

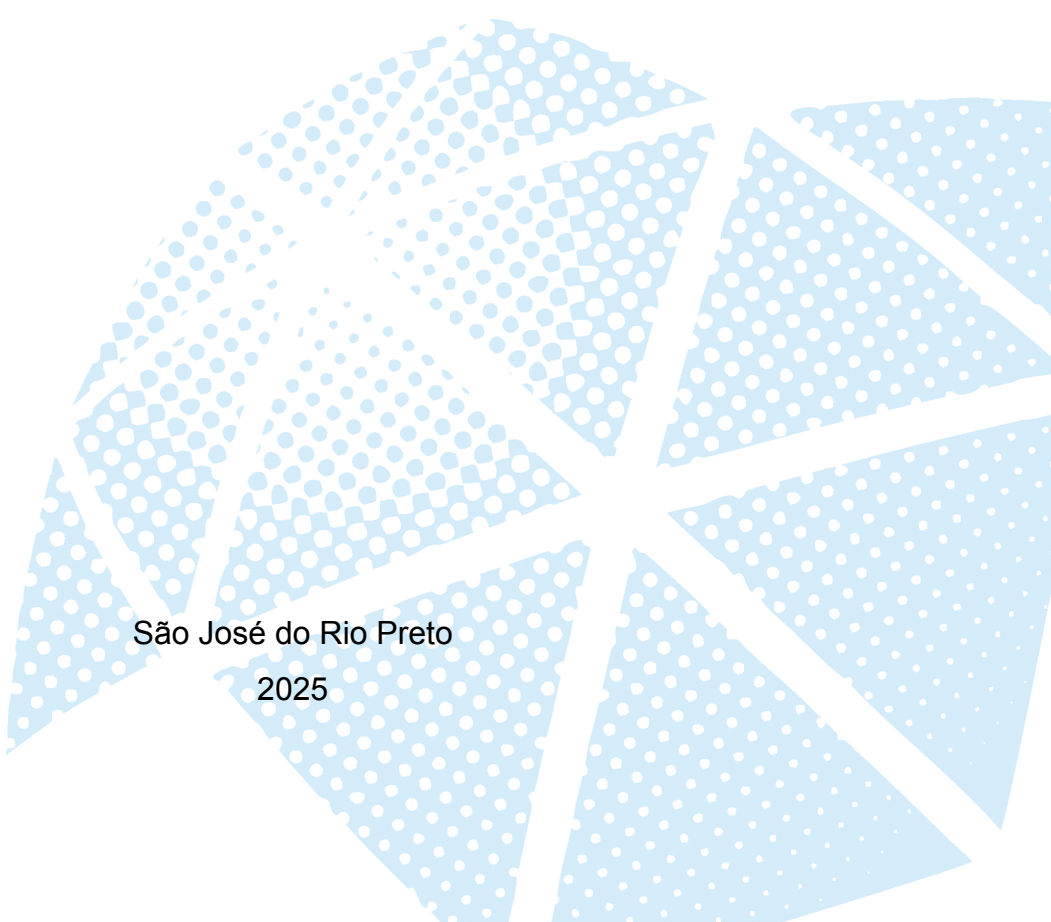
UNIVERSIDADE ESTADUAL PAULISTA - UNESP
Departamento de Letras Modernas - Campus de São José do Rio Preto - Ibilce

Carolina Tavares de Carvalho

**CORPUS SKILLS DEVELOPMENT IN DDL TRAINING FOR (FUTURE)
TEACHERS**
a Brazilian and Vietnamese parallel

São José do Rio Preto

2025



Carolina Tavares de Carvalho

**CORPUS SKILLS DEVELOPMENT IN DDL TRAINING FOR (FUTURE)
TEACHERS**

a Brazilian and Vietnamese parallel

Doctoral Thesis presented to Universidade Estadual Paulista (UNESP), Departamento de Letras Modernas, São José do Rio Preto - Ibilce, to obtain the academic degree Ph.D. in Linguistics Studies.

Advisor: Prof^a. Dr^a. Paula Tavares Pinto.

São José do Rio Preto

2025

C331c Carvalho, Carolina Tavares de
Corpus skills development in DDL training for (future)
teachers : a Brazilian and Vietnamese parallel / Carolina
Tavares de Carvalho. -- São José do Rio Preto, 2025
130 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (UNESP),
Instituto de Biociências Letras e Ciências Exatas, São José do
Rio Preto
Orientadora: Paula Tavares Pinto

1. Linguística de Corpus. 2. Data-Driven Learning. 3.
Educational Technology. 4. Corpus Skills. 5. Teacher
Education. I. Título.

IMPACTO POTENCIAL DESTA PESQUISA

Esta pesquisa fortalece a formação de professores de inglês no Brasil e no Vietnã ao desenvolver letramento de corpus para DDL, aprimorar a análise de planos de aula (TIOI adaptado) e propor diretrizes e ferramentas alinhadas às realidades locais, como a Skill 19 e o DDL Checker, com foco em equidade e sustentabilidade.

POTENTIAL IMPACT OF THIS RESEARCH

This research advances English teacher education in Brazil and Vietnam by building corpus literacy for DDL, refining a lesson-plan analysis framework (adapted from TIOI), and offering context-sensitive supports, such as Skill 19 and an AI-assisted DDL Checker, to promote equity and sustainable, practical uptake.

Carolina Tavares de Carvalho

CORPUS SKILLS DEVELOPMENT IN DDL TRAINING FOR (FUTURE)

TEACHERS

a Brazilian and Vietnamese parallel

Doctoral Thesis presented for obtaining the title of Linguistics Studies Ph.D. in the Postgraduate Program in Linguistics Studies of São Paulo State University "Júlio de Mesquita Filho", São José do Rio Preto's Campus.

Scholarship: CAPES: 88887.923810/2023-00 / 88887.716716/2022-00

Advisor: Prof^a. Dr^a. Paula Tavares Pinto

Data da defesa: 04/12/2025

Comissão Examinadora:

Prof^a. Dr^a. Paula Tavares Pinto

UNESP - Departamento de Letras Modernas - Campus de São José do Rio Preto - Ibilce

Prof. Dr. Eduardo Batista da Silva

UEG - Universidade Estadual de Goiás - Campus Sudeste

Prof^a. Dr^a. Emiliana Fernandes Bonalumi

UFR - Instituto de Ciências Humanas e Sociais / Universidade Federal de Rondonópolis

Prof^a. Dr^a. Sandra Mari Kaneko Marques

UNESP - Departamento de Letras Modernas - Campus de Araraquara - FCLAr

Prof^a. Dr^a. Talita Serpa

UNESP - Departamento de Estudos Linguísticos e Literários - Campus São José do Rio Preto - Ibilce

São José do Rio Preto
2025

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to the University of Queensland for welcoming me as a sandwich PhD student, and to Dr Peter Crosthwaite, whose support and intellectual generosity were central to this experience. A very special thank you to Fran Spinelli, who received me in Brisbane with open arms and helped me navigate every challenge—academic and personal alike. Truly, without Fran, many things could have gone awry. My gratitude also extends to everyone at the University of Queensland and across Brisbane, especially the Brazilian community, who contributed to an environment of kindness, rigour, and collaboration.

I am sincerely grateful to the qualification committee, to CAPES for the scholarship, and to PPGE – the Graduate Program in Linguistic Studies at IBILCE/UNESP for their unfailing support throughout this journey. My most profound appreciation goes to my mother, my family, my wife, and our pets, whose unwavering love and companionship have sustained and strengthened me.

To my research group, thank you for the stimulating discussions, generous collaborations, and steady encouragement that enriched every stage of this work.

A very special and heartfelt thank you to my advisor, Prof. Dr Paula Tavares Pinto. Without her, none of this would have been possible. Her guidance, patience, and profound expertise shaped each step of this research. Beyond academic brilliance, Paula's encouragement, kindness, and steadfast support were a constant source of motivation and inspiration. She has been more than a mentor; she has been a role model in both professional and personal spheres, showing me how to be a better researcher and how to navigate the challenges of academia with integrity and resilience. I am incredibly fortunate to have had her as my advisor.

To IBILCE/UNESP, thank you for providing an academic home and an environment conducive to research and learning.

Finally, I am deeply grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) for funding this research (Grant Codes 88887.923810/2023-00 and 88887.716716/2022-00).

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“Everything that has a beginning has an end”.
(The Matrix Revolutions, 2003)

ABSTRACT

Teaching English in developing countries requires innovative pedagogical approaches and meaningful engagement with digital resources in an age of globalized communication. Data-Driven Learning (DDL) — the incorporation of corpora and corpus-based tools into classroom practices — has emerged as a promising method to enhance language instruction worldwide. However, few studies have explored how in-service teachers in under-resourced environments develop corpus literacy and integrate DDL into their daily teaching routines. This thesis addresses that gap by comparing two distinct case studies: (i) a professional development course for public school English teachers in Brazil and (ii) a bilingual DDL teacher education program for teachers in Vietnam. The overarching aim of this thesis is to support English language teaching in developing countries by helping teachers work with DDL through the development of corpus skills and the adaptation of the approach to diverse educational contexts. More specifically, the study investigates how these educators acquire corpus skills, translate them into lesson activities, and respond to challenges related to infrastructure and workload. Using interviews, lesson-plan analyses, self-assessment surveys, and a mixed-methods parallel case study design, this research captures teachers' perspectives and classroom adaptations. In evaluating the lesson plans, the study employs and adapts the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) to assess the extent and quality of DDL integration, and proposes a refined framework for systematically analysing corpus-based pedagogical interventions. The findings reveal strong teacher enthusiasm but a notable gap in actual classroom implementation, primarily due to limited digital infrastructure, heavy workloads, and insufficient pedagogical scaffolding. To confront these challenges, the study introduces an AI-assisted DDL Checker tool. It proposes "Skill 19," a new corpus skill added to Peres-Paredes's (2020) proposal, designed to guide teachers in creating targeted prompts. These enhancements bolster corpus literacy and foster effective use of DDL resources. The implications highlight the necessity for ongoing mentorship, institutional support, and context-sensitive professional development to ensure sustainable DDL integration in resource-limited settings. Additionally, the enhanced TIOI (Harris; Grandgenett; Hofer, 2010) framework provides a more

comprehensive method for future research to assess lesson plans, facilitating better investigation and effective implementation of data-driven approaches.

Keywords: data-driven learning; corpus tools; teacher education; TIOI; educational technology; corpus skills.

RESUMO

Em uma era de comunicação globalizada, o ensino de inglês em países em contextos em desenvolvimento exige não apenas abordagens pedagógicas inovadoras, mas também um envolvimento significativo com recursos digitais. A Aprendizagem Baseada em Dados (Data-Driven Learning, DDL) — que envolve a incorporação de corpora e ferramentas baseadas em corpus às práticas de sala de aula — vem se destacando como um método promissor para aprimorar o ensino de línguas em escala mundial. No entanto, poucos estudos têm investigado como professores em serviço, atuando em ambientes com recursos limitados, desenvolvem habilidades relacionadas a corpora e, subsequentemente, integram o DDL em suas rotinas docentes. Este trabalho busca suprir essa lacuna ao comparar dois estudos de caso distintos: (i) um curso de desenvolvimento profissional para professores de inglês da rede pública no Brasil e (ii) um programa de treinamento bilíngue em DDL para professores no Vietnã. O objetivo geral desta tese é apoiar o ensino de inglês em países em desenvolvimento, auxiliando professores a trabalhar com DDL por meio do desenvolvimento de habilidades de corpus e da adaptação da abordagem a diferentes contextos educacionais. O objetivo é investigar de que forma esses educadores adquirem habilidades de uso de corpus, transformam essas habilidades em atividades de aula e lidam com desafios ligados à infraestrutura e às demandas de trabalho. Por meio de entrevistas, análises de planos de aula, questionários de autoavaliação e um delineamento de estudo de caso paralelo com métodos mistos, esta pesquisa documenta as perspectivas e adaptações dos professores em sala. Na avaliação dos planos de aula, o trabalho emprega — e adapta — o Instrumento de Observação de Integração Tecnológica (TIOI) (Harris; Grandgenett; Hofer, 2010) para aferir a extensão e a qualidade da integração do DDL, propondo uma estrutura refinada para analisar sistematicamente intervenções pedagógicas baseadas em corpus. Os resultados revelam entusiasmo por parte dos professores, mas um déficit considerável na implementação prática em sala de aula, em grande parte devido à infraestrutura digital restrita, à sobrecarga de trabalho e à falta de suporte pedagógico adequado. Para enfrentar esses obstáculos, o estudo apresenta uma ferramenta de DDL assistida por IA (DDL Checker) e propõe a “Habilidade 19”, além das 18 propostas por Pérez-Paredes (2020), um novo conjunto de competências em corpus projetado para orientar os professores na criação de

prompts mais direcionados. Essas inovações reforçam a alfabetização em corpus e promovem o uso efetivo de recursos de DDL. As implicações ressaltam a necessidade de cursos de atualização, apoio institucional e formação profissional adaptada ao contexto, garantindo a sustentabilidade da integração de DDL em cenários com recursos limitados. Além disso, a versão aprimorada do TIOI (Harris; Grandgenett; Hofer, 2010) oferece um método mais abrangente para futuras pesquisas avaliarem planos de aula, facilitando investigações mais consistentes e a implementação eficaz de abordagens direcionadas por dados.

Palavras-chave: aprendizagem baseada em dados; ferramentas de corpus; formação de professores; TIOI; tecnologia educacional; habilidades de corpus.

LIST OF FIGURES

Figure 1 – Comparative overview of English teacher education and DDL implementation in Brazil and Vietnam	41
Figure 2 – Brazil × Vietnam data-collection flowchart	43
Figure 3 – Methodological Timeline	49
Figure 4 – CBLP training procedure	53
Figure 5 – Methodological sequence adopted in the Vietnamese course, outlining the main steps followed from initial steps to course implementation and evaluation	54
Figure 6 – Corpora post-course perceptions (Q01–Q17)	64
Figure 7 – Course evaluation (Q01–Q12)	65
Figure 8 – Most interesting Modules	66
Figure 9 – Most liked Tools	66
Figure 10 – Frequency of top content words in interviews 1 (Teacher, Student and School)	68
Figure 11 – Frequency of top content words in interviews 2 (Training, Plan and Lesson)	69
Figure 12 – Frequency of top content words in interviews 3 (Tool, Skill, Activity, and Versatext)	71
Figure 13 – Trainees' initial perspectives (Map part I)	75
Figure 14 – Trainees' initial perspectives (Map part II).	77
Figura 15 – Trainees' understanding of corpora	79
Figure 16 – Trainees' corpus search abilities	80
Figure 17 – Trainees' Analyzing concordance lines	80
Figure 18 – Trainees' perceptions of the advantages of corpora	81
Figure 19 – Trainees' perceptions of the advantages of corpora	82
Figure 20 – Trainees' future intention to implement CBLP	82
Figure 21 – Trainees' perspectives after the course (Part A)	84
Figure 22 – Trainees' perspectives after the course (Part B)	87

Figure 1 – Comparative overview of English teacher education and DDL implementation in Brazil and Vietnam	41
Figure 23 – DDL Checker interface	102
Figure 24 – Example of interaction with DDL Checker	102
Figure 25 – DDL Checker explaining the concept of absolute and relative frequency	103
Figure 26 – DDL Checker suggesting how to approach simple past in the classroom using DDL	104
Figure 27 – A seven-step procedure for developing any of the 18 corpus skills	108

LIST OF TABLES

Table 1 – Corpus Skills	29
Table 2 – Program Schedule and Plan	46
Table 3 – Technology Integration Observation Instrument (TIOI)	55
Table 4 – Adapted TIOI	57
Table 5 – Mean table of adapted TIOI rating for each lesson plan	73
Table 6 – Number and % of corpus activities per lesson plan component (n=19)	88
Table 7 – Mean TIOI ratings for each submitted lesson plan	93
Table 8 – Brazil and Vietnam TIOI results	99

SUMMARY

1	INTRODUCTION	15
1.1	LANGUAGE: THE EVER-EXPANDING FRONTIER	15
1.2	SANDWICH PhD (CAPES-PRINT): BRIDGING BRAZIL, AUSTRALIA, AND VIETNAM	18
1.3	CONTEXT AND PROBLEM STATEMENT	20
1.4	RESEARCH QUESTIONS AND OBJECTIVES	24
1.5	RELEVANCE AND CONTRIBUTION OF THE STUDY	25
1.6	THESIS STRUCTURE (ROADMAP)	26
2	LITERATURE REVIEW	28
2.1	THEORETICAL BACKGROUND	28
2.2	PREVIOUS STUDIES ON DATA-DRIVEN LEARNING AND CORPUS-BASED PEDAGOGY	31
2.3	RESEARCH GAPS	32
2.4	BRAZIL AND VIETNAM'S HISTORICAL CONTEXT	34
2.4.1	Teacher Education in Brazil: Policy Ambitions, Institutional Diversity, and Pedagogical Innovation	34
2.5	TEACHER EDUCATION IN VIETNAM: HISTORICAL LEGACIES, INSTITUTIONAL STRUCTURES, AND PEDAGOGICAL TENSIONS	36
2.6	COMPARATIVE REFLECTIONS: VIETNAM AND BRAZIL IN DIALOGUE WITH DDL	39
3	RESEARCH METHODS	42
3.1	OVERALL APPROACH	42
3.2	DATA COLLECTION	43
3.2.1	Instruments and Procedures	44
3.2.2	Qualitative Analysis	44
3.3	CASE 1 (BRAZIL) - A DDL PROGRAM FOR IN-SERVICE TEACHER	45
3.3.1	Methodological sequence used in Brazil	45
3.3.1.1	<i>Participants</i>	49
3.3.2	Instruments for data collection	49
3.3.3	CASE 2 (VIETNAM): TEACHER EDUCATION MODEL IN VIETNAM	50
3.3.4	METHODOLOGICAL SEQUENCE USED IN VIETNAM	51
3.3.5	Data Collection Methods	54
3.4	DATA ANALYSIS	56
3.4.1	Case 1 (Brazil)	56
3.4.2	Lesson plans	57
3.4.3	Case 2 (Vietnam)	59
3.5	ETHICAL CONSIDERATIONS	60
4	RESULTS	62
4.1	RESULTS FROM THE BRAZILIAN STUDY	63
4.1.1	Interviews	67

4.1.2	Lesson Plans	72
4.2	RESULTS FROM THE VIETNAMESE STUDY	75
4.2.1	Interviews	75
4.2.2	Post-course survey	78
4.2.3	Lesson Plan	87
5	DISCUSSION	98
5.1	COMPARISON OF KEY FINDINGS	98
5.2	ADDRESSING THE GAPS: AI AND NEW CORPUS SKILLS	100
5.3	IMPLICATIONS FOR THEORY AND PRACTICE	106
6	CONCLUSIONS AND FUTURE WORK	115
	REFERENCES	118
	APPENDIX A - THE ETHICS COMMITTEE'S APPROVAL (PLATAFORMA BRASIL)	
	APPENDIX B - REPOSITORY LINK	
	DADOS CURRICULARES	

1 INTRODUCTION

This chapter introduces the study's context and rationale and outlines how the thesis is anchored in the lived realities of English language teaching in developing contexts. It begins by framing language learning as an ever-expanding frontier (Section 1.1) and then situates the research trajectory through the sandwich PhD experience that bridged Brazil, Australia, and Vietnam (Section 1.2). Next, it presents the context and problem statement (Section 1.3), followed by the research questions and objectives (Section 1.4), and the study's relevance and contribution (Section 1.5). The chapter closes with a roadmap of the thesis structure (Section 1.6).

1.1 LANGUAGE: THE EVER-EXPANDING FRONTIER

Language is not a fixed code but a living, evolving frontier, shaped by culture, thought, and interaction. In classrooms across the world, especially in resource-limited contexts, language teachers often navigate complexity with limited tools. Yet within these challenges lies immense potential. This research is guided by the belief that Corpus Linguistics (CL) and Data-Driven Learning (DDL) can offer teachers and learners new ways of engaging with language, ones that are more autonomous, evidence-based, and context-sensitive. It is an effort to make corpus-based approaches more accessible, to understand what skills educators need, and to explore how DDL can be meaningfully integrated into diverse teaching realities.

My postgraduate journey began when I faced challenges with a group of B1-level learners who struggled with listening comprehension. Their skills did not match the expectations set by the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). Despite their engagement, a mismatch existed between the syllabus expectations and their actual language abilities.

As I tried to help them, I eventually stumbled upon Corpus Linguistics (CL), which led me to choose it as the focus of my Master's research. What began as a methodological choice soon became a turning point in how I understood language teaching. Corpora are not just repositories of linguistic data but representations of

living language: they make visible authentic patterns of use, the voices of real speakers, and the subtle yet powerful ways in which language reflects social reality. By engaging with these patterns, language teaching can not only increase awareness and accuracy but also promote inclusion, acceptance, and a stronger sense of belonging.

Language is conceived not only as a system of signs or a communicative tool, but as an archive of shared cognitive and cultural patterns that can be accessed through empirical observation. Language connects us to other people and cultures, serving as a pathway to empathy, understanding, and personal growth. It is how we access information, share our experiences, and discover new ways of expressing ourselves. Drawing on a light interpretation of the Sapir-Whorf hypothesis (Sapir, 1921; Whorf, 1956), the languages we speak shape how we experience the world, what we can name, describe, and even feel. Learning a new language, therefore, is not just about acquiring vocabulary or grammar, but about expanding the boundaries of one's world. It offers various ways to experience the world (UNESCO, 2014).

This perspective aligns with CL, which, as Berber Sardinha (2004) notes, is the branch of Applied Linguistics devoted to collecting and analysing large, digitally available corpora of spoken and written texts. By taking an explicitly empirical stance, CL lets us examine how often words occur, which words typically precede or follow them, and the probability that they appear in specific contexts (Berber Sardinha, Delfino, and Rampaso, 2017). What emerges from these analyses is the realization that linguistic choices are anything but random; they reveal stable, community-wide preferences.

This observation aligns with Renouf and Sinclair's (1991) idiom principle, which argues that speakers select lexical items based on pre-established, phraseological patterns rather than relying solely on abstract grammatical rules. In this view, language users draw from a repertoire of familiar combinations, some of which are so deeply embedded that they feel almost automatic.

The CL concept can be linked with Jungian archetypes. Jung (2016) described archetypes as primordial behavioural templates anchored in the collective unconscious. Likewise, CL uncovers what might be termed discursive archetypes, recurrent linguistic patterns that structure how speakers habitually think, categorise, and act. Corpora can make visible culturally entrenched linguistic unconscious: combinations and collocations that guide expression beneath the level of conscious

choice. By exposing these patterned regularities, CL does more than chart frequencies; it reveals a cognitive and cultural scaffolding that shapes everyday language use.

Corpus-based and corpus-driven approaches revealed a more profound insight: the potential of corpora to promote learner autonomy. When students are encouraged to explore patterns, test hypotheses, and derive rules from real data, they not only expand their language awareness but also develop a more critical and reflective attitude toward learning. Over time, however, it became clear that this potential was often overlooked in classrooms, particularly in developing countries like ours, which must constantly strive to compete with more technologically and economically advanced nations in an increasingly globalised world.

This underuses did not stem from a lack of interest but from a complex web of constraints that made corpus-based methodologies appear distant or inaccessible, including perceived technical complexity, cultural and pedagogical barriers, entrenched teaching beliefs, rigid curriculum demands, bureaucratic pressures, and limited infrastructure.

Within this context, DDL emerged as a particularly relevant approach. First proposed by Tim Johns (1986), DDL encourages a shift in the teacher's role from transmitter of content to facilitator of inquiry. Rather than providing ready-made answers, teachers are invited to supply learners with linguistic evidence that supports independent exploration. DDL proposes that students engage with authentic language data and take responsibility for drawing their conclusions. As Berber Sardinha (2004) points out, DDL develops learner autonomy by positioning the student as a language researcher and promoting an inductive learning process grounded in real use. One of its most recognisable practices is the use of concordance lines, which allow learners to observe patterns of meaning and form in context, cultivating both analytical thinking and linguistic awareness.

After completing my Master's, I turned my attention more directly to the challenges teachers face when adopting DDL in real classrooms. Many had never heard of concordancers, and those who had often felt overwhelmed by statistical outputs or by interfaces that did not seem designed with teachers in mind. Rather than a lack of pedagogical interest, there was a sense that these tools belonged to a highly specialised, academic domain. I wanted to help demystify them and make them more accessible to educators. This context prompted me to reflect on several

key issues: What kinds of skills do teachers need to implement DDL effectively in everyday practice? How can its integration be made smoother and less intimidating? And how can corpus tools and methodologies be adapted to culturally and materially diverse contexts, such as Brazil and Vietnam?

These issues are the starting point of my doctoral research. My perspective on language pedagogy has evolved from viewing corpora as powerful linguistic resources to understanding them as pedagogical instruments that, when humanized, localized, and simplified, can make a difference where it matters most: in everyday teaching practice. To that end, this thesis explores not just the use of corpora but the development of corpus literacy, the ability to engage with, interpret, and creatively apply corpus data for pedagogical purposes.

LC matters not just for what it reveals about language, but for what it enables in education. It is a lens through which language becomes visible, and a bridge between research and practice, especially in settings where resources are scarce, but the desire to teach and learn is abundant. This exploration of language as an ever-expanding frontier (personal, pedagogical, and empirical) continues to guide the trajectory of this research and the mission it seeks to fulfil.

1.2 SANDWICH PhD (CAPES-PRINT): BRIDGING BRAZIL, AUSTRALIA, AND VIETNAM

Undertaking the sandwich PhD at the University of Queensland (UQ) through CAPES PrInt was more than a change of postal code; it connected three educational ecosystems and made this comparative project feasible in practice. By funding targeted international collaboration, CAPES PrInt created the conditions for an intercultural research itinerary in which a training model first developed in Brazil was refined in Australia and later adapted with a Vietnamese partner. In practice, the programme turned institutional plans into day-to-day cooperation, allowing research questions to be tested, adjusted, and brought back into sharper focus.

An essential element in this trajectory was the mentorship and network provided by Prof. Peter Crosthwaite, together with my supervisor, Prof. Paula Tavares Pinto. Their prior collaborations with Brazilian and Vietnamese teams, although previously separate, positioned them as a natural bridge, aligning people, tools, and timelines and opening the door for my placement at the University of Queensland

(UQ). The UQ is a private institution, whereas UNESP is public, and the partnership was grounded in work exchange, co-development of materials, and mutual academic enrichment rather than fee-for-service arrangements. This allowed the collaboration to focus on substance, corpus-based pedagogy, teacher support, and pragmatic integration of DDL into real classrooms.

Chronologically, the Brazilian course came first and functioned as a living prototype. It allowed us to consolidate aims, instruments, and rubrics (for example, corpus-skills emphasis and lesson-plan production and observation) and to document what teachers actually did when DDL entered their weekly routines. The sojourn in Australia then operated as a research-and-development phase. Access to the UQ resources and sustained discussions with the team supported targeted adjustments, tightening the training sequence, clarifying task scaffolds, and testing materials for transferability. By the time we reached the Vietnamese course, the package was not simply transplanted but recontextualised. We preserved the pedagogical core while adapting delivery and support to local routines, including bilingual touchpoints, more explicit step-by-step modelling, and closer alignment between workshop input and lesson-plan output.

This sequencing helps explain the methodological asymmetries across the two courses. The Brazilian iteration prioritized teacher creativity and local adaptation, consistent with a culture of pedagogical autonomy. That invited a broader exploratory bandwidth, but it also risked dispersion and uneven technical depth. The Vietnamese implementation, informed by the Australian interlude, moved toward higher procedural explicitness and tighter alignment among aims, tools, and evidence.

Rather than signalling inconsistency, the differences reflect iterative design in motion: we learned with Brazil, refined with UQ, and operationalized with Vietnam. In comparative terms, this lends internal validity to our analysis, since we can attribute contrasts not only to context, but also to the planned evolution of the training model.

The academic dividends of CAPES PrInt were twofold. Substantively, the program enabled a richer evaluation of how DDL travels, identifying components that remain robust across settings, such as concordance-driven noticing and the principled selection of tools. It also helped distinguish those that require contextual tailoring, including language of delivery, the level of guidance, and the pacing of hands-on/hands-off tasks. Methodologically, it strengthened the study's credibility:

shared instruments, consistent analytic frames, and staged refinement across sites yielded a coherent basis for comparison, rather than two unrelated case studies assembled after the fact.

Finally, there was the human side, which was just as important in a project about teaching and learning. Living and working across these three settings helped me develop an everyday intercultural literacy that would have been hard to build at a distance: understanding how colleagues schedule around institutional demands, how teachers narrate their own professional growth, and how students respond when corpus evidence unsettles what “sounds right”.

This experience broadened my repertoire, intellectually and personally. It deepened my empathy, sharpened my sense of what is feasible in real schools, and reinforced my conviction that DDL succeeds not through uniformity but through thoughtful decontextualisation.

Therefore, the sandwich PhD, via CAPES PrInt, UQ mentorship, and the bridge-building of Prof. Peter Crosthwaite together with Prof. Paula Tavares Pinto, did not merely facilitate mobility; it structured a cycle of design, refinement, and context-sensitive implementation. That cycle both justifies and illuminates the methodological differences between the Brazilian and Vietnamese courses, and it is precisely what enriched the comparative analysis at the heart of this thesis.

1.3 CONTEXT AND PROBLEM STATEMENT

In an increasingly globalized world, English has emerged as a lingua franca for international business, academia, and cross-cultural communication (Jenkins, 2013). This status is particularly evident in the corporate sphere: while only about 5.1% of the Brazilian population aged 16 or older has some knowledge of English (British Council, 2014), as many as 91% of executives surveyed in 77 countries consider English essential for business (British Council, 2014). This gap between the general population’s English proficiency and the demands of the global marketplace underscores the urgency of providing effective English language education, especially in contexts where resources and training opportunities are often limited.

International organizations, such as UNESCO, have also emphasized the importance of learning additional languages for broader social, cultural, scientific,

economic, and academic access (UNESCO, 2014). In practical terms, it is no longer sufficient to rely solely on one's native language, given the increasing interconnectivity of diverse communities. Consequently, language learners are expected to develop a range of 21st-century skills to succeed in both academic and professional settings, including:

- (1) Learning and Innovation (critical thinking, problem-solving, communication, collaboration, creativity, and innovation);
- (2) Digital Literacy (understanding and effectively using digital tools and resources); and
- (3) Career and Life Skills (flexibility, adaptability, initiative, self-direction, and social engagement).

