1984-6398 · DOI: 10.1590/1984-6398202549797

Revista Brasileira de Linguística Aplicada

Thematic Dossier

Data-Driven Learning in Brazilian Educational Contexts: An Analysis of a Teacher Training Program Following BNCC's Guidelines

Aprendizagem Movida por Dados no contexto educacional brasileiro: estudo sobre um curso de formação para professores à luz da BNCC

Paula Tavares Pinto*
*Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) | São José do Rio Preto | SP | BR
paula.pinto@unesp.br
https://orcid.org/0000-0001-9783-2724

Carolina Tavares de Carvalho*
*Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) | São José do Rio Preto | SP | BR
carolina.tavares@unesp.br
https://orcid.org/0000-0001-9446-6394

Adriane Orenha-Ottaiano*
*Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) | São José do Rio Preto | SP | BR
adriane.ottaiano@unesp.br
https://orcid.org/0000-0001-8417-5120

Franciele Spinelli**
**Universidade de Queensland (UQ) | Brisbane | AU
spinelli.franciele@gmail.com
https://orcid.org/0000-0002-4923-329X

Talita Serpa*
*Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) | São José do Rio Preto | SP | BR
talita.serpa@unesp.br
https://orcid.org/0000-0003-3324-9593

Peter Crosthwaite**
**Universidade de Queensland (UQ) | Brisbane | AU
p.cros@uq.edu.au
https://orcid.org/0000-0002-7687-5538

Abstract: This article discusses the results of a teacher training program on Data Driven Learning (DDL) for Brazilian teachers of English. The program's main objective was to help teachers improve their language teaching practices through digital technologies to promote autonomous learning with authentic linguistic data. Over two months, English teachers completed seven modules of the program, each covering a DDL tool and a set of corpus skills. At the end of the program, five participating teachers developed lesson plans that integrated the tools and skills they had learned. Some of these teachers later participated in a semi-structured interview. Technology Integration Observation Instrument (TIOI) evaluation of the lesson plans and tabletop analysis of the interviews show that, although teachers considered it important to use DDL tools in their classes, they did not feel entirely confident implementing them. This highlights the need for continuous support and professional development when it comes to integrating DDL into language teaching.

Keywords: data-driven learning (DDL); teacher training program; lesson plans.

Resumo: Este artigo apresenta os resultados de um curso sobre a Aprendizagem Movida por Dados (DDL) para professores brasileiros de língua inglesa. O objetivo foi auxiliá-los a aprimorar o ensino de línguas com o apoio de tecnologias digitais para promover um aprendizado

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

autônomo com dados linguísticos autênticos por parte de seus alunos. Ao longo de dois meses, os professores concluíram sete módulos, cada um abrangendo uma ferramenta de DDL específica, e desenvolveram planos de aula integrando essas ferramentas ao seu conteúdo. Alguns dos professores participaram de entrevistas semiestruturadas. Uma avaliação do Instrumento de Observação de Integração de Tecnologia (TIOI) dos planos de aula e uma Tabela de Análises de Dados Principais das entrevistas mostram que, embora os professores considerassem importante o uso das ferramentas de DDL em suas aulas, eles não se sentiam completamente seguros para implementá-las. Este fator chama a atenção para a necessidade de um apoio contínuo ao desenvolvimento profissional quando se refere à integração do DDL na prática profissional.

Palavras-chave: aprendizagem movida por dados (DDL); programa de formação para professores; planos de aula.

1 Introduction

Contemporary perspectives on 21st-century education highlight the need for students to learn how to use digital technologies proficiently to meet the needs of the current virtual world, thus requiring qualified teachers to respond to this demand (Prensky, 2009). However, there is a gap among teachers who need specific training to meet this demand (Mishra; Koehler, 2006; Kenski, 2012; Moran, 2015).

Today's learners are expected to develop skills and learning dispositions essential for success in society and the workplace, including (i) learning and innovation, (ii) digital literacy, and (iii) career and life (Horton, 2007; Binkley *et al*., 2012). These skills are required in professional careers and are crucial in international business and educational environments where English is widely used due to its prominence in the world (Jenkins, 2012; 2013). However, English learners, especially from non-native English-speaking countries, may not have fully developed these skills, which may limit their access to linguistic resources necessary for their professional endeavors. According to the *Demand for English Learning in Brazil* survey (British Council, 2014), only 5.1% of the Brazilian population over 16 years of age has some knowledge of English (British Council, 2014).

One way to improve English proficiency development in Basic and Higher Education settings would be to integrate technology for learners to explore language use in different contexts (British Council, 2014). However, a lack of access to language data and training in language data analysis for teaching and learning is the reality in many Latin American contexts, including Brazil, given the large socioeconomic differences within the population. Access to technological and digital tools, such as computers, tablets, and smartphones, is also

particularly scarce in the Public Teaching System, where students from lower socioeconomic backgrounds are concentrated – a scarcity well-documented by Pretto (1999), Behrens (2001), and Mattar (2013). According to the Brazilian report *Relatório Guia Edutec - Diagnóstico do Nível de Adoção de Tecnologia nas Escolas Públicas Brasileiras em 2022* (CIEB, 2022), most public schools in Brazil have limited access to digital educational resources, and while infrastructure has improved, there are insufficient resources for all students, even blackboards, books, and chalk. This is a pressing problem in many Brazilian classrooms, further complicating efforts to integrate more advanced technological tools. As noted by Sbrogio (2021), teachers in Brazil face a significant shortage of traditional teaching apparatus, and require time and resources to learn and integrate digital tools into their teaching practices.

Despite the efforts of educational institutions and dedicated teachers, significant obstacles still hinder the successful implementation of teacher training programs that foster technological integration into pedagogical practices. In Brazil, teachers often attain only basic digital skills, which limits their ability to explore and integrate technologies into their teaching routines. Educational institutions do not possess the resources to provide students with access to computer-based programs and software. As a result, the few devices available are operated only by the teacher, leaving learners passively receiving knowledge rather than actively engaging with their educational progress (Sbrogio, 2021; Leite, 2021; De Lima e Loureiro, 2016; Moran, 2013). As Haddad (2007) suggests, there is a need for continuous education that encourages experimentation and active teaching methodologies with a deliberate focus on the use of technology. These methodologies should align with the pedagogical goals of schools and the national education system's curriculum.

The efforts to implement technology as part of language teaching practices are currently supported and documented by the National Curricular Common Base (*Base Nacional Comum Curricular* – BNCC, Brazil, 2017). BNCC is related to the Brazilian Federal Constitution (1988) and the Law of Directives and Bases of National Education –CNE-CP Resolution 2/2015 (Da Silva, 2020), both of which demand the creation of balanced national bases with specific content that reflects Brazil's ethnic, geographic, and cultural diversity (Silva and Santos, 2018). Regarding the relationship between English Language Learning and Technology, the BNCC presents a chapter on "Languages and its Technologies", focusing on how the use of computational tools, hypertext, multimedia, interactive resources, and social networks can expand autonomy, protagonism, and authorship in the practices of different languages by students. As pointed out by Garcia (2020), the competencies discussed by BNCC related to (i) scientific, critical, and creative thinking; (ii) communication; and (iii) digital culture have a strong potential link with corpus linguistics and its pedagogical application in Data-Driven Learning (DDL). DDL requires language to be investigated through scientific empiricism with the help of technological tools and computer software programs (Berber Sardinha, 2004). BNCC also discusses the development of competencies and skills related to the critical and responsible use of digital technologies, given that the successful incorporation of classroom DDL would involve a process of critical, meaningful, reflective, and ethical communication referent to accessing and disseminating linguistic information, producing knowledge, solving problems, and exercising protagonism (Crosthwaite, 2020). In this light, a DDL program tailored to the needs of Brazilian English teachers, in line with BNCC guidelines and values, was proposed. The *Innovative Pedagogical Practices for English Language Teaching* program was conducted in an online format over seven modules (30 hours), covering pedagogical

applications of DDL; corpus tools, including VersaText, Sketch Engine, SKELL, and TradCorpus; and hands-on/off activities.

