GLOTTODIDACTICA LII/1 (2025)
ADAM MICKIEWICZ UNIVERSITY PRESS POZNAŃ
https://doi.org/10.14746/gl.2025.52.1.7

Maria Ammari
Adam Mickiewicz University, Poznań, Poland

Advancing ESP instruction through DDL: A structured training framework

Abstract. This study explores the urgent need for corpus-based English instruction within English for Specific Purposes (ESP), an area with untapped potential despite its growing importance. The limited use of corpora in classrooms, due to shortcomings in teacher training, underscores the need for strategies that enhance teacher proficiency in corpus use, addressing the unique linguistic needs of learners. It proposes an integrated training framework, extending Carter and McCarthy's (1995) '3 Is', further expanded to '4 Is' by Flowerdew (2009), to a more comprehensive training framework of '6 Is' to facilitate a balance between inductive and deductive learning. The paper highlights the necessity of specialised training to empower teachers with the skills for effective Data-Driven Learning implementation in ESP teaching, aiming to improve the overall educational experience by integrating corpus insights into language instruction. Adopting a theoretical approach, this study synthesises insights derived from a specialised corpus and DDL pedagogy to develop a structured training framework. It provides a roadmap for ESP educators, equipping them with the necessary competencies to effectively integrate corpora into instruction. By offering a structured methodology, the study contributes to ongoing discussions on ESP teacher development and the normalisation of corpus-based teaching practices.

Keywords: corpus linguistics, data-driven learning, ESP teacher training, inductive and deductive learning, 6 I's training framework.

1. INTRODUCTION

The increasing importance of English for Specific Purposes (ESP), as a tailored approach to language instruction, designed to respond to learners' unique linguistic needs and objectives (Dudley-Evans & St. John 1998), sets it apart from conventional English Language Teaching (ELT) methods (Hyland 2002). Although these needs are activity-oriented, ESP[1] assumes that such activities inherently involve and rely on specific registers, genres, and their corresponding language, which students must be able to use effectively so as to perform these

[1] Although the specific-purposes approach is not confined to the English language, this paper focuses primarily on the teaching of English.

activities (Dudley Evans & St. John 1998: 4). Hence, *language* in ESP is seen as a defining feature that stems from ESP's primary focus on the learner's unique requirements, determining the course's content and aim (Richards & Schmidt 2002: 181), specified through needs analysis (Dudley Evans & St. John 1998: 4).

Long's suggestion to "view every course as involving specific purposes" (Long 2005: 19) plays a distinctive role in ESP, critical in an era where the role of English as a global academic lingua franca demands specialized language competencies that conventional ELT might not sufficiently cater to. Additionally, the heightened relevance of ESP is driven by the rise in international student numbers and the English language's increasing dominance in academia, underscoring, in turn, the tailored educational experience it offers, aligning with the practical language usage requirements of its audience (Hyland 2002). The British Council and its affiliated organisation, Studyportals, has published a global review, The *Changing Landscape of English-Taught Programmes*, on December 7, 2021, identifying 27,874 complete MA and BA degree programmes taught in English outside the big four English-speaking study destination nations. It is thus evident why English academic discourse has nowadays "become the language of literacy" (Halliday & Martin 1993: 11) and has been "established as the world's leading language for the dissemination of academic knowledge" (Hyland 2004: ix).

Concurrently, though, this customized, "needs-driven" approach of ESP, as highlighted by De Chazal (2014: 5), leads to a broad spectrum of purposes within the ESP discipline, with varying objectives and goals each time. Consequently, this approach necessitates language educators to venture into unfamiliar academic and occupational domains, akin to ESP students, and to critically reflect on whether teaching the language practices of target discourse communities effectively meets learners' objectives (Belcher 2006: 2). This has generated an increasing amount of research to provide insights into the language used in academic discourse and the distinct registers it comprises, showing how specific vocabulary and related grammatical structures define different types of discourse (Coxhead & Byrd 2007: 130). These distinct characteristics of ESP's specificity and diversity have been able to be efficiently served through corpora.

The advent of corpus linguistics and the potential of corpus analysis through corpora have provided researchers and ESP teachers with a window into the *language-in-use* (see; Coxhead & Byrd 2007). This approach allows them to access and evidence data that can assist them to focus on the ways in which language is actually used for communication. By using corpora, ESP teachers can thus validate their linguistic intuitions that, in turn, often reveal unexpected research avenues (Partington 1998: 1), embodying Higgins's (1988) "serendipity" principle (Higgins, as cited in Partington 1998: 1) and highlighting corpora's dual role in both teaching and research.