Aligning language education with these broader skill sets can be challenging, particularly because it requires critical engagement: teachers must not only retrieve linguistic information, but also question its applicability in specific pedagogical or research contexts.

In Brazil, for instance, access to digital tools remains uneven, constraining the development of English proficiency in many public schools (British Council, 2014; CIEB, 2022). Although recent educational policies, such as the *Base Nacional Comum Curricular* (BNCC) (BRASIL, 2017), emphasize the role of computational tools and interactive resources in enhancing student autonomy and critical thinking, much work remains to equip teachers with digital competencies when technological resources are scarce or unevenly distributed.

However, Data-Driven Learning (DDL) and other corpus-based approaches offer a way to merge language instruction with digital literacy and to promote autonomy and engagement among learners (Crosthwaite, Luciana, and Wijaya, 2021). By exposing students to authentic linguistic data and guiding them through hands-on exploration, teachers can help develop language proficiency and critical thinking skills. Yet, effectively implementing DDL requires well-trained educators, institutional support, and alignment with national guidelines, as seen in the Brazilian BNCC (Brasil, 2017) and related documents (Garcia, 2020).

The rapid expansion of digital technologies in the 21st century has heightened the need for teachers to effectively integrate these technologies into their instructional practices (Prensky, 2009; Crosthwaite, Luciana, and Wijaya, 2023). In the context of English language education, these technologies often manifest in

corpus-based language teaching (CBLT) or DDL, which rely on extensive collections of authentic texts and specialized tools to foster empirical language exploration (Boulton and Vyatkina, 2021). Within DDL, a key component is the set of 18 corpus skills (Pérez-Paredes, 2020), foundational competencies that enable educators and researchers to effectively harness corpora. These skills encompass understanding the importance and interpretation of frequency, recognizing textual features and concordance lines¹, and applying collocations and register awareness².

They also include more practical proficiencies, such as building or using an existing corpus, performing comparisons across corpora, employing statistical tests, and transcribing or annotating data. Finally, these skills are required to implement DDL in everyday practice (Berber Sardinha, 2004).

The contexts of Brazil and Vietnam share significant similarities, as both are developing countries facing comparable challenges, such as insufficient infrastructure (e.g., schools, materials, and internet access) and limited resources for teacher education. These structural barriers hinder the effective implementation of innovative teaching methodologies, such as DDL.

In both countries, education is widely perceived as a key avenue for social mobility, with families making considerable sacrifices to secure better educational opportunities for their children. Nevertheless, systemic issues persist. Brazil struggles with a public education system that is often underfunded and lacks political prioritization. In contrast, Vietnam's education system remains heavily reliant on rote memorization, which may not align well with the exploratory nature of DDL (Warren, 2017). Furthermore, both countries exhibit gaps in integrating digital resources into pedagogical practices, making it difficult for teachers to exploit the potential of corpus-based learning in their classrooms fully (Chung *et al.*, 2024; Pinto *et al.*, 2023).

At the same time, Brazil and Vietnam present distinct cultural and educational dynamics. Brazilian teachers are frequently encouraged to adopt creative and innovative teaching approaches, while Vietnamese teachers prioritize adherence to established rules and guidelines, which may limit their willingness to experiment

¹ According to Berber Sardinha (2004), a concordance line consists of a list of occurrences of a specific item in a corpus, where the searched word is positioned at the center and surrounded by its original context.

² According to Berber Sardinha (2004), collocations are habitual co-occurrences of words that exhibit a strong statistical association within a language corpus, aiding in natural and idiomatic language use. Register awareness, on the other hand, refers to the ability to recognize and appropriately use language variations depending on the context, such as formal, informal, academic, or conversational settings.

with new methodologies (Chung *et al.*, 2024; Crosthwaite, Ningrum, and Schweinberger, 2022). These pedagogical orientations shape how DDL is introduced and adapted in each context, influencing both the perceived feasibility of corpus-based instruction and the extent to which teachers feel empowered to implement it.

Despite these differences, comparing Brazil and Vietnam offers valuable insights into how DDL can be implemented in developing countries facing similar structural constraints, such as limited opportunities for professional development, unequal access to digital tools, and curricular rigidity. Understanding how these challenges manifest across distinct socio-educational settings enables a more comprehensive analysis of the conditions that facilitate or hinder the adoption of corpus-based approaches.

In this sense, making DDL truly usable for busy teachers in resource-limited contexts requires not only conceptual advocacy but also pragmatic scaffolding that translates corpus literacy into concrete classroom practices. In this thesis, these scaffolds are operationalised through the design of DDL teacher-centred courses and the proposal of Skill 19 (Developing Effective Prompts for Corpus Exploration).

The course offers practical, low-cost supports such as stepwise routines for querying, interpreting, and recontextualising corpus evidence; ready-to-adapt task templates and checklists; and teacher-facing guidance on formulating effective prompts for corpus exploration. Such mechanisms can help bridge the research–practice gap, reduce cognitive load during planning, and enable both hands-off and hands-on pathways that fit different levels of student autonomy and infrastructure.

Throughout this thesis, terms such as *teacher education*, *professional development*, *course/program*, and *training* are used to describe processes of teacher learning and support. The literature in applied linguistics and language education does not always draw sharp boundaries between these labels, and the present study therefore uses them with some overlap. Where relevant, *teacher education* refers to the broader field, and *professional development* to in-service provision, while *course/program* and *training* may refer to the more procedural components of the initiatives examined.

1.4 RESEARCH QUESTIONS AND OBJECTIVES

The overarching aim of this thesis is to support English language teaching in developing countries by helping teachers work with DDL through the development of corpus skills and the adaptation of the approach to diverse educational contexts. More specifically, it pursues three interconnected objectives: (1) to design a DDL course for language teachers that foregrounds corpus skills and digital tools, to better prepare them to use DDL in their own classes; (2) to examine how the different contexts of Brazil and Vietnam shape the implementation of DDL in language classrooms, considering structural, cultural, and institutional factors; and (3) to outline strategies that make DDL more manageable for teachers in developing countries by addressing concrete challenges, responding to recurrent doubts, and improving the integration of DDL into everyday practice. Together, these aims aim to help narrow the English proficiency gap while equipping teachers with the skills expected in 21st-century education.

To address these aims, the research is organised around the following research questions (RQs) and hypotheses (Hs):

RQ1: How can a DDL course focusing on corpus skills and digital tools be designed to empower language teachers in developing countries to implement the DDL approach effectively?

H1: A structured, step-by-step DDL course addressing contextual challenges (e.g., limited technological resources and large class sizes) will significantly increase teachers' readiness and confidence to implement DDL.

H2: Teachers in developing countries will find a guided approach—offering explicit instructions and practical examples—more accessible and impactful than a loosely defined or purely theoretical course.

RQ2: How do structural, cultural, and institutional factors in Brazil and Vietnam influence the implementation of DDL in language classrooms, and what similarities or differences emerge between the two contexts?

H3: Despite cultural variations, structural constraints (e.g., technology gaps, funding, and administrative burdens) will ultimately have a more significant impact on DDL implementation than cultural norms in both contexts.

RQ3: In light of the challenges identified for teachers implementing DDL, how might the newly proposed Skill 19—“Developing Effective Prompts for Corpus Exploration”—help address these limitations and further support DDL integration?

H4 (Propositional): If teachers adopt Skill 19—focusing on creating prompts that encourage critical exploration of corpus data—they could mitigate other corpus skill gaps, thereby strengthening DDL implementation in resource-constrained settings.

Taken together, these research questions and hypotheses frame the investigation of (i) how a DDL course can be designed for developing-country contexts, (ii) how structural and cultural factors shape DDL adoption in Brazil and Vietnam, and (iii) how new strategies – including Skill 19 – might simplify and enhance DDL integration in resource-limited environments.

1.5 RELEVANCE AND CONTRIBUTION OF THE STUDY

This research is significant because it addresses gaps in the literature on DDL and teacher education regarding its implementation in developing countries. While numerous works have explored DDL in higher education or well-resourced contexts (Boulton and Vyatkina, 2021), fewer have focused on in-service teachers working in settings where infrastructure constraints and limited training impede the integration of corpus-based methods. A bibliometric analysis of two decades of Scopus-indexed publications reveals a lack of research explicitly addressing these under-resourced realities, particularly in teacher education and practical classroom implementation (Crosthwaite; Ningrum; Schweinberger, 2022). In Brazil, teachers often face limited digital resources and large class sizes (Cieb, 2022). In Vietnam, studies in both urban and semi-rural regions suggest that structured teacher professional development is lacking, and many educators remain uncertain about how to adapt corpus-based approaches to their specific classrooms (Chung *et al.*, 2024). By analysing these two settings side by side, the thesis fills a comparative gap. It offers contextualised insights into how structural, cultural, and institutional factors can either facilitate or hinder DDL adoption.

Beyond contributing to theoretical discussions on corpus linguistics and language pedagogy, this thesis also proposes concrete ways to support in-service teachers. It does so in two main respects: first, by suggesting improvements to DDL

courses tailored to resource-limited environments, and second, by advancing Skill 19: “Developing Effective Prompts for Corpus Exploration”, as an extension of the 18 corpus skills proposed by Pérez-Paredes (2020) in *Corpus Linguistics for Education*. Rather than treating prompt design as a peripheral technique, Skill 19 frames it as a pedagogical competence in its own right, guiding teachers in designing corpus-based activities that encourage more independent, critically engaged learners. Although Pérez-Paredes, Curry, and Aguado Jiménez (2025) discuss prompt design at the intersection of corpus and AI literacies, this dimension has not yet been formalised as a structured corpus skill within existing frameworks. In this sense, the proposal for Skill 19 builds on previous work on teacher education and corpus literacy (Crosthwaite; Ningrum; Schweinberger, 2021) while directly addressing implementation challenges reported by educators in developing-country contexts.

Policymakers, curriculum designers, and teacher trainers may leverage these findings to create sustainable, context-sensitive professional development programs. Consequently, this work advances the broader goal of bridging the English proficiency gap and equips educators with corpus-based methodologies adaptable to diverse socioeconomic realities.

1.6 THESIS STRUCTURE (ROADMAP)

This thesis³ is organized into six chapters, each addressing different aspects of the research process and findings:

Chapter 1 – Introduction

Chapter 1 presents the overall background and rationale of the study (Section 1.1), the international sandwich PhD context linking Brazil, Australia, and Vietnam (Section 1.2), the research problem (Section 1.3), the research questions and objectives (Section 1.4), and the relevance and contribution of the study (Section 1.5). Section 1.6 outlines the structure of the thesis.

Chapter 2 – Literature Review

Chapter 2 provides the theoretical and contextual foundations. It discusses corpus linguistics, DDL, teacher cognition, and technology integration (Section 2.1), reviews previous research on DDL and corpus-based pedagogy (Section 2.2), and

3 Portions of this study have been published in the article 'Walking the Walk? (Mis)alignment of EFL Teachers' Self-Reported Corpus Literacy Skills[...]' (Chung et al., 2024) and Data Driven Learning in Brazilian Educational Contexts: An Analysis of a Teacher Training Program Following BNCC's Guidelines (Pinto *et al.*, 2025).

identifies research gaps (Section 2.3). It then presents the historical and educational contexts of Brazil and Vietnam (Sections 2.4 and 2.5) and closes with comparative reflections on how both systems create conditions for DDL (Section 2.6).

Chapter 3 – Research Methods

Chapter 3 describes the research design and methodological procedures. It explains the overall approach (Section 3.1), the data collection instruments and procedures for both cases (Sections 3.2 and 3.3), and the analytical procedures used for interviews and lesson plans (Section 3.4). Section 3.5 addresses ethical considerations.

Chapter 4 – Results

Chapter 4 presents the empirical findings. It reports the results from the Brazilian in-service DDL programme, based on interview data and lesson plans (Section 4.1), and the findings from the Vietnamese professional development course, including interviews, survey data, and lesson plans (Section 4.2).

Chapter 5 – Discussion

Chapter 5 brings the findings into dialogue with the theoretical framework. It compares key results across the Brazilian and Vietnamese cases (Section 5.1), discusses how new corpus skills and AI-based tools can extend DDL in teacher education (Section 5.2), and outlines theoretical and practical implications (Section 5.3).

Chapter 6 – Conclusions and Future Work

Chapter 6 revisits the research questions, summarises the main contributions and limitations of the study, and proposes directions for future work, especially regarding corpus literacy and DDL-oriented teacher education in diverse sociocultural contexts.

With this roadmap, each chapter builds upon the previous one, offering a structured and comprehensive exploration of the research problem, theoretical underpinnings, methodology, and findings. The study ultimately culminates in actionable insights applicable to both academic and practical domains.

2 LITERATURE REVIEW

This chapter reviews the literature that frames the study and clarifies the analytical lenses used to interpret DDL in teacher learning across Brazil and Vietnam. It begins by outlining the theoretical background, with attention to DDL, corpus literacy, and corpus based language pedagogy (Section 2.1). It then surveys previous research on DDL and corpus based pedagogy, focusing on teacher development and classroom implementation (Section 2.2), before identifying the main research gaps this thesis addresses (Section 2.3). The chapter next situates the discussion in the historical and educational contexts of Brazil and Vietnam (Sections 2.4 and 2.5), with a closer look at teacher education trajectories and institutional conditions in each setting (Section 2.4.1). It closes by bringing the two contexts into dialogue through comparative reflections on how DDL is shaped by shared constraints and local pedagogical logics (Section 2.6).

2.1 THEORETICAL BACKGROUND

The Data-Driven Learning (DDL) approach is an inductive pedagogical method that analyses authentic language data from corpora using digital tools such as concordancers (JOHNS, 1991). At its core, DDL aims to foster corpus literacy, a set of skills essential for interpreting and applying linguistic database results in the classroom (Pérez-Paredes, 2020; Ma *et al.*, 2023).

These competencies: understanding frequency data, identifying textual features, analysing concordance lines, and recognizing collocations (Boulton and Vyatkina, 2021), offer educators and researchers means to integrate corpus-based insights into language teaching. Pérez-Paredes (2020) identifies 18 skills required for effective engagement with corpus linguistics. Within this framework, the teacher creates an environment where learners can actively develop discovery strategies that promote autonomous exploration of language patterns (John, 1988). Table 1 outlines these eighteen skills in detail.

Table 1 - Corpus Skills

Number of skill	Corpus Skills
01	why frequency matters
02	how to interpret frequency
03	understanding text types: register basics
04	understanding textual features and textual data
05	using an existing corpus
06	reading concordance lines
07	handling frequencies
08	understanding collocations
09	basic corpus design features
10	comparing two corpora
11	statistical tests
12	transcribing your data
13	annotating and querying your data: using your 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

Source: Pérez-Paredes (2020, p.18).

Closely related to DDL, corpus-based language pedagogy (CBLP) focuses on integrating corpus linguistics technology into classroom instruction to enhance language learning (Ma *et al.*, 2021). This pedagogical model aligns with Technological Pedagogical Content Knowledge (TPACK) (Koehler and Mishra, 2005), emphasizing the effective combination of content, pedagogy, and technology for successful implementation in educational settings. Corpus literacy is considered a prerequisite for this approach, as it involves understanding how corpora are structured and analysed and transforming corpus findings into meaningful teaching materials (Mukherjee, 2006; Callies, 2016).

UNESCO's (2014) educational guidelines highlight linguistic diversity and multilingualism as key components in fostering intercultural dialogue and improving education quality. Educators can expose students to various linguistic varieties by incorporating DDL into language teaching, promoting a critical understanding of language use across different contexts (O'Keefee *et al.*, 2007). In developing countries like Brazil and Vietnam, where access to technology remains a challenge, the structured implementation of DDL can bridge gaps in digital literacy and enhance pedagogical outcomes despite infrastructural limitations.

Ongoing professional learning plays a crucial role in the successful application of DDL. Previous studies indicate that corpus literacy development has varying effects on teachers' adoption of DDL, depending on their prior experience and institutional contexts (Ma *et al.*, 2023). Common challenges include the complexity of corpus analysis tools and the difficulty of translating corpus insights into practical classroom activities (Bennet *et al.*, 2023). To address these issues, initiatives such as professional development courses and digital tools can help bridge the gap between theoretical knowledge and classroom practice.

Taken together, this body of work also suggests the value of mixed-methods designs in DDL research, since survey data can capture teachers' self-reported perceptions and intentions, while qualitative accounts and artefacts such as lesson plans can provide more contextualised evidence of pedagogical reasoning and design. For this reason, the present thesis follows a mixed-methods logic by combining quantitative and qualitative sources and integrating them in analysis to examine convergences and tensions across data strands (Dawadi; Shrestha; Giri, 2021).

Recent research suggests that the successful adoption of DDL depends not only on teachers' technical expertise in corpus literacy but also on contextual factors such as institutional support, available infrastructure, and teachers' epistemological beliefs regarding technology in language education (Kim *et al.*, 2011; Karlsen; Monsen, 2020). Understanding these factors is essential for optimising DDL implementation and ensuring its effectiveness as a pedagogical approach.

2.2 PREVIOUS STUDIES ON DATA-DRIVEN LEARNING AND CORPUS-BASED PEDAGOGY

Over the past decades, DDL and CBLP have been the focus of various empirical studies (Boulton, 2012; Daskalovska, 2015; Crosthwaite, Luciana, Wijaya, 2023; Ma *et al.*, 2023), particularly in teacher development and classroom implementation. Research in this field has sought to assess the effectiveness of DDL in different educational contexts and identify the challenges teachers and students face when integrating corpus linguistics into language instruction.

One key area of study involves teacher professional development in corpus literacy. Several studies have explored how pre-service and in-service teachers develop corpus skills and integrate them into their teaching practices (Heather; Helt, 2012; Léńko-Szymńska, 2014; Zareva, 2016; Latif, 2021; Ma *et al.*, 2022; Schmidt, 2022; Bennet; Dhonnchadha, 2023). For instance, Heather and Helt (2012) used a case study approach to analyze how six pre-service teachers improved their corpus literacy skills throughout a semester-long course. Similarly, Leńko-Szymańska (2014) investigated a teacher development program in Poland and found that, while trainees valued corpus-based activities, they often struggled to design effective lesson plans due to a lack of familiarity with corpus tools.

Another critical aspect of previous research is the assessment of corpus literacy and its link to classroom practice. Ma *et al.* (2023) proposed a theoretical framework for assessing corpus literacy among language teachers, using survey data from 183 participants to categorize corpus skills into five subcategories: (1) understanding corpus structure, (2) searching for linguistic patterns, (3) analysing concordance lines, (4) recognizing the pedagogical advantages of corpora, and (5) acknowledging the limitations of corpus use. While their study reported a positive correlation between corpus literacy and teachers' intention to implement CBLP, it also

highlighted that some skills, particularly corpus search and analysis, required further reinforcement to ensure practical uptake.

In addition, longitudinal studies have examined the sustainability of corpus-based pedagogy beyond initial professional development. Latif (2021) analyzed corpus literacy development in Saudi Arabian EFL teachers, finding that while enthusiasm for DDL was initially high, many participants did not continue using corpora in their teaching two years after the training. Similar findings emerged in studies conducted in different educational settings, including Vietnam (Le and Song, 2018), Iran (Dashtestani, 2014), and Norway (Karlsen and Monsen, 2020). These studies suggest that contextual factors, such as technological infrastructure, institutional support, and teachers' epistemological beliefs, significantly influence the integration of corpus-based approaches into long-term teaching practices.

The literature also points to challenges in assessing the practical application of corpus literacy in teacher education contexts. Ma *et al.* (2023) contribute to this area by analysing teachers' lesson plans, a common strategy in corpus-based pedagogy research. As in most studies, including this one, such analyses provide an informative snapshot of how corpus-informed ideas are translated into planning. However, fully capturing the consistency and evolution of corpus-based classroom practice would require longitudinal designs that are logistically demanding and therefore still relatively uncommon. Additionally, research has yet to fully determine how corpus-based instruction impacts student learning outcomes, as most studies focus on teacher perceptions rather than empirical evidence of student progress (Farr, 2008; Ma *et al.*, 2022).

Overall, previous studies indicate the potential of DDL and CBLP to improve language teaching by equipping teachers with data-driven analytical skills. However, they also highlight persistent challenges in developing corpus literacy, achieving long-term adoption, and translating corpus insights into classroom activities. These challenges lead to significant research gaps, which the present study aims to address.

2.3 RESEARCH GAPS

Despite extensive research on DDL and corpus-based language pedagogy, several gaps remain, particularly in the development of corpus literacy,

pedagogical integration, and the contextual barriers faced in developing countries. This study aims to contribute to this discussion by examining how teachers in Brazil and Vietnam develop corpus literacy skills and apply DDL-based practices in their educational settings.

One key research gap concerns the practical challenges teachers face when transitioning from corpus literacy training to actual classroom implementation. While studies have explored teachers' initial perceptions of corpus-based approaches (Lénko-Szymanska, 2014; Zareva, 2016), less attention has been given to how educators adapt DDL to their specific teaching contexts after the course. By analyzing lesson plans and teacher feedback, this study provides insights into how teachers design corpus-based activities and what factors influence their pedagogical decisions.

Another gap relates to how DDL can be effectively integrated into teacher professional development programmes in developing countries. Existing research has primarily focused on well-resourced educational contexts where access to corpus technology and digital infrastructure is readily available (MA *et al.*, 2021). However, in countries such as Brazil and Vietnam, where technological constraints and institutional barriers may limit the adoption of DDL, additional research is needed to identify feasible implementation strategies (Dashtestani, 2014; Latif, 2021). By comparing these two contexts, this study examines how teachers navigate challenges related to digital access, curriculum constraints, and pedagogical adaptation.

Furthermore, although previous studies have identified technical corpus skills as a prerequisite for practical DDL application, there remains a lack of focus on pedagogical strategies that bridge the gap between corpus analysis and classroom instruction. This study proposes Skill 19: Developing Effective Prompts for Corpus Exploration, an extension of Pérez-Paredes' (2020) corpus skills framework, which addresses how teachers can design prompts that guide learners in critically engaging with corpus data.

Finally, although teacher perspectives on corpus-based pedagogy have been documented, less research has explored how teachers refine—or struggle to implement—DDL practices after initial professional development. In this study, Brazilian teachers' follow-up reflections collected 1 year after the DDL literacy program indicated that none of the interviewees had fully incorporated DDL into their

classroom practice. This finding shows the challenges of moving from training to actual pedagogical practice, particularly in contexts where technological limitations, curriculum constraints, and a lack of institutional support may hinder the adoption of corpus-based methods.

This study analyses lesson plans and teacher feedback to examine the extent to which DDL methods are adjusted, adopted, or ultimately abandoned in different teaching environments. Its insights contribute to a better understanding of the pedagogical and institutional barriers that affect teachers' ability to integrate DDL. The study highlights the need for more structured support mechanisms, tailored resources, and ongoing training to ensure corpus-based approaches translate into sustainable teaching practices.

This study aims to advance corpus-based pedagogy in teacher education by offering practical recommendations for DDL implementation, digital tool integration, and pedagogical innovation in under-resourced educational contexts. In line with the thesis's overarching aim of supporting English language teaching in developing countries through corpus skills and context-sensitive adaptation, DDL is treated here as a practical pathway for operationalising CBLP in everyday teaching.

2.4 BRAZIL AND VIETNAM'S HISTORICAL CONTEXT

2.4.1 Teacher Education in Brazil: Policy Ambitions, Institutional Diversity, and Pedagogical Innovation

Brazilian English teacher education cannot be understood apart from the broader conditions of basic education, which remain marked by inequality and fragmentation. Basic education, organized into early childhood, primary (*ensino fundamental*), and secondary (*ensino médio*), is predominantly public and state-managed. Yet, despite constitutional guarantees of equal access and quality, disparities remain in infrastructure, teacher allocation, and learning outcomes across regions and social groups. Studies repeatedly show that many schools, especially in peripheral and rural areas, operate in precarious conditions, with fragile buildings, scarce teaching materials, and limited technological resources. These constraints affect not only student learning but also teachers' working conditions and their opportunities for professional development. Garcia (2024) notes that Brazilian

educational policies have historically oscillated between expanding access and improving quality, often without confronting the structural roots of inequality that undermine both agendas.

In this context, teacher education policies have also suffered from discontinuity and decentralisation. Although national programmes like the Plano Nacional de Formação de Professores da Educação Básica (PARFOR) and institutional grants like the Programa Institucional de Bolsa de Iniciação à Docência (PIBID) have sought to promote more inclusive and practice-oriented training models, implementation tends to be uneven and highly dependent on institutional initiative and local support networks. These dynamics are crucial to understanding the limits and possibilities for adopting innovative practices such as DDL in teacher education programmes.

Teacher education is primarily offered through Licenciatura programmes at both public and private universities. These three- or four-year undergraduate degrees typically include coursework in English language, linguistic theory, pedagogy, and supervised teaching practice. However, their structure, content, and quality vary widely across regions and institutions, often reflecting Brazil's deep-rooted disparities in access to educational resources (Lima and Gimenez, 2020).

A significant policy milestone in recent years was the publication of the Base Nacional Comum Curricular (BNCC) in 2017. This national curriculum framework outlines key learning objectives for all subjects and levels of basic education. Regarding English language teaching, the BNCC positions the language not only as a tool for international communication but also as a means to promote digital citizenship, intercultural understanding, and lifelong learning (Brasil, 2017). It calls for shifts in teaching practices toward communicative approaches and learner autonomy—aims that are closely aligned with DDL principles.

In practice, however, many teacher education programmes remain anchored in traditional methodologies. Grammar-focused teaching, translation exercises, and reliance on textbooks still dominate many courses, creating a gap between what future teachers are officially expected to do and the tools they actually encounter during training (Paiva, 2019). In this context, DDL has been explored in a limited number of universities and research groups as a promising way to narrow that gap. By bringing learners into contact with authentic language use through corpora, DDL opens the door to inductive learning, critical reflection on language, and attention to real-world usage patterns (Berber Sardinha, Delfino, and Rampaso,

2017).

Despite these challenges, Brazilian educators have long demonstrated considerable creativity and pedagogical agency. In contexts where official materials are scarce or outdated, teachers frequently develop their own resources, collaborate with peers, and reflect critically on their practices to find locally meaningful solutions (Lima and Gimenez, 2018). This culture of improvisation can be an entry point for DDL, especially if corpus tools are introduced in ways that are adaptable, intuitive, and responsive to the specific needs of teachers and learners.

Teacher education in Brazil operates within a field of tensions: between innovation and inertia, policy and practice, standardisation and local diversity. In this study, DDL is discussed through a corpus-based language pedagogy (CBLP) lens, foregrounding how corpus evidence is pedagogically mediated and translated into lesson design. The integration of DDL in this context requires more than technical training; it calls for a reflective pedagogy that empowers teachers to question, explore, and reconstruct language with their students. Understanding the realities of Brazilian teacher education is therefore essential to developing DDL practices that are feasible, equitable, and genuinely transformative.

2.5 TEACHER EDUCATION IN VIETNAM: HISTORICAL LEGACIES, INSTITUTIONAL STRUCTURES, AND PEDAGOGICAL TENSIONS

English-language teachers' formation in Vietnam is the product of a layered historical, ideological, and institutional influence. These include the country's colonial past, the divergent paths taken by North and South Vietnam during the Cold War, and ongoing efforts to modernise the higher education system under global pressures.

Early twentieth-century educational movements already reflected attempts to reframe traditional values through modern pedagogical lenses. One key example is *Phụ nữ tân văn*, a widely circulated women's periodical published in Saigon between 1929 and 1934. Aimed at a growing literate urban readership, the magazine combined essays, letters, and fictional narratives to promote ideas of gender reform, education, and civic engagement. It provided a platform for writers and intellectuals to debate the roles of women in a changing society, encouraging readers to envision new forms of domesticity, morality, and modern citizenship. By promoting a model of

the “modern woman,” Phụ nữ tân văn integrated Confucian values (i.e., family-centred, hierarchical moral principles rooted in East Asian Confucianism that emphasise duty, harmony, and gendered virtues) with reformist ideals of education and social participation (Phùng and Vũ, 2024). These efforts positioned education and educators as central drivers of social reform, with women often cast as symbolic agents of this transformation.

Following independence and national reunification in 1975, teacher education became part of broader state-led strategies to construct a unified national identity. As Le *et al.* (2018) explains, language policies in higher education were shaped by competing ideologies. The North embraced a monolingual, socialist framework, while the South retained more multilingual, market-oriented practices. Although the state has since consolidated control, these dual legacies continue to influence teacher identity, pedagogical style, and curriculum design across the country.

Currently, foreign language teacher education is primarily offered at the undergraduate level through a four-year Bachelor of Education in English Language Teacher Education, regulated by the Ministry of Education and Training (MOET). This programme typically includes modules on linguistics, English language skills, pedagogy, and a classroom-based practicum. However, Nguyen, Walker, and Hoang (2023) highlight that although MOET sets a national curriculum framework, implementation varies across institutions, particularly between public and private universities. The standard programme comprises approximately 120-130 credit units, combining theoretical coursework in methodology and assessment with practical components. Students are also expected to achieve a defined level of English proficiency, typically B2 or C1 on the CEFR scale, by the end of their studies.

One of the most ambitious initiatives in this domain was *The National Foreign Language Project 2020*, known as Project 2020. This policy aimed to align language education standards in Vietnam with international benchmarks, most notably by integrating the CEFR into curriculum design, teaching goals, and assessment practices. It also encouraged universities to revise teacher education programmes to include modules on learner autonomy, language assessment, and the use of technology in teaching.

In this context, the integration of Information and Communication Technology (ICT) became a central component. In educational settings, ICT refers to

the use of digital tools such as computers, internet-based resources, interactive whiteboards, and mobile applications to enhance teaching and learning. For English teachers in training, this includes learning to use educational software, digital corpora, online dictionaries, and virtual learning platforms. MOET's policy aimed to ensure that future teachers were not only digitally literate but also capable of applying these tools in student-centred pedagogical contexts. Nevertheless, the implementation of ICT training has been uneven, with institutions reporting gaps in infrastructure, limited faculty expertise, and a general lack of confidence among pre-service teachers in using such tools effectively in real classrooms (Chung *et al.*, 2024).