2 Data-Driven Learning (DDL) and Corpus-Based Language Pedagogy

DDL represents an inductive pedagogical approach focused on the authentic language data analysis (as found in language corpora) through digital tools, such as concordancers (Johns, 1991). At the core of DDL lies the acquisition of corpus literacy skills, which are essential for interpreting extensive linguistic databases (Pérez-Paredes, 2020; Ma *et al*., 2023). These skills range from interpreting frequency data and understanding textual features to analyzing concordance lines,[1] identifying collocations, and engaging in critical reading (Boulton; Vyatkina, 2021). Through the application of DDL tools, both educators and learners can explore relevant data (language information) to enrich the teaching and learning process (Crosthwaite, 2020; Le Foll, 2021). According to Johns (1986, 1988), while the learner's role is to "uncover" the foreign language, the language teacher's role is to create an environment where the learner can cultivate their discovery strategies.

The use of DDL offers potential strategies to assist educators in adequately preparing their students for future challenges, aligning closely with the guidelines established by UNESCO (2014). Firstly, these guidelines underscore linguistic diversity as a fundamental human asset, highlighting the enrichment that multilingualism brings to individuals and societies. Secondly, they recognize multilingualism as promoting dialogue, mutual understanding, tolerance, and respect for identities, essential in our increasingly multicultural and multilingual world. Through DDL, educators can promote the learning of multiple languages more effectively, ensure high-quality linguistic training, and foster intercultural competence among students (O'Keeffee, McCarthy and Carter, 2007; Braun, 2007; Crosthwaite, 2024), thereby supporting these international guidelines. Therefore, allowing Brazilian students to access technology through DDL techniques might help them become more connected with the ideas of a globalized world that they may not so easily be able to reach.

Teachers require appropriate training to select and use the plethora of DDL tools available. DDL can serve as a valuable resource in this respect, involving instructing teachers on corpus learning techniques, proficiency in handling corpus technology (corpus tools), designing and implementing corpus-based activities, and enhancing instructions on how to use these materials effectively. Similarly, learners need to be equipped with strategies for effectively acquiring a foreign language through corpus analysis, fostering curiosity and speculation in their learning journey (Johns, 1991, p. 2) and, most importantly, nurturing their motivation to engage with the language in this new digital environment. If DDL is to be successfully integrated, training teachers in corpus-based language pedagogy is therefore essential.

Ma *et al*. (2021, p. 2) define corpus-based language pedagogy as "the ability to integrate corpus linguistics technology into classroom language pedagogy to facilitate language teaching". In corpus-based language pedagogy, the focus is on using electronic corpora to design, build, and analyze language samples through concordance lines. This approach also involves developing activities and materials for teaching and learning based on the corpus data.

[1] A concordance line is a list of occurrences of a specific item in a corpus, where the word searched is in the center and surrounded by its original context (Berber Sardinha, 2004).

In order to apply DDL in class, students need to develop some Corpus Skills, such as being able to interpret linguistic information from linguistic databases. This can be done through user-friendly platforms, such as Versa Text and SKELL for DDL tasks (see detailed information about other tools in Section 3). Péres-Paredez (2020, p.18) defined the concept of Corpus Skills, which should be taught to language learners and are listed as follows:

> (1) why frequency matters; (2) how to interpret frequency; (3) understanding text types: register basics; (4) understanding textual features and textual data; (5) using an existing corpus; (6) reading concordance lines; (7) handling frequencies; (8) understanding collocations; (9) basic corpus design features; (10) comparing two corpora; (11) statistical tests; (12) transcribing your data; (13) annotating and querying your data: using your own annotation taxonomy; (14) understanding keywords; (15) researching nouns and noun phrases; (16) looking at the lexicon of a corpus: n-grams; (17) complex searches; (18) remaining critical.

The selection of appropriate tools also holds significant importance for corpus-based pedagogy. Goulart, Kostromitina, and Klein (2024, p. 209) proposed some criteria for choosing web-based collocation tools, emphasizing key factors, such as their accessibility (being free), corpus-based nature, and functionality, in enabling word searches. These criteria ensure that educators can effectively leverage digital resources in line with the principles of corpus-based language pedagogy in order to enhance language instruction and learning experiences.

In this study, a teacher training program, entitled *Innovative Pedagogical Practices for English Language Teaching: Student as Language Detectives and Learning Protagonists*, was proposed to train teachers in DDL and help them develop corpus-based materials. This paper presents this training program, which was conducted for in-service language teachers from various states in Brazil and discusses their feedback after the program. The following research questions guided the study:

1. How did participants incorporate what they learned during the program into their lesson plans, and how much of the DDL approach were they able to use in their classes?
2. What were participants' perceptions of the tools and skills they learned after being introduced to the DDL approach?

3 A DDL Program for In-Service Teachers

The *Innovative Pedagogical Practices for English Language Teaching: Student as Language Detectives and Learning Protagonists* teacher training program aimed to provide professional training to public school English teachers in Brazil. The program focused on helping these teachers 1) familiarize themselves with some digital tools commonly used in corpus linguistics, 2) reflect on how to implement these tools in their pedagogical practices so that students could play an active role in their language learning process, 3) learn how to better select tools according to their teaching context, and 4) become more pedagogically motivated towards the DDL approach for language teaching and learning.

The 30-hour teacher training program was administered over a two-month period and consisted of seven modules. It was conducted online, combining synchronous and asynchronous activities to accommodate a diverse group of educators. Five modules were delivered synchronously via Zoom by different instructors, each focusing on different corpus-based tools and skills (Pérez-Paredes, 2020). The two asynchronous modules were part of the *Improving your Writing Through Corpora* online course, designed to help foreign language learners enhance their academic writing using corpus tools (Crosthwaite, 2019, 2020). Three extra consultation sessions were also offered to address participants' questions about the program activities and the final assessment (i.e., developing a lesson plan using the tools and skills learned throughout the program). Table 1 provides a summary of the training program schedule and plan, as well as the corresponding corpus skills as presented by Pérez-Paredes (2020).

Table 1 – Program Schedule and Plan

Module	Topic	Tool	Corpus skills
1	Introduction: Pedagogical applications of DDL and introduction	VersaText and Sketch Engine	1, 5, 6, 8, 15
2	Improving your writing through corpora, Part 1	Several tools	1, 2, 3, 4, 5, 8, 9, 10, 12
3	Improving your writing through corpora, Part 2	Several tools	1, 2, 3, 4, 5, 8, 9, 10, 12
4	Linggle and Skell: Educational applications	Linggle, SKELL	6, 15
5	VocabProfile: Educational applications	VocabProfile	1, 2, 4, 7
6	PlatCol and online English collocations Platform: Educational applications	PlatCol platform	1, 2, 8, 17
7	Hands-on/off activities (Classroom) e TradCorpus (pedagogical application)	TradCorpus	1, 2, 4, 5, 8, 9, 10, 12, 18

Source: the authors.