Despite the widespread adoption of corpora by researchers globally, only a small number of language teachers have managed to overcome the barriers to directly use corpora in their teaching, even though a few have shown enthusiastic acceptance of this approach (Frankerberg-Garcia 2010: 475, see also Leńko-Szymańska 2017). Instead, a reluctance by teachers to use corpora in language instruction has been reported in research (Mukherjee 2004; Romer 2009; Tribble 2015, as cited by Leńko-Szymańska 2017).

This paper aims to inspire ESP teachers, and ESP teacher trainers by highlighting the advantages of Data-Driven Learning (hereforth DDL) and familiarizing them with contemporary tools and resources. It outlines practices to empower teachers for the proficient delivery of DDL activities to ESP students. By adopting the newly proposed "6 Is" framework and the accompanying strategies and practices, ESP teachers and trainers can improve their instructional efficacy, enhance their capacity to deliver engaging and effective DDL-based lessons, and ultimately enrich the language learning experience for their students. Considering the various factors influencing DDL activities' effectiveness, the study is meticulously designed around key inquiries, including the optimal approach to ESP teacher training and the selection of appropriate tools and DDL activities.

In light of the challenges faced by ESP teachers in adopting corpus-based instruction, this study seeks to answer the following research questions:

1. What are the key barriers preventing ESP teachers from effectively incorporating corpus-based resources into their instruction?
2. How can a structured "6 Is" framework enhance the efficacy of ESP teacher training in DDL?

By positioning these inquiries within the broader context of ESP instruction and corpus linguistics, this study aims to provide actionable insights for improving teacher training methodologies and advancing ESP education through data-driven approaches.

2. LITERATURE REVIEW

2.1. Corpora, DDL and the learner's role

The unique contribution of corpus analysis to language education extends beyond merely offering a new technological tool; it introduces a transformative philosophy for understanding language (Partington 1998: 1). It empowers students to proficiently sift through and interpret linguistic corpora, thus bolstering their autonomous learning skills. Over twenty-five years ago, Johns (1991a, 1991b), recognized as the "father" of Data-Driven Learning (DDL), underscored

the importance of DDL as a learner-centred approach, famously suggesting that "every student [become] a Sherlock Holmes." This metaphor emphasizes *learner agency* (a concept that has gained popularity in educational research during the past twenty years[2]), encouraging students to participate in their own learning, through active learning engagement, autonomy, and self-regulation[3], thus facilitating discovery of salient language patterns and promoting the noticing of these patterns, as highlighted by Boulton and Cobb (2017).

DDL activities may be divided into *direct* or *indirect* types of corpus usage. Direct use involves the direct access of a corpus by learners through corpus query software and demands training. The indirect one typically entails teacher-curated concordances, where the teachers pre-select corpus data ready for learner mediation, without necessitating direct corpus interaction by learners or any related training. Through both types of activities students *actively engage* with authentic use of language and analyse keyword lists, read concordances, read collocate information and visual graphs/charts. As Johns (1991b: 2) proposed, the language learner becomes "a research worker whose learning needs to be driven by access to linguistic data, hence the name data driven learning." Thus, Johns' (1991a: 29) aspiration was "to provide adequate opportunities for students to raise problems and queries," through "inductive strategies developed in the classroom," that should be equally applicable outside the classroom, however, so that students "survive and learn by themselves" (Johns 1991a: 29).

As far as the learner's role is concerned, beyond the earlier mentioned notions of *learner agency* and *active engagement*, according to Aston (2001), "the greatest attraction of corpora for language pedagogy is their potential for autonomous learning" (Aston 2001: 41). Such *autonomy* fosters a transformation in language students, evolving them into conscious learners who actively engage with and reflect upon their learning processes. This heightened consciousness though can be beneficial, as it enables learners to tailor their educational experiences to their personal needs and goals, leading to more efficient and impactful language acquisition. Aligning with the above principles and attributes of DDL, Vyatkina, and Boulton (2017) highlight its role in promoting noticing and raising awareness within L2 instruction, aiming ultimately to "foster greater independence and improved language competencies over time" (Han & Shin 2017: 173–174).

This innovative approach of DDL not only equips learners with the tools to independently explore language patterns, but also instils a profound understanding of language use in real-world contexts, delving into grammatical struc-

[2] For an overview of the role of agency in educational theory and practice Biesta and Tedder (2007).

[3] Self-regulation refers to learners' ability to actively manage their own learning by setting goals, monitoring progress, and adjusting strategies as needed to achieve desired outcomes.

tures, word meanings, and various language facets by engaging with extensive amounts of authentic linguistic data. The effectiveness of DDL across various teaching and learning contexts has been documented by Crosthwaite and Boulton (2024). However, Crosthwaite and Boulton (2024) caution against viewing DDL as a universal solution or panacea for all educational scenarios, underscoring the need for empirical validation through more longitudinal research to fully ascertain its impact.