Despite these reforms, many institutions continue to report a disconnect between policy goals and classroom realities (Do, Le, and Vu, 2023; Nguyen *et al.*, 2023; Ho and Dimmock (2010). Teachers and students alike remain under pressure from standardised testing regimes, which often incentivise exam preparation over communicative competence. As a result, even as CEFR descriptors are embedded into curricula, the dominant classroom practices frequently remain focused on memorisation and textbook use. The persistent gap between policy and classroom enactment highlights the value of teacher-facing procedural supports such as ready-to-adapt lesson templates, step-by-step task sequences, and alignment checklists — that reduce planning load, make the use of evidence explicit in the lesson artefacts, and link each activity to declared outcomes and assessment criteria (Ho; Dimmock, 2010).

Vietnamese pedagogical culture, which values hierarchy, standardization, and exam results, may initially seem at odds with exploratory approaches like DDL. Yet this same structure can provide an effective scaffold for integrating corpus-based tools into existing curricula. When paired with explicit, classroom-ready guidance on procedures and timing, corpus-based activities can be embedded in structured lesson plans or blended learning environments, facilitating gradual adoption that aligns with institutional expectations. As Kelley and Sasges (2024) argue, Vietnamese modernity operates through selective and strategic adaptation, which they describe as “vernacular modernity.” This flexible orientation may offer pathways for meaningful localization of DDL.

Teacher education, therefore, represents a dynamic negotiation between institutional constraints and the pursuit of pedagogical innovation. Understanding

how teachers are trained and how they navigate the demands of their profession is essential for developing corpus-based methodologies that are both contextually appropriate and pedagogically impactful. Any implementation of DDL in Vietnam must engage with these realities, not as barriers to overcome, but as conditions that shape what is possible.

2.6 COMPARATIVE REFLECTIONS: VIETNAM AND BRAZIL IN DIALOGUE WITH DDL

Although geographically and culturally distinct, Vietnam and Brazil share several challenges and opportunities in integrating DDL into the training of English language teachers. Both countries are marked by deep educational inequalities, ambitious national curriculum reforms, and a desire to participate more actively in the global academic and economic spheres (Nguyen *et al.*, 2023; Lima and Gimenez, 2020). Yet, the way these elements materialize in teacher education reveals important differences in institutional structure, pedagogical culture, and epistemological orientation.

One of the most evident similarities lies in the mismatch between policy and practice. In both Vietnam and Brazil, national frameworks such as MOET's CEFR-aligned initiatives under Project 2020 and Brazil's BNCC (Brasil, 2017) articulate progressive goals regarding learner autonomy, critical thinking, and the integration of digital tools. These ambitions, on paper, align well with DDL's principles. However, classroom implementation remains uneven. Vietnamese institutions often struggle with rigid curricula and exam-focused teaching cultures (Le, 2024), while Brazilian programmes grapple with resource disparities and a persistent reliance on traditional methodologies (Paiva, 2019). In both contexts, reforms tend to be absorbed into pre-existing structures rather than catalyse deep pedagogical shifts.

A second point of convergence is the limited exposure to corpus linguistics in pre-service teacher education. While a small number of institutions in both countries have begun experimenting with corpus tools and DDL-based materials (Berber Sardinha, Delfino and Rampaso, 2017; Viana and Lu, 2021), such practices remain far from mainstream. Teachers in both contexts frequently complete their undergraduate studies without formal training in corpus analysis or in the pedagogical strategies needed to translate corpus data into classroom practice.

Despite these parallels, key divergences emerge when examining how each country's educational culture shapes the potential for DDL integration. In Vietnam, teacher education is centrally regulated, and although this may initially limit autonomy, it also opens the possibility for large-scale, top-down implementation. The institutional alignment fostered by MOET, particularly under Project 2020, could enable corpus-based methods to be embedded across national curricula once they are formally endorsed (Nguyen *et al.* 2023). Brazil, on the other hand, has a decentralised, heterogeneous system that offers greater flexibility at the local level. This can be an advantage in institutions where educators are engaged in corpus-based research or innovation, but it also makes national coordination more difficult (Lima and Gimenez, 2020).

Another relevant contrast is the role of pedagogical improvisation and agency. Brazilian educators often navigate under-resourced classrooms with remarkable adaptability. This culture of resilience supports grassroots experimentation with DDL and other exploratory methods, even when formal training is lacking (Pinto *et al.*, 2025). Vietnamese educators, by contrast, work within more hierarchical systems where policy plays a stronger role in shaping practice. While this can inhibit spontaneous experimentation, it also makes targeted professional development programmes more scalable and coherent once they are in place (Le, 2024).

Finally, cultural perceptions of language learning shape how DDL is received and adapted. In Brazil, the emphasis on communicative competence and critical pedagogy provides a conceptual bridge to the student-centred and discovery-oriented nature of DDL (Frankenberg-Garcia, 2018). In Vietnam, pedagogical culture tends to privilege authoritative knowledge and examination performance. Nevertheless, scholars such as Kelley and Sasges (2024) argue that Vietnamese modernity is not rigid but strategic, guided by what they call vernacular modernity. This adaptive disposition suggests that DDL, when appropriately framed and scaffolded, can align with local expectations for rigour and control.

To synthesize the key similarities and differences discussed above, Figure 1 below presents a comparative overview of how English teacher education systems in Brazil and Vietnam align with the principles and implementation of Data-Driven Learning. This visual summary highlights dimensions such as curriculum regulation, teacher autonomy, access to resources, and pedagogical orientation. While both

contexts face structural and institutional barriers to DDL integration, each also offers distinctive pathways for innovation: Brazil through decentralized agency and local improvisation, and Vietnam through coordinated policy-driven reform.

Figure 1 – Comparative overview of English teacher education and DDL implementation in Brazil and Vietnam

Aspect	Brazil	Vietnam
National Curriculum	BNCC (2017), decentralised implementation	MOET-led CEFR integration (Project 2020)
Teacher Training Regulation	Decentralised across institutions	Centrally regulated by MOET
Infrastructure Gaps	High (especially in rural areas)	Variable, improving in urban areas
Exposure to DDL	Limited, isolated initiatives	Limited, small-scale pilots
Curricular Flexibility	High at institutional level	Low; standardised across institutions
Implementation Style	Bottom-up, teacher-driven	Top-down, policy-driven
Innovation Pathway	Grassroots experimentation	Systematic implementation via policy
Cultural Orientation	Communicative, learner-centred	Hierarchical, exam-oriented

Source: Adapted from Lima and Gimenez (2020); BNCC (2017); MOET Project (2020).

The integration of DDL in teacher education, therefore, is not merely a technical decision but one that reflects how knowledge is valued, distributed, and legitimised within national systems. While both Brazil and Vietnam face structural and pedagogical barriers, each offers unique entry points for DDL innovation. Vietnam may benefit from coordinated implementation supported by institutional alignment, while Brazil's pedagogical culture enables flexible, locally relevant experimentation.

The comparative analysis presented in this chapter underscores the need to treat DDL not as a one-size-fits-all methodology, but as a pedagogical approach that must be recontextualised according to local values, educational structures, and teacher agency. By examining the convergences and divergences between Brazil and Vietnam, this study sets the stage for a deeper investigation into how teachers internalise and adapt corpus-based strategies in real-life educational settings. The following chapter presents the research methods used to empirically explore these dynamics.

3 RESEARCH METHODS

This chapter outlines the research methodology adopted in this thesis, which investigates the development of corpus literacy skills in teacher education contexts in Brazil and Vietnam. The project uses a comparative case study design, examining two teacher education settings in parallel to explore how teachers engage with DDL techniques and integrate them into their pedagogical practices. Data collection in Brazil occurred approximately 9 months before the Vietnam program. During the sandwich PhD in Australia, the design was refined collaboratively with partners, leading to targeted adjustments in instruments, delivery, and bilingual facilitation for the Vietnam case.

Differences in timing and data-collection instruments, such as Brazil's interviews conducted about a year after the program, while Vietnam's data were collected immediately after the course, were interpreted in light of each site's practical constraints and aims and incorporated into the analysis as part of the research context.

The chapter begins by explaining the overall methodological approach and design logic (Section 3.1), before detailing the data collection process and instruments used across both cases (Section 3.2), including procedures, qualitative analysis, and strategies for maintaining comparability (Sections 3.2.1 to 3.2.3). It then presents the two cases in turn, describing the Brazil program, its participants, and its methodological sequence (Sections 3.3 and 3.3.1-3.3.2), followed by the Vietnam case and its training model, sequence, and data collection methods (Sections 3.3.3-3.3.5). The chapter next outlines the data analysis procedures for each case, including the analysis of lesson plans (Sections 3.4-3.4.3), and concludes by addressing ethical considerations (Section 3.5).

3.1 OVERALL APPROACH

The research employs a qualitative case study approach and quantitative analysis elements to assess teacher engagement with corpus-based pedagogical strategies. The study explores teachers' perceptions, skill development, and implementation challenges when integrating DDL into their teaching practices.

A parallel case study design was chosen to compare two course contexts:

Case 1 (Brazil): A professional development course for in-service teachers in Brazilian public schools.

Case 2 (Vietnam): A structured bilingual program for teachers working in a university-affiliated language centre in Vietnam.

The study examines these two cases to identify patterns, challenges, and opportunities for integrating DDL in developing educational contexts.

3.2 DATA COLLECTION

This section describes how data were gathered across the two cases. A mixed-methods design combined surveys, production of corpus-based lesson plans, independent rubric-based evaluation, and interviews. Each case followed a fixed sequence that moved from baseline profiling to training activities, and then to evidence of use and perceptions. Instruments were tailored to local aims, while the core assignment and rating procedures remained constant to ensure meaningful cross-case comparison. Figure 2 presents a side-by-side flowchart of the Brazilian and Vietnamese sequences and serves as a visual guide for the subsections that follow.

Figure 2 - Brazil × Vietnam data-collection flowchart

Dimension	Brazil	Vietnam
Setting & delivery	30h PD; 7 modules; Zoom + 2 IWTC; workshops in Portuguese + consultations.	Multi-campus bilingual programme; IWTC modules → workshops (EN/VI).
Participants	n=21; in-service teachers; multi-state; mixed public/private.	n=48; teacher trainees; ~81% attendance; primary EFL (Starters–KET).
Instruments	Baseline survey; module self-assessments; post perceptions; interviews (PT).	Baseline survey; post corpus-literacy survey (Ma et al.); IPR interviews (VI→EN).
Lesson-plan task & rating	Complete lesson built on corpora; TIOI (2 raters; reconcile).	Lesson plan with corpus tasks; TIOI (2 raters; reconcile).
Qualitative analysis	Tabletop thematic analysis + LancsBox for salient items/contexts.	In Vivo Coding → thematic mind-map with exemplar quotes.

Source: Prepared by the author (2024).

The two cases differed in contextual features that shaped how data were gathered and interpreted. Brazil delivered the program in Portuguese, administered a post-program perceptions survey immediately after the course, and conducted follow-up interviews approximately 1 year later to assess longer-term feasibility and uptake. Vietnam delivered bilingual workshops in English and Vietnamese and

administered the Ma *et al.* (2023) corpus-literacy survey along with interviews guided by the Interview Protocol Refinement (IPR) approach within the same window to document immediate responses. These timelines and instruments reflected local aims and logistics; they were taken into account in the comparative analysis, while the core assignment (design of a corpus-based lesson) and its double rating with the adapted TIOI (Harris; Grandgenett; Hofer, 2010) served as a daily basis for comparison. Further details of each sequence, instrument, and rating procedure are presented in the following sections.

3.2.1 Instruments and Procedures

We used an adapted version of the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) with four dimensions: (i) curriculum goals × technology, (ii) instructional strategies × technology, (iii) technological selection, and (iv) overall fit. The rubric was extended to capture the learner's role as "language detective," the corpus skills targeted (e.g., reading concordance lines, collocations, frequency), and the alignment of assessment with corpus activities. Two calibration rounds were conducted on the sample lesson, plans, and anchor notes were recorded. All plans were double-rated independently; discrepancies of more than 1 point were discussed to consensus, with an audit trail. Outputs included dimension scores, an overall fit level (no fit / partial fit / fit / strong fit), and qualitative notes.

Surveys comprised, in Brazil, a baseline profile, module self-assessments, and a post-program perceptions questionnaire (four-point Likert plus open items); in Vietnam, a post-course corpus-literacy survey (Ma *et al.* 2023).

Interviews were semi-structured in Portuguese in Brazil (~45 min, ~one-year follow-up) and guided by the Interview Protocol Refinement approach in Vietnam (same window as the post-survey). Full item banks, response scales, interview guides, and the lesson-plan template are provided in Appendix B.

3.2.2 Qualitative Analysis

Brazil followed a Tabletop procedure (Saldaña, 2016): transcription and anonymisation; clustering of excerpts to surface candidate categories; corpus-based

clusters, expanded context and examples. For Vietnam, In Vivo coding preserved participants' own terms, which were grouped into categories and themes and mapped by frequency/co-occurrence with theme maps including exemplar quotations. Two or three short excerpts per case illustrate how codes were applied and why alternatives were rejected.

3.2.3 Comparability across cases

Comparisons were anchored in a common core task (design of a corpus-based lesson plan) and the same double-rating TIOI (Harris; Grandgenett; Hofer, 2010) procedure. Outcome dimensions included lesson-plan quality (TIOI) (Harris; Grandgenett; Hofer, 2010), readiness and intentions (surveys), and barriers/facilitators (interviews). Context variables made explicit included the timing of data collection (Brazil $\approx$: 1-year follow-up; Vietnam: immediate), setting and teacher profile, learner level (BNCC/CEFR), delivery mode, language of delivery, and infrastructure.

3.3 CASE 1 (BRAZIL) - A DDL PROGRAM FOR IN-SERVICE TEACHER

We introduced the Brazilian case by tracing the learner's journey through the program. The route began with a baseline survey, continued with self-paced Improving Writing through Corpora (IWTC) modules, and moved into workshops and targeted consultations. Participants then designed a corpus-based lesson, which was evaluated by two independent raters using an adapted version of the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010). Immediate reactions were recorded in a post-program perceptions survey, and interviews conducted about a year later offered a longer view of feasibility and uptake. Figure 3, previously shown, depicts this route and provides a visual guide for the discussion that follows.

3.3.1 Methodological sequence used in Brazil

As a starting point, a teacher development program titled Innovative Pedagogical Practices for English Language Teaching: Students as Language Detectives and Learning Protagonists was designed to provide professional development for public school English teachers in Brazil (Pinto *et al.*, 2025). The program aimed to (1) familiarize teachers with digital tools commonly used in corpus linguistics, (2) support reflection on how to implement these tools so that students take an active role in their language learning, (3) improve teachers' selection of tools for their teaching contexts, and (4) strengthen their pedagogical commitment to a data-driven learning (DDL) approach.

The 30-hour teacher education programme ran over two months and consisted of seven modules. It was conducted online, combining synchronous and asynchronous activities, to accommodate a diverse group of educators. Eligibility prioritised in-service English-language teachers holding a degree; one final-year undergraduate in Letras was admitted as an exception. Applicants from varied teaching contexts (public and private; basic and higher education) were welcomed. Different instructors delivered five modules synchronously via Zoom, each focusing on a different corpus-based tool and skill (Pérez-Paredes, 2020). The two asynchronous modules to be completed by the course participants were part of the *Improving Your Writing through Corpora Online Course*, designed to help second language learners enhance their academic writing using corpus tools (Crosthwaite, 2019; 2020). Table 2 summarizes the course schedule and plan.

Table 2 - 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	Lingle and Skell: Educational applications	Lingle, 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: Prepared by author (2024).

The tools frequently used by the course instructors were taught in the program, from the most user-friendly to the most complex ones, which are described as follows:

1. VersaText (Thomas, 2020) is a web-based tool that allows learners of English to explore the language of a single text. It functions as a single-text corpus by lemmatising the input and providing part-of-speech information. As a result, teachers and students can generate word clouds and examine different word categories through concordance lines.

2. Sketch Engine (Kilgarrieff et al., 2014) is a corpus analysis platform that works with authentic collections of texts to help users identify typical patterns of language use as well as rarer, unusual, or emerging forms. It also supports broader text analysis and text mining applications.

3. SKELL (Baisa; Suchome, 2014) is a user-friendly tool designed to help students and teachers quickly check whether, and how, particular words or phrases are used by English speakers. Teachers and learners can use it to

search for collocations, synonyms, and contextualised examples of a target word, and it also allows users to generate word clouds of related vocabulary.

4. Linggle (Boisson, 2013) is similar to the tools described above and functions as a search engine that supports writers in checking word choice, usage patterns, and collocations in order to produce more natural-sounding texts.

6. VocabProfile (Cobb, 2004) is a simple web-based program for analysing the vocabulary load of a text. It enables users to identify which vocabulary items occur in a particular text or set of texts and to view them in relation to frequency-based lists, such as the General Service List and the Academic Word List (GSL/AWL), as well as simplified lists derived from the British National Corpus.

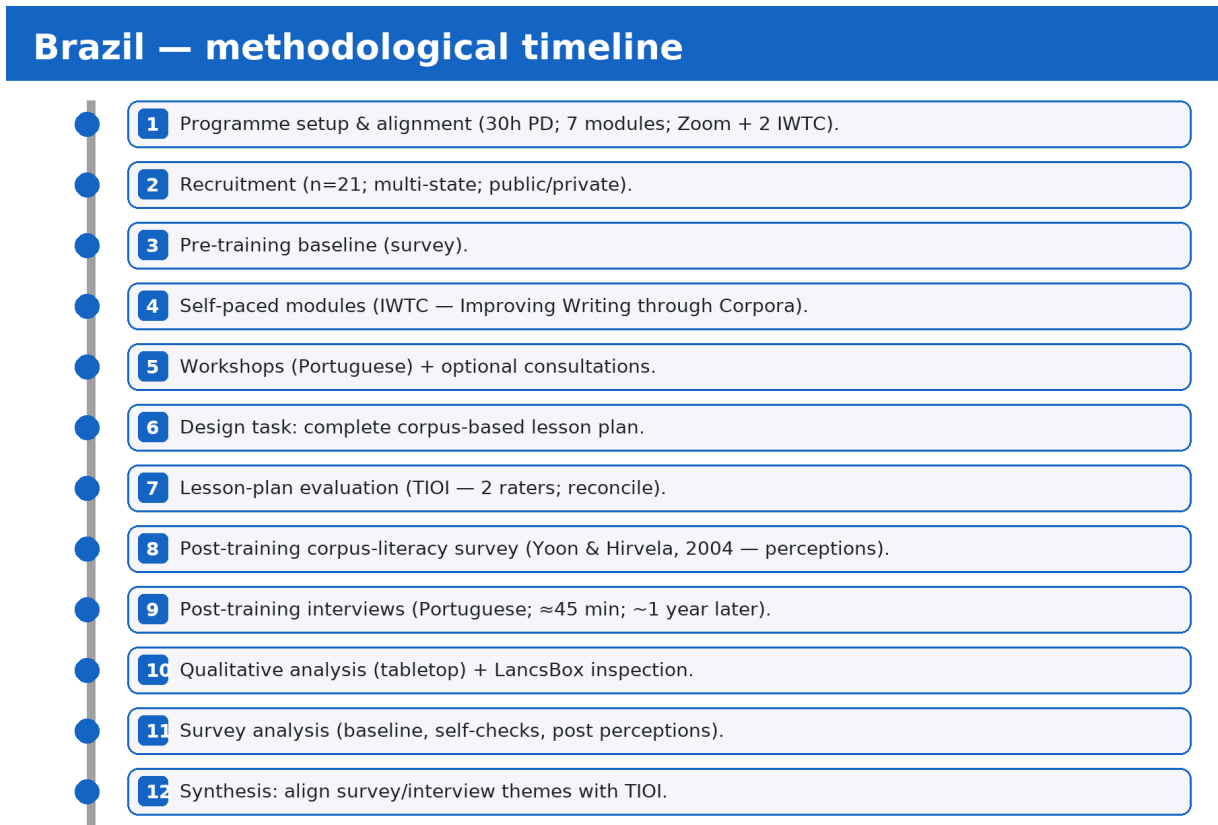
7. PlatCol (Orenha-Ottaiano, 2021) PlatCol (Orenha-Ottaiano, 2021) supports collocation learning and translation in English/Portuguese, boosting proficiency and more natural language use.

8. TradCorpus (Serpa *et al.*, 2023) is a platform with specialized corpora and glossaries that users can access if they are working with English and Portuguese specialized language and literature.

In addition to the scheduled meetings for each module, three extra consultation sessions were offered to address participants' questions about the program activities and the final assessment, which required them to develop a lesson plan using the tools and skills learned throughout the program.

Figure 3 presents the methodological sequence adopted in the Brazilian course, outlining the main steps followed from course design to implementation and evaluation.

Figure 3 - Methodological Timeline



Source: Prepared by author (2024).

3.3.1.1 Participants

A total of 21 teachers from five Brazilian states participated in the corpus literacy course, including São Paulo (57.5%), Paraná (22.5%), Goiás (12.5%), Rio de Janeiro (5%), and Bahia (2.5%). All participants hold at least an undergraduate degree, and one participant is currently completing his bachelor's degree and is ready to work as an English language teacher. Regarding their professional backgrounds, participants worked in primary, secondary, and higher education, with 62.8% in public high schools and 32.6% in private institutions.

3.3.2 Instruments for data collection

Participants completed a baseline questionnaire before the programme began to collect demographic information (gender, English proficiency, educational background, place of origin, and teaching context). After each module, they filled in a

short self-assessment on their familiarity with the corpus tools introduced, challenges encountered, and the feasibility of using the tools in their own classrooms. Each self-assessment also included open-ended items for feedback and suggestions to inform subsequent seminars.

To capture overall perceptions at the end of the programme, participants completed a final questionnaire adapted from Yoon and Hirvela (2004). This elicited views on (1) the programme as a whole, (2) the development of corpus-based skills, (3) the usefulness of each module and tool, and (4) satisfaction with the instructors and programme structure. Full item banks and response scales for the baseline, module self-assessments, and post-programme questionnaire appear in Appendix B, alongside a permanent link to the project repository containing the raw instrument files. For transparency, an archived copy of the post-course form used in practice is also included (Appendix B, Repository Link 1).

Upon program completion, participants were asked to complete a final assignment that required elaborating on a lesson plan using the pedagogical DDL tools and corpus skills learned during the program modules. Participants were provided with a lesson plan template. They were expected to include the subject to be taught, the linguistic framework used, the educational context, the proficiency level of the target students, 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 interview protocol was developed from prior DDL studies and adapted to our context; the complete guide (“Interview Questions.docx”) is available in Appendix B.

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

3.3.3 CASE 2 (VIETNAM): TEACHER EDUCATION MODEL IN VIETNAM

The Vietnam case is introduced by tracing the pathway teachers followed through the program. The route began with a baseline survey, continued with

self-paced Improving Writing through Corpora (IWTC) modules, and moved into bilingual workshops in English and Vietnamese with focused analysis of sample DDL lesson plans. Participants then designed a corpus-based lesson using a shared template that was evaluated with an adapted version of the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010); by two independent raters. A post-course corpus-literacy survey (Ma *et al.*, 2023) and interviews guided by the Interview Protocol Refinement (IPR) approach were conducted in the same window to capture immediate responses and feasibility. Figure 4 depicts this sequence and provides a visual guide for the discussion that follows.

3.3.4 METHODOLOGICAL SEQUENCE USED IN VIETNAM

Chung *et al.* (2024) conducted this part of the study as a multi-campus collaboration between the Victory Informatics and Foreign Language Centre of Tra Vinh University, the University of Queensland, and UNESP, bringing together expertise from different institutions to strengthen teacher education in corpus-based language pedagogy.

The research site, located in Tra Vinh, Vietnam, is a Level 2 city with a population of approximately 100,000. The centre's 2021–2025 Strategic Plan prioritises integrating teaching and learning technology and digitising instructional activities to maintain its leadership in foreign language education locally and extend its influence across the Mekong River Delta.

In line with this agenda, the centre offers in-service teacher development focused on Computer-Assisted Language Learning (CALL) for young learners, delivered by permanent centre staff, British Council experts, and university faculty from partner institutions. Since 2014, the centre has also been a Cambridge Authorised Exam Centre, offering structured Cambridge English courses for learners aged 4 to 15, which often complement school-based English programmes at parents' request.

The study's participants included local teacher trainees, permanent centre staff, and university faculty members. Training on corpora and DDL was specifically designed for those teaching Cambridge English levels Starters (pre-A1), Movers (A1), Flyers (A2), and KET (A2), ensuring its direct applicability to their teaching

practice. A total of 48 participants took part in the study, with an average attendance rate of 81% throughout the training.

Expanding on Chung *et al.* (2024) earlier description, this bilingual training initiative was developed as a joint effort between the three partner institutions, combining online modules with in-person workshops to engage participants in corpus fundamentals and lesson plan design in both English and Vietnamese.

Researchers from the University of Queensland and UNESP participated remotely, delivering key training components and supporting teachers in developing their DDL skills through virtual sessions. The program adopted a holistic bilingual approach, integrating corpus literacy course, pedagogical resources, and face-to-face sessions.

The course began with self-paced online modules, in which participants explored corpora fundamentals, corpus searches, and data-driven lesson planning using preexisting digital resources (*Improving Writing through Corpora*, Crosthwaite, 2020). Teachers then analysed sample DDL lesson plans sourced from educational repositories (Pinto *et al.*, 2023), allowing them to observe real-world applications of corpus-based pedagogy.

Following the online modules, participants attended interactive training workshops, initially conducted in English via Zoom and reinforced through face-to-face Vietnamese sessions. This bilingual format was essential due to the participants' varying levels of English proficiency, ensuring that they could fully engage with the content while receiving additional support in their native language (Simsek, 2023).

The structured course design enabled teachers to progressively build their corpus literacy skills, while blended learning strategies provided flexibility and support. Figure 4 provides a detailed breakdown of the course stages and instructional materials.

Figure 4 - CBLP training procedure

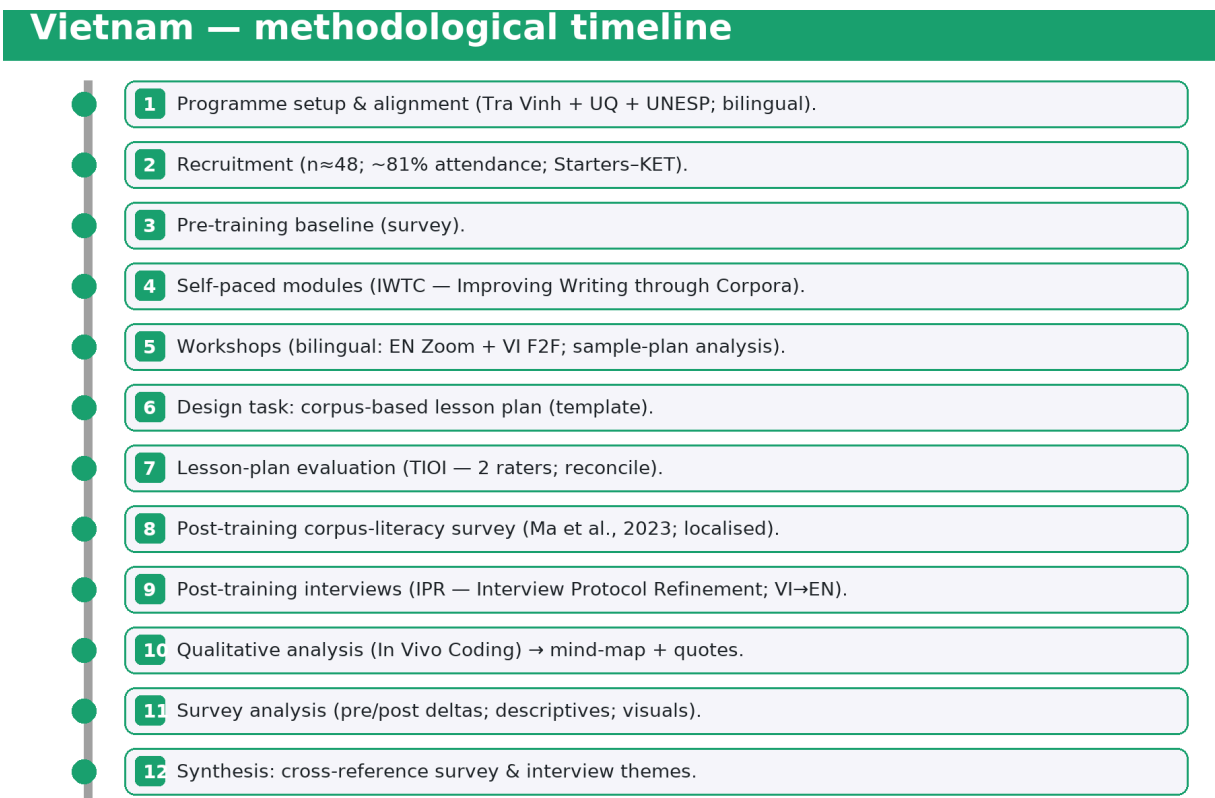
Activity	Period (from/to)	Topic	Online Learning	Face to Face Activity
Pre-workshop Activities				
Pre-workshop Activity	27/02 - 03/03	Understanding corpora, corpus searches (Modules 1-2 of an online course)*	Online materials provided	2.5h in-person session in Vietnamese, Q&A and translation
Pre-workshop Activity	03/03 - 10/03	Analyzing corpus-based activities (Pinto et al., 2023)	Online readings/materials	Additional in-person activities
CBLP Training Workshops				
Workshop 1: Intro to Corpora	03/04	Frequency, collocation, lexis & grammar searches	1h Zoom workshop; design activity (part 1)	2h in-person Q&A session
Workshop 2: Corpus Ideas	04/04	Ideas for corpus-based teaching	1h Zoom workshop; design activity (part 2)	2h in-person Q&A session
Lesson Planning Feedback 1	05/04	Feedback on lesson plans (Starters & Movers)	1h Zoom workshop	2h in-person Q&A session
Lesson Planning Feedback 2	06/04	Feedback on lesson plans (Movers, Flyers & KET)	1h Zoom workshop	2h in-person Q&A session
Post-Workshop Activities				
Post-Workshop Activity	10/04 – 21/04	Designing a lesson plan with corpus-based activities	Final lesson plan creation (2 weeks, pair/group work)	(Discussion/feedback if applicable)

* Improving writing through corpora: Data-driven learning? <https://edge.edx.org/courses/course-v1:UQx+SLATx+2019/about>

Source: adapted from Chung *et al.*, (2024).