The tools used and taught in the program, from the most user-friendly to the most complex, are described as follows:

1. VersaText (Thomas; Baisa, 2018) is a web-based tool that allows learners of English to investigate the language used in a single text. It processes the text as a small corpus, providing lemmatization and part-of-speech information. With these features, both teachers and students can generate word clouds and explore how different categories of words appear within concordance lines.
2. Sketch Engine (Sketch Engine, 2015) is a tool for analyzing authentic text collections (corpora), enabling users to quickly identify typical language patterns as well as rare, unusual, or emerging usages. It is also designed to support text analysis and text mining tasks.

3. SKELL (Sketch Engine, 2015) is an intuitive resource for English learners and teachers, designed to verify how words or phrases are authentically used by native speakers. It allows users to explore collocations, synonyms, and contextualized examples of a given search term, as well as to generate word clouds of semantically related items.
4. Linggle (Linggle, 2025) is a corpus-informed search engine that assists users in evaluating lexical choices, collocations, and usage patterns, thereby supporting the production of more natural and idiomatic writing.
5. VocabProfile (Cobb, 2015) is an online tool that analyzes the lexical composition of texts, providing insights into their vocabulary load. Users can identify the distribution of words across frequency lists such as the GSL (General Service List), AWL (Academic Word List), or simplified versions of the British National Corpus lists.
6. PlatCol (Grupo de Banco de Dados, 2021) is a digital platform of collocations dictionaries developed to enhance the learning and translation of collocations. Its primary goal is to foster proficiency in English and Portuguese, encouraging learners to use both languages with greater naturalness and accuracy.
7. TradCorpus (Corpora, 2019) is a platform with specialized corpora and glossaries that users can have access to if they are working with specialized English and Portuguese language and literature.

3.1 Participants

A total of 21 teachers from five Brazilian states participated in the training program, including São Paulo (57.5%), Paraná (22.5%), Goiás (12.5%), Rio de Janeiro (5%), and Bahia (2.5%). All participants held at least an undergraduate degree, with only one participant having completed his Bachelor's degree during the program while working as an English language teacher. In terms of their professional backgrounds, participants worked in elementary, secondary, and higher education, with 62.8% working in public high schools and 32.6% in private institutions.

3.2 Instruments for Data Collection

Participants completed a pre-program questionnaire to gather their demographic information, including gender, English proficiency level, study background, place of origin, and teaching context/location. After each module, participants completed a self-assessment questionnaire focused on their familiarity with corpus-based tools learning, any challenges they encountered while using the tools, and the feasibility of integrating these tools into their teaching context. An open-ended question was also included in the questionnaires for participants to provide feedback and suggestions for the upcoming seminars. Upon concluding the program, participants completed a final questionnaire, adapted from Yoon and Hirvela (2004), on their general perceptions regarding: 1) the overall teacher training program, 2) their development of corpus-based skills, 3) the effectiveness of each module and corpus-based tools, and 4) their satisfaction with instructors and the program structure.

Upon completing the program, participants were asked to complete a final assignment, which consisted of elaborating a lesson plan using the pedagogical DDL tools and corpus skills learned during the program modules. Participants were provided with a lesson plan template from Santos (2024), as found in Appendix A. They were expected to include the subject to be taught, the linguistic framework used, the educational context, the target students' proficiency level, the mode of delivery, and the assessment format (i.e., online, face-to-face, or hybrid). Three participants who had completed the lesson plan agreed to participate in a post-program semi-structured online interview one year later.

The interviews took approximately 45 minutes each, were audio-recorded, and conducted in Portuguese so that participants would feel more confident to express themselves in their native language. The interview questions asked participants to reflect on their experiences with the program, whether and how they were able to implement the lesson plan they had developed, and any experiences they had had, sharing their knowledge about DDL with fellow teachers.

3.3 Data Analysis

3.3.1 Interviews

Translated into English and then analyzed through a tabletop methodology and using ##LancsBox to answer the second research question regarding participants' perceptions of the tools and skills learned through the DDL approach. The tabletop methodology (Saldaña, 2016) was used to systematically organize the interview data, involving the arrangement of the interview transcriptions on a tabletop and the grouping of information into categories to discern patterns more clearly. By identifying the frequency of specific pieces of information, this categorization helps to recognize the emerging themes or concepts.

##LancsBox version 6 (Brezina; Weill-Tessier; McEnery, 2021) was used to support the tabletop analysis and to explore the linguistic data and corpora from the interviews. The initial step involved extracting the ten most frequent words from the interview transcriptions. Each word was then thoroughly investigated within its context in the interviews. This analysis enabled the categorization of data based on word frequency, facilitating the identification and examination of relevant themes and concepts.

3.3.2 Lesson Plans

Participants' lesson plans represented the final assessment of the teacher training program. The submitted plans were assessed and analyzed using an adapted version of the Technology Integration Observation Instrument (TIOI) proposed by Harris, Grandgenett, and Hoferet (2010) (see Table 2). TIOI facilitated a comprehensive assessment of the integration of technological tools within the lesson plans, ensuring alignment with contemporary pedagogical methodologies and instructional objectives. The framework also helped to evaluate how well teachers understood the key concepts of DDL covered in the program. Each lesson plan was individually rated by two members of the research team, who met later on to discuss their final scores. The mean of these results is reported in this paper. The lesson plans

evaluation helped to answer the first research question regarding how teachers incorporated what they learned into their lesson plans and the degree to which they could apply the DDL approach in their classes. Table 2 describes the adapted TIOI used in the analysis.

Table 2 – Adapted from Harris, Grandgenett, and Hoferet (2010). Technology Integration Observation Instrument (TIOI)

Criteria	4	3	2	1
Lesson goals	Lesson goals are strongly aligned with the proposed activities.	Lesson goals are aligned with the proposed activities.	Lesson goals are partially aligned with the proposed activities.	Lesson goals are not aligned with the proposed activities.
Teaching context (if the lesson proposed is suitable for the instructional context)	The lesson plan is very suitable for the teaching context.	The lesson plan is suitable for the teaching context.	The lesson plan is partially suitable for the teaching context.	The lesson plan is not suitable for the teaching context.
DDL tool selected for the lesson plan (Tools selection)	Technology selection(s) are exemplary, given the lesson goals and instructional context.	Technology selection(s) are appropriate, given the lesson goals and instructional context	Technology selection(s) are marginally appropriate, given the lesson goals and instructional context	Technology selection(s) are inappropriate given the lesson goals and instructional context
Instructional strategies for DDL tool use (if the instructions for the use of the technology are clear)	Technology use optimally supports instructional strategies.	Technology use supports instructional strategies.	Technology use minimally supports instructional strategies.	Technology use does not support instructional strategies.
Students' role as detectives of the language (if a problem was presented to students and they need to use a tool to solve this problem)	The lesson plan strongly focuses on the student as a detective of the language.	The lesson plan focuses on the student as a detective of the language.	The lesson plan partially focuses on the student as a detective of the language.	The lesson plan does not focus on the student as a detective of the language.
Corpus skills (if the corpus skill chosen is developed in the lesson plan)	Corpus skills are exceptionally developed in the lesson plan.	Corpus skills are developed in the lesson plan.	Corpus skills are minimally developed in the lesson plan.	Corpus skills are not developed in the lesson plan.
Assessment	The assessment is strongly aligned with the lesson and the target corpus skills.	The assessment is aligned with the lesson and the target corpus skills.	The assessment is partially aligned with the lesson and the target corpus skills.	The assessment is not aligned with the lesson and the target corpus skills.
Fit (overall)	Content, instructional strategies, and technology fit together strongly within the lesson plan.	Content, instructional strategies, and technology fit together within the lesson plan.	Content, instructional strategies, and technology fit together somewhat within the lesson plan.	Content, instructional strategies, and technology do not fit together within the lesson plan.