2.2. Transforming ESP education: The impact of corpora and DDL

The emergence of corpora has fundamentally transformed the landscape of language learning, marking a significant shift in the methodologies employed in linguistic education. Innovations in corpus linguistics have enabled a more empirical and data-driven approach (Johns 1991a), granting learners and educators access to vast databases of authentic language use. In turn, this access facilitates a deeper understanding of linguistic patterns, usage, and variation, offering nuanced insights into the complexities of language in ESP and EAP contexts.

Corpora occupy a central role in ESP, serving as invaluable teaching aids, learning tools, and reference resources (Boulton 2012: 261). ESP is widely regarded as particularly suited to corpus-based teaching and learning (Gavioli 2005: 14) due to its contextual relevance. Bennett (2010: 11) further highlights this point, stressing that "ESP is probably one of the most obvious and pointed applications of corpus linguistics." As a result, this brings DDL into close alignment with the particular linguistic requirements of learners within their respective areas of study or professional practice, perfectly echoing ESP's commitment to relevance and specificity. Additionally, this alignment resonates with Benesch's (2001) call for a transition to ESP 'rights analysis' to enhance inclusivity, Casanave's (2002) urge on the necessity for adaptive change to effectively integrate new members into academic and professional communities, and Cadman's (2002: 85) proposal to redefine EAP as 'English for Academic Possibilities', thereby expanding its scope and purpose.

Corpus-based / -driven studies offer hands on access to authentic materials and examples, and may indeed shed light on "important aspects of a text or text collection that may go unnoticed otherwise" (Römer & Wulff 2010: 101), but also highlight aspects of academic language use that are underrepresented (Chen 2010; Pho 2008) as well as validate or challenge intuitive assumptions about learners' language difficulties (Rundell & Granger 2007). Hence, by focusing on learning directly from linguistic data, the DDL approach could potentially revolutionize language learning and teaching in specialized domains of ESP.

However, the potential of DDL to revolutionize language learning and teaching in specialized domains of ESP remains largely untapped. Empirical research in the realm of computer-assisted language learning (CALL) has predominantly focused on university students in general-purpose language classes. This trend underscores Gillespie's (2020) broader critique of empirical studies in CALL, highlighting a gap in research specifically targeting the application and effectiveness of DDL within ESP contexts. Gillespie (2020) suggests that this approach can be highly rewarding for ESP teachers working with overly confident students reluctant to accept corrections to erroneous knowledge. By exposing students to concordance tools teachers can encourage them to independently recognize and correct their mistakes, potentially leading to more effective learning.

This observation aligns with Chambers and O'Sullivan's (2004: 168) assertion that corpus consultation is "good for unlearning errors." Moreover, O'Sullivan (2007) enumerates an extensive list of corpus-related skills that are essential for effectively engaging with and analysing language data, further emphasizing the potential of DDL to enhance ESP teaching practices. These skills include "predicting, observing, noticing, thinking, reasoning, analysing, interpreting, reflecting, exploring, making inferences (either inductively or deductively), focusing, guessing, comparing, differentiating, theorizing, hypothesising, and verifying" (O'Sullivan 2007: 277).

O'Sullivan's (2007) detailed enumeration highlights the multifaceted cognitive processes involved in working with corpus data, ranging from initial observation to in-depth analysis and validation of linguistic patterns and hypotheses which are indispensably relevant to ESP/EAP writing needs. While it is crucial to acknowledge that the DDL approach to learning "may seem rather time-consuming for a single word enquiry" (Boulton & Cobb 2017: 349) for both students and teachers, Boulton and Cobb (2017) emphasize that the significance lies in the process itself. They argue that this approach results in "increased language sensitivity, noticing, induction, and ability to work with authentic data" (Boulton & Cobb 2017: 349).

2.3. Teachers and DDL

Advancements in educational research and psychology over the past three decades have established a robust theoretical foundation for new teaching practices. These trends have "changed the teacher's role from that of knowledge-transmitter to consultant, guide, coach, and/or facilitator" (Chong 2016). However, the role of an ESP teacher encompasses even more facets. Thus, Swales (1985) aptly uses the term 'ESP practitioner' to better capture this breadth – extending

beyond traditional teaching to include needs analysis, syllabus design, materials development, and evaluation (Hutchinson & Waters 1987: 157). This expanded perspective reflects the diverse responsibilities of ESP educators, emphasizing the complexity and wide-ranging functions of their role beyond the conventional classroom setting.