Figure 5 provides an overview of the step-by-step methodological sequence followed in the Vietnam case, serving as a visual guide to the procedures described below.

Figure 5 - Methodological sequence adopted in the Vietnamese course, outlining the main steps followed from initial steps to course implementation and evaluation



Source: Prepared by the author (2024).

Taken together, the sequence shows how baseline profiling, guided practice (self-paced modules and bilingual workshops), and an authentic design task (corpus-based lesson planning) were combined with post-course instruments (survey and interviews) and systematic analysis procedures to generate comparable evidence of corpus-literacy development. This organization is particularly relevant to the cross-case design of the thesis, as it clarifies which stages were shared with the Brazilian context and which were shaped by local constraints—most notably the bilingual delivery format and the timing of post-course data collection.

3.3.5 Data Collection Methods

Lesson planning

Trainees were asked to produce sample lesson plans using a template containing at least one corpus-based activity. Detailed feedback on the suitability and feasibility of the CBLP plan was provided for each plan.

Trainees then produced a final lesson plan for evaluation by the researchers using the Technology Integration Observation Instrument (Harris; Grandgenett; Hofer, 2010), previously used to analyse DDL lesson plans in Crosthwaite *et al.*, (2021).

As previously detailed in Chung *et al.* (2024), we employed the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) to assess each lesson plan in this study.

The TIOI uses a four-point scale (from 1 = lowest to 4 = highest) to examine technology integration across four key dimensions: (1) Curriculum Goals and Technologies, (2) Instructional Strategies and Technologies, (3) Technology Selection, and (4) overall 'Fit' among these components, as you can see in Table 3.

Because the same dataset and analytical framework are used here, our application of the TIOI (Harris; Grandgenett; Hofer, 2010) is entirely consistent with the earlier publication.

Table 3 - Technology Integration Observation Instrument (TIOI)

Criteria	4	3	2	1
Curriculum Goals & Technologies (Curriculum-based technology use)	Technologies selected for use in the instructional plan are strongly aligned with one or more curriculum goals.	Technologies selected for use in the instructional plan are aligned with one or more curriculum goals.	Technologies selected for use in the instructional plan are partially aligned with one or more curriculum goals.	Technologies selected for use in the instructional plan are not aligned with any curriculum goals.
Instructional Strategies & Technologies (Using technology in teaching/learning)	Technology use optimally supports instructional strategies.	Technology use supports instructional strategies.	Technology use minimally supports instructional strategies.	Technology use does not support instructional strategies.
Technology Selection(s) (Compatibility with curriculum goals & instructional strategies)	Technology selection(s) are exemplary, given curriculum goal(s) and instructional strategies.	Technology selection(s) are appropriate, but not exemplary, given curriculum goal(s) and instructional strategies.	Technology selection(s) are marginally appropriate, given curriculum goal(s) and instructional strategies.	Technology selection(s) are inappropriate, given curriculum goal(s) and instructional strategies.
"Fit" (Content, pedagogy and technology)	Content, instructional strategies and technology fit together strongly within the instructional plan.	Content, instructional strategies and technology fit together within the instructional plan.	Content, instructional strategies and technology fit together somewhat within the instructional plan.	Content, instructional strategies and technology do not fit together within the instructional plan.

Source: Adapted by the author (2024) from Harris, Grandgenett, Hofer (2010).

Two researchers rated each lesson plan individually, then met to discuss any discrepancies in scores before deciding on the final score for each criterion.

An initial pre-course survey was administered via Google Forms to collect demographic information and assess participants' prior corpus experience. The results showed that most teachers (79.2%) had not previously used DDL or corpora in their teaching.

For the post-course survey (also administered via Google Forms), this part of the study adapted the CBLP questionnaire from Ma *et al.* (2023) to better fit the target trainees. In the Search subfactor (S2), “keywords” was changed to “keywords and expressions” to avoid confusion with the specialised corpus-linguistics meaning of keywords. Similarly, in Analysis (A2 and A3), “keyword” was replaced with “searched target item”. Finally, one open-ended item was added—“Do you think the corpus approach will work well in Vietnam? Please explain”—to capture contextual factors shaping future CBLP uptake.

As Chung *et al.*, (2024) reported, 48 teachers participated in the project, with an overall attendance rate of 81%. They completed the pre-training survey, and 43 trainees responded to the post-course survey.

3.4 DATA ANALYSIS

3.4.1 Case 1 (Brazil)

Interviews

The interview recordings were transcribed in Portuguese and later translated into English. To systematically organize and analyze the interview data, the research team used an adapted version of the tabletop methodology (Saldaña, 2016). This methodology involved arranging the interview transcriptions on a tabletop, enabling the grouping of information into categories to more clearly discern patterns. By identifying the frequency of specific pieces of information, this categorization helps recognize emerging themes or concepts.

LancsBox version 6 (Brezina, Weill-Tessier and McEnery, 2021) supported the tabletop analysis and explored the linguistic data and corpora from the interviews. The initial step was to extract the 10 most frequently occurring words from the interview transcriptions. Each word was then thoroughly investigated within its context in the dataset. This analysis allowed for categorizing the data based on word frequency, facilitating the identification and examination of relevant themes and concepts.

The tabletop and LancsBox analyses focused on answering the second research question, which concerned participants' perceptions of the tools and skills learned through the DDL approach.

3.4.2 Lesson plans

In this thesis, the term lesson plan is used as an operational category to refer to a written outline or blueprint that organises a class session, typically bringing together learning objectives, materials, activities, assessment, and instructional strategies. This operationalisation is consistent with how the term emerged in the interviews, where participants described planning as the careful alignment of goals, procedures, and evaluation within a specific teaching unit.

The lesson plans participants completed and submitted as the final assessment of the corpus literacy program were evaluated and analysed using an adapted version of the evaluation framework Technology Integration Observation Instrument (TIOI) proposed by Harris, Grandgenett, and Hoferet (2010) (see Table 3). TIOI facilitated a comprehensive assessment of the integration of technological tools within the lesson plans, ensuring alignment with contemporary pedagogical methodologies and instructional objectives. It also helped evaluate how well teachers understood the key concepts of DDL throughout the program. The lesson plans' evaluation helped answer part of 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 4 describes the adapted TIOI used in our analyses:

Table 4 - Adapted 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's goals and instructional context.	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 detective 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 language detective.	The lesson plan focuses on the student as a detective of the language.	The lesson plan partially focuses on the student as a language detective.	The lesson plan does not focus on the student as a language detective.
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 firmly fit within the lesson plan.	Content, instructional strategies, and technology fit together within the lesson plan.	Content, instructional strategies, and technology fit within the lesson plan.	Content, instructional strategies, and technology do not fit within the lesson plan.

Source: Adapted from Harris, Grandgenett, Hofer, (2010).

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, during our analytical process in our instructional context, we identified the need for additional criteria beyond those covered by the original framework. The analysis of the lessons is grounded in the content and activities presented within each lesson. The criteria for demonstrating various levels of appropriateness were based on those of Chung *et al.* (2024). Therefore, as shown in Table 4, we adapted the framework by augmenting and refining specific criteria. Specifically, we added considerations such as specifying lesson goals, delineating the teaching context, identifying the DDL tools used, articulating instructional strategies for integrating DDL tools, elucidating students' roles as language detectives, emphasizing skills development, and formulating an assessment framework tailored to the lesson plan's instructional objectives.

Two members of our research team individually rated each lesson plan. They later met to discuss their final scores, and the mean of these results was used in this work.

3.4.3 Case 2 (Vietnam)

Interviews

Chung *et al.*, (2024) conducted a series of interviews before and after the course to investigate participants' experiences with DDL. The pre-workshop interviews involved 13 participants, our permanent teachers and nine guest instructors, while the post-course interviews included 20 participants, featuring four permanent teachers, 14 guest instructors, and two university staff members. Only those who had submitted their lesson plans for review were invited to participate in the post-course interviews. The pre-workshop interviews aimed to gauge participants' prior knowledge and use of corpora, as well as their expectations for the training. In contrast, the post-course interviews concentrated on how participants intended to integrate corpora into their teaching practices and the potential challenges they might encounter. Due to scheduling constraints, the pre-workshop interviews occurred after participants completed the online training modules, but before they attended the workshops. Consequently, their responses reflected some prior exposure to corpora rather than a complete beginner perspective.

To ensure a structured and effective interview process, Chung et al. (2024) adopted the Interview Protocol Refinement (IPR) framework proposed by Castillo-Montoya (2016). This framework follows four key stages: (1) aligning interview questions with research objectives, (2) structuring an inquiry-based dialogue, (3) refining the interview guide based on feedback, and (4) testing the interview process. Interviews began with an overview of the study's goals, followed by open-ended questions addressing participants' perceptions, benefits, and challenges related to DDL and corpora use, both before and after the corpus literacy course. Researchers used follow-up and probing questions throughout the discussions to encourage more in-depth responses. All interviews were conducted in Vietnamese, and the responses were translated into English for analysis.

Chung *et al.* (2024) applied the In Vivo Coding method (Saldana, 2016) to analyse the collected interview data, a qualitative technique that extracts codes directly from participants' verbatim responses. This process involves identifying keywords or short phrases used by participants, which are then converted into codes that serve as organizational units for categorizing data into thematic groups. A defining characteristic of In Vivo Coding is its commitment to preserving participants' original expressions, ensuring that their voices remain central throughout the analysis. By preserving the authenticity of their language, this approach enables a deeper, more nuanced understanding of their lived experiences, thoughts, and emotions.

The resulting codes were subsequently structured into a comprehensive mind map that visually represented the frequency and distribution of themes raised by participants (Chung *et al.*, 2024). The mind map is divided into sections: the first identifies the central theme under investigation, the second presents subthemes that emerged from the interviews, the third lists recurring words within each theme along with their frequency of occurrence across the interviews, and the final section contains direct quotations from participants, preserving their original statements. This structure allows for a clear and systematic representation of how participants articulate their perspectives, highlighting key patterns and insights derived from the data.

3.5 ETHICAL CONSIDERATIONS

Ethical approval for this study was obtained from the ethics committees in Vietnam, Australia, and Brazil, ensuring compliance with international research standards and institutional guidelines. The study adhered to the following ethical measures to protect participant rights and ensure responsible research practices:

- Informed consent: Before participating in the study, all participants signed consent forms acknowledging their voluntary participation and the research's purpose.
- Anonymity and confidentiality: Participants' names, institutional affiliations, and any identifying details were anonymized to safeguard their identities.
- Voluntary participation: Teachers were informed that their participation was voluntary and that they could withdraw at any time without consequences.
- Data security: All research data were securely stored in encrypted databases, with access restricted to authorized researchers.

Given the sensitive nature of self-reported data, particularly regarding challenges in implementing DDL, special care was taken to ensure participants felt comfortable sharing their experiences without fear of professional or institutional repercussions. This study was submitted for ethical review and received approval from the Comitê de Ética em Pesquisa (Plataforma Brasil) under CAAE: 283225923.9.0000.5466 (Appendix A), ensuring compliance with rigorous ethical standards. This multi-country ethical approval framework fostered a safe and transparent research environment for all participants.

4 RESULTS

This chapter presents the findings from the two case studies conducted in Brazil and Vietnam, focusing on the impact of the DDL courses on teachers' corpus literacy skills, pedagogical adaptations, and implementation challenges. The results are organized into two sections: the Brazilian study (4.1) and the Vietnamese study (4.2).

This research adopts a comparative case study design, drawing on established discussions of case study traditions in education and applied linguistics (Yazan, 2015). It also follows a mixed-methods rationale, combining quantitative and qualitative sources to enable triangulation across data strands and a more contextualised account of teacher development and pedagogical uptake (Dawadi; Shrestha; Giri, 2021).

Both case studies examined how teachers engaged with corpus-based tools, integrated (or attempted to integrate) these resources into their lesson planning, and what factors influenced their ability to apply DDL strategies in practice. The data were analysed based on survey responses, lesson plan evaluations, and post-course interviews, enabling a comparison of teachers' experiences across two distinct educational contexts.

The chapter opens with the Brazilian case (Section 4.1), using interview accounts to trace how participants interpreted the course and the possibilities of corpus-based work in their everyday teaching (Section 4.1.1). It then turns to the lesson plans to examine how these ideas were translated, or resisted, at the level of pedagogical design (Section 4.1.2). The second half of the chapter shifts to the Vietnamese case (Section 4.2), where interviews provide context for teachers' responses to the bilingual programme and their reported learning trajectories (Section 4.2.1). Survey results are then used to capture broader post-course perceptions and self-assessed corpus literacy patterns (Section 4.2.2), before the lesson plans offer a final view of how DDL uptake was represented in planned classroom practice (Section 4.2.3).

4.1 RESULTS FROM THE BRAZILIAN STUDY

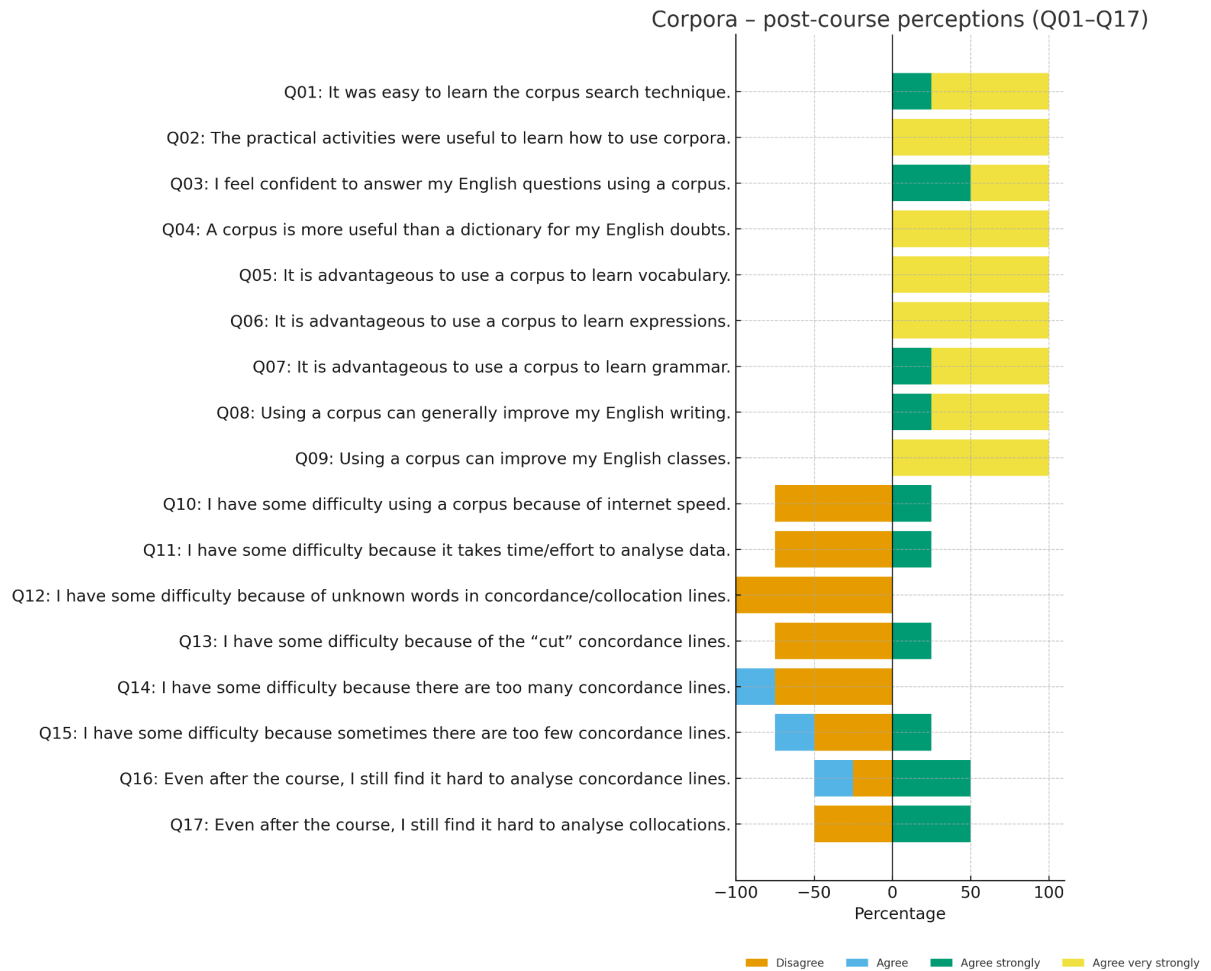
Post-course survey (Brazil)

This subsection reports Brazilian participants' post-course perceptions of corpus use, course delivery, preferred modules, and preferred tools. The post-course questionnaire comprised four sections: (1) Corpora—post-course perceptions (17 Likert-type items) gauging perceived usefulness of corpora for vocabulary, expressions, grammar, writing, and self-help, as well as common barriers such as time/effort, internet speed, unknown words, the “cut” nature of concordance lines, and the number of lines returned; (2) Course evaluation (12 Likert-type items) assessing satisfaction with synchronous and asynchronous methodology, organisation, clarity of explanations, and responsiveness of instructors/organisers; (3) Modules—most interesting (single-choice item) identifying which part of the pathway participants found most engaging; and (4) Tools—most liked (single-choice item) capturing preferences among the corpus tools introduced. Likert items used a six-point agreement scale (Disagree very strongly, Disagree strongly, Disagree, Agree, Agree strongly, Agree very strongly).

For readability, the charts display only the response options selected by at least one participant; categories with zero responses were omitted. Results are presented with diverging-bar and bar charts, followed by short interpretations that highlight implications for DDL implementation.

Figure 6 summarises the next section of the survey, the course-evaluation items (Q01–Q17). The underlying dataset (Excel file) for this figure is available in Appendix B, which provides a Google Drive link to all research data.

Figure 6 - Corpora post-course perceptions (Q01–Q17)

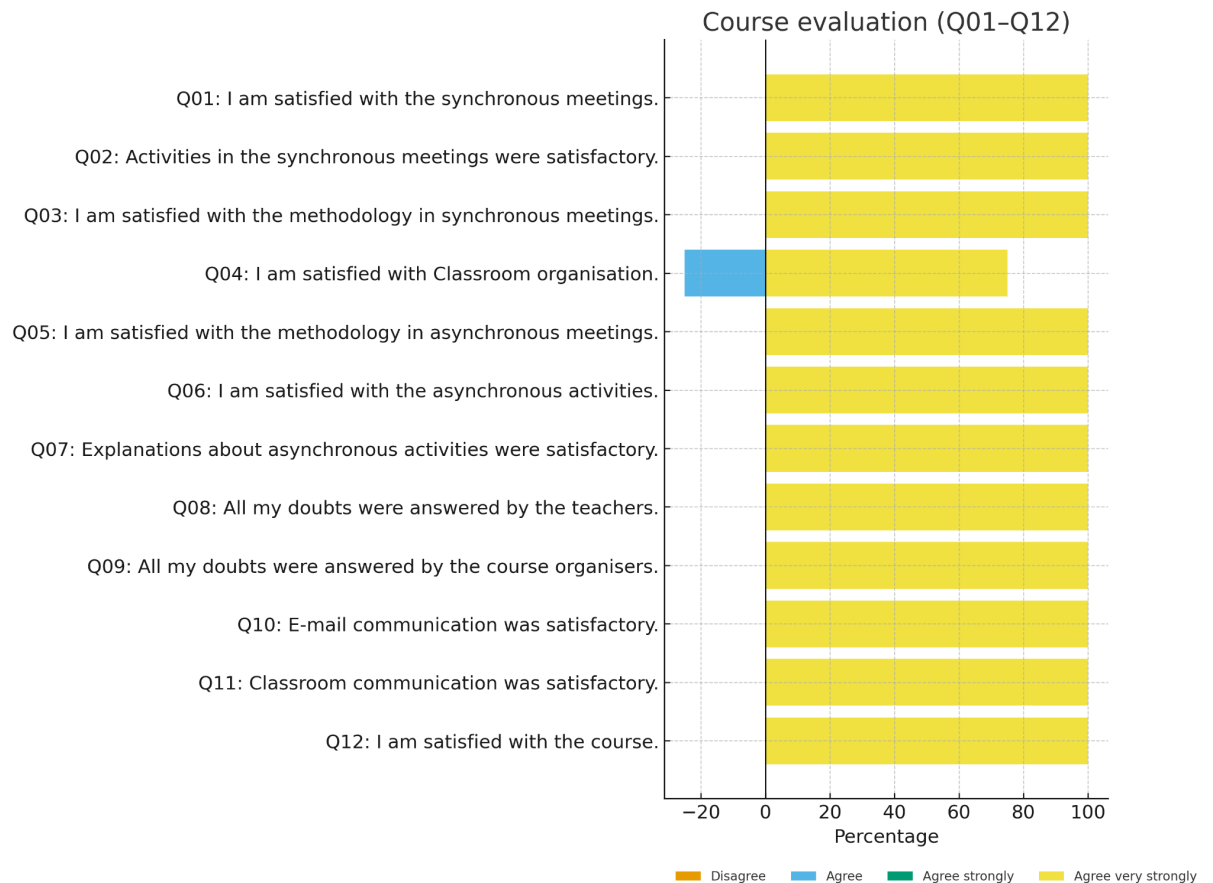


Source: Prepared by the author (2024).

Overall, participants agreed that corpora are useful for vocabulary, expressions, grammar, and writing, and several reported feeling confident using them to answer English-usage questions. At the same time, a notable minority still struggled with key aspects of concordance work. Reported challenges clustered around (a) the time/effort required to analyse data, (b) internet speed, (c) unknown words in concordance/collocation lines, (d) the “cut” nature of lines, and (e) the volume of lines returned (too many or too few). Even after the course, some participants still found it hard to interpret concordance lines and collocations, suggesting the need for additional scaffolding (e.g., worked examples, smaller samples early on, and brief interpretation checklists). Figure 7 summarizes the next section of the survey, the course-evaluation items (Q01–Q12). The underlying

dataset (Excel file) for this figure is available in Appendix B, which provides a Google Drive link to all research data.

Figure 7 - Course evaluation (Q01–Q12)



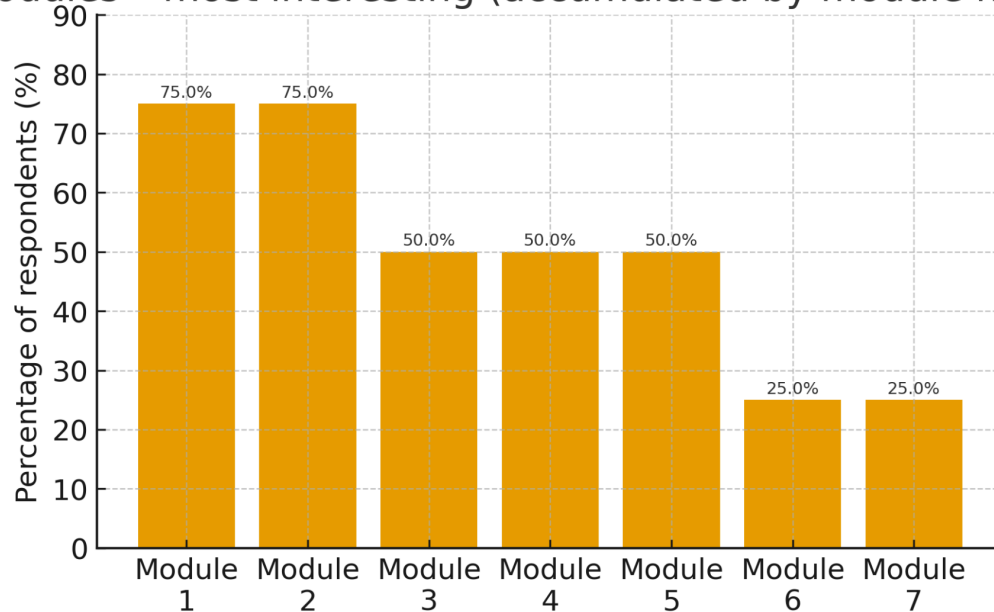
Source: Prepared by the author (2024).

Satisfaction was very high across components: synchronous meetings, the methodology of both synchronous and asynchronous activities, explanations for asynchronous work, and the responsiveness of instructors and organisers.

Overall satisfaction with the course was uniformly high, indicating that delivery and support met participants' expectations. The next graph, Figure 8, displays the distribution of selections for the "most interesting module(s)" question.

Figure 8 - Most interesting Modules

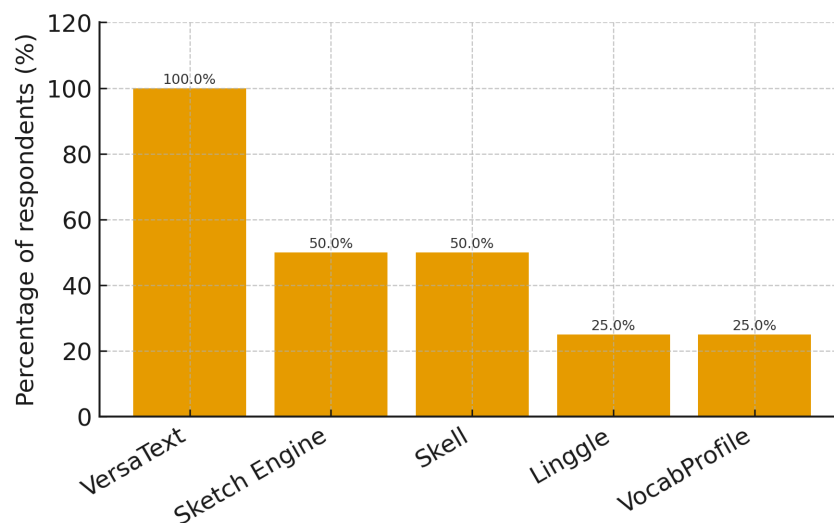
Modules – most interesting (accumulated by module number)



Source: Prepared by the author (2024).

Preferences were broadly balanced, with a consistent emphasis on the introductory workshop plus the first two writing-focused modules. This pattern is consistent with the Brazilian delivery: an instructor-led introduction to DDL and the online pathway was followed by self-paced study, during which Modules 1 and 2 of Improving Your Writing through Corpora (IWTC) were required, while completion of subsequent modules was optional. Finally, Figure 9 shows the tools rated as “most liked.”

Figure 9 - Most liked Tools



Source: Prepared by the author (2024).

VersaText emerged as the clear favorite, followed by moderate interest in Sketch Engine and SkELL, and lower but present interest in Linggle and VocabProfile. VersaText likely stood out because it is highly user-friendly, uses color to highlight patterns, and can be operated with just a few clicks, giving it a more playful, classroom-ready feel. This low barrier to entry and immediate visual feedback align with the needs of busy in-service teachers and help explain its top ranking.

Building on these survey insights, future iterations could simplify concordance and collocation work with brief worked examples, smaller starter datasets, and quick-reference guides. They could also strengthen procedural scaffolding with stepwise checklists from goal-setting to interpretation, and could expand low-stakes practice, so teachers rehearse analysis before full classroom integration. Lesson-plan supports could be refined through templated sequences and rubrics that make alignment among objectives, corpus tasks, and assessment more explicit, while tool pathways could be broadened to suit both low-tech contexts and richer platforms.

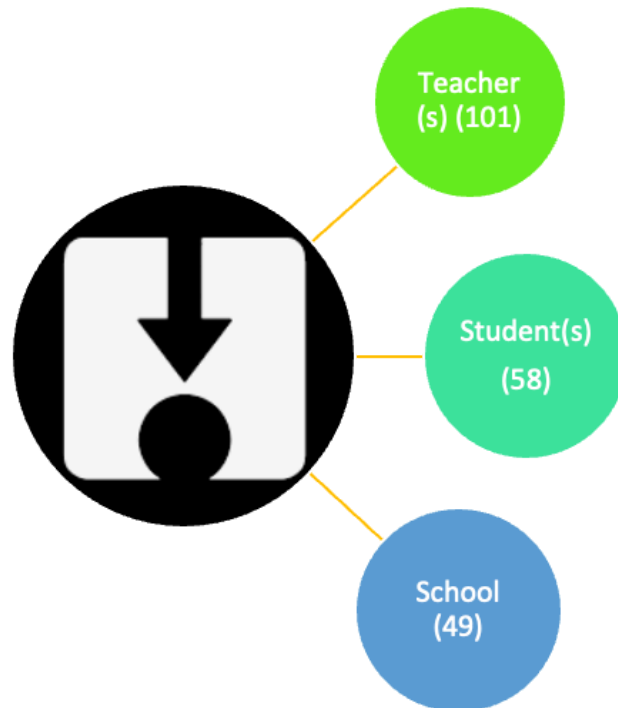
4.1.1 Interviews

We used LancsBox to identify the most frequent content words that would lead us to the most relevant themes discussed in the interviews, following Saldaña's (2016) approach. We worked with LancsBox v6.x (Brezina *et al.*, 2021), generated a wordlist by “types”, and filtered for content words only.

Furthermore, we exported the list and verified the top items against the interview contexts to avoid false positives (e.g., names, fillers). Our analytic logic links frequency to theme: recurring lexical items across all three interviewees indicate shared concerns/priorities, which we then interpreted through In Vivo Coding (Saldaña, 2016) to move from counts to meaning.

The most frequent words generated by the tool were: teacher/s (101); student/s (58); school (49); training (44); plan (44); lesson (40); tool (31); activity (25); and Versatext (24). Figure 10 shows the first top 3 words used in the interview.

Figure 10 – Frequency of top content words in interviews 1 (Teacher, Student and School)



Source: Prepared by the author (2024).