Source: Adapted from Harris, Grandgenett e Hoferet, 2010 Technology Integration Observation Instrument (TIOI).

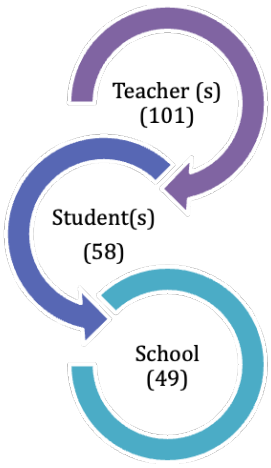
It is important to note that the original TIOI framework proposed by Harris, Grandgenett, and Hofer (2010) includes the dimensions of Curriculum Goals & Technology, Instructional Strategies & Technology, and Technology Selection and Fit. However, the analytical process within the instructional context of the study revealed the need to add further criteria not covered by the original framework. The analysis of the lessons is grounded in the content and activities presented within each lesson. The criteria used to demonstrate various levels of appropriateness were based on the guidelines used by Chung *et al*. (2024). Therefore, the framework was adapted, as shown in Table 2, by augmenting and refining certain criteria. Specifically, considerations such as specifying lesson goals, delineating the teaching context, identifying DDL tools used, articulating instructional strategies for DDL tool integration, elucidating students' roles as language detectives, emphasizing corpus skills development, and formulating an assessment framework tailored to the lesson plan's instructional objectives.

4 Results

4.1 Interviews

As explained in the methodology, #LancsBox was used to identify the most frequent content words, showing the most relevant themes discussed in the interviews, following Saldaña's (2016) approach. The most frequent words were: teacher/s (101), student/s (58), school (49), training (44), plan (44), lesson (40), tool (31), activity (25), and Versatext (24). We grouped these words in three figures, 1, 2 and 3, and presented some excerpts of the interviews to illustrate the teacher's opinions about the content they had seen in the course based on these words.

Figure 1 – Number of times each word was mentioned by the teachers who were interviewed



Source: the authors.

Regarding "Teachers", perceptions varied, with some expressing concerns over reduced proficiency levels post-pandemic, with one of them acknowledging a perceived decline in his own proficiency. As stated by Participant 3: "*Low knowledge, lower than before the pandemic changed the teacher's profile a lot eh I don't know*".

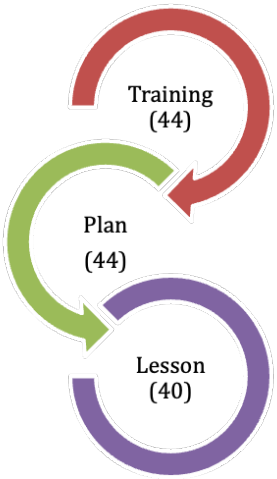
However, amidst these concerns, there was recognition of the critical importance of educators being adequately prepared and versatile. In the participants' view, teachers are expected to incorporate new tools and methodologies into their instructional approaches, indicating a commitment to ongoing professional development and adaptability.

When participants discuss the theme "Students", they mean active participants in the learning process, engaging in group activities and demonstrating critical thinking skills. The emphasis is on empowering students to think critically and apply their knowledge effectively. This is evidenced by students' ability to activate prior knowledge and engage in meaningful discussions to deepen their understanding of the material. See extract from Participant 3:

> *What is an adverb? Not that he has to know the nomenclature, but to understand why things are right. Why is the adverb there? Why does it have so much adverb? Because the author wanted it?*

Regarding the "School" system, concerns were raised about the level of preparedness and proficiency level among teachers. There was an acknowledgement of the need for ongoing support and training to ensure quality education delivery and professional development initiatives.

Figure 2 – Number of times each word was mentioned by the teachers who were interviewed



Source: the authors.

According to the participants, "training" emerges as a pivotal element in preparing teachers to integrate new tools and methodologies into their teaching practices adeptly. They emphasized the need to equip educators with the requisite skills and knowledge to enhance their instructional strategies and effectively support student learning. This finding indicates the significance of ongoing professional development initiatives in ensuring educators remain in tune with best practices and pedagogical advancements.

Participant 2 explains the broader goal of training, not only in language pedagogy, but also as an institutional objective. They also highlighted the importance of disseminating the knowledge to other teachers.

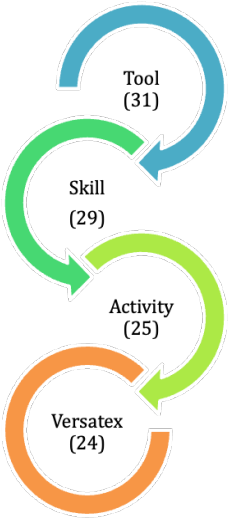
> *For example, this year the (educational) secretary is giving a program for the advanced level. It is giving a program of methodology and curriculum "in action" of the English language. Basic training curriculum in action for those who are in the pedagogical nucleus. And then I go to school, I promote this. I send it by email, I put it in the (WhatsApp) group of English teachers, I put it in the (WhatsApp) group of coordinators and I get to speak to the teachers, two, (or) three times.*

The term "lesson plan," as encountered in the interviews, refers specifically to the meticulously crafted outline that educators develop to structure their teaching activities within a particular class session. This encompasses detailed arrangements for teaching materials, learning objectives, activities, assessments, and instructional strategies tailored to a specific lesson or instructional unit. The primary emphasis lies in aligning the lesson plan with educational objectives, student requirements, and instructional methods to foster effective teaching and learning experiences. Challenges associated with devising comprehensive and engaging lesson plans are also deliberated upon, suggesting the significance of deliberate planning in facilitating student understanding and engagement.

In contrast, the term "plan" is employed more broadly within the educational context and professional development spheres, encompassing a spectrum of instructional design, curriculum development, and strategic decision-making aspects. It extends beyond individual lesson preparation to encompass overarching plans for school enhancement, curriculum coherence, or professional advancement initiatives.

Notably, this distinction may have arisen in the interviews because two of the three interviewees also held roles as coordinators. One of them had shown Versatext to her team and explained how the tool showed the relevant words within a text. However, she was not able to implement this tool in her school due to her lack of experience and confidence in dealing with it.

Figure 3 – Number of times each word was mentioned by the teachers who were interviewed



Source: the authors.

Regarding aspects of effective teaching practices, focusing on the integration of educational "tools", "skills" development, and engaging "activities", the participants unanimously recognize the value of incorporating innovative tools, such as Versatex, into their teaching practices to enhance vocabulary learning and foster student engagement. "Versatex" is lauded for its effectiveness in facilitating vocabulary acquisition and creating interactive learning environments. This highlights the importance of leveraging technology to cultivate dynamic classrooms and optimize student learning experiences.