This transformation is crucial, as it redefines the ESP teacher's role to include scaffolding and facilitating corpus analysis experiences, especially significant for learners. Yoon and Jo (2014: 113) highlight that teacher guidance is instrumental in creating DDL-friendly environments, enabling students to engage effectively with linguistic data. Although Tim Johns (1991a: 1) emphasizes that "at the heart of the approach is the use of the machine not as a surrogate teacher or tutor, but as a rather special type of informant" (1991a: 1), the teacher's assistance as a facilitator guiding students through targeted advice or "focused tips" is crucial in leading "students through the data discovery and interpretation" of DDL instructional approach (Corino & Onesti 2019: 2). Rather than overtly dispensing knowledge explicitly and directly, educators in DDL take on the roles of "research directors and collaborators," creating opportunities for students to independently seek solutions and derive meanings (Corino & Onesti 2019: 2).

Additionally, Gavioli (2005: 15) argues that the particularity of concordance-type data means it must be treated as "*samples* rather than *examples* of language," requiring careful interpretation and analysis. However, learners should be *guided* to be able to discover the foreign language (Johns 1991), to be able to "[identify] recurrences and [infer] patterns which appear in some way typical of certain contexts" (Gavioli 2000: 129). Moreover, Bernardini (2004) highlights that discovery learning fosters a "supportive, non-authoritarian environment." This is especially true in ESP settings, where the dynamic interplay between a student's emerging disciplinary literacy, an educator's ability to scaffold domain-specific language skills, and the integration of data-driven teaching approaches may enhance the overall learning process, and thus prove particularly valuable in specialized contexts.

In this context, another crucial role ESP teachers assume in the context of DDL is that of the learning expert – Bernardini (2004) underscores this shift, noting that "the teacher acts as a learning expert rather than a language expert" (Bernardini 2004: 28). As such, this exchange of knowledge between student and teacher enriches the learning experience, shifting the teacher's role from an authority figure, dictating the learning process to a collaborator, guiding the process – a perspective that suggests a move towards a more collaborative and interactive educational environment and aligning seamlessly with the goals of DDL education.

2.3.1. Corpora integration challenges in language teaching

Although language instructors are increasingly aware of the benefits of self-regulation and autonomy that Data-Driven Learning (DDL) offers to students, many remain reluctant to employ corpora in their teaching due to a variety of either perceived, or actual barriers (Leńko-Szymańska 2014a). This reluctance was highlighted in a survey conducted by Mukherjee (2004) among language teachers in Germany, which revealed a stark contrast between the optimism of corpus linguists' regarding the pedagogical value of corpus tools and their real-world application in English language teaching. One key reason for this disconnect appears to be a lack of awareness among instructors about how linguistic databases can be effectively utilized in the classroom (Tribble 2015; Römer 2009, 2010; Mukherjee 2004).

To address this discrepancy, knowledge of corpus linguistics and corpora applications in teaching have begun to be integrated into the curriculum of language departments at both undergraduate (BA) and postgraduate (MA) levels, as well as within English for Specific Purposes (ESP) programs. However, Leńko-Szymańska succinctly argues that this experience "does not automatically imply that they know how to apply corpora in their teaching" practices (Leńko-Szymańska 2014a). ESP teachers, in particular, face unprecedented challenges in integrating "new forms of digital literacies" into classroom activities, as noted by Bloch (2012: 390). The mere inclusion of digital tools in the classroom is insufficient; hence, effective adaptation of new digital literacies necessitates ESP teachers to evaluate "the nature of the literacy […] and the type of authorship that is best supported by each technology" (Bloch 2012: 390). Thus, adopting an informed and reflective approach is crucial for successful implementation.

Mukherjee (2004) advocated for a collaborative effort to popularize corpus linguistics and unlock its language pedagogy potential. The intricacy of corpus literacy skills, however, encompassed in corpus linguistics, complicates things (in enhanced DDL language learning for both students and teachers). Heather and Helt (2012: 417) describe corpus literacy as "the ability to use the technology of corpus linguistics to investigate language and enhance the language development of students" (2012: 417). Drawing from Mukherjee (2006) and Dalton-Puffer[4] (2014), Callies (2016: 395) summarizes the subcomponents of corpus literacy into (1) understanding corpus linguistics fundamentals, (2) searching corpora and analysing corpus data with software tools, (3) interpreting corpus data, and

[4] Dalton-Puffer, C. (2014). *Corpus Linguistics in language teacher education*. (Plenary lecture given at the 14th Klagenfurt Conference on Corpus Based Applied Linguistics [CALK 14], 25–27 September 2014). University of Klagenfurt, Austria.

(4) using corpus findings to generate educational materials and activities. This set of skills is crucial for foreign language educators to effectively integrate corpus resources into teaching strategies.