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

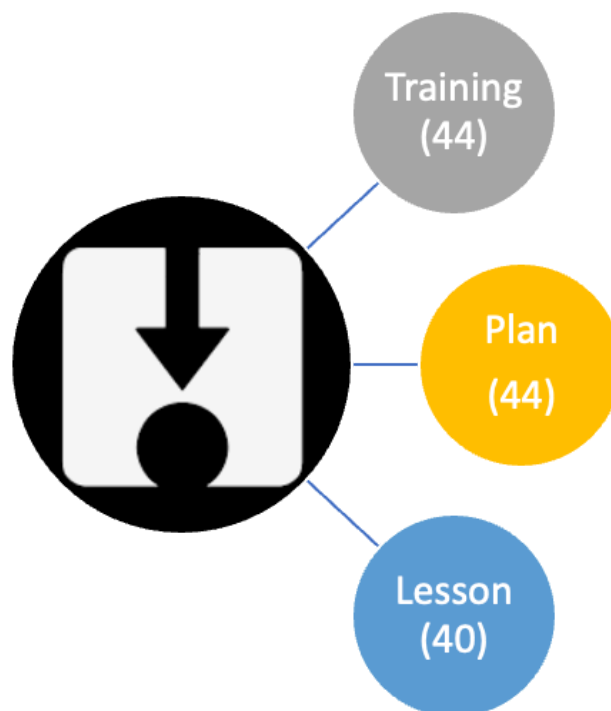
However, amidst these concerns, there was recognition of the critical importance of educators being adequately prepared and versatile. In their view, teachers are expected to incorporate new tools and methodologies into their instructional approaches, indicating a commitment to ongoing professional development and adaptability. This aligns with the high salience of teacher(s) in the wordlist, signalling a cross-participant focus on the profession's evolving demands and on workload/time-management as enabling or constraining factors.

"Students" are viewed as 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. We can see it in this 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?”*

Here, the prominence of student(s) in the frequency profile supports an interpretation of learners as agents, an insight we retained when framing the analytic categories and when linking tasks to observable student behaviours (e.g., argumentation, peer explanation). Concerns were raised about the level of preparedness and proficiency among teachers regarding the "school" system. The need for ongoing support and training to ensure quality education delivery is acknowledged. Flexibility and adaptability are identified as crucial factors in addressing challenges within the school environment, showing the importance of continuous professional development initiatives. The repeated mention of school also indicates that structural/institutional conditions (infrastructure, policies, time) are not merely background variables, but active drivers shaping what teachers can implement. As evidenced in the quote below by participant 2: *“Most schools have gotten rid of their computer labs”*. Figure 11 shows the next 3 most frequent words.

Figure 11 - Frequency of top content words in interviews 2 (Training, Plan and Lesson)



Source: Prepared by the author (2024).

According to the participants interviewed, “training” is an important element in preparing teachers to adeptly integrate new tools and methodologies into their teaching practices. The emphasis is on equipping educators with the requisite skills and knowledge to enhance their instructional strategies and effectively support student learning. This fact highlights the significance of ongoing professional development initiatives in ensuring educators remain abreast of best practices and pedagogical advancements.

Participant 2 statement emphasised the broader goal of training, not just in language pedagogy, but as an institutional objective, and highlighted the importance of disseminating the knowledge gained from the training to other teachers. The excerpt below illustrates how Participant 2’s statement frames dissemination as an institutional responsibility:

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 (Interview with participant 2, 2023).

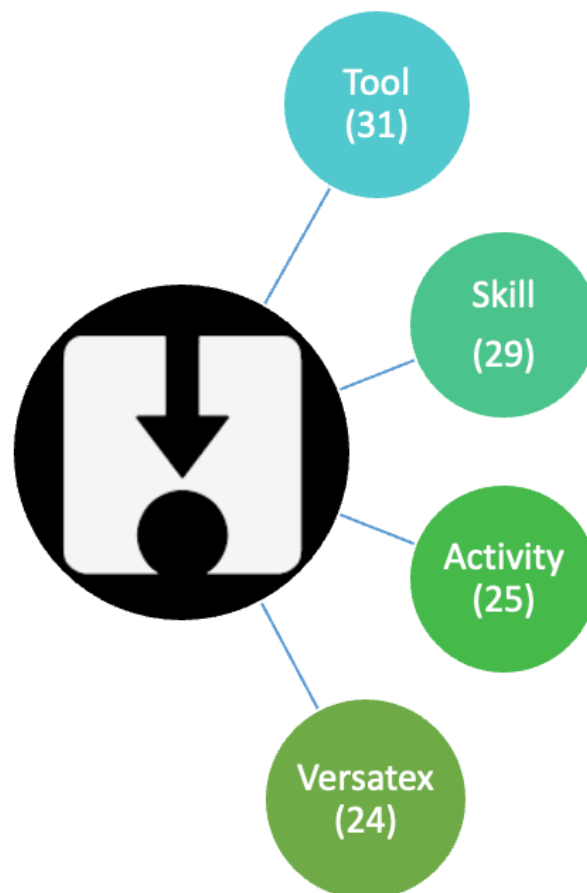
The term “lesson plan,” as encountered in the interviews, refers specifically to the meticulously crafted outline or blueprint that educators develop to structure their teaching activities within a particular class session. This encompasses detailed arrangements for teaching materials, learning objectives, 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 discussed, underscoring the importance of deliberate planning in facilitating student understanding and engagement.

In contrast, participants used the term plan more broadly in the interviews, particularly when referring to school-based and professional development contexts. In their accounts, plan could denote instructional design, curriculum-oriented decisions, and wider strategic actions. Rather than being limited to preparation for a

single lesson, it was often invoked to refer to overarching aims such as school improvement, curriculum coherence, or professional advancement initiatives.

Notably, this distinction may have arisen in the interviews because two of the three interviewees also served as coordinators. One of them had shown Versatext to her team and explained how the tool highlights relevant words in a text. However, she was unable to implement this tool at her school due to her lack of experience and confidence in using it. To make this contrast visible to the reader, we explicitly treat plan as the macro-level construct (strategic/curricular) and lesson plan as the micro-level artefact (class-session design with objectives, tasks, and assessment), an operational distinction that reappears in our coding. Figure 12 presents the last 3 most frequent words.

Figure 12 - Frequency of top content words in interviews 3 (Tool, Skill, Activity and Versatext)



Source: Prepared by the author (2024).

The interview highlights the importance of integrating educational “tools, “skills development, and engaging “activities” into effective teaching practices. It also

highlights that educators unanimously recognise 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. Thus, the interview highlights the importance of leveraging technology to cultivate dynamic classrooms and optimise student learning experiences.

Skills development emerges as a priority in educational planning and instruction. Teachers emphasise the importance of cultivating 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 prioritise 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 (Interview with participant 1, 2023).

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 maximising instructional effectiveness.

4.1.2 Lesson Plans

Table 5 presents the mean ratings of the lesson plans developed by participants using the adapted Technology Integration Observation Instrument (TIOI).

Table 5: 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)	Assessment	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: Prepared by the author (2024).

A careful look into each of the criteria showed that the explicit instructions for the use of DDL tools were somewhat unclear to be followed and understood by students. As an example of misalignment, excerpts 1 and 2 are labelled as "does not fit" since the instructions in the activities seem to lack information about their goals. We see this fact in the following examples:

1. *Go to SKELL, write the word that you don't know in 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 categorised as "fit together" because it is clear how the teacher guides students to use DDL tools. Another example is Lesson 4, in which the lesson goal seamlessly integrates with the planned activities, demonstrating a cohesive relationship between the overall goal and the instructional strategies employed.

4. General goal: *To present strategies for identifying 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;* iii) *Instigate language discovery (with a focus on vocabulary exercises) through data-driven activities* (Lesson 4).

It is also important to note that, as Examples 5 and 6 demonstrate, some evaluation procedures focus on evaluating students' overall performance rather than measuring their 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 specialised collocations in the field. The following skills will be worked on: understanding collocations and keywords (SKILLS 8 and 14)⁴ and manipulating frequencies (SKILL 7)⁵.* (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 their frequency of use, with examples to support the findings. The teacher explained it as follows:

⁴ Skill 8: understanding collocations. Skill 14: understanding keywords.

⁵ Skill 7: handling frequencies.

8. In activity 03, SKELL is explored as a way of expanding 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 analyses 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.

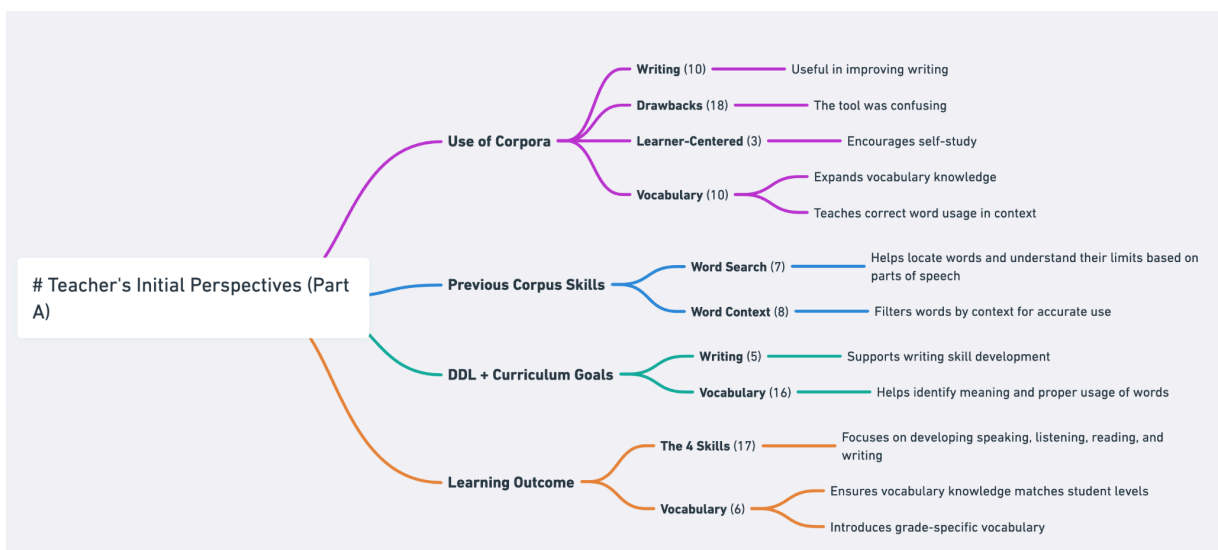
4.2 RESULTS FROM THE VIETNAMESE STUDY

4.2.1 Interviews

Pre-workshop interviews

The pre-training interviews, analysed using the In Vivo Coding method (Saldana, 2016), provided insights into teachers' expectations and concerns regarding corpus use in the classroom. The mind map created was then divided in two to facilitate the analyses. As illustrated in Figures 13 and 14, trainees discussed potential applications of corpora in lesson planning, the alignment of DDL with curriculum goals, and the skills they aimed to develop. While enthusiasm for corpus-based learning was evident, trainees also expressed challenges and uncertainties regarding its implementation (Chung *et al.*, 2024).

Figure 13 – Trainees' initial perspectives (Map part I)



Source: Adapted from Chung *et al.*, (2024).

As illustrated in Figure 13, trainees frequently discussed the potential applications of corpora in their teaching, with a particular focus on vocabulary acquisition and writing development. Many viewed corpora as valuable tools for expanding students' vocabulary and for introducing words in context, thereby promoting a data-driven approach to language learning (Chung *et al.*, 2024).

Vocabulary (10 mentions): Some trainees saw corpora as a useful resource for building students' vocabulary, helping them learn word meanings and usage in a structured way.

Example (1): *"I will use corpora to provide them with vocabulary and meaning; they will approach from the simplest, most basic words, and then, for example, they can explore self-learning."*

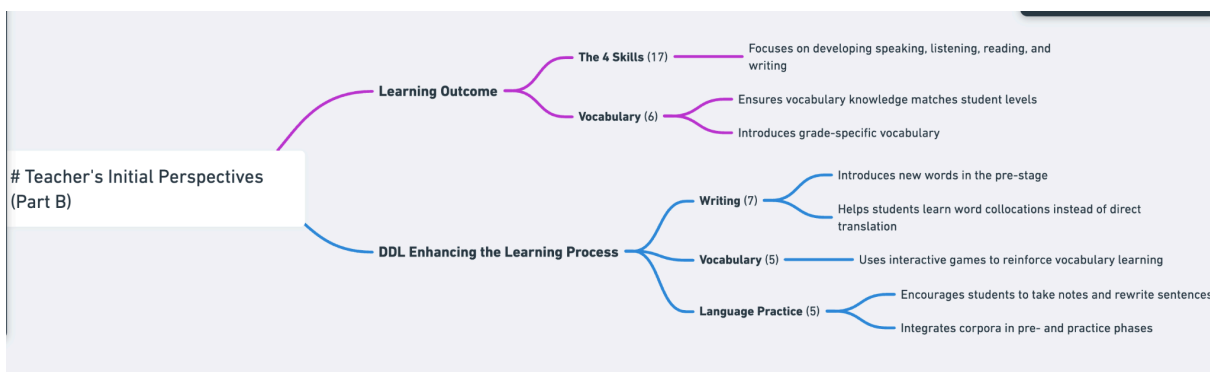
Writing (10 mentions): Trainees also linked corpus use to writing skill development, particularly in sentence structuring and word classification.

Example (2): *"The learning outcome, [...] the requirement of writing skills; when applying to writing in sentences, students must know what kind of word it is, so they can use corpora."*

The Four Skills (17 mentions): Some trainees anticipated that DDL could support all four language skills—listening, speaking, reading, and writing—particularly in standardized assessments.

Example (3): *"I see the Starters' learning outcome is based on the children's taking a test based on the Cambridge format, assessed by 4 skills: listening, speaking, reading, writing, so maybe DDL can help with this."*

Figure 14 – Trainees' initial perspectives (Map part II).



Source: Adapted from Chung *et al.*, (2024).

As illustrated in Figure 14, discussions focused on expected learning outcomes and how DDL could enhance the learning process. Trainees frequently discussed how DDL could contribute to overall language proficiency, particularly in supporting the four skills (listening, speaking, reading, and writing) and vocabulary development.

The Four Skills (17 mentions): Many trainees saw corpora as a resource for enhancing multiple language skills beyond just writing, aligning with standardized test preparation.

Vocabulary (6 mentions): Some participants highlighted the potential of corpora to introduce grade-specific vocabulary and ensure that students' lexical knowledge matches their learning levels.

Trainees also anticipated that DDL would help structure learning activities, making vocabulary acquisition and writing more effective.

Writing (7 mentions): Several trainees saw corpora as a tool for introducing new words, helping students understand word collocations rather than relying on direct translation.

Example: *"Introducing new words in the pre-stage helps students learn word collocations instead of direct translation."*

Vocabulary (5 mentions): Some trainees suggested using interactive games to reinforce vocabulary learning through corpus-based exercises.

Language Practice (5 mentions): Others discussed integrating corpora across different phases of lesson planning, particularly in pre-lesson activities and practice exercises.

Example: *"Students are encouraged to take notes, rewrite sentences, and integrate corpora in pre- and practice phases."*

The pre-training interviews (Chung *et al.*, 2024) indicate that trainees were optimistic about DDL's potential to enhance multiple language skills, particularly in vocabulary instruction and writing development. They envisioned corpora as a tool for reinforcing lexical knowledge, promoting interactive learning, and supporting structured language practice.

4.2.2 Post-course survey

The post-course survey was based on Ma *et al.* (2023) theoretical framework for assessing corpus literacy, which categorizes corpus-related competencies into five key areas: 1 - Understanding corpus structure; 2 - Searching for linguistic patterns; 3 - Analyzing concordance lines; 4 - Recognizing the pedagogical benefits of corpora; and 5 - Acknowledging their limitations. Our analysis demonstrates that Vietnamese primary EFL teachers exhibited strong self-reported proficiency across these dimensions, reinforcing the predicted link between corpus literacy and the intention to implement CBLP (Chung *et al.*, 2024).

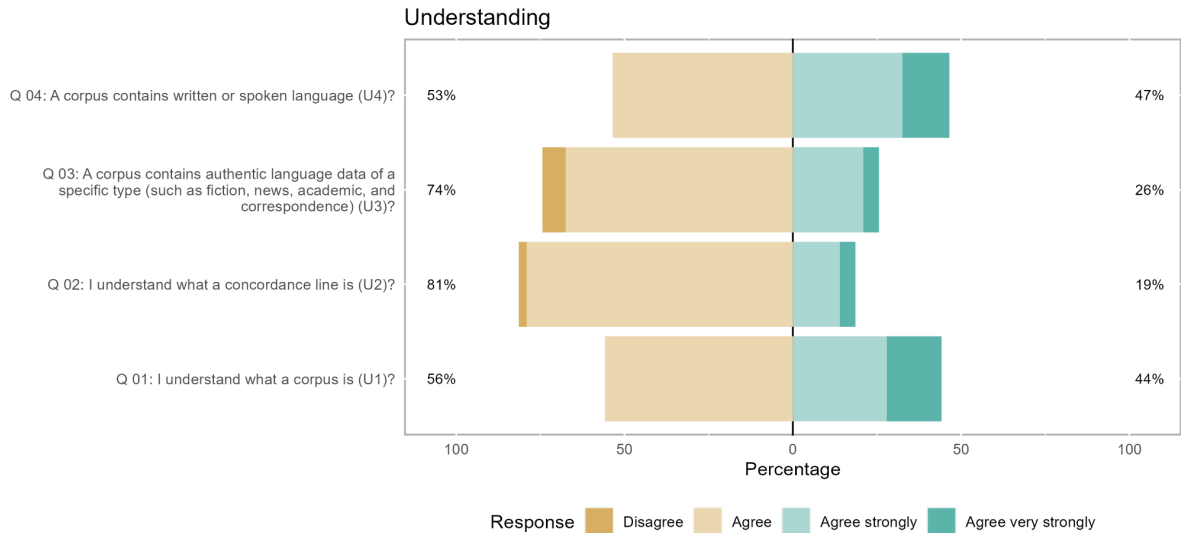
Figure 15 illustrates trainees' responses to the "Understanding" component of Ma *et al.* (2023) corpus literacy survey. The questionnaire was structured using self-questioning prompts, where statements were framed as questions (e.g., "I understand what a corpus is?") to encourage participants to actively reflect on their comprehension rather than simply recall information.

Each question in this section concludes with (U), indicating its association with "Understanding" as a core component of corpus literacy. This tagging system is used consistently across the instrument (e.g., (S) for "Search" and (A) for "Analysis"), providing a clear structure for organising responses and comparing results across components. Participants responded using a six-option agreement scale (Disagree very strongly, Disagree strongly, Disagree, Agree, Agree strongly, Agree very strongly). For readability, Figure 15 displays only the response options selected by at least one trainee; categories with zero responses were omitted.

Overall, trainees reported a strong post-course understanding of corpora across the four items, indicating that, from their perspective, the training reinforced their foundational grasp of what corpora are, how they are structured, and how they

can be used in both linguistic research and language teaching. This self-assessment suggests that participants experienced the course as effective in consolidating the conceptual knowledge required for corpus-based pedagogy (Chung *et al.*, 2024).

Figure 15: Trainees' understanding of corpora



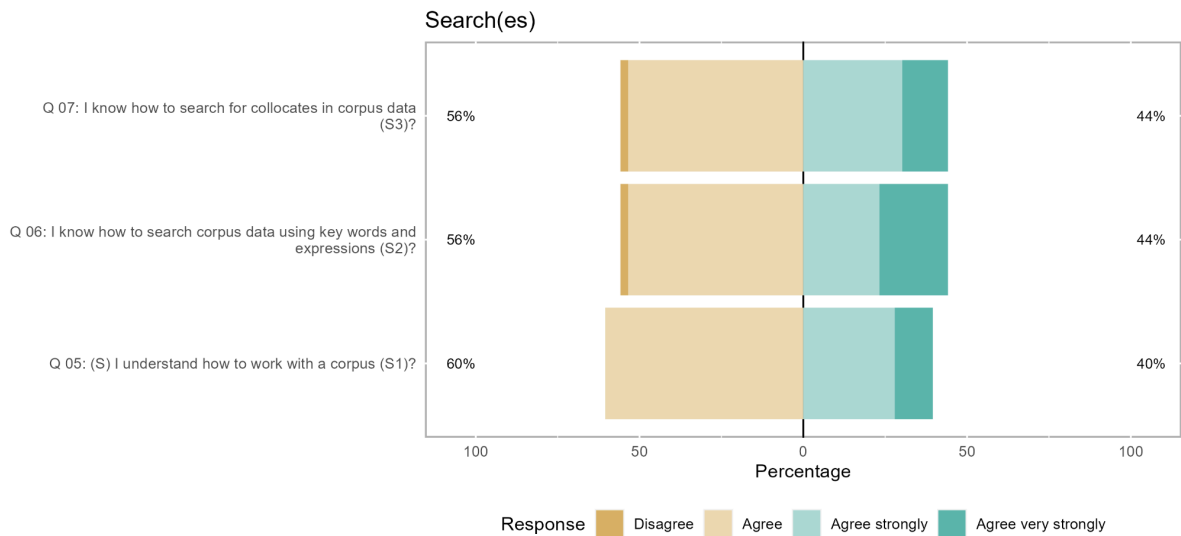
Source: adapted from Chung *et al.* (2024).

Figure 16 explores trainees' responses regarding their ability to query corpora (Search). Similar to their reported understanding of corpus structure, participants were overwhelmingly confident in their ability to search corpora for collocates, keywords, and expressions.

This finding suggests that trainees grasped the theoretical aspects of corpus querying and acquired practical skills in navigating corpus tools to retrieve meaningful linguistic data.

The strong self-reported proficiency in this area highlights the effectiveness of hands-on training exercises that enabled participants to practice formulating search queries and interpreting results (Chung *et al.*, 2024).

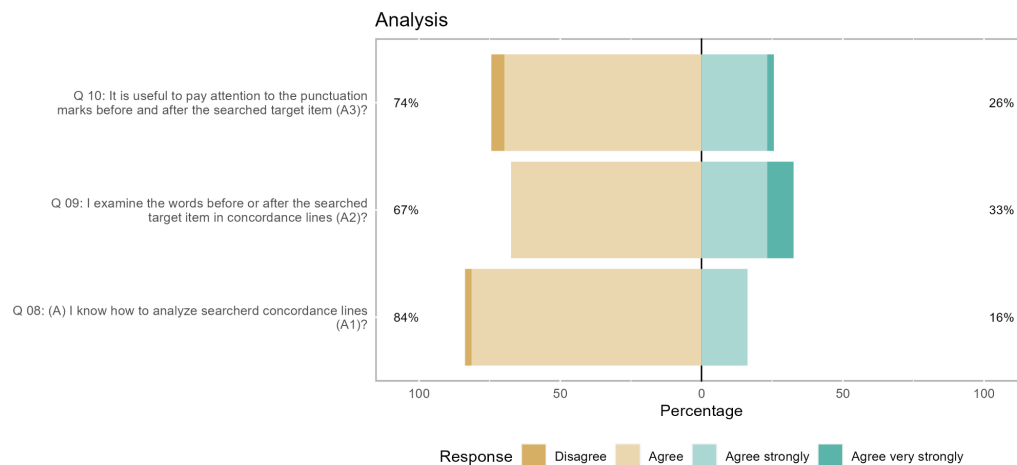
Figure 16: Trainees' corpus search abilities



Source: adapted from Chung *et al.* (2024).

Figure 17 shows that the ability to analyse concordance lines is a crucial skill for applying corpus findings in language pedagogy. Participants reported high confidence in reading and interpreting concordance lines, demonstrating an understanding of how to extract useful vocabulary patterns, phraseology, and syntactic structures from corpus data. Their self-assessment indicates that they are prepared to apply this skill in instructional settings, using concordance lines to support students in discovering authentic language patterns. The training's emphasis on direct engagement with corpus tools likely contributed to this positive outcome, reinforcing Ma *et al.* (2023) categorization of corpus literacy skills (Chung *et al.*, 2024).

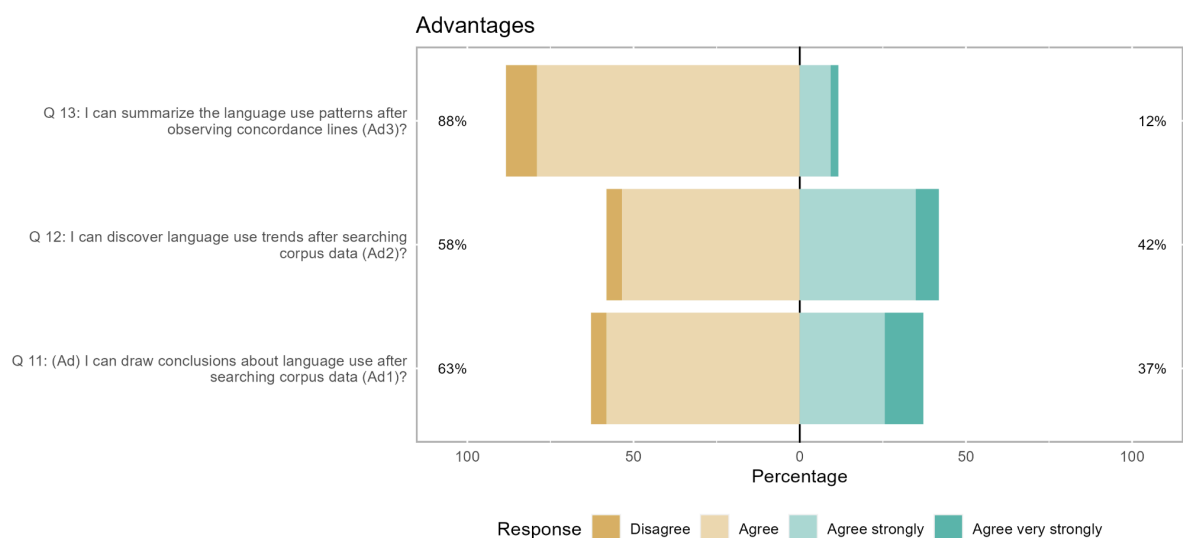
Figure 17 – Trainees' Analyzing concordance lines



Source: adapted from Chung *et al.* (2024).

Figures 18 and 19 detail trainees' perceptions of the pedagogical advantages of corpora, revealing agreement of over 90% across all six survey questions in this category. In Figure 18, we can see that participants widely recognized that corpora provide valuable insights into real-world language use, enabling learners to observe authentic linguistic patterns and develop data-driven insights into language structure and use. Additionally, they acknowledged that corpus-based activities can enhance language awareness and foster learner autonomy. This high level of agreement underscores a key outcome of the training: participants think that not only they developed technical corpus skills but also gained an appreciation for how corpora can enrich language teaching (Chung *et al.*, 2024).

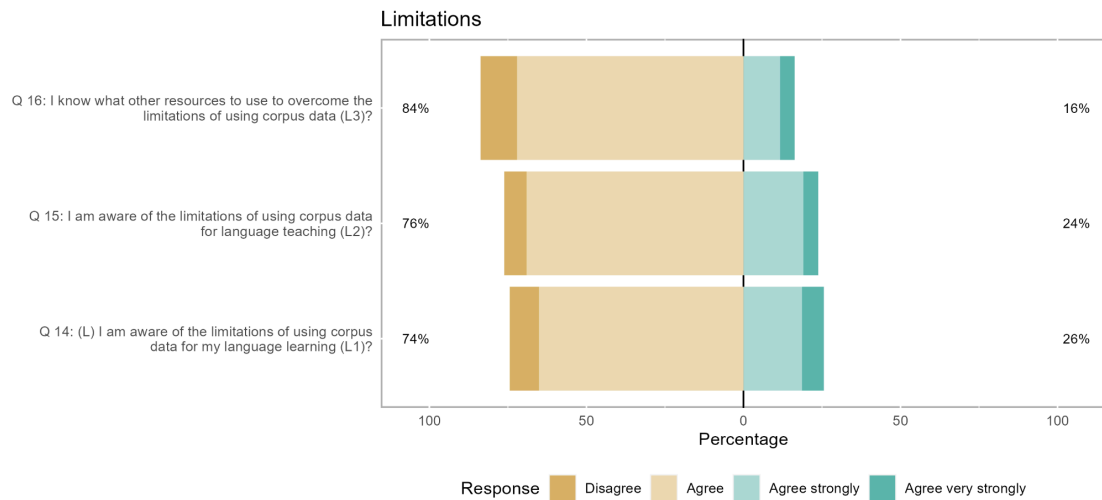
Figure 18: Trainees' perceptions of the advantages of corpora



Source: adapted from Chung *et al.* (2024).

Figure 19 shows that besides recognizing the benefits, participants also demonstrated awareness of the challenges associated with corpus use. The responses showed a similar level of agreement regarding the limitations of corpora in CBLP. Teachers acknowledged that working with corpora may require additional training, that corpus searches can sometimes yield overwhelming amounts of data, and that not all corpora are equally helpful for classroom contexts. Notably, the answers also identified strategies to mitigate these challenges, such as selecting pedagogically relevant corpora and supplementing corpus analysis with other instructional materials (Chung *et al.*, 2024).

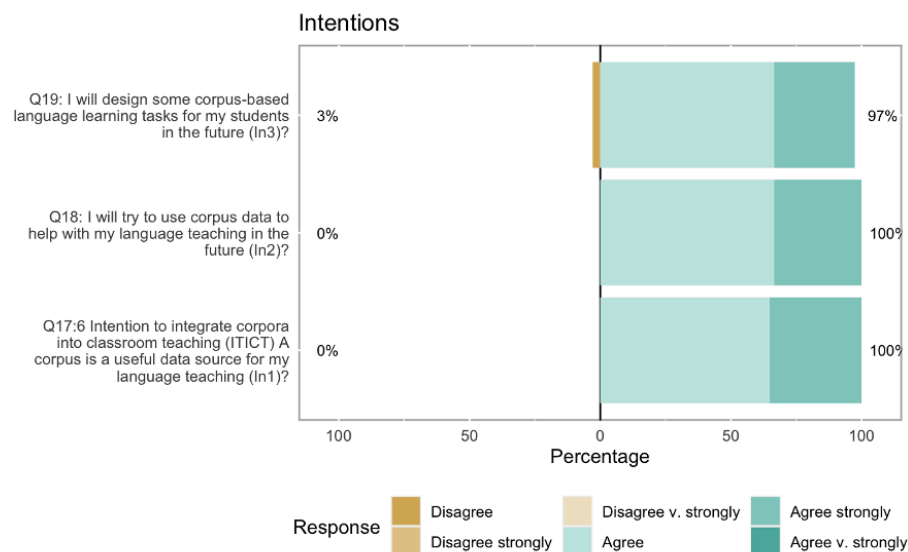
Figure 19 – Trainees’ perceptions of the advantages of corpora



Source: adapted from Chung *et al.* (2024).