Skills development emerges as a priority in educational planning and instruction. Teachers emphasize the importance of fostering critical thinking, effective communication, and real-world application skills among students. Rather than attempting to address all skills simultaneously, there is a consensus on the need to prioritize specific skills for development. Participant 1 expressed her concern about how to assess 30 students regarding the skills outlined in her lesson plan.

> [...] *to develop these skills of Concordance Lines, Reading Concordance lines* [...] Like [...]*how one thing connects to the other. I think what I meant with this form of evaluation that I used is that I care to see the participation of the student, right? Because, for example,* [...] *if the class has 30 students so they could do reading activity in groups we divide them into groups of five*.

Engaging activities are identified as essential components for promoting student participation and comprehension in the classroom. Teachers design activities that encourage collaboration, critical thinking, and the practical application of knowledge, thereby fostering dynamic and interactive learning environments. The integration of technology, such as Versatex, is viewed as a valuable resource for enhancing these activities and maximizing instructional effectiveness.

4.2 Lesson Plans

Table 3 shows the mean rating of the lesson plans developed by the participants, using the adapted TIOI.

Table 3 – Mean table of adapted TIOI rating for each lesson plan

Lessons	Lesson Goals	Teaching context (if the lesson proposed is suitable for the instructional context)	DDL tool selected for the lesson plan (Tools selection)	Instructional strategies for DDL tool use (if the instructions for the use of the technology are clear)	Students' role as detective of the language (if a problem was presented to students and they need to use a tool to solve this problem)	Corpus skills (if the corpus skill chosen is developed in the lesson plan)	Assessement	Fit (overall)	TIOI Adp. Level
Lesson 01	2	3	1	1	1,5	1	1,5	2	Fits somewhat
Lesson 02	1,5	1,5	2	2	2,5	1	1	1,5	Does not fit
Lesson 03	3,5	4	4	3,5	4	4	3	4	Fits together strongly
Lesson 04	3,5	3	3,5	2	3,5	3	3	3,5	Fits together
Lesson 05	1,5	2,5	2	1,5	2	1,5	1,5	1,5	Does not fit
Average	2,4	2,8	2,5	2	2,7	2,1	2	2,5	Fits somewhat

Source: the authors.

1. *Go to SKELL, write the word that you don't know on the search bar and click on "Examples". (Lesson 04).*
2. *The class watches the video of the song, then accesses the platform https://versatext.versatile.pub/. They click on input, paste the lyrics, then click on wordcloud. They select the grammatical classes: noun, adjective, and verb. (Lesson 05).*

Some activities were well-guided and could play an important role in encouraging a discussion of linguistic characteristics. For instance, in Example 3, the teacher's guidance helps students to observe collocational dynamics within language, thereby deepening their awareness of lexical conventions and their frequency of use:

3. [...] *teacher then promotes a discussion in which he explains that although some words are also used in the corpus, they may occur less frequently, which is an indicator of the conventionality of collocations in the language. In this way, the students can come to the conclusion that although it is possible to form different sentences, the frequency shows that some words / expressions / collocations are already conventionalized in the language and in the area in question (20 min).* (Lesson 05).

In that case, the instructions are categorized as "fit together", since it is clear how the teacher is guiding students to use DDL tools. Another example is seen in Lesson 4, in which the lesson goal seamlessly integrates with the planned activities, demonstrating a cohesive relation between the overall goal and the instructional strategies employed.

4. *General goal: To present strategies to identify keywords and cognates through data-driven activities.*
   Specific goals: i) Promote the gathering of students' prior knowledge and the triggering of schemas through pre-reading exercises; ii) Promote students' contact with tools adapted from Corpus Linguistics; and iii) Instigate language discovery (with a focus on vocabulary exercises) through data-driven activities. (Lesson 04).

Examples 5 and 6 demonstrate that some of the **evaluation procedures** focused on evaluating students' overall performance rather than measuring student's learning of lesson objectives.

5. *Students will be assessed on their participation, their research at home (of the lyrics of the chosen songs) and their subsequent oral presentation in class.* (Lesson 02).
6. *In general, students will be assessed on their productions, records, participation, and development of the activities carried out during the lesson.* (Lesson 03).

Regarding the **development of corpus skills**, although their importance was acknowledged in the lesson design, their acquisition was not consistently addressed. Notably, Lesson 03 provided a good example of it by explicitly stating the intended skill development and the corresponding instructional approach:

7. [...] *Using vocabulary related to the medical field in English in order to practice the skills of understanding and producing specialized collocations in the field. The following skills will be*

worked on: understanding collocations and keywords (SKILLS 8 and 14)[2] *and manipulating frequencies (SKILL 7).*[3] (Lesson 03).

Furthermore, one participant described his/her approach to developing corpus skills, which included using SKELL to improve vocabulary learning. In this exercise, collocations and modifiers were examined with a focus on how frequently they were used, with examples to back it up provided. The teacher explained it as follows:

8. I*n activity 03, SKELL is explored as a way to expand the vocabulary on the topic, using different collocations/modifiers. At this stage, the teacher points out that the words appear according to their frequency of use, as shown in the pictures of the examples (20 min)*. (Lesson 05).

The examples above highlight how DDL can be successfully incorporated into the classroom to help students build their corpus abilities and learn language autonomously by questioning information provided by corpus tools.

5 Discussion

This study aimed to discuss ways to present the DDL approach (Johns, 1991; Boulton; Vyatkina, 2021) as a way to complement school syllabuses that suggest the use of technology in English classes for Elementary and High School students. To do so, the BNCC's guidelines were analyzed to check where and how digital technologies were mentioned to be used by school teachers in their classrooms. One of BNCC's focused on the use of computational tools and technologies to support students' autonomy and protagonism in their learning. To meet that end, we chose some of the most user-friendly corpus tools to be taught to language teachers during our course so that they would integrate these technological components into their own teaching programs and school syllabuses.

After one year, three course participants were interviewed again to understand their opinions about the use of corpus tools and check if they had included these tools in their teaching. Regarding the first research question on how participants had incorporated what they learned during the program into their lesson plans and the extent to which they could apply the DDL approach in their classes, participants demonstrated high engagement during the initial program modules. They actively applied their knowledge to create lesson plans that integrated DDL tools, such as VersaText and Sketch Engine, and to show that they were able to align these tools with their lesson objectives effectively. However, as the interview showed, many faced challenges in applying these tools in their classrooms. While they could design lesson plans incorporating DDL methodologies, the daily integration of these tools into classroom teaching proved to be difficult to implement. There are still few studies on the use of DDL integrated into school syllabuses in Brazil, which include Dutra and Marques (2023) first attempt to encourage teachers to use #LancsBox in their classes and Marques' study (2024) on the use of hands-off concordance lines activities with children, aged 9 to 11

[2] Skill 8: understanding collocations. Skill 14: understanding keywords.

[3] Skill 7: handling frequencies.

years, showing positive results in the use of authentic language in the initial levels of the elementary school. However, it is important to say that Marques worked with one language teacher who was familiar with corpus linguistics, which was a crucial point for the successful result. The teachers who participated in the teacher training program discussed in this study course, by contrast, had little or no knowledge of corpus linguistics beforehand.