Moreover, Mukharjee (2004) highlighted that teachers must first understand and appreciate the value of corpus data in addressing their teaching challenges before adopting more sophisticated, learner-centred activities. Challenges such as syllabus integration and the limited involvement and contributions of non-researcher language teachers in corpus-based practices remain significant barriers to normalizing corpus use. As Pérez-Paredes (2022: 36) notes, further theorization is required for DDL and corpora to have a meaningful impact on mainstream second language education. Additionally, researchers are identified as the primary stakeholders in DDL utilization, a conclusion backed by evidence that 94% of surveyed studies occurred in university settings, where researchers likely benefit from more straightforward access to data samples (Pérez-Paredes 2022).

The role of ESP teachers is crucial, particularly in university environments where collaboration with fellow researchers is readily facilitated (unless of course they are involved in research themselves). The need for effective application of corpus data in teaching and advancing DDL practices towards *normalisation* requires both ESP teachers and researchers to make concerted efforts in training and introducing other ESP or language teachers to corpora and corpus literacy. This paper outlines strategies for incorporating corpus data into language teaching to enhance corpus literacy training and promote DDL *normalization*, enhancing corpus literacy training. By addressing the gap identified by Pérez-Paredes (2019) between theoretical corpus linguistics and practical classroom application, it proposes tools, resources, and strategies to empower teachers in effectively integrating corpora into their instructional practices.

2.3.2. Normalization

Easing the *integration* of corpus linguistics into teachers' lessons is closely related to the concept of normalization, initially defined by Bax (2003: 23) as "the stage when the technology becomes invisible, embedded in everyday practice and hence normalised." Later, Bax (2011) revisited this definition, refining the concept. Bax's implicit hypothesis on the concept of normalization posits that "technology has reached its fullest possible effectiveness in language education […], as a valuable element in the language learning process," when it becomes an unobtrusive, integrated part of the learning process.

Pérez-Paredes et al. (2022) emphasize in their study that although approximately 70% of language teachers (in Spain and the UK) incorporate online plat-

forms or web-based services into their teaching, only a small number are familiar with using L1 corpora or learner corpora in language instruction. Moreover, Leńko-Szymańska (2017) observed a one-semester course within an MA program, conducted over five consecutive years from 2011 to 2016, and concluded that it was insufficient for trainees to adequately develop technical Corpus linguistics and pedagogical skills. The course spanned thirteen to fifteen 90-minute class sessions and covered three modules throughout the course[5], however trainees still did not manage to become autonomous corpus users, and educators proficient in corpus application (Leńko-Szymańska 2017: 234).

In Bax's terms (2003, 2011), this suggests that the trainees did not achieve normalization. The technology had neither been seamlessly integrated into language education nor reached its fullest potential effectiveness, failing to become an essential component of the learning process. According to Bax (2003: 22–23), the integrated CALL approach epitomizes normalization, where technology is omnipresent and unobtrusively woven into daily educational activities, transcending its status as a topic of debate. However, Bax (2003: 22) asserts that achieving this integration requires treating computers as an essential, though not central, component of the learning environment, meaning they should complement rather than dominate language lessons.

2.4. Corpus tools usage in DDL language education

As previously highlighted, teachers play a crucial role in facilitating students' learning through the use of corpora and DDL. However, recent studies show no or low use of corpora. Vyatkina and Boulton (2017: 67) explain that there are two DDL formats that teachers can employ for pedagogical interventions. DDL exploration can be either *hands-off* – involving indirect applications through teacher-prepared corpus-based materials – or *hands-on* – engaging students in direct exploration of corpora. In regards to autonomy, teacher-constructed concordance tasks (e.g. Vincent 2013, as cited in Charles 2022) offer less autonomy, whereas discovery learning (e.g. Bernardini 2002, as cited in Charles 2022) promotes greater student autonomy. It could also present any combination of the two styles (Vyatkina & Boulton 2017: 67).

Flowerdew (2009) and Kennedy (2008) highlight the underexplored question of which resource – corpus, grammar guide, or dictionary – is most effective for addressing specific linguistic inquiries. For instance, Kennedy (2008) notes that

[5] Corpora in Foreign Language Teaching, offered by the Institute of Applied Linguistics at the University of Warsaw.

dictionaries might better clarify differences between terms like 'tall' and 'high' than corpora. However, Bernardini (2004: 43) highlights that teachers often either fail to recognize semantic differences altogether or, when they do, are unable to articulate these nuances and their usage effectively. This issue becomes even more pronounced for learners whose native language does not distinguish between such terms, as seen with Chinese students struggling with 'tall' and 'high'.