Finally, Figure 20 illustrates the overwhelmingly positive responses regarding trainees’ intention to implement corpus-based language pedagogy in their future teaching. Nearly all participants expressed a strong commitment to applying DDL principles in their classrooms, supporting Ma *et al.* (2023) argument that increased corpus literacy correlates with a greater willingness to integrate corpora into teaching practice. This finding highlights the effectiveness of the training in fostering both confidence and motivation among participants, positioning them to incorporate corpus-based approaches into their instructional repertoire (Chung *et al.*, 2024).

Figure 20 – Trainees’ future intention to implement CBLP



Source: adapted from Chung *et al.* (2024).

Our analysis demonstrates that Vietnamese primary EFL teachers reported strong proficiency across these dimensions, reinforcing the predicted link between corpus literacy and the intention to implement corpus-based language pedagogy (CBLP). However, when comparing their self-assessments with their lesson plans and the practical application of these concepts, this confidence was not fully substantiated. Participants faced challenges in adapting and integrating DDL into their lesson plans, indicating a gap between perceived competence and actual implementation. This discrepancy highlights the need for further support and pedagogical scaffolding to facilitate the effective incorporation of corpus-based approaches into classroom practice.

Post-course interviews

Similar to the pre-course interviews, the post-course interviews were analyzed using the in-vivo coding method (Saldana, 2016). This approach enabled an in-depth understanding of trainees' perceptions through their own words, ensuring consistent data interpretation and facilitating the identification of recurring themes.

As illustrated in Figure 21, the post-course interviews (Chung *et al.*, 2024) revealed that trainees frequently discussed integrating corpus-based activities into their lesson planning, particularly for pattern identification. The significance of this approach, as outlined in Kennedy and Miceli (2010), was reinforced during training workshops. Trainees noted how corpus tools contributed to grammar accuracy and vocabulary development, ultimately making lesson design more structured and effective.

Pattern Finder (14 mentions): Many trainees highlighted the potential of corpora for identifying word patterns, supporting students in distinguishing subject-verb agreement, and recognizing collocations.

Example (1): *"The children do and distinguish which subject is used with 'want' and which subject is used with 'wants'; the subject he, she, it, they, we, I, the child must use the words 'like' with or without -s; there are many combinations behind the word 'break' and students can find the words that come after 'break up.'"*

Grammar Accuracy (11 mentions): Some trainees reported using corpus tools like Versatext and Corpus Mate to analyze grammar structures and refine student expressions.

Example (2): *"I use Versatext to analyze Nouns, Verbs, and Adjectives; Often students will ask me 'go visit' or 'go to visit', so [...] I show them examples in the Word sketch; I went to Corpus Mate to check the phrases that students used to see which is more accurate."*

Vocabulary Enhancement (8 mentions): Most trainees viewed corpora as valuable for vocabulary instruction, particularly in helping students explore word associations and improve lexical choices.

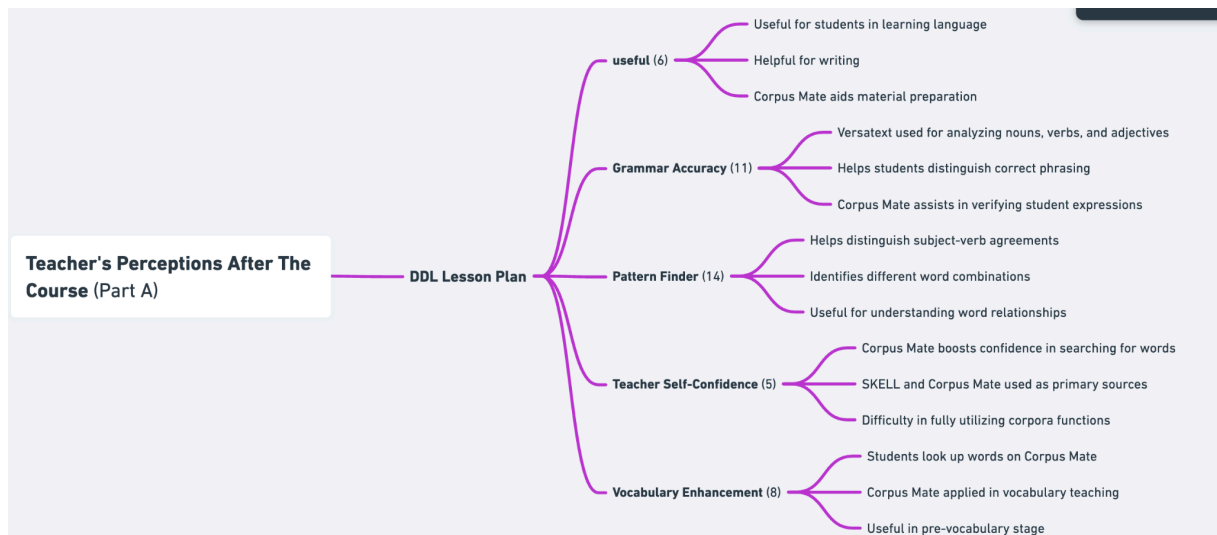
Example (3): *"The second activity is for students to use Corpus Mate to apply new vocabulary; then students will use Corpus to look up which words will go with those words; I only use Corpus Mate for teaching vocabulary."*

Teacher Self-Confidence (5 mentions): Some participants acknowledged an increase in their confidence in using corpora, yet others reported initial struggles in fully utilizing corpus functionalities.

Example (4): *"Corpus Mate boosts confidence in searching for words."*

Example (5): *"SKELL and Corpus Mate used as primary sources."*

Figure 21 – Trainees' perspectives after the course (Part A)



Source: Adapted from Chung *et al.*, (2024).

In Figure 22, the post-course interviews (Chung *et al.*, 2024) provided insights into how trainees approached the practical application of corpora in teaching, particularly the contrast between hands-on and hands-off DDL approaches. Participants emphasized the importance of structured implementation, highlighting different strategies based on their students' needs and levels of autonomy.

Hands-Off Approach (13 mentions): Many trainees preferred introducing corpus-based learning through pre-selected materials and printed worksheets. This allowed students to familiarize themselves with corpus-informed patterns before engaging directly with corpus tools.

Example (6): *"I will implement the hand off part first because the students are young, so I will do it first to let them understand; I will pre-select and I will print it out for students; designing hands off, prepare first at home and then go to class and distribute."*

Hands-On Approach (15 mentions): As students became more independent, teachers considered transitioning to hands-on DDL, guiding learners to explore corpora directly and identify linguistic patterns autonomously.

Example (7): *"I give students a link to Corpus Mate so that, when needed, students can access it, students can look the word up by themselves; it will be better than for students to do it themselves; I use hands on DDL, I will first instruct [students] how to use Corpus first. Then in the examples such as past simple or prepositions, students will look it up and search for prepositions."*

Use-Based Learning (6 mentions): Some trainees highlighted the role of corpora in helping students understand form, meaning, and usage through contextualized examples.

Example (8): *"For some children who are good and smart, they will certainly exploit more, discover a lot of things by themselves; I find Corpus Mate simpler to be used for KET learners, generally suitable for these learners; there is this corpus, it can help the students to discover for themselves, to find out the results to absorb is a good thing, an interesting thing."*

Practice (11 mentions): Teachers frequently used corpus tools to reinforce word structure understanding and support vocabulary application in writing tasks.

Example (9): *"While students prepare their answers, they can look it up on Corpus Mate to check their grammar and vocabulary points using 'compare results'. In short, students go to Corpus Mate to check their answers; As for the practice part, they will use SKELL, they will look for context, then they will check if the past simple is correct or not."*

Usefulness for Students (20 mentions): Many trainees recognized the educational benefits of corpora, particularly their ability to foster self-learning and independent exploration among students.

Example (10): *"I find that the attitude-behaviour part is that they're more interested and make fewer grammatical mistakes; I saw that they were very active, everyone was doing well; Students were very passionate, telling me 'please play again'."*

Despite the overwhelmingly positive reception, trainees also identified barriers to effectively implementing DDL in their teaching contexts.

Lack of IT Infrastructure (Example 11): A recurring concern was the lack of technological resources, particularly in rural areas, which limited students' access to corpus tools.

Example (11): *"As you know, the students living in the countryside don't have books as well as some other things so they do not have enough facilities to study; the problem is that the facilities in the district schools can certainly not be used, only in a few schools with adequate conditions."*

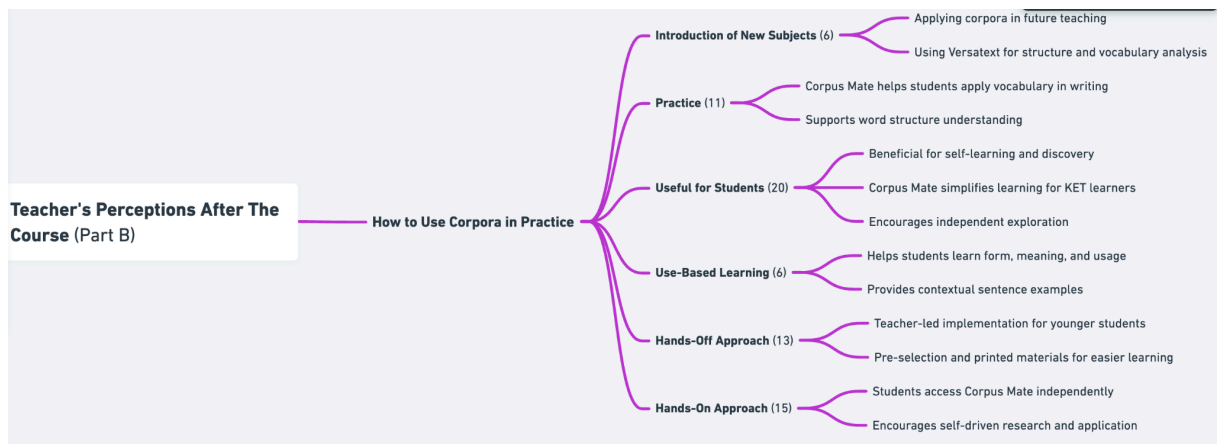
Teacher Mindset (Example 12): Some trainees noted that teachers' resistance to change was a significant challenge, as many doubted their students' ability to work with corpora.

Example (12): *"I see the only problem I have in mind is that teachers will always have the mindset that their students won't be able to do it; the main obstacle is the teacher because if the teacher has a change in thinking, the students will change accordingly."*

Confusion Regarding DDL Tools (Example 13): A few participants admitted to struggling with the overwhelming amount of data available in corpora, finding it difficult to filter and apply information effectively.

Example (13): *"I myself don't know it all yet and just looking for it, I'm not very interested; The data is immense. For example, just search for a word, it comes out a lot, it doesn't know which one it takes."*

Figure 22 – Trainees' perspectives after the course (Part B)



Source: Adapted from Chung *et al.* (2024).

The post-course interviews provided valuable insights into the trainees' experiences with corpus-based learning, highlighting both its pedagogical benefits and implementation challenges. Teachers recognized the potential of DDL for lesson planning, grammar instruction, and vocabulary development, as well as for fostering student autonomy and engagement. The structured integration of hands-off and hands-on DDL approaches enabled a gradual introduction of corpus tools, accommodating varying levels of student independence.

However, barriers to adoption remained evident, particularly regarding technological limitations, teacher readiness, and the complexity of corpus tools. While some teachers expressed increased confidence in using corpora, others struggled with the overwhelming amount of data and the time required for preparation. Additionally, limited access to IT infrastructure in certain schools posed significant obstacles to widespread DDL implementation.

4.2.3 Lesson Plan

As reported by Chung *et al.* (2024), 19 lesson plans were analysed, yielding 58 corpus-based activities, most of which were concentrated in the 'presentation' stage. Table 6 describes the elements of the trainee's lesson plan template.

Table 6 – Number and % of corpus activities per lesson plan component (n=19)

Lesson	Warm Up/ Review	Introduction	Presentation	Guided Practice	Evaluation	Total
1	1		1	1		3
2			1	1		2
3		1		2		3
4			1	3	1	5
5			1		1	2
6			1	4		5
7				2		2
8			1			1
9	1	1	1	1	1	5
10			1	3		4
11			1	3		4
12			1	3		4
13			1	1		2
14	2		1	1		4
15						
16			1	2		3
17			3	3	1	7
18				1		1
19				1		1
Total	4	2	16	32	4	58
Average %	6.90%	3.45%	27.59%	55.17%	6.90%	100.00%

Source: from Chung *et al.*, (2024).

Most activities involving corpora were found in the ‘presentation’ and ‘guided practice’ sections of the lesson plan, with trainees rarely including corpora in the other sections of the plan. Examples of corpus-based activities, taken verbatim from the relevant section of these lesson plans, are provided below.

Enabling Skills

Students will be able to use “some and any” correctly and know how to use DDL to check the usage of “some and any”
--

Use VERSATEXT in studying the grammar structure of present continuous tense.
--

Descriptions of the skills corpora could promote in the lesson plans were almost entirely absent (hence we did not include them in Table 3). In the two examples above, the trainees instead indicated the linguistic features to be acquired together with a (vague) description of how the selected tool (either CorpusMate/Versatext) could be used, e.g. ‘check the usage’, ‘study the grammar structure’.

Warm-up / Review

<i>Ask students (sts) to match the nouns on the left column with photos on the right column</i>

<i>- guide sts how use CorpusMate to check vocabulary</i>
--

Corpora were also rarely used in the ‘warm up/review’ section of the lesson plans. In this example, the trainee(s) produced a match task involving pictures with target vocabulary. The use of corpora is limited to ‘checking vocabulary’, albeit with little indication of the intended procedure, tool functions or instructional strategy involved in completing the task. This was considered as a ‘poor’ fit using the TIOI.

Introduction

Show the game and give instructions.

- *Get students to play the game individually*
- *Ask students to use Corpus Mate (Hands-on) to check the answers*
- *Instruct students how to check*

1. tell a lie or say a lie

*Check: * a lie*

2. worry about or worry for

*Check: worry **

Check: worry Pre

3. love playing or love play

*Check: love **

Check: love V

4. mind people asking or mind people ask

*Check: mind someone **

Check: mind someone V

Corpora were also rarely used in the 'introduction' section of the lesson plans. In this example, the trainee(s) use the CorpusMate platform, with students asked to check their answers to an introductory 'game' that explores alternatives for common expressions (e.g. tell vs. say a lie).

Ask sts to work in pairs: using hands-on DDL, CorpusMate

- 1. think of five interesting facts about your school.*
- 2. read Jake's description. What is his favourite subject? (Page 27)*
- 3. read the description again and answer the questions. (Page 27)*
- 4. find the conjunctions and, but and or in the description. Then use them to complete the explanations below. (Page 28)*
- 5. choose the correct conjunction in each sentence. (Page 28)*

Ask sts to use CorpusMate to look up for conjunctions (and, but, or) and some new words.

Here, the trainees use CorpusMate to consolidate a previous reading activity involving conjunctions as the target language for acquisition, as well as any unfamiliar vocabulary from the reading passage.

Presentation

The teacher will show pictures of people partying. Students will be asked with the following questions:

What time do people in your country usually hold a party?

What time do you usually go to a party?

What are the people doing in the pictures?

What are some party activities you know?

The teacher will now present and discuss the present continuous tense. Rules and sample sentences will be shown. Sample sentences may be lifted from corpus mate. From the previous activity, the teacher will ask the students to give again the verbs that they have mentioned on the activity in the present continuous tense.

Corpora were more frequently used to 'present' linguistic information to students. In this example, the teacher uses a standard present-practice-produce (PPP) approach to the presentation of verbs in present continuous tense, using hands-off concordances taken from CorpusMate.

Guided practice

Task 5:

- The teacher asks students to reference the sample text in part A and use the ideas they discussed in Task 4 in order to write a paragraph of 100 words about their school board game.

- The teacher asks students to use CorpusMate to check vocabulary like phrasal verbs, collocations, and how words are properly used while they are doing the task (pattern hunting).

Task 6: The teacher models or shows students how to look up the words "participate", "take*", etc. in CorpusMate to see what words will go after them to describe the school game.*

- The teacher monitors and offers students assistance if needed.

Corpora were also frequently employed in teacher-led guided practice activities within the lesson plan structure. The example above shows how the trainees incorporated CorpusMate into a structured paragraph writing task, with the teacher monitoring and providing assistance to students during this process.

Evaluation / Assignment

Self-correction (5'): *ask sts to check their first draft individually, resolve grammatical or lexical mistakes and errors by using compare results, pattern finder.*

In CorpusMate, query the word(s) they need, click Distribution by topic to check the Written/spoken distribution of documents]

Corpora were rarely used in the lesson plans at this stage. In this example, the trainees employ CorpusMate (together with accompanying guidance on which functions to use) to produce a short piece of writing across multiple drafts, using the corpus tool to resolve errors for subsequent revision (Chung *et al.*, 2024).

While trainees were generally able to incorporate corpus-based activities into their lesson plans, their usage remained largely confined to the 'presentation' and 'guided practice' stages, leaving other potential applications underexplored. However, the effectiveness and depth of these attempts warrant further examination. As shown in Table 7, the mean ratings of the submitted lesson plans, assessed using the TIOI (Harris; Grandgenett; Hofer, 2010) rubric, revealed a low overall alignment with the established evaluation criteria (Chung *et al.*, 2024). Despite the overwhelmingly positive self-reported perceptions documented in the Post-course survey (Ma *et al.* 2023), a closer analysis of the lesson plans demonstrated limited integration of DDL principles, with only four lesson plans achieving a TIOI (Harris; Grandgenett; Hofer, 2010) score of 3 and just one plan reaching a score of 4. This disparity between perception and practice, analyzing the quality of implementation and the challenges that may have hindered more comprehensive integration of corpus-based learning strategies (Chung *et al.*, 2024).

Table 7: Mean TIOI ratings for each submitted lesson plan

Lesson plan	Curriculum Goals & Technologies	Instructional Strategies and Technologies	Technologies Selection (s)	FIT	TIOI Level
1	1	1,25	1,5	1,5	Does not fit
2	1,5	1,5	1,25	1,25	Does not fit
3	1,5	2,25	2,5	2,25	Fits somewhat
4	2,5	3	3,5	3	Fits together
5	2,5	2,5	3	2,5	Fits somewhat
6	1	1	1,5	1,25	Does not fit
7	1,5	2,5	2,5	2,5	Fits somewhat
8	1,5	2,25	2	2,25	Fits somewhat
9	1,75	4	4	3	Fits together
10	1,5	1,5	1,75	1,75	Does not fit
11	1	1	1,25	1,25	Does not fit
12	2	2,75	2	2,75	Fits somewhat
13	1,5	2,75	3	2,75	Fits somewhat
14	1,5	3	3	3	Fits together
15	1	2,5	2,25	2,25	Fits somewhat
16	1,5	2,25	3	2,5	Fits somewhat
17	4	4	4	4	Fits together strongly
18	2,75	2,25	2,25	2	Fits somewhat
19	2,5	3	3	3	Fits together
Average	1.86	2.50	2.61	2.46	

Source: Chung *et al.*, (2024).

Notably, trainees rarely connected their use of corpora within their lesson plan learning objectives, 'enabling skills', or even the section of the template marked

‘corpus literacy skills’, resulting in poor scores for ‘curriculum goals and technologies’. Most entries followed the example below, simply listing the corpus tools rather than the skills involved:

Learning Objectives: *Students will be able to recognize, name some common actions and say what they and other people are doing at the moment.*

Enabling Skills: *Listening, Speaking.*

Corpus literacy (skills): *SKELL, Versa text*

In the higher scoring lesson plans, trainees explicitly outlined how the corpus tools would be used to address which linguistic targets, although, as mentioned, this practice was rare.

Learning Objectives: *Students will be able to:*

1. *identify the verbs in present continuous tense*
1. *use Versatext to recognize present continuous structure*
1. *learn Vocabulary related to telling time*
1. *use present continuous terms to talk about activities in progress*

Enabling Skills:

Article Reading – Reading Comprehension Skills

Vocabulary Enhancement - Party Vocabulary

Corpus literacy (skills):

Use VERSATEXT in studying the grammar structure of present continuous tense.

Overall, the trainees performed better for both ‘instructional strategies and technologies’ and ‘technology selections’ criteria, with scores between the ‘minimally supports’ and ‘supports’ rating for the former, and ‘marginally appropriate’ and ‘appropriate but not exemplary’ for the latter. For “instructional strategies...” a higher TIOI (Harris; Grandgenett; Hofer, 2010) example is seen below, including the linguistic target, the corpus data (in hands-off format) and the proposed queries to be conducted, although more guidance as to how to derive the ‘correct form’ could have been provided:

EXERCISE 1: Write “want” or “wants”

**Hands off DDL: give Sts results from Corpusmate and ask them find out: the correct form of want/ wants with singular and plural nouns.*

In the next example, the trainee has produced a viable attempt to help students understand the differences between written and spoken registers. The trainee clearly indicates the learning goal (self-correction), the tool (CorpusMate), and the step-by-step procedure and functions of the corpus tool to be used. This indicates a high level of ‘fit’ between goals, instruction, and technology, as per the TIOI (Harris; Grandgenett; Hofer, 2010).

Self-correction (5’): ask sts to check their first draft individually, resolve grammatical or lexical mistakes and errors by using compare results, pattern finder. In CorpusMate, query the word(s) they need, click Distribution by topic to check the Written/spoken distribution of documents

Many trainees however simply listed the corpus tool to be used without accompanying information as to the linguistic target or query parameters, often in the form of using corpora to ‘check vocabulary/grammar’:

- T asks Ss to work in pairs, ask and answer the question “What do you have in your pencil case?”*
- T guide Ss to use Corpus mate to check the vocabulary and use them correctly.*
- T invites some pairs to orally present in front of the class.*
- T gives feedbacks*

- Write the sts’ answers on the board - Underline the words “some” and “any” - Give more examples with “some and any” - Introduce sts how to use Corpus Mate on-hand - Divide the class into 4 groups - Ask sts to give feedback - Conclude the usage of “some and any” and explain more if necessary

Regarding ‘technology selections’, higher scoring trainees were able to identify how a given corpus tool’s main functions would enable them to complete the task at hand, as exemplified in the following example:

The teacher shows students how to put their first paragraph in the versatext, click wordcloud to detect the repeated words. After that, the teacher advises students to search for synonyms in thesaurus to avoid repetition mistakes.

However, the trainees often selected corpora that would have been inappropriate for the target age group/proficiency, such as the BAWE corpus within SketchEngine. While they were trained briefly on the use of this tool as part of the training materials, that they believed it would be suitable for the target students was disappointing:

*Have Ss read the text and underline several new vocabularies:
Species, population, become extinct, be at risk of, organization, affect, destroy
Show Ss how to use BAWE corpus to check the meaning of new vocabularies
Show Ss some examples of sentences including ‘species’ to familiarize them with using BAWE corpus.
Give a mini sentence completion exercise to make sure students’ understanding of the new vocabularies.*

Although the participants reported feeling prepared to integrate DDL into their classrooms, the analysis of lesson plans tells a different story. The findings indicate that while teachers successfully incorporated corpus-based elements in some stages — particularly presentation and guided practice — they struggled to embed corpus methodologies consistently across all instructional components. Most activities lacked clear scaffolding, leaving students with tools but without structured guidance on how to analyse and interpret corpus data effectively. Furthermore, corpus-based learning was rarely applied in the warm-up, review, and assessment stages, highlighting a disconnect between theoretical knowledge and its practical application. This suggests that, despite their perceived readiness, teachers faced

significant challenges in translating DDL principles into coherent, pedagogically sound lesson designs, reinforcing the need for further structured training and sustained support to bridge this gap.

5 DISCUSSION

This chapter discusses the key findings from the Brazilian and Vietnamese case studies, comparing teacher experiences, implementation challenges, and the feasibility of DDL integration. It also introduces practical solutions to address corpus literacy gaps, including the DDL Checker AI assistant and the proposal for Skill 19: Developing Effective Prompts for Corpus Exploration. Additionally, it examines the pedagogical implications of these findings for teacher development programs and corpus-based pedagogy. The study also proposes modifications to the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) to better assess the depth of corpus integration in language teaching, ensuring that both technological and pedagogical aspects are effectively evaluated. The discussion lays the foundation for the conclusions and future directions presented in the next chapter.

The chapter first brings the two cases into dialogue, highlighting convergences and contrasts in how teachers understood DDL, where they struggled, and what enabled uptake across contexts (Section 5.1). It then turns to the gaps identified in the findings and presents the proposed responses, with particular attention to AI supported scaffolding through the DDL Checker and the addition of Skill 19 as a prompt focused corpus competency (Section 5.2). Finally, it draws out the implications for theory and practice, considering what these results suggest for teacher learning, program design, and the evaluation of corpus based pedagogy through tools such as the adapted TIOI framework (Section 5.3).

5.1 COMPARISON OF KEY FINDINGS

Findings from Brazil and Vietnam reveal both shared patterns and distinct challenges in implementing DDL in teacher education contexts. One of the most significant findings concerns the gap between teachers' self-reported corpus literacy skills and their actual implementation ability. In Vietnam, teacher trainees expressed high confidence in their understanding of corpus tools but struggled to effectively integrate them into lesson plans.

The lesson plan analysis (see Table 8) showed that most plans "somewhat fit" with the instructional goals but lacked a strong pedagogical foundation, aligning with findings from Ma *et al.* (2023). The proposed modifications to the TIOI (Harris;

Grandgenett; Hofer, 2010) aim to address this issue by offering a more precise framework for evaluating how teachers align corpus-based activities with curriculum goals and instructional strategies. These refinements ensure that corpus literacy is assessed not just in terms of familiarity with tools, but also in terms of practical application and integration into effective teaching methodologies.

Table 8 - Brazil and Vietnam TIOI results

Criteria (Original TIOI)	Vietnam (TIOI Original)	Vietnam Fit Classification	Criteria (Adapted TIOI)	Brazil (Adapted TIOI)	Fit Classification
Curriculum Goals & Technologies	1.86	Does not fit	Lesson Goals	2.4	Fits somewhat
Instructional Strategies and Technologies	2.50	Fits somewhat	Teaching Context	2.8	Fits somewhat
Technology Selection	2.61	Fits somewhat	DDL Tool Selection	2.5	Fits somewhat
Overall Fit	2.46	Fits somewhat	Instructional Strategies	2.0	Fits somewhat
xxx	xxx	xxx	Students' Role as Detective	2.7	Fits somewhat
xxx	xxx	xxx	Corpus Skills	2.1	Fits somewhat
xxx	xxx	xxx	Assessment	2.0	Fits somewhat
xxx	xxx	xxx	Overall Fit	2.5	Fits somewhat
Average	2.36	Fits somewhat	Average	2.43	Fits somewhat

Adapted from Technology Integration Observation Instrument (TIOI), (Harris, Grandgenett, Hofer, (2010).

In Brazil, participants showed similar enthusiasm for corpus-based teaching but faced greater difficulties translating knowledge into classroom practice. The lesson plan evaluation using the adapted TIOI (Harris; Grandgenett; Hofer, 2010) framework (Table 9) indicated that while teachers selected appropriate tools, they struggled with instructional strategies and aligning corpus-based activities with BNCC guidelines. This suggests that beyond corpus literacy, teachers require more structured guidance on how to effectively implement corpus methodologies in their teaching context.

Despite the potential benefits of DDL, both Brazil and Vietnam faced significant contextual and institutional obstacles that hindered the long-term integration of corpus-based pedagogy into classroom practice.

In Vietnam, limited internet access and a general lack of familiarity with digital tools posed major challenges. While teachers engaged actively with corpus-based methods during training, sustaining these practices independently proved difficult without the necessary technological infrastructure and continued support. The rigid curriculum structure and emphasis on standardized assessment further discouraged teachers from experimenting with new methodologies in their classrooms.

Meanwhile, in Brazil, teachers enjoyed greater autonomy in lesson planning, but this flexibility came with its own set of challenges. Heavy administrative workloads, scarce technological resources, and institutional resistance to adopting new tools made it difficult for educators to consistently apply DDL techniques. Unlike their Vietnamese counterparts, Brazilian teachers were not bound by strict curriculum requirements, yet they struggled with a heavy workload and the need for technical training to confidently implement corpus-based approaches.

These findings align with Chung *et al.* (2024), who highlight teacher reluctance to adopt CALL tools due to a combination of logistical constraints, technological limitations, and professional workload. The reality in both countries suggests that access to corpus tools alone is not enough—without institutional backing and well-defined implementation strategies, teachers face considerable barriers to adopting new methodologies in their daily practice.

5.2 ADDRESSING THE GAPS: AI AND NEW CORPUS SKILLS

As we analysed the data from both Brazil and Vietnam, it became increasingly clear that one of the major obstacles to implementing DDL in real classrooms was the gap between theory and practice. While teachers demonstrated engagement with corpus-based methodologies during training, their ability to apply these concepts in a meaningful and sustained way was often hindered by technical difficulties, lack of structured support, and limited professional development opportunities.

Recognizing these challenges and considering the rapid advancements in artificial intelligence (AI) for education, we sought to develop a practical solution to bridge this gap. This led to the creation of the DDL Checker, an AI-powered assistant designed to help teachers implement corpus-based approaches more effectively. The

tool was conceived to provide immediate, individualized assistance to teachers struggling with corpus literacy, ensuring that they could receive guidance beyond the initial phase of the course.

At the time of its development, the DDL Checker was built using the ChatGPT API, allowing it to provide real-time explanations of key corpus concepts, assist teachers in formulating effective corpus queries, and guide lesson planning by suggesting contextually relevant corpus-based activities. However, since the API operates on a pay-per-character model, the bot⁶ remains a proof-of-concept rather than a fully deployed solution. Due to these financial constraints, the project is currently positioned as an exploratory initiative for future research rather than an immediately scalable tool. Nevertheless, it serves as an example of how AI can support language teachers in integrating DDL methodologies into their pedagogical practices.