This information emphasizes the importance of ongoing assistance and professional development for teachers as they bridge the gap between academic knowledge and practical implementation in their teaching approaches. Addressing these challenges is crucial to enhancing participant learning outcomes and promoting the successful integration of DDL in language teaching practices among in-service teachers in Brazil.

Considering the second research question regarding participants' perceptions of the tools and skills they learned after being introduced to the DDL, participants' feedback was mostly positive. They recognized the value of incorporating innovative tools like VersaText into their teaching practices to enhance vocabulary learning and foster student engagement. Participants recalled appreciating the hands-on experience with various DDL tools during the program, which allowed them to apply their newly acquired skills, such as strategies for identifying keywords and understanding collocations, and how it offered practical ways to integrate critical thinking skills into their practice, a crucial skill highlighted in the BNCC that could receive greater emphasis the next time we offer a new DDL course.

The lesson plans were assessed as "somewhat fitting" in the adapted TIOI rating. Participants excelled in selecting appropriate tools aligned with their lesson goals but encountered significant challenges in practically applying them to their lesson plans. The analysis of the lesson plans produced during the course showed the diverse approaches taken by participants in integrating DDL tools into their instructional strategies. While some lesson plans demonstrated strong alignment with the program objectives and effective use of technological tools, others highlighted areas for improvement, e.g., Instructional strategies for DDL tool and assessment. Some of the reasons to be explored in the future have to do with the level of technological knowledge and teaching approaches used by these teachers. We believe that those who were more used to using online tools in class could more easily implement DDL activities in their lesson plans.

The interviews conducted later with three of the five participants revealed that two held leadership positions within their schools, likely contributing to their dedication and completion of the program. A notable aspect revealed by these interviews was the lack of practical application of the lesson plans in their classrooms. Despite their enthusiasm and theoretical understanding, participants struggled to integrate the corpus-based activities effectively into their teaching practices. This finding aligns with Chung *et al*. (2024), where participants also faced difficulties in implementing the designed lesson plans in actual classroom settings. Both studies, despite being conducted in different educational contexts, highlight common challenges, such as limited resources, insufficient support, and the complexity of corpus tools, which hinder the practical application of DDL. Similar teaching programs, such as that discussed in this study could certainly be supported by the use of specific teaching materials, such as those recently published by LeFoll (2021), Goulart and Veloso (2023), and Pinto *et al*. (2023), where teachers will find ready-to-use lesson plans and instructions to be used in their language classes.

6 Conclusions

The DDL program for Brazilian English teachers aimed to enhance their pedagogical practices with the integration of corpus-based tools and techniques. Throughout the program, participants engaged in hands-on learning experiences with various DDL tools, such as VersaText, Sketch Engine, SKELL, Linggle, VocabProfile, PlatCol, and TradCorpus, to develop their corpus skills.

Some limitations faced during the program should be considered, such as the noticeable decline in attendance, which could be attributed to the time of synchronous classes in the evening, after long workdays. The end of their school semester certainly influenced their participation as well. This suggests the need to offer shorter crash courses or alternate workshops throughout the semester, focusing on individual tools at a time. These challenges are common in Brazil's educational context and are part of the learning process, and the efforts of those who persevered should be acknowledged.

One year after completion of the program, interviews with participants revealed valuable reflections on their experiences, practical application of learned tools, implementation of developed lesson plans, and knowledge dissemination about DDL to their peers. These insights underscored the importance of providing teachers with the necessary skills and resources to effectively incorporate DDL methodologies into their teaching practices.

In conclusion, the use of specific DDL teaching materials, such as those provided by Goulart and Veloso (2023) and Pinto *et al*. (2023), could be tested in secondary school language classes in the Brazilian context, taking into consideration the BNCC's guidelines. Results would point out how young language learners in Brazil deal with digital technologies, such as DDL tools, and if language acquisition would be as positive as that presented by Marques (2024) with elementary school students. Finally, promoting a collaborative learning environment and equipping educators with practical skills, such as the corpus skills presented here, can contribute even more to the professional development of teachers and the advancement of DDL practices in the Brazilian educational context.

Funding

The authors acknowledge the support of the Global Strategy and Partnerships Seed Funding Project (University of Queensland), CAPES/Print Program, the Coordination for the Improvement of Higher Education Personnel (Process: 307287/2021-1) and Fapesp (Process no. 2022/05908-0).

Ethics Committee Report

The research was approved by São Paulo State University Ethics Committee on October 15th, 2024 under the number CAAE: 83225923.9.0000.5466.

References

BERBER SARDINHA, A.C. *Lingüística de corpus*. São Paulo: Editora Manole Ltda, 2004.

BINKLEY, M. *et al*. Defining Twenty-First Century Skills. *In*: GRIFFIN, P.; CARE, E. (eds.). *Assessment and Teaching of 21st Century Skills*. Springer: Dordrecht, 2012. p. 17–66.

BOULTON, A.; VYATKINA, N. Thirty Years of Data-Driven Learning: Taking Stock and Charting New Directions Over Time. *Language, Learning and Technology*, v. 25, n. 3, p. 66–89, 2021. DOI: https://doi.org/10.64152/10125/73450.

BRASIL. *Base nacional comum curricular*. Brasília: MEC, 2017.

BRASIL. *Constituição da república federativa do Brasil*. Brasília: Senado Federal, 1988.

BEHRENS, M. A. A aprendizagem por projetos num paradigma emergente. *In*: MORAN, J. M.; MASETTO, M.; BEHRENS, M. A. *Novas tecnologias e mediação*. 4. ed. Campinas: Papirus, 2001. p. 67-132.

BREZINA, V.; WEILL-TESSIER, P.; MCENERY, A. *#LancsBox x*. Lancaster: Lancaster University, 2021. v. 6.

BRITISH COUNCIL. *Demandas de aprendizagem de inglês no Brasil*. São Paulo: British Council, 2014. Available at: https://www.britishcouncil.org.br/sites/default/files/demandas_de_aprendizagempesquisacompleta.pdf. Accessed on: 5 Jul. 2022.

BRAUN, S. Integrating Corpus Work into Secondary Education: From Data-Driven Learning to Needs-Driven Corpora. *ReCALL*, v. 19, n. 3, p. 307-328, 2007. DOI: https://doi.org/10.1017/S0958344007000535.

CENTRO DE INOVAÇÃO PARA A EDUCAÇÃO BRASILEIRA. *Relatório guia edutec - diagnóstico do nível de adoção de tecnologia nas escolas públicas brasileiras em 2022*. São Paulo: CIEB, 2022. E-book.

CORPORA. *TRADCorpus*: Tradução e Corpora On-line para Formação de Competências Profissionais. São José do Rio Preto: UNESP, [2019?]. Available at: https://sites.google.com/unesp.br/tradcorpus/corpora. Accessed on: Jun. 2024.

COOB, T. *Compleat Lexical Tutor*: For data-driven Language Learning on the Web [computer program]. [S.l.: s.n], 2015. Available at: https://www.lextutor.ca/. Accessed on: Jun. 2024.

CROSTHWAITE, P. Data-Driven Learning and Younger Learners: Introduction to the Volume. *In*: CROSTHWAITE, P. (ed.). *Data-Driven Learning for the Next Generation*: Corpora and DDL for Pre-Tertiary Learners. London: Routledge, 2019. p. 1-10.