Consequently, this highlights a significant research gap in understanding how resources, learning processes, and teaching activities interact and influence learning goals and outcomes (Bernardini 2004: 31). Consequently, this reveals a significant research gap in understanding how resources, learning processes, and teaching activities interact and influence learning goals and outcomes (Bernandini 2004: 31). Addressing this gap requires further exploration to delineate the optimal applications of linguistic resources for effective language learning.

3. TOOLS AND RESOURCES FOR TEACHERS

This section is devoted to informing about some state-of-the-art tools and resources that are user-friendly and easily accessible, enabling the choice of the most appropriate ones for teaching objectives and maximizing the educational advantages of corpora and DDL in the classroom. In DDL, concordancing serves as a pivotal tool for exposing learners to authentic language use, enabling them to explore and analyse linguistic patterns directly from real-world texts. The most common format for concordance lines is KWIC[6] (Key Word In Context). It is a computer-generated index which displays a selected word or phrase in the middle of the display, *the node* [italics – M.A.] with the text forming its context on either side. Beyond concordancing, DDL utilizes various tools including corpus query software (e.g. AntConc, Sketch Engine, LacsBox), collocation dictionaries (e.g. Ozdic), frequency list generators (Just the word, SkeLL, WebCorp), keyword analysis and n-gram tools (SketchEngine), corpus platforms (CorpusMate, WebCorp), online corpora (e.g. British National Corpus, COCA) accessible in english-corpora.org, all aimed at facilitating the exploration of authentic language use and patterns.

Collocation dictionaries, such as Ozdic, simplify the process of finding common word combinations. Concordancers, such as Just the word and SkeLL help learners understand word usage in context, with WebCorp offering real-time

[6] The term KWIC traces back to Hans Peter Luhn's (1896–1964) work in the late 1950s at IBM, where he developed indexing methods for information retrieval, leading to the alignment of keywords (Stubbs 2007: 318). The adoption of KWIC concordances by linguists quickly followed, with concordance packages available by the mid-1960s (Stubbs 2007: 318).

corpus data from the web. CorpusMate, Fraze.it and Flax stand out for their unique exploration features. A prominent feature of CorpusMate is generating a concordance for a specific discipline (e.g. science, culture, arts, music, chemistry, education, law, etc.). It offers a "compare results" between disciplines functionality and a "pattern finder" displaying left or right context of the word in search. Fraze.it offers learners a wealth of options, among them, results of authentic sentences from British newspapers, as well as a function of video outcomes of TED lectures. The Flax online tool highlights essential components of academic texts, including academic vocabulary, key concepts, common collocations, and lexical bundles (Wu, Fitzgerald & Franken 2019), leveraging large corpora like the British National Corpus and COCA for deep linguistic insights. Unique in its offering, Compleat Lexical Tutor (v.8) provides a suite of tools for ESL educators, focusing on effective vocabulary instruction through various online resources tailored for both teaching and research purposes.

For advanced users engaged in corpus analysis, software like AntConc provides robust text analysis capabilities, while Sketch Engine offers comprehensive keyword analysis tools. LancsBox appeals for its user-friendly interface for corpus exploration, and WordSmith Tools specializes in detailed linguistic investigations.

4. TOWARDS A TRAINING MODEL FOR TEACHERS

Developing a DDL course on corpus usage for educators is inherently complex; DDL is at the intersection of corpora, teachers, and learners, representing a collaborative and interactive space where learners actively engage with corpora. In this environment, learners may receive varying levels of guidance, from minimal to substantial, depending on their proficiency, with the most advanced reaching a stage of autonomy. The DDL setting facilitates a dynamic, symbiotic as well as reciprocal relationship that optimizes language learning by harnessing "corpus-linguistics skills and get to know various corpus resources in their foreign language" (Leńko-Szymańska 2017: 234). Essentially, the goal is to equip teachers with the skills to efficiently teach corpus-based methods to their students, fostering both "computer and cognitive skills" essential for corpus exploitations, but also recognizing its value in learning, thus improving the educational process. As Breyer (2011) highlights, if teacher trainees can grasp "corpora's potential for their own learning" (Breyer 2011: 230), then this may intrinsically motivate them to incorporate corpus analysis in their teaching careers, equipping them to navigate and tackle the challenges inherent in this method.