To ensure that the DDL Checker could effectively assist teachers, it was trained using frequently asked questions and common difficulties encountered by teachers and students when first learning about DDL. These questions were compiled from observations of DDL course programs and direct input from participants. Additionally, the bot's knowledge was enriched through a curated dataset of academic articles on corpus linguistics, teacher professional development, and data-driven language-learning methodologies. By integrating these sources, the AI assistant was designed to address both conceptual and practical concerns, making it a valuable tool for teachers at different levels of corpus literacy.

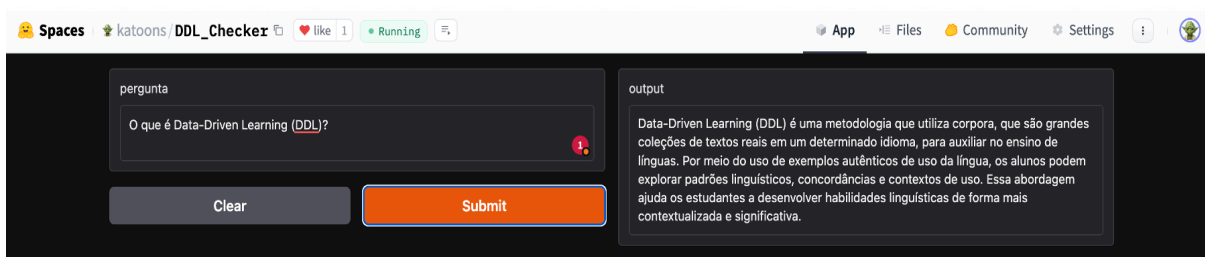
One of the key features of the DDL Checker is its ability to generate practical examples of how DDL can be implemented in the classroom. Teachers can ask specific questions about corpus analysis, concordance lines, collocations, or lesson planning, and the bot provides clear, structured guidance. These responses often include step-by-step instructions, example lesson plans, and suggested corpus tools, helping educators visualize the application of corpus methodologies in real teaching scenarios.

The DDL Checker interface (Figure 23) was designed to be simple and user-friendly, ensuring accessibility for teachers with varying levels of technological expertise. The interface provides a chat-based interaction model, allowing teachers

⁶ A bot is an AI-driven software application that automates tasks or simulates human-like interactions, often using natural language processing (NLP) to respond to user inputs in real time.

to submit queries and receive instant responses tailored to their specific teaching contexts. Currently, the DDL Checker supports both English and Brazilian Portuguese, allowing educators to interact in their preferred language. In the future, we aim to expand its capabilities to support all languages, making it a truly universal tool for corpus-based teaching.

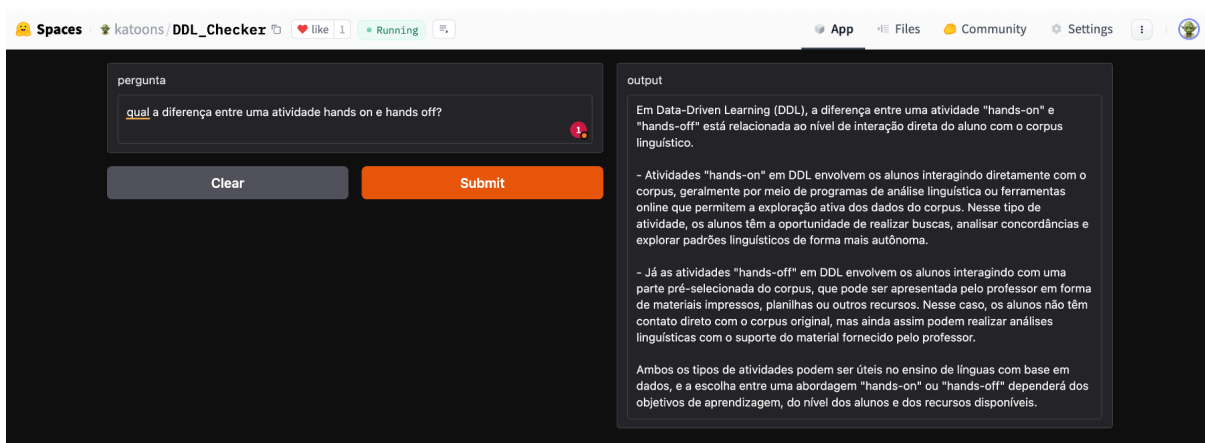
Figure 23 - DDL Checker interface



Source: Prepared by the author (2024).

Through its interactive design, the DDL Checker allows teachers to refine their DDL implementation strategies (as seen in Figure 24) by offering personalized recommendations and explanations in real time. Unlike static instructional materials, this AI-driven tool can adapt to the teacher's individual needs, ensuring that guidance is relevant to their specific classroom challenges.

Figure 24 – Example of interaction with DDL Checker

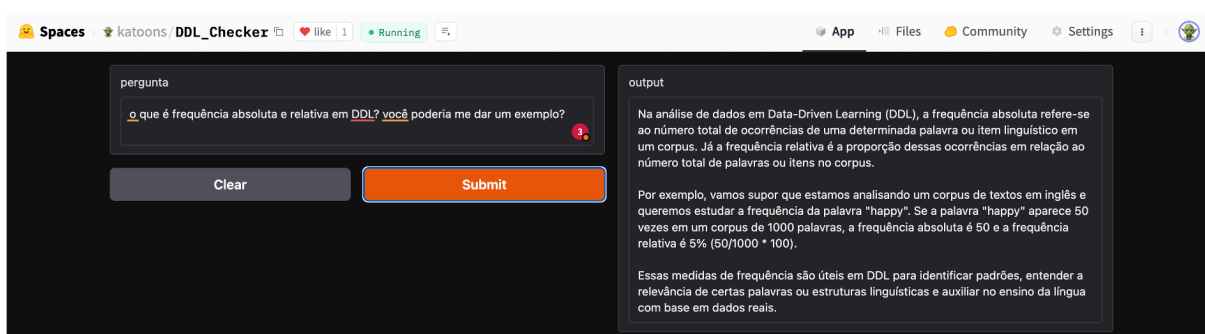


Source: Prepared by the author (2024).

Another advantage of the DDL Checker is its ability to help teachers overcome challenges in formulating corpus searches. Many educators struggle to construct

effective search queries for corpus tools, limiting their ability to extract meaningful data and apply statistical concepts. In Figure 25, the AI assistant explains the concepts of absolute and relative frequency. It can also suggest optimized search queries, guiding teachers in refining their corpus searches and better interpreting linguistic patterns.

Figure 25 – DDL Checker explaining the concept of absolute and relative frequency

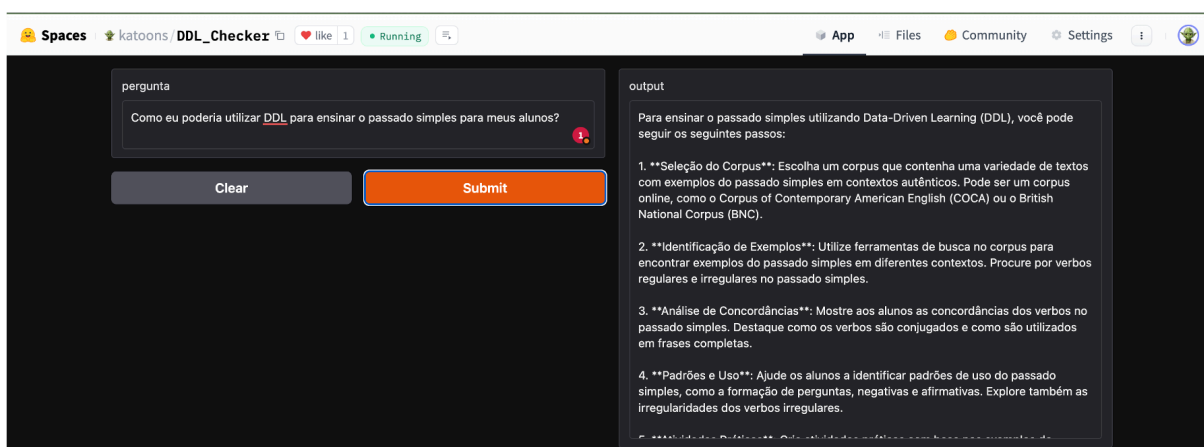


Source: Prepared by the author (2024).

Although still in its initial development phase, the DDL Checker represents a significant step towards integrating AI-driven support into corpus literacy training, particularly in resource-limited educational settings. It is important to highlight that it is still a prototype. Future research will focus on refining and expanding the tool, exploring alternative AI models that could enhance accessibility and sustainability, and examining its effectiveness in real classroom environments.

The ongoing development of AI-driven educational tools, such as the DDL Checker, underscores the potential for technology to bridge pedagogical gaps and provide teachers with the resources to confidently apply corpus-based methodologies in their teaching practices. In Figure 26, we can see an example of how the tool can help professors to approach their classroom subject using DDL.

Figure 26 – DDL Checker suggesting how to approach simple past in the classroom using DDL



Source: Prepared by the author (2024).

To further enhance the effectiveness of AI-driven solutions like the DDL Checker, it is crucial to ensure that teachers possess the necessary corpus literacy skills to engage meaningfully with these tools. One of the challenges identified in our study is that many educators struggle with formulating effective queries, both when searching within corpora and when interacting with AI-based educational tools. This issue underscores the need for a more structured approach to developing search and query skills, ensuring that teachers can maximize the benefits of both corpus tools and AI-driven assistance. In response, we propose adding a new corpus skill to bridge this gap and enhance teachers' ability to effectively explore linguistic data.

As part of this research, we also identified a missing skill in the current corpus literacy framework that is essential for effectively using corpus tools and AI-assisted learning. Formulating precise search queries is a fundamental part of corpus analysis, and teachers often struggle with constructing prompts that yield meaningful results. Similarly, AI-assisted tools such as the DDL Checker rely heavily on well-structured prompts to generate useful responses. To bridge this gap, we propose Skill 19: Developing Effective Prompts for Corpus Exploration.

This new skill focuses on helping teachers and learners create, adapt, and utilize prompts that guide them in critically exploring corpus data. By refining their ability to construct effective queries, teachers can optimize their corpus searches, improve lesson planning, and enhance their students' engagement with corpus-based materials. The DDL Checker directly contributes to this process by

assisting teachers in formulating and refining prompts, reinforcing their understanding of corpus exploration strategies.

This skill plays a crucial role in bridging both technical and pedagogical gaps. While existing corpus skills primarily focus on either technical corpus analysis or critical interpretation of corpus data, Skill 19 connects these two domains by applying technical knowledge to pedagogical practice. Additionally, it addresses one of the core challenges identified in our study: teachers' struggles with corpus implementation due to a lack of expertise in crafting effective queries. By equipping educators with the ability to develop well-structured prompts, this skill directly supports the integration of DDL in classrooms, making corpus-based learning more accessible and effective.

Beyond enhancing teacher proficiency, this skill also contributes to student autonomy. By improving their ability to create effective queries, teachers empower students to explore corpus data independently, fostering self-directed learning and critical engagement with linguistic patterns. Encouraging students to formulate their own corpus-based research questions further strengthens their analytical thinking skills, making them more active participants in their language-learning process.

The integration of Skill 19 into corpus literacy training highlights the evolution of corpus linguistics in education and the need for new pedagogical approaches that accommodate the growing role of AI-assisted learning tools. Future research should explore how this skill can be systematically incorporated into corpus literacy programs, ensuring that teachers are equipped with the necessary competencies to maximize the potential of both AI and corpus-based methodologies.

The DDL Checker offers AI-powered support to bridge the gap between theoretical knowledge and practical application in corpus-based teaching, assisting teachers with queries, explanations, and lesson planning. Building on this, including Skill 19: Developing Effective Prompts for Corpus Exploration further reinforces the importance of equipping teachers with the necessary competencies to engage effectively with corpus data.

Although currently a proof-of-concept, the DDL Checker and the proposed new corpus skill demonstrate the potential for AI-driven tools and refined pedagogical approaches to enhance teacher autonomy, sustain professional development beyond corpus literacy programs, and facilitate more effective integration of corpus-based approaches. Future research will optimize the tool's functionality, expand its dataset,

and evaluate its impact on real-world teaching practices, ensuring that AI continues to serve as a practical and scalable solution for supporting corpus literacy in language education.

5.3 IMPLICATIONS FOR THEORY AND PRACTICE

The findings from this study contribute to both the theoretical development of CBLP and its practical implementation in teacher education programs, particularly in developing countries. More broadly, they also speak to a well-established concern in teacher education: the persistent gap between what teachers learn conceptually and what becomes sustainable classroom practice. This theory–practice tension has long been described as a perennial issue in teacher education, precisely because professional knowledge is not simply “applied” but needs to be reconstructed over time through situated work and reflection (Korthagen, 2010; Freeman, 2018). In language teacher education, this reinforces the importance of carefully designed pedagogies and support structures that connect professional learning to the realities of teaching contexts (Johnson; Golombek, 2018).

This dynamic helps interpret both cases examined here. In Vietnam, participants expressed high confidence in their understanding of corpus tools, yet their lesson plans suggested difficulties in translating that knowledge into coherent DDL-informed pedagogy. In Brazil, teachers reported similarly positive orientations towards corpus-based approaches, but post-course accounts and implementation evidence pointed to barriers that limited sustained uptake. Seen through a theory–practice lens, these patterns suggest that corpus literacy development, while necessary, often requires additional pedagogical mediation and contextual support for teachers to move from learning about tools to using them purposefully in classroom design (Korthagen, 2010; Freeman, 2020).

This research reinforces previous studies that emphasize the need for comprehensive corpus literacy development (Pérez-Paredes, 2020; Ma *et al.*, 2023), but it also reveals that technical corpus skills alone are insufficient. Teachers require pedagogical strategies to align corpus tools with their instructional goals and overcome practical challenges arising from curriculum constraints, resource limitations, and institutional resistance. To address these problems, this study discusses two proposals:

1 - A seven-step procedural framework for systematically developing the 18 established corpus skills (Pérez-Paredes, 2020), ensuring teachers move deliberately from establishing learning contexts to formulating hypotheses.

2 - Skill 19: Developing Effective Prompts for Corpus Exploration, a competency focused on refining unstructured questioning into purposeful, iterative query formulation.

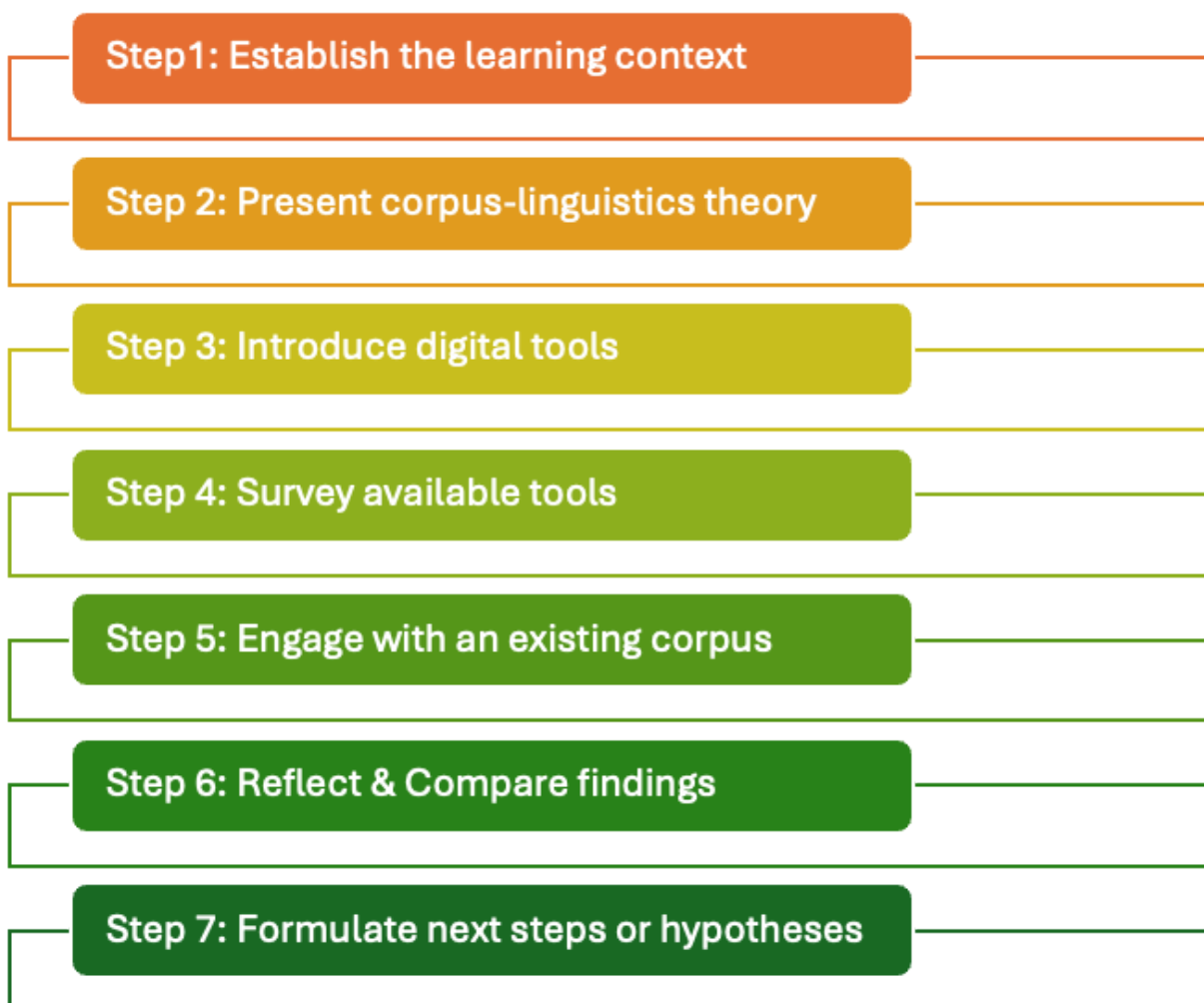
Skill 19 directly addresses the gap between technical corpus literacy and pedagogical application. While teachers and students already engage in informal inquiry when using corpora, this questioning often lacks systematic refinement or alignment with learning objectives. Skill 19 equips educators with strategies to transform vague prompts (e.g., “study prepositions”) into targeted queries (e.g., “Which prepositions follow ‘interested’ in academic texts?”), enhancing their ability to interact with corpus data and AI-assisted tools, such as the DDL Checker.

Complementing Skill 19, the seven-step framework provides a structured pathway for cultivating corpus skills (e.g., frequency analysis, collocation identification). For example, teachers might use the framework to guide students through analysing discourse markers in research articles: establishing context (Step 1), introducing corpus tools like Sketch Engine (Step 3), and refining hypotheses based on corpus findings (Step 7). This model ensures adaptability to diverse educational environments while maintaining a focus on pedagogical relevance.

By integrating Skill 19 with the seven-step framework, this study proposal enhances existing corpus literacy training while simultaneously introducing a new skill for the AI era. Together, they provide teachers with the tools to navigate technical and pedagogical challenges, ensuring corpus methodologies translate meaningfully into classroom practice.

To illustrate how such an approach can be systematically implemented, we first present a general seven-step procedure for cultivating any of the 18 corpus skills (Pérez-Paredes, 2020). This model aims to help teachers move through a deliberate sequence—from establishing context and exploring relevant tools to reflecting on corpus outputs—while leaving room for adaptation to different educational environments. Figure 27 presents a seven-step procedure for developing any corpus skills.

Figure 27 – A seven-step procedure for developing any of the 18 corpus skills



Source: Prepared by the author (2024).

Step 1 - Establish the Learning Context: Before introducing corpus tools, it is essential to define the instructional goals. Teachers must consider:

- What specific language features do students struggle with?
- How does the lesson fit into the broader curriculum?
- What are the students' language proficiency levels and learning needs?

For example, if a teacher notices that students misuse modal verbs, this step ensures that corpus tools are selected to address this issue specifically. Without this preliminary stage, corpus exploration risks becoming unstructured and ineffective.

Step 2 - Present Corpus Linguistics Theory: teachers need a foundational understanding of corpus linguistics after establishing the learning context. This stage cover basic corpus concepts, such as frequency, collocations, and concordance lines.

- The role of corpus tools in providing authentic language data.
- The importance of DDL as an approach that encourages students to analyze real-world language use.

For instance, before using Sketch Engine, teachers should demonstrate how frequency lists help distinguish common collocations (e.g., "make a decision" vs. "do a decision"). If teachers skip this conceptual foundation, corpus tools may appear confusing or disconnected from traditional teaching methods.

Step 3 - Introduce Digital Corpus Tools: at this stage, teachers and students explore specific corpus platforms, such as:

Sketch Engine – A tool that provides in-depth collocation analysis.

SKELL – A user-friendly corpus designed for language learners.

VersaText – A tool used for text comparison and keyword extraction.

For example, if students struggle with preposition use, the teacher might introduce SKELL's concordance feature, showing how words like "interested" appear in context (e.g., "interested in" vs. "interested on"). The key challenge here is ensuring that teachers feel comfortable navigating these platforms before applying them in class.

Step 4 - Survey and Compare Available Tools: With multiple corpus tools available, teachers should evaluate:

- Which platform best suits their teaching objectives?
- Whether the tool allows customized searches (e.g., specific collocations, registers, or grammatical structures).
- The technological accessibility of the tool in their school setting.

For example, if a school has limited internet access, teachers might prioritize offline corpus databases or smaller web-based tools like SKELL. Comparing options helps teachers make informed decisions and ensures that corpus tools align with their teaching constraints.

Step 5 - Engage with an Existing Corpus: teachers and students now explore corpora through hands-on activities. This involves:

- Conducting frequency counts to analyze word usage.

- Using collocation tools to study patterns in authentic texts.
- Extracting keywords in context (KWIC) to understand lexical relationships.

For example, in a writing class, a teacher might ask students to analyze the use of academic discourse markers (e.g., "however," "therefore") in research articles. Students would use Sketch Engine to examine how often these markers appear and in what contexts. This stage transforms corpus learning into an interactive discovery process.

Step 6 - Reflect and Compare Findings: teachers and students must now evaluate:

- Whether corpus insights match initial lesson goals.
- What patterns or surprises emerge from corpus analysis?
- How these findings can inform classroom instruction.

For example, after analysing modal verbs in a learner corpus, students might realize that "must" is rarely used in conversational English, despite its prominence in textbooks. This type of reflection helps teachers refine their teaching materials, ensuring that lessons reflect authentic language use.

Step 7 - Formulate Next Steps and Hypotheses: the final step encourages teachers to:

- Plan follow-up lessons incorporating corpus-based discoveries.
- Develop student-centred DDL activities for continued exploration.
- Propose new research questions for further investigation.

For instance, if students struggle with academic writing, the teacher might ask them to analyse passive voice usage in research articles and apply their findings in their next essay. This iterative process ensures that corpus skills extend beyond isolated lessons and become part of a long-term learning strategy.

To demonstrate how teachers can systematically develop Skill 19—the ability to create and refine effective prompts for corpus queries—we can apply a structured approach that emphasizes "conversing" with the corpus. This skill is not about memorizing steps or passively using tools; rather, it is about learning how to ask the right questions, interpret data, and refine queries to uncover meaningful linguistic patterns. By focusing on prompt creation, teachers and students alike can develop the ability to interact with corpus data in a purposeful and insightful way.

The process begins with establishing a clear linguistic objective, such as exploring how prepositions interact with specific adjectives or analysing the

frequency of modal verbs in different registers. This step is crucial because it helps students understand what they want from the data. For example, instead of a vague goal like "study prepositions," students might formulate a specific question like "Which prepositions most commonly follow the adjective 'interested'?" This shift from rote memorization to inquiry-based learning is at the core of Skill 19, as it encourages students to think critically about language use and develop targeted questions that guide their investigations. The goal is not just to use the corpus, but to learn how to "talk to" it effectively.

Once the learning context is established, teachers introduce foundational concepts in corpus linguistics, such as collocations, concordance lines, and frequency distributions. These concepts help students understand how corpus tools process and display linguistic data, enabling them to interpret results effectively. Teachers also demonstrate how different platforms, such as Sketch Engine, SKELL, or COCA, provide unique insights depending on the structure of the prompt. For instance, a prompt like "Let's find collocations of 'decision' in academic texts" might yield different results in Sketch Engine compared to a more straightforward query like "Let's find examples of 'decision' which may or may not use academic text. This line of thought makes students ponder what corpora are and which ones they should use. This step is not about mastering the tools but understanding how to tailor prompts to extract the most relevant data from the best source.

Hands-on engagement with corpus tools is where students refine their ability to "converse" with the corpus. Starting with an initial query, such as "interested" to explore prepositional patterns, students often discover that their first attempts yield irrelevant or incomplete results. This underscores the importance of iterative refinement, where teachers guide students in adjusting their prompts—perhaps by adding context or narrowing search parameters—to yield more accurate and meaningful data. For example, a broad prompt like "Let's find examples of 'however'" might be refined to "Let's find examples of 'however' in the first sentence of a paragraph." This teaches students to think critically about how their prompts shape the results they obtain, and to adapt their queries based on the data they encounter. It's about learning how to "ask better questions" as they interact with the corpus.

As students analyse corpus findings, teachers encourage them to reflect on how well their prompts worked and whether the results align with their initial objectives. For instance, students might discover that "must" is rarely used in spoken

English compared to "have to," challenging their textbook-based assumptions. This reflective practice reinforces the value of inquiry-driven learning and helps students refine their prompt-writing skills for future investigations. By comparing corpus findings with their intuition and textbook explanations, students develop a deeper understanding of language patterns and the power of well-crafted prompts in uncovering linguistic insights. This reflection helps students internalize "conversing" with the corpus.

Finally, students synthesize their findings into practical applications, such as creating corpus-informed lesson plans, designing collocation-based activities, or initiating student-driven research projects. For example, if students analysed passive voice in academic writing, they might apply their findings to revise their own essays or propose a new question, such as "How does passive voice usage vary across disciplines?" This step encourages students to think beyond isolated lessons and develop more sophisticated prompts for future investigations, such as "Compare passive voice usage in humanities versus science research articles." By integrating prompt creation into their learning process, students take ownership of their exploration and develop a nuanced understanding of how language works in real-world contexts. This ability to formulate effective queries not only enhances their engagement with corpus data but also prepares them to interact meaningfully with AI-driven tools. Building on this, the DDL Checker demonstrates the potential of AI-powered solutions to provide scalable, cost-effective support for teachers, particularly in developing contexts. Rather than merely assisting with corpus tool navigation, the DDL AI tool could be more active in designing and refining corpus-based activities. For this to work, teachers must have a previous DDL theoretical framework to critically assess AI outputs, distinguishing proper instructions from misleading ones or so-called "hallucinations."

Future teacher professional development initiatives could explore blended learning models that combine structured, face-to-face corpus literacy instruction with on-demand AI support. This approach would empower teachers to navigate technical and pedagogical challenges in real time and reinforce their ability to effectively apply corpus-based methodologies in diverse classroom settings.

Additionally, this study highlights the need for institutional support when introducing DDL methodologies in schools. Teachers in Brazil and Vietnam expressed concerns about curriculum restrictions, assessment pressures, and the

lack of institutional incentives for innovation. Successful implementation of corpus-based pedagogy requires top-down support from educational policymakers, ensuring that DDL is not perceived as an isolated experiment but rather an integrated component of the language education curriculum.

Integrating AI into corpus linguistics raises an important question: If AI can assist with search formulation, do educators still require the 18 corpus skills outlined by Pérez-Paredes (2020)? A superficial analysis might suggest that AI can replace many of these skills, particularly in frequency analysis, keyword searches, and collocation identification. Suppose an AI assistant can instantly retrieve frequency distributions, compare corpora, or annotate linguistic data while explaining the corpus concepts to the user. Does this reduce the need for traditional corpus training? While AI undoubtedly enhances accessibility and efficiency in corpus exploration, the question remains whether teachers can critically engage with AI-generated data without a solid understanding of corpus methodology.

Our findings suggest that AI does not eliminate the need for critical corpus literacy; on the contrary, it makes it even more essential. Educators who lack foundational knowledge of corpus methodology risk misinterpreting AI-generated results or blindly relying on automated recommendations without understanding their limitations. This is where Skill 19: Developing Effective Prompts for Corpus Exploration can help. By refining the ability to formulate precise and meaningful queries, educators maintain agency over corpus exploration, ensuring that AI tools act as enhancements rather than replacements.

At present, AI has not yet been fully integrated into corpus tools such as Sketch Engine or LancsBox. However, future developments will likely allow direct AI-driven corpus querying, where users could use natural language questions, such as *"What is the distribution of the word clock in Sherlock Holmes corpus?"*, and receive precise corpus-based insights without needing to master wildcard searches or advanced corpus syntax. This type of natural language interaction with corpus tools would represent a significant advancement in implementing DDL, as it would simplify how teachers and students engage with corpus data, making corpus exploration more intuitive and accessible.

Moreover, our findings suggest that existing models of corpus literacy assessment may need revision. Many studies rely on self-reported data to evaluate teachers' corpus skills and perceptions (Heather and Helt, 2012; Ma *et al.*, 2023), but

our analysis of lesson plans and post-course interviews revealed discrepancies between what teachers believe they can do and how they actually implement DDL techniques. A multi-layered assessment model, incorporating structured lesson plan analyses, classroom observations, and longitudinal follow-ups, would provide more accurate insights into how corpus literacy translates into teaching practice.

The proposed modifications to the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) seek to enhance the assessment of corpus-based teaching activities by focusing on the depth of corpus engagement, the effectiveness of instructional strategies, and the alignment of corpus-based activities with broader curricular objectives. By implementing these changes, the adapted TIOI (Harris; Grandgenett; Hofer, 2010) offers a more comprehensive framework for evaluating the integration of corpus tools into teachers' pedagogical practices.

Finally, this study highlights the critical role of continuous pedagogical support and technological assistance in helping teachers integrate DDL into their instructional practices. While our course programs introduced participants to corpus tools, many struggled to implement them long-term, particularly in low-resource educational settings. Once teachers develop foundational corpus literacy skills, such as interpreting frequency data, identifying collocations, and analysing concordance lines, AI-driven tools like the DDL Checker can serve as pedagogical guides, supporting the application of DDL methodologies while mitigating AI's limitations.