CROSTHWAITE, P. Taking DDL Online: Designing, Implementing and Evaluating a SPOC on Data-Driven Learning for Tertiary L2 Writing. *Australian Review of Applied Linguistics*, v. 43, n. 2, p. 169-195, 2020. DOI: https://doi.org/10.1075/aral.00031.cro.

CROSTHWAITE, P. (ed.). *Corpora for Language Learning*: Bridging the Research-Practice Divide. London: Taylor & Francis, 2024.

DA SILVA, J. M. N.; NUNES, V. G. C. Formação continuada docente: uma análise a partir da Lei de Diretrizes e Bases da Educação Nacional (LDB/1996) e das Diretrizes Curriculares Nacionais para a Formação Inicial e Continuada de Professores da Educação Básica (Resolução CNE-CP 2/2015). *Research, Society and Development*, v. 9, n. 8, p. e353985150-e353985150, 2020. DOI: https://doi.org/10.33448/rsd-v9i8.5150

DE LIMA, L.; LOUREIRO, R. C. A Aprendizagem Significativa do conceito de Tecnodocência: integração entre Docência e Tecnologias Digitais. *Revista Novas Tecnologias na Educação*, v. 14, n. 1, 2016. DOI: https://doi.org/10.22456/1679-1916.67335.

DOS SANTOS, L. V. *Atividades de leitura em língua inglesa direcionadas por dados*: o uso de um corpus com a temática dos ODS. 2024. Thesis (Master´s degree in Linguistics) – Faculdade de Letras e Ciências, Universidade Estadual Paulista, Araraquara, 2024.

DUTRA, D. P.; MARQUES, A. L. Getting Started with Digital Literacy Using Pedagogic Corpora in Young Learners' English Classes. *The ESPecialist*, [S. l.], v. 44, n. 2, p. 121–144, 2023. DOI: https://doi.org/10.23925/2318-7115.2023v44i2e61842.

GARCIA, W. D. *Fanfictions, linguística de corpus e aprendizagem direcionada por dados*: tarefas de produção escrita com foco no uso autêntico de língua e atividades que visam à autonomia dos alunos de letras em analisar preposições. 2020. Dissertation (Master's in Linguistics) – Instituto de Biociências, Letras e Ciências Exatas (IBILCE), Universidade Estadual Paulista, São José do Rio Preto, 2020.

GOULART, L.; VELOSO, I. *Corpora in English Language Teaching*: Classroom Activities for Teachers New to Corpus Linguistics. Open Educational Resource, Montclair State University, 2023. DOI: https://doi.org/10.5281/zenodo.7900420.

GOULART, L.; KOSTROMITINA, M.; KLEIN, J. From Corpus to Classroom: Evaluating Web-Based Tools to Teach Collocations. *In*: SARMENTO, S.; REBECHI, R.; MATTE, M. L. (org.). *English for Academic Purposes*. Porto Alegre: Editora UFRGS, 2024. p. 206-242.

GRUPO DE BANCO DE DADOS (GBD). *Multplatcol*: dicionário de colocações. São José do Rio Preto: UNESP, [2020?]. Available at: http://www.institucional.grupogbd.com/dicionario/sobre?locale=en. Accessed on: Jun. 2024.

HADDAD, S. A educação continuada e as políticas públicas no Brasil. *Revej@*:Revista de Educação de Jovens e Adultos. v. 1, p. 27-37, ago. 2007

HARRIS, J.; GRANDGENETT, N.; HOFER, M. Testing a TPACK-Based Technology Integration Assessment Rubric. In: SOCIETY FOR INFORMATION TECHNOLOGY & TEACHER EDUCATION INTERNATIONAL CONFERENCE, 2010, Association for the Advancement of Computing in Education (AACE). *Anais...* Chesapeake: AACE, 2010. p. 3833-3840.

HORTON JR, F. W. *Understanding Information Literacy*: A Primer. Edited by the Information Society Division. Paris: UNESCO, 2007.

JENKINS, J. English as a Lingua Franca from the Classroom to the Classroom. *ELT journal*, v. 66, n. 4, p. 486-494, 2012. DOI: https://doi.org/10.1093/elt/ccs040

JENKINS, J. *English as a Lingua Franca in the International University*: The Politics of Academic English Language Policy. London: Routledge, 2013.

JOHNS, T. Micro-Concord: A Language Learner's Research Tool. *System*, v. 14, n. 2, p. 151-162, 1986. DOI: https://doi.org/10.1016/0346-251X(86)90004-7.

JOHNS, T. Whence and Whither Classroom Concordancing? *In*: BONGAERTS, T. *et al*. (eds.). *Computer Applications in Language Learning*. Berlin: Foris, 1988.

JOHNS, T. F. Should You Be Persuaded: Two Samples of Data-Driven Learning Materials. *In*: JOHNS, T. F.; KING, P. (ed.). *ELR Journal*, v. 4, p. 1-16, 1991.

KENSKI, V. M. *Educação e tecnologias*: O novo ritmo da informação. 8. ed. Campinas: Papirus, 2012.

KOEHLER, M. J. *et al*. The Technological Pedagogical Content Knowledge Framework. *In*: SPECTOR, J.; MERRILL, M.; ELEN, J.; BISHOP, M. J. (ed.). *Handbook of Research on Educational Communications and Technology*. New York: Springer, 2014. p. 101-111.

LE FOLL, E. *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom*: A Step-By-Step Guide for (Trainee) Teachers Using Online Resources. Queensland: Pressbook, 2021.

LEITE, B. S. Tecnologias digitais e metodologias ativas: quais são conhecidas pelos professores e quais são possíveis na educação?. *VIDYA*, v. 41, n. 1, p. 185-202, 2021. DOI: https://doi.org/10.37781/vidya.v41i1.3773.

LINGGLE. *Linggle*: Diverse English learning platform. [*S.l.*: *s.n*], 2025. Available at: https://linggle.com/. Accessed on: Jun. 2024.

MA, Q. *et al*. Review of Corpus Tools for Vocabulary Teaching and Learning. *Journal of China Computer-Assisted Language Learning*, v. 1, n. 1, p. 177-190, 2021. DOI: https://doi.org/10.1515/jccall-2021-2008.

MA, Q. *et al*. Teachers' Perceived Corpus Literacy and Their Intention to Integrate Corpora into Classroom Teaching: A Survey Study. *ReCALL*, v. 35, n. 1, p. 19-39, 2023. DOI: https://doi.org/10.1017/S0958344022000180

MARQUES, A. L. S. P. *Science and Geography Lexicons in English for Learners in Ensino Fundamental 1*: A Corpus-Based Investigation. 2024. Dissertation (Ph.D. in Linguistics) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2024.

MATTAR, J. Aprendizagem em ambientes virtuais: teorias, conectivismo e MOOCs. *Teccogs*: Revista digital de tecnologias cognitivas, São Paulo, n. 7, p. 21-40, 2013.

MORAN, J. M. Mediação pedagógica: desafios que as tecnologias digitais nos trazem. *In*: MORAN, J. M. *Novas tecnologias e mediação pedagógica*. Campinas: Papirus, 2013. p. 30-35.

MORAN, J. M. A integração das tecnologias na educação. *Revista Brasileira de Informática na Educação*, v. 23, n. 1, p. 1-8, 2015.

MISHRA, P.; KOEHLER, M. J. Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, v. 108, n. 6, p. 1017-1054, 2006.