The proposed model is based on previous theoretical underpinnings and research by Kennedy and Miceli (2017), Quinn (2015), Carter and McCarthy (1995), Flowerdew (2009), Crosthwaite et al. (2021) and McEnery and Xiao (2011). Kennedy and Miceli (2017) introduce two notions to help the learner cultivate effective corpus use – "to enrich the content and language of a text through what we call *pattern-hunting*, and to edit a text for lexico-grammatical accuracy through *pattern-refining*" (Kennedy & Miceli 2017: 93, italics in original). By introducing students to pattern-hunting, Kennedy and Miceli (2017) teach students to explore a "corpus for ideas and language patterns – i.e. groups of words – to borrow, adapting them as necessary" (Kennedy & Miceli 2017: 94), which is exploratory in nature. Later in their apprenticeship students are introduced to *pattern-refining* work, meant to teach "ways of editing [...] text for lexico-grammatical accuracy" (Kennedy & Miceli 2017: 94). The pattern-refining stage involves open-ended questions, and is problem-solving in nature. The technique almost always involves searching for a word or combination of words (Kennedy & Miceli 2017: 94), with the goal of enhancing their students' awareness and recognition of language often being composed of "prefabricated chunks" sourced "from a kind of mental database of formulaic language" (Lewis 1996, as cited in Kennedy & Miceli 2017: 94).

Quinn (2015), in her scholarly work, proposes an 'introduction' to initiate an L2 training course as beneficial for someone's training, underscoring the value that a structured introductory phase could offer to demystify corpus usage and facilitate its integration into educational practices. Quinn (2015) and Boulton (2009) suggest that not all teachers and students are knowledgeable regarding corpora and corpus consultation, and thus lack of sufficient training poses a major barrier (Breyer 2006; Boulton 2009). However, Sinclair (2004) suggests that "with only a modest few hours of orientation", "both teacher and student can make use of a corpus right away" (Sinclair 2004: 288). In Quinn's (2015: 165) research, teachers are guided "in preparing intermediate L2 writers for learner concordancing, so they can offer students an alternative reference" to the traditional dictionary usage. In her L2 DDL training course, Quinn suggests initiating the course by introducing the students to basic corpora knowledge. Quinn (2015) thus divides the training course into two distinct parts: Stage 1 and Stage 2. The initial five lessons focus on General Learner Training, covering the essence of what a corpus is and the rationale behind its use and also preparing students with paper-based activities, and starting online corpus searches. This stage provides the basics of corpus referencing (Quinn 2015: 166), and may potentially assist any learner comprehend the rationale behind corpus consultation, thus making the whole process more conscious, while motivating attendance and anticipation for what follows. The next ten classes focus on writing, practising essay revisions,

through corpus practices and teaching students how to use concordancing for essay improvement.

English language teachers at large may face challenges and apprehension in integrating concordancing tools into language learning (Boulton 2009). While corpora are a staple in research settings, their application in the classroom is less common, and teachers might lack basic knowledge of corpus linguistics (Boulton 2009). Lack of such specialised knowledge, though, is critical for ESP teachers, who could greatly enhance their disciplinary instruction and lexico-grammatical nuances through corpora. Corpora can aid the L2 writing process by providing support at the discourse level offering a broader context, "in contrast to the isolated dictionary entries that many intermediate writers rely on" (Quinn 2015: 165). However, teachers unfamiliar with concordancing may hesitate to use it, due to concerns about their lack of expertise or perceived misalignment with their instructional methods. This highlights a gap between the potential benefits of corpus tools and their practical application in language education. The present theoretical model, thus suggests the incorporation of a structured 'introduction', where any trainee is familiarized with corpora through inquiries such as 'What does a corpus consist of?', 'What do corpus data look like?', 'What information can be learnt?', and 'How does a corpus compare to bilingual dictionaries?' exploring basic but crucial corpus notions (Quinn 2015). Subsequently, they can engage in analysis, contrasting the insights obtainable from both a bilingual dictionary and a corpus, as discussed in Quinn's study (2015).

Further, Carter and McCarthy (1995) devised the "three Is" strategy (Illustration-Interaction-Induction) as a more effective alternative to the traditional "three Ps" (Presentation-Practice-Production), intended for teaching spoken grammar, emphasizing the use of real data to navigate the subtleties of conversational English, such as ellipsis and topicalization. Their analysis of pedagogical grammars revealed inconsistencies in the treatment of grammatical features, from adequate to entirely omitted. Advocating for real spoken data to enhance classroom language awareness and inductive learning, they highlighted that the only prerequisite for such practices is the need for curiosity, access to authentic data, and a drive for discovery in language education (Carter & McCarthy 1995).