6 CONCLUSIONS AND FUTURE WORK

This study aimed to support language teaching in developing countries by helping teachers effectively implement the Data-Driven Learning (DDL) approach, thereby developing corpus skills and adapting to diverse educational contexts. To achieve this, we designed a structured DDL course focusing on corpus tools and digital literacy, empowering teachers to integrate DDL methodologies into their classrooms. Additionally, we examined how the differing educational landscapes of Brazil and Vietnam affect DDL implementation, taking into account structural, cultural, and institutional factors. Finally, we proposed strategies to simplify DDL adoption by addressing specific challenges, clarifying common doubts, and improving integration practices.

To accomplish these objectives, we drew on the theoretical framework of Corpus Linguistics, referencing key works such as Pérez-Paredes (2020), Berber Sardinha (2004; 2012), Boulton and Vyatkina (2021), and Crosthwaite *et al.* (2020). We employed a mixed-methods approach, analysing teacher interviews, lesson plans, and self-assessment surveys to assess the impact of the DDL course. The study also introduced and refined the Technology Integration Observation Instrument (TIOI) (Harris; Grandgenett; Hofer, 2010) to better evaluate corpus-based lesson plans. Furthermore, we developed the DDL Checker, an AI-assisted tool designed to address teachers' challenges in applying corpus methodologies.

The findings provide support for our research hypotheses:

- A structured DDL course significantly increased teachers' confidence and preparedness to implement corpus-based teaching, especially when challenges such as technological limitations and large class sizes were addressed.
- Guided instruction with explicit steps and practical examples was found to be more effective than a purely theoretical or open-ended course.
- Structural constraints, including limited technological resources, administrative workload, and funding gaps, had a greater impact on DDL adoption than cultural factors in both Brazil and Vietnam.
- The introduction of Skill 19: Developing Effective Prompts for Corpus Exploration may help mitigate corpus-skill gaps and facilitate more effective DDL application in resource-constrained settings.

This study also highlighted two key pathways for developing corpus skills. First, a seven-step method was proposed to systematically guide teachers in implementing corpus skills, ensuring a structured approach from setting learning goals to evaluating corpus outputs. Second, an AI-assisted approach using the DDL Checker allowed teachers to generate and refine corpus-based prompts through automated suggestions, providing scalable support for those unfamiliar with corpus analysis.

Beyond these pedagogical contributions, this research proposed modifications to the TIOI framework (Harris; Grandgenett; Hofer, 2010), making it more suitable for evaluating corpus-based lesson plans by incorporating criteria for assessing corpus engagement depth, query formulation quality, and alignment with learning objectives. These refinements offer a more precise tool for researchers and educators in language teaching.

While this study addressed the research questions and produced findings broadly consistent with the proposed hypotheses, several limitations should be acknowledged. First, the absence of longitudinal evidence limits our ability to determine whether DDL adoption is sustained over time. Long term follow ups are therefore needed to examine continued implementation and impact. Second, technological constraints, particularly the costs associated with API based AI tools, created practical barriers and point to the need to test more affordable alternatives. Finally, institutional resistance also shaped adoption, as curriculum restrictions and limited incentives for pedagogical innovation constrained teachers' room to experiment with corpus based pedagogy. Further work should investigate strategies to build institutional support and policy conditions that enable DDL at scale. Quantitative analyses may also be extended, where appropriate, through inferential procedures to assess whether observed patterns are statistically robust across contexts.

Building on these findings, future research will focus on:

1. Refining the DDL Checker, integrating additional corpus training resources, and optimizing its accessibility.
2. Expanding the study to additional developing countries to better understand the contextual factors influencing DDL adoption.
3. Developing scaffolded lesson plan templates tailored to different teaching needs, facilitating a smoother integration of corpus-based activities.

This study reinforces the potential of DDL as a transformative approach to language teaching, particularly in under-resourced educational settings. However, teachers require continuous training, structured resources, and institutional backing for their effective implementation. Integrating AI-driven tools like the DDL Checker, coupled with the refinement of corpus literacy training and the proposed TIOI (Harris; Grandgenett; Hofer, 2010) modifications, represents a promising pathway for bridging the gap between theory and practice. As research in this field advances, ensuring that teachers in developing countries have the tools and knowledge to fully embrace corpus based pedagogy will be crucial to its long-term success.

REFERENCES

- BAISA, V.; SUCHOMEL, V. Skell: web interface for English language learning. In: HORÁK, A.; RYCHLÝ, P. (ed.). **Proceedings of Recent Advances in Slavonic Natural Language Processing**. Karlova Studánka, Czech Republic, 2014.
- BANDURA, A. **Social learning theory**. Eaglewood Cliffs, NJ: Prentice-Hall, 1977.
- BENNETT, C.; DHONCHADHA, E. U. Becoming corpus literate: an in-service EFL teacher education framework for integrating corpora into EFL teaching. **Applied Corpus Linguistics**, v. 3, n. 1, 2023.
- BERBER SARDINHA, A. C. **Linguística de corpus**. São Paulo: Manole, 2004.
- BINKLEY, M. et al. Defining twenty-first century skills. In: GRIFFIN, P.; CARE, E. (ed.). **Assessment and teaching of 21st century skills**. Dordrecht: Springer, 2012. p. 17–66.
- BOISSON, J.; KAO, T.; WU, J.; YEN, T.; CHANG, J. Linggle: a web-scale linguistic search engine for words in context. In: **Proceedings of Association for Computational Linguistics (Conference System Demonstrations)**, Sofia, Bulgaria, 2013. p. 139–144.
- BOULTON, A. Data-driven learning for younger learners: obstacles and optimism. In: CROSTHWAITE, P. (ed.). **Data-driven learning for the next generation: corpora and DDL for pre-tertiary learners**. London: Routledge, 2019. p. 14–20.
- BOULTON, A.; VYATKINA, N. Thirty years of data-driven learning: taking stock and charting new directions over time. **Language Learning & Technology**, v. 25, n. 3, p. 66–89, 2021.
- BOULTON, A. What data for data-driven learning?. **EuroCALL Review**, v. 20, n. 1, p. 23–27, 2012.
- BRASIL. **Base nacional comum curricular**. Brasília, DF: MEC, 2017.
- BRASIL. **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, Centro Gráfico, 1988.
- 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.
- BREZINA, V.; WEILL-TESSIER, P.; McENERY, A. **LancsBox**. Lancaster: University, 2021.
- BRITISH COUNCIL. **Demandas de aprendizagem de inglês no Brasil**. São Paulo: British Council, 2014. Available: https://www.britishcouncil.org.br/sites/default/files/demandas_de_aprendizagempesquisacompleta.pdf. Accessed at em: 5 jul. 2022.

CALLIES, M. Towards corpus literacy in foreign language teacher education: using corpora to examine the variability of reporting verbs in English. In: KREYER, R.; SCHAUB, S.; GULDENRING, B. A. (ed.). **Angewandte Linguistik in Schule und Hochschule** [Applied Linguistics in Secondary School and at University]. Berlin: Peter Lang, 2016. p. 391–415.

CASTILLO-MONTOYA, M. Preparing for interview research: the interview protocol refinement framework. **Qualitative Report**, v. 21, n. 5, p. 811–831, 2016. Available: <http://nsuworks.nova.edu/tqr/vol21/iss5/2>. Accessed at: 10 nov. 2023.

CENTRO DE INOVAÇÃO PARA A EDUCAÇÃO BRASILEIRA (CIEB). **Relatório Guia Edutec** – diagnóstico do nível de adoção de tecnologia nas escolas públicas brasileiras em 2022. São Paulo: CIEB, 2022.

CHUNG, T.; CROSTHWAITE, P.; CAO, H.; DE CARVALHO, T. Walking the walk? (Mis)alignment of EFL teachers' self-reported corpus literacy skills and their competence in planning and implementing corpus-based language pedagogy. **TESOL Quarterly**, v. 58, 2024.

COBB, T.; FREE, P. **The complete lexical tutor**. TESL-EJ, v. 8, n. 3, 2004. Available: <https://www.tesl-ej.org/ej31/m2.html>. Accessed at: 14 sept 2025.

CROSTHWAITE, P. (ed.). **Corpora for language learning: bridging the research–practice divide**. London: Taylor & Francis, 2025. No prelo.

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.; LUCIANA; WIJAYA, D. Exploring language teachers' lesson planning for corpus-based language teaching: a focus on developing TPACK for corpora and DDL. **Computer Assisted Language Learning**, v. 36, n. 7, p. 1392–1420, 2021.

CROSTHWAITE, P.; NINGRUM, S.; SCHWEINBERGER, M. Research trends in corpus linguistics: a bibliometric analysis of two decades of Scopus-indexed corpus linguistics research in arts and humanities. **International Journal of Corpus Linguistics**, v. 28, n. 3, p. 344–377, 2022. DOI: <https://doi.org/10.1075/ijcl.21072.cro>.

CROSTHWAITE, P.; STELL, A. It helps me get ideas on how to use my words: primary school students' initial reactions to corpus use in a private tutoring setting. In: CROSTHWAITE, P. (ed.). **Data-driven learning for the next generation: corpora and DDL for pre-tertiary learners**. London: Routledge, 2019. p. 150–170.

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, 2020.

DASHTESTANI, R. Exploring English as a foreign language (EFL) teacher trainers' perspectives on challenges to promoting computer literacy of EFL teachers. **JALT CALL Journal**, v. 10, n. 2, 2014.

DASKALOVSKA, N. Corpus-based versus traditional learning of collocations. **Computer Assisted Language Learning**, v. 28, n. 2, p. 130–144, 2015.

DAWADI, S.; SHRESTHA, S.; GIRI, R. A. Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. **Journal of Practical Studies in Education**, v. 2, n. 2, p. 25–36, 2021. DOI: 10.46809/jpse.v2i2.20. Available: https://www.researchgate.net/publication/349663713_Mixed-Methods_Research_A_Discussion_on_its_Types_Challenges_and_Criticisms. Accessed at: 13 dez. 2025.

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.

DECKER, P.; NATHAN, B. **Behavior modeling training**. New York, NY: Praeger Publishers, 1985.

DO, T.; LE, T.; VU, X.. Innovating Vietnamese education in the context of Industrial Revolution 4.0: a systematic review. **International Journal of Current Science Research and Review**, v. 6, n. 10, out. 2023. DOI: 10.47191/ijcsrr/v6-i10-11. Available: <https://app.dimensions.ai/details/publication/pub.1164796601>. Accessed at em: 15 jun. 2025.

DUTRA, D. P.; MARQUES, A. L. Getting started with digital literacy using pedagogic corpora in young learners' English classes. **The ESpecialist**, v. 44, n. 2, 2023.

FARR, F. Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users. **Language Awareness**, v. 17, n. 1, p. 25–43, 2008. DOI: <https://doi.org/10.2167/la414.0>.

FRANKENBERG-GARCIA, A. **New directions in DDL for EAP writers**. BAWE Quicklinks, 2018.

FREEMAN, D. Arguing for a knowledge-base in language teacher education, then (1998) and now (2018). **Language Teaching Research**, v. 24, n. 1, p. 5–16, 2020. DOI: 10.1177/1362168818777534.

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**. Dissertação (Mestrado em Linguística Aplicada) – Universidade Estadual Paulista, 2020.

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.

GOULART, L.; VELOSO, I. **Corpora in English language teaching**: classroom activities for teachers new to corpus linguistics. Montclair: Montclair State University, 2023.

HADDAD, S. A educação continuada e as políticas públicas no Brasil. **Revista de Educação de Jovens e Adultos**. v. 1, n. 8, 2007.

HALLINGER, P. Making education reform happen: is there an ‘Asian’ way? **School Leadership and Management**, v. 30, n. 5. DOI: <10.1080/13632434.2010.513457>. Accessed at: 7 jun. 2025.

HARRIS, J.; GRANDGENETT, N.; HOFER, M. Testing a TPACK-based technology integration assessment rubric. In: MADDUX, C. D.; GIBSON, D.; DODGE, B. (ed.). **Research highlights in technology and teacher education**. Chesapeake, VA: Society for Information Technology & Teacher Education (SITE), 2010. p. 323–331.

HEATHER, J.; HELT, M. Evaluating corpus literacy training for pre-service language teachers: six case studies. **Journal of Technology and Teacher Education**, v. 20, n. 4, 2012.

HO, L.; DIMMOCK, C. Changing teachers’ beliefs and practices towards learner-centred education: experiences and lessons from Vietnam’s education system reforms. **Practice**, v. 5, n. 1. 2010. DOI: <10.1080/25783858.2023.2177191>. Available: <https://app.dimensions.ai/details/publication/pub.1155332266>. Accessed at em: 15 jun. 2025.

HORTON JR, F. W. **Understanding information literacy**: a primer. 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.

JENKINS, J. **English as a lingua franca in the international university**: the politics of academic English language policy. London: Routledge, 2013.

JENKINS, J. Should you be persuaded: two samples of data-driven learning materials. In: JOHNS, T. F.; KING, P. (ed.). **Classroom concordancing**. ELR Journal, v. 4, p. 1–16, 1991.

JENKINS, J. Whence and whither classroom concordancing? In: BONGAERTS, T. et al. (ed.). Computer applications in language learning. Berlin: Foris, 1988.

JOHNSON, K. E.; GOLOMBEK, P. R. Informing and transforming language teacher education pedagogy. **Language Teaching Research**, v. 24, n. 1, p. 116–127, 2020. DOI: 10.1177/1362168818777539.

JOHNS, T. Micro-concord: a language learner’s research tool. **System**, v. 14, n. 2, 1986.

KARLSEN, P. H.; MONSEN, M. Corpus literacy and applications in Norwegian upper secondary schools: teacher and learner perspectives. **Nordic Journal of English Studies**, v. 19, n. 1, 2020. DOI: <https://doi.org/10.35360/njes.500>.

KELLEY, L. C.; SASGES, G. **Vietnam over the long twentieth century**: Becoming modern, going global. Springer Nature, 2024.

KENNEDY, C.; MICELI, T. An evaluation of intermediate students' approaches to corpus investigation. **Language Learning & Technology**, v. 5, n. 3, 2001.

HEATHER, J.; HELT, M. Evaluating corpus literacy training for pre-service language teachers: Six case studies. **Journal of Technology and Teacher Education**, v. 20, n. 4, p. 415-440, 2012.

KILGARRIFF, A. et al. The Sketch Engine: ten years on. **Lexicography**, v. 1, n. 1, 2014.

KIM, C. *et al.* Teacher beliefs and technology integration: corpus literacy and applications in Norway. **Teaching and Teacher Education**, v. 29, 2011.

KOEHLER, M. J.; MISHRA, P. What happens when teachers design educational technology? the development of technological pedagogical content knowledge. **Journal of Educational Computing Research**, v. 32, n.8 2005. DOI: <https://doi.org/10.2190/0EW7-01WB-BKHL-QDYV>.

KORTHAGEN, F. A. J. The relationship between theory and practice in teacher education. In: PETERSON, Penelope; BAKER, Eva; McGAW, Barry (eds.). **International Encyclopedia of Education**. Oxford: Elsevier, 2010. v. 7, p. 669–675.

LATIF, M. M. A. Corpus literacy instruction in language teacher education: investigating Arab EFL student teachers' immediate beliefs and long-term practices. **ReCALL**, v. 33, n. 1, p. 34–48, 2021. DOI: <https://doi.org/10.1017/S0958344020000129>.

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.

LE, N.; SONG, J. TPACK in a CALL course and its effect on Vietnamese pre-service EFL teachers. **Asian EFL Journal**, v. 9, n. 1, 2018.

LEITE, B. S. Tecnologias digitais e metodologias ativas: quais são conhecidas pelos professores e quais são possíveis na educação? **Rev.Vidya**, v. 41, n. 1, 2021.

LENKO-SZYMAŃSKA, A. Is this enough? a qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. **ReCALL**, v. 26, n. 2, p. 260–278, 2014. DOI: <https://doi.org/10.1017/S095834401400010X>.

LEŃKO-SZYMAŃSKA, A. **Training teachers in data-driven learning**: tackling the challenge. 2017.

LIMA, D. C.; GIMENEZ, T. **English language teacher education in Brazil**: current debates and challenges. Campinas: Pontes Editores, 2020.

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, 2023.

MA, Q.; MEI, F. Review of corpus tools for vocabulary teaching and learning. **Journal of China Computer-Assisted Language Learning**, v. 1, n. 1, 2021.

MA, Q.; YUAN, R.; CHEUNG, L. M. E.; YANG, J. **Teacher paths for developing corpus-based language pedagogy**: a case study. *Computer Assisted Language Learning*, publicação avançada online, 2022. Available at: <https://doi.org/10.1080/09588221.2022.2040537>.

MARQUES, A. L. S. P. **Science and geography lexicons in English for learners in ensino fundamental 1**: a corpus-based investigation. 2024. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2024.

MATTAR, J. Aprendizagem em ambientes virtuais: teorias, conectivismo e MOOCs. Teccogs: **Revista Digital de Tecnologias Cognitivas**, v. 8, n. 7, 2013.

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

MUKHERJEE, J. Corpus linguistics and language pedagogy: the state of the art—and beyond. In: BRAUN, S.; KOHN, K.; MUKHERJEE, J. (ed.). **Corpus technology and language pedagogy**: new resources, new tools, new methods. Berlin: Peter Lang, 2006. p. 5–24.

NGUYEN, H. T. M. *et al.* Developing professional capacity for Content Language Integrated Learning (CLIL) teaching in Vietnam: tensions and responses. **Language and Education**, v. 37, n. 6, 2023. DOI: <10.1080/09500782.2023.2260374>. Available: <https://app.dimensions.ai/details/publication/pub.1164496486>. Accessed at em: jun. 2025.

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. **From corpus to classroom**: language use and language teaching. Cambridge: Cambridge University Press, 2007.

ORENHA-OTTAIANO, A. *et al.* Corpus-based methodology for an online multilingual collocations dictionary: first steps. **Proceedings of eLex**, p. 1–28, 2021.

PAIVA, V. L. M. O. Abordagens e metodologias de ensino de inglês na escola pública. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 19, n. 1, p. 45–66, 2019.

PÉREZ-PAREDES, P. **Corpus linguistics for education**: a guide for research. London: Routledge, 2020.

PÉREZ-PAREDES, P.; CURRY, N.; AGUADO JIMÉNEZ, P. Integrating critical corpus and AI literacies in applied linguistics: a mixed-methods study. **Computer Assisted Language Learning**, p. 1-27, 2025.

PHÙNG, T; VŨ, Đăng M. An educational regime of truth for social reform in late colonial Vietnam. In: KELLEY, Liam C.; SASGES, Gerard (Org.). **Vietnam over the long twentieth century: becoming modern, going global**. Singapore: Palgrave Macmillan, 2024. p. 79–100.

PINTO, P. T. *et al.* Data-Driven Learning in Brazilian Educational Contexts: An Analysis of a Teacher Training Program Following BNCC's Guidelines. **Revista Brasileira de Linguística Aplicada**, v. 25, n. 1, p. e49797, 2025.

PINTO, P. T. *et al.* **Using language data to learn about language: a teachers' guide to classroom corpus use**. Australia: The University of Queensland, 2023. DOI: <https://doi.org/10.14264/3bbe92d>. Accessed at: 05 set. 2023.

PRENSKY, M. H. Sapiens digital: from digital immigrants and digital natives to digital wisdom. **Innovate: Journal of Online Education**, v. 5, n. 3, 2009.

PRETTO, N. Education and technological innovation: a look at Brazilian public policies. **Revista Brasileira de Educação**, n. 11, 1999.

ROSE, D.; MARTIN, J. R. **Learning to write, reading to learn: genre**. Sheffield, UK: Equinox, 2012.

SALDAÑA, J. **The coding manual for qualitative researchers**. 3. ed. London: SAGE, 2016.

SAPIR, E. **Language: an introduction to the study of speech**. New York: Harcourt, Brace and Company, 1921.

SARDINHA, T. B.; DELFINO, M.C. N.; RAMPASO, M. Preparação de material didático para ensino de línguas com base em corpora. **The ESpecialist**, [S. l.], v. 38, n. 1, 2017. DOI: 10.23925/2318-7115.2017v38i1a12. Disponível em: <https://revistas.pucsp.br/index.php/esp/article/view/32224>. Acesso em: 27 sept. 2025.

SCHMIDT, N. Unpacking second language writing teacher knowledge through corpus-based pedagogy training. **ReCALL**, v. 35, n. 1, p. 40–57, 2022. Available: <https://doi.org/10.1017/S0958344022000106>.

SERPA, T.; PINTO, P. T.; SBROGIO, R. O.; CAMARGO, D. C. O ambiente digital Tradcorpus: proposta de repositório de dados de corpora on-line para o ensino de línguas estrangeiras e de tradução. **Trabalhos em Linguística Aplicada**, v. 62, n. 1, 2023.

SILVA, M. V.; SANTOS, J. M. C. T. **A BNCC e as implicações para o currículo da educação básica**. In: I CONGRESSO NACIONAL DA DIVERSIDADE (CONADIS), 2018, Natal. Anais [...]. Natal: Conadis, 2018.

ŞİMŞEK, T. *et al.* **Integration of corpus literacy consultation into language teacher education: course design, implementation, and evaluation**, 2023.

THE MATRIX REVOLUTIONS [Filme]. Direção: **Lilly Wachowski**; Lana Wachowski. Estados Unidos: Warner Bros. Pictures; Village Roadshow Pictures; Silver Pictures, 2003. 129 min, son., color.

THOMAS, J. E. **Two web-based tools for learning language from language.**, 2020. Available: <http://www.versatile.pub/versatext-info.html>. Accessed at: 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).

VIANA, V.; LU, L.. Corpus linguistics and continuous professional development: participants' prior knowledge, motivations and appraisals. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 29, n. 2, p. 1485–1527, 2021.

WARREN, J. **Cultures of development: Vietnam, Brazil and the unsung vanguard of prosperity**. New York: Routledge, 2017.

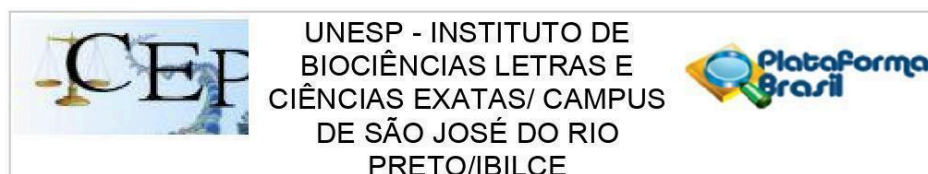
WHORF, B. L. **Language, thought, and reality: Selected writings of Benjamin Lee Whorf**. Edited by John B. Carroll. Cambridge, MA: MIT Press, 1956.

YAZAN, B. Three approaches to case study methods in education: Yin, Merriam, and Stake. **The qualitative report**, v. 20, n. 2, p. 134–152, 2015. Accessed at: 29 nov. 2025.

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.

ZAREVA, A. Incorporating corpus literacy skills into TESOL teacher training. **ELT Journal**, v. 71, n. 1, 2016. Available at: <https://doi.org/10.1093/elt/ccw045>.

APPENDIX A - THE ETHICS COMMITTEE'S APPROVAL (PLATAFORMA BRASIL)



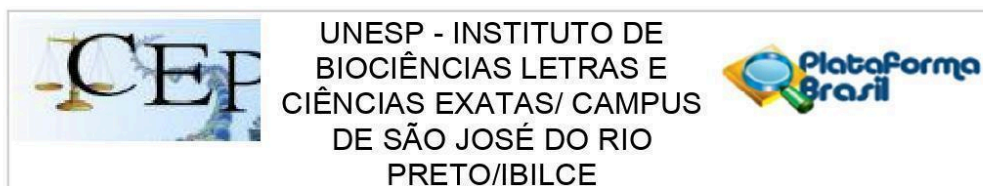
Continuação do Parecer: 7.158.124

Os relatórios parciais deverão ser encaminhados anualmente, contando a partir desta data, bem como o relatório final, conforme modelo em nossa página: <https://www.ibilce.unesp.br/#!/comite/etica-em-pesquisa/relatorio-projeto/>

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2022777.pdf	11/10/2024 13:55:16		Aceito
Outros	Carolina_carta_respostaassinado.pdf	11/10/2024 13:53:25	CAROLINA TAVARES DE CARVALHO	Aceito
Outros	Corrigido_Instrumento_de_coleta_de_dados.docx	11/10/2024 13:52:52	CAROLINA TAVARES DE CARVALHO	Aceito
Projeto Detalhado / Brochura Investigador	Corrigido_04_Carolina_Tavares_Projeto_Plataforma_Brasil.docx	11/10/2024 13:51:53	CAROLINA TAVARES DE CARVALHO	Aceito
Cronograma	Corrigido_Cronograma.docx	11/10/2024 13:51:05	CAROLINA TAVARES DE CARVALHO	Aceito
Folha de Rosto	FR_Carolina_Tavares_Final.pdf	16/09/2024 12:29:01	CAROLINA TAVARES DE CARVALHO	Aceito
Cronograma	Cronograma.docx	11/09/2024 16:31:51	CAROLINA TAVARES DE CARVALHO	Aceito
Projeto Detalhado / Brochura Investigador	02_Carolina_Tavares_Projeto_Plataforma_Brasil.docx	11/09/2024 16:14:31	CAROLINA TAVARES DE CARVALHO	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_2.docx	11/09/2024 16:11:06	CAROLINA TAVARES DE CARVALHO	Aceito
Outros	Instrumento_de_coleta_de_dados.docx	22/03/2023 04:44:03	CAROLINA TAVARES DE CARVALHO	Aceito

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro: JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município:** SÃO JOSE DO RIO PRETO
Telefone: (17)3221-2567 **Fax:** (17)3221-2409 **E-mail:** cep.ibilce@unesp.br



Continuação do Parecer: 7.158.124

Objetivo da Pesquisa:

Foram avaliados e contemplados no parecer anterior.

Avaliação dos Riscos e Benefícios:

Foram avaliados e contemplados no parecer anterior.

Comentários e Considerações sobre a Pesquisa:

Foram avaliados e contemplados no parecer anterior.

Considerações sobre os Termos de apresentação obrigatória:

Foram avaliados e contemplados no parecer anterior.

Recomendações:

Nada a declarar.

Conclusões ou Pendências e Lista de Inadequações:

Lista de pendências atendidas/adequadas

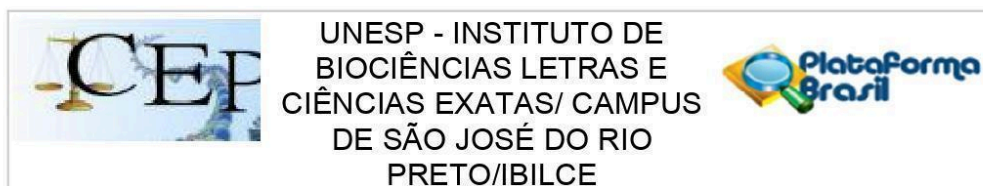
Considerações Finais a critério do CEP:

Deliberação ad referendum do Comitê de Ética em Pesquisa do IBILCE/UNESP, de 15 de outubro de 2024.

Dispõe sobre a análise do Projeto de Pesquisa CAAE: 83225923.9.0000.5466, "Desenvolvimento das Habilidades de corpus (HACs) na formação de professores de Língua Inglesa", da pesquisadora Carolina Tavares de Carvalho.

Tendo em vista a urgência que o assunto requer e o parecer favorável da relatora, manifesto pela aprovação, ad referendum do Comitê de Ética em Pesquisa, do Projeto de Pesquisa CAAE: 83225923.9.0000.5466, "Desenvolvimento das Habilidades de corpus (HACs) na formação de professores de Língua Inglesa", da pesquisadora Carolina Tavares de Carvalho.

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro: JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município:** SAO JOSE DO RIO PRETO
Telefone: (17)3221-2567 **Fax:** (17)3221-2409 **E-mail:** cep.ibilce@unesp.br



UNESP - INSTITUTO DE
BIOCIÊNCIAS LETRAS E
CIÊNCIAS EXATAS/ CAMPUS
DE SÃO JOSÉ DO RIO
PRETO/IBILCE



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Desenvolvimento das Habilidades de corpus (HACs) na formação de professores de Língua Inglesa

Pesquisador: CAROLINA TAVARES DE CARVALHO

Área Temática:

Versão: 2

CAAE: 83225923.9.0000.5466

Instituição Proponente: Instituto de Biociências Letras e Ciências Exatas/ Campus de São José do

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 7.158.124

Apresentação do Projeto:

Trata-se de atendimento às pendências referente ao projeto de pesquisa intitulado “Desenvolvimento das Habilidades de corpus (HACs) na formação de professores de Língua Inglesa”, sob a responsabilidade da pesquisadora Carolina Tavares de Carvalho.

Lista de Pendências:

No projeto, incluir a descrição detalhada e cronograma de todos os instrumentos e procedimentos de coleta de dados que serão utilizados durante o período em que será ministrado o curso de formação de professores (contexto de coleta). Essas informações estão minimamente explicitadas no TCLE. No entanto, no projeto, não está claro exatamente quais instrumentos de coleta serão utilizados, em quais momentos e com quais propósitos.

Resposta:

Conforme solicitado, incluímos detalhes específicos sobre os instrumentos de coleta de dados que serão utilizados durante o curso de formação de professores, detalhando o uso de questionários no Google Forms e os momentos específicos em que cada questionário será aplicado.

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro: JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município:** SAO JOSE DO RIO PRETO
Telefone: (17)3221-2567 **Fax:** (17)3221-2409 **E-mail:** cep.ibilce@unesp.br

APPENDIX B - REPOSITORY LINK

Repository Link: Thesis data

DADOS CURRICULARES

IDENTIFICAÇÃO	
	Carolina Tavares de Carvalho 12/06/1981
Nacionalidade	brasileira
Nome em citações bibliográficas:	Carvalho, Carolina Tavares de De Carvalho, Carolina Tavares Tavares de Carvalho, Carolina
Currículo Lattes	http://lattes.cnpq.br/9577321159874381
ORCID	https://orcid.org/0000-0001-9446-6394
outro identificador: ResearchGate	https://www.researchgate.net/profile/Carolina-De-Carvalho-3?ev=hdr_xprf