PÉREZ-PAREDES, P. *Corpus Linguistics for Education*: A Guide for Research. London: Routledge, 2020.

PINTO, P. T. *et al*. *Using Language Data to Learn About Language*: A Teachers' Guide to Classroom Corpus Use. Brisbane: The University of Queensland, 2023. DOI: https://doi.org/10.14264/3bbe92d.

PRETTO, N. Education and Technological Innovation: A Look at Brazilian Public Policies. *Revista Brasileira de Educação*, v.4, n. 11, p. 75-85, 1999.

SBROGIO, R. O. *Design e ensino-aprendizagem: entre slides e formação de professores*. 2021. Dissertation (Ph.D. in Media and Technology) – Faculdade de Arquitetura, Artes, Comunicação e Design, Universidade Estadual Paulista (Unesp), Bauru, 2021.

SALDAÑA, J. *The Coding Manual for Qualitative Researchers*. 3. ed. London: SAGE, 2016.

SKETCH ENGINE: *Create and search a text corpus*. [*S.l.*: *s.n*], [2015?]. Available at: https://www.sketchengine.eu/. Accessed on: Accessed on: Jun. 2024.

SKETCH ENGINE. *SKELL*: examples and collocations for learners of English. [*S.l.*: *s.n*], [2015?]. Available at: https://www.sketchengine.eu/skell/. Accessed on: Accessed on: Jun. 2024.

THOMAS, J.; BAISA, V. V*ersaText Information and Resources*. Versatile.pub, 2018. Available at: https://www.versatile.pub/versatext-info.html. Accessed on: 29 Jun. 2024.

UNESCO. Executive Board: Hundred and ninety-fourth Session. *Language teaching in education systems*, Paris, 31 March 2014. (194 EX/29. Item 29 of the provisional agenda).

YOON, H.; HIRVELA, A. ESL Student Attitudes Toward Corpus Use in L2 Writing. *Journal of second language writing*, v. 13, n. 4, p. 257-283, 2004. DOI: https://doi.org/10.1016/j.jslw.2004.06.002.

History

Date of submission: 30/06/2024.
Date of approval: 14/08/2025.

Responsible Editor

Andréa Machado de Almeida Mattos, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais/MG, Brasil. Lattes: http://lattes.cnpq.br/7749222257907067, ORCID: https://orcid.org/0000-0003-3190-7329, e-mail: andreamattos@ufmg.br.

Authors' Contribution

Paula Tavares Pinto focused on planning and pedagogical analysis, linking the training program to BNCC guidelines. Adriane Orenha-Ottaiano and Talita Serpa contributed with methodological grounding in Corpus Linguistics and the adaptation of DDL tools to the Brazilian context. Franciele Spinelli and Peter Crosthwaite, from the University of Queensland, provided international support, bringing insights from recent research on DDL and its application in teacher education. Carolina Tavares de Carvalho developed the empirical data analysis, particularly in evaluating lesson plans and interviews, highlighting the challenges faced by teachers and the need for continuous support in implementing corpus-based practices.

Declaration of Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability Statement

The data is not public but it may be available upon request. Please contact the authors.

Use of AI

AI was used to review the English language in some paragraphs of this article.

Reviews

As part of the commitment made by the Brazilian Journal of Applied Linguistics to Open Science, the journal publishes the reviews issued regarding its published works, when authorized by all parties involved.

Appendix A - Santos (2024)

PLANO DE AULA	
Atividade	**Língua**
An introduction to Keywords and Cognates	Inglesa
Contexto educacional	**Nível/série**
Estratégias de leitura em língua inglesa	Nível básico (A1 e A2)

Modalidade	Aulas estimadas	Duração da aula
On-line	1	1h30min

Habilidade de Corpus
6. read concordance lines; 14. understanding Keywords.

Objetivo(s) e/ou habilidade(s)
Objetivo geral: Apresentar as estratégias de identificação de palavras-chave e cognatos por meio de atividades direcionadas por dados. Objetivos específicos: i) Promover o levantamento de conhecimento prévio dos alunos e acionamento de esquemas por meio de exercícios de pré-leitura; ii) Promover o contato dos alunos com ferramentas adaptadas de Linguística de Corpus; iii) Instigar o descobrimento da língua (com foco para exercícios de vocabulário) por meio de atividades direcionadas por dados.

Conteúdo/Objeto(s) do conhecimento

Estratégias de leitura:
Keywords;
Cognates.

Materiais/ferramentas

Apresentação de Slides (Powerpoint);
Ferramenta de videoconferência (Google Meet);
Ferramentas de Linguística de Corpus para o ensino (VersaText e Skell);
Artigo Climate change mitigation opportunities based on carbon footprint estimates of dietary patterns in Peru (Vázques-Rowe; Larrea-Gallegos; Gilardino, 2017).

Desenvolvimento

Atividade de Compreensão Escrita direcionada por dados para o momento de pré-leitura
Introdução às Estratégias de leitura – palavras-chave e cognatos
Tipo: Hands on e hands off
Ferramentas utilizadas: VersaText e Skell

Reading Strategies – Class 1: An introduction to Keywords and Cognates

A professora inicia a aula apresentando em um slide a seguinte nuvem de palavras:

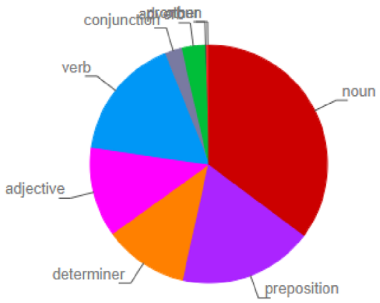


Em seguida, lança para a turma as perguntas a seguir:

What do you see in the picture?
Do you know these words?
What are the words familiar to you?

Então, propõe que os alunos analisem o gráfico, contido nos slides, gerado pelo Profiler da ferramenta Versatext, para analisarem a classe morfológica das palavras da nuvem.

Take a look at the graph below and find out the morphological class of the words.



Depois, a professora apresenta nos slides os exercícios hands on abaixo. Sugere-se que sejam dados alguns minutos para que os alunos tentem responder. Esse é o momento em que os alunos acessam a ferramenta Skell de seus computadores.

a) Go to Skell, write the word that you don't know on the search bar and click on "Examples". Write down and say two sentences that helped you understand the meaning of the word.

_____

_____.

b) Can you create two sentences using some of the words from the wordcloud? *Examples: GHG emissions cause impacts to health. /Food production needs policies to avoid environmental impacts.*

Depois de conferir e discutir as respostas dos alunos, a professora informa que lerão juntos a introdução de um artigo científico e propõe o seguinte exercício:

c) If the title of its paper were one of the sentences you have created, what could be the probable topic?

_____

_____.

Posteriormente, apresenta-se o título do artigo, Climate change mitigation opportunities based on carbon footprint estimates of dietary patterns in Peru, e faz-se as seguintes perguntas:

What scientific area does this paper belong to?
What kind of researcher could have written it?
By looking at the title, what is the text about?
Are there any words that helps you infer this information? Write them down.

Depois de discutir as respostas com os alunos, a professora pergunta se eles entendem o título do artigo e se sabem dizer o que são palavras-chave. Nesse momento, explicita-se que os exercícios haviam direcionado os alunos para que se atentassem para a identificação de palavras-chave, palavras recorrentes que sumarizam as ideias de um texto.