Moreover, McEnery and Xiao (2011) suggest that the "three Is" strategy, initially devised by Carter and McCarthy (1995) for teaching spoken grammar, could be effectively extended to encompass broader language education contexts. Additionally, McEnery and Xiao (2011: 36) report on the corresponding features between Carter and McCarthy's (1995) "three Is" teaching approach and Johns' (1991b) triptych framework of Data-Driven Learning (DDL) – observation (of concordanced evidence), classification (of salient features) and generaliza-

tion (of rules) – with a shared emphasis on inductive reasoning, illustrating approaches that promote active, evidence-based learning processes. Hence, the congruence of these models lies not just in their sequential stages – observing concordanced data, classifying salient features, and generalizing rules in DDL, paralleling illustration, interaction, and induction in the "three Is" – but also in their foundational principles, rooted in the methodologies' mutual advocacy for an empirical approach to language learning. As such, both Carter and McCarthy's (1995) and Johns' (1991) approaches encourage learners to derive linguistic rules from authentic data, fostering deeper cognitive engagement and a more nuanced understanding of language patterns. Consequently, this dynamic, learner-focused methodology, which emphasizes knowledge construction through direct engagement with authentic linguistic evidence, proves exceptionally suited to the intricacies of ESP and EAP lexicogrammar. This alignment not only facilitates a deeper exploration of language patterns but also critically enhances learners' ability to navigate and apply complex linguistic structures within their specific academic or professional contexts.

Furthermore, Carter and McCarthy's (1995) model was enriched by Flowerdew's (2009) proposition. Flowerdew (2009: 407) suggested an optional fourth "I" is needed between interaction and induction, that of "intervention," allowing "for finer-tuning of corpus queries." The stage of intervention can be crucial in the process of inducing phraseological tendencies, addressing students' difficulties with the complex semantic relationships revealed through concordancing, as it provides students with hints or clues. This added stage then facilitates a smoother transition across the inductive-deductive continuum (Flowerdew 2009: 407). While we acknowledge that corpora are valuable for phraseological inquiries, the language bridging lexis and grammar is not always straightforwardly accessible from traditional resources like grammars or dictionaries, and this is exactly where intervention can serve in assisting students in making connections between meanings, which is especially crucial for those who may not have the advanced language skills required to independently decipher the nuanced semantic relationships within corpus data (Flowerdew 2009: 407–408). Consequently, according to Flowerdew (2009: 408), this revised approach recognizes challenges, especially for novice speakers or those with limited linguistic proficiency, in navigating the intricate interplay of lexis and grammar; aspects that may not always be clearly or explicitly conveyed in dictionaries.

Expanding on Quinn (2015), Carter and McCarthy's (1995) and Flowerdew's (2009) frameworks in DDL, the introduction of a sixth "I," "integration," could significantly enhance teacher training courses. This concept is suggested for the systematic blending of corpus insights into the broader teaching context,

ensuring these insights support the objectives of language education. Hence "integration" aims to contextualize corpus findings within the curriculum – not as an isolated component, but as a coherent part of language teaching – thereby reinforcing the connection between empirical language analysis and seamless practical language use, as well as enhancing the relevance and application of corpus studies to real-world language usage. The integration stage was conceptualized and incorporated in response to findings from an initial corpus-based analysis, which identified a significant gap in existing DDL frameworks (see Figure 1). The corpus was compiled using WebBootCaT in Sketch Engine with the following seed words: "corpus-linguistics," "corpus," "Data-Driven Learning," "instruction," "teacher training," and "ESP." It was then manually cleaned to retain only relevant articles. This preliminary corpus, compiled to examine patterns of DDL implementation, revealed the lack of a structured and systematic integration phase in teacher training models. While researchers increasingly emphasize embedding DDL into curricula and English language instruction, integration as an explicit, scaffolded phase remains largely absent from existing frameworks. These findings underscore the need for a systematic approach to integration, which will be further examined in a forthcoming article (Ammari, forthcoming). By addressing this oversight, the integration phase enhances the progression of training by bridging theoretical constructs with practical implementation, thereby ensuring a cohesive and sustainable application of DDL within educational settings. As mentioned earlier, integration into normalization, as outlined by Bax (2003, 2011), occurs when technology becomes an invisible, seamlessly integrated part of educational practice, fully enhancing language learning. This transition requires that corpus linguistics and other digital tools are embedded within daily teaching activities, making them indispensable, without, however, overwhelming the pedagogical context. Achieving such integration, as demonstrated by the limited uptake of corpus tools among language teachers (Pérez-Paredes et al. 2018) and the challenges faced by trainees in becoming proficient users (Leńko-Szymańska 2017), is crucial for technology to reach its full potential in language education.

Thus, by positioning *integration* as a core component of teacher training focused on DDL, this approach emphasizes the shift from simple corpus data exploration data to its practical and integrated seamless application in teaching, aiming for a pedagogical model that merges corpus linguistics with diverse teaching methods to both deepen students' language understanding and enhance instructional approaches. It is hereby suggested that such a strategy would advocate a comprehensive pedagogical framework that integrates corpus linguistics with diverse instructional techniques, and bridges a gap between theory and practice, thereby enriching students' language comprehension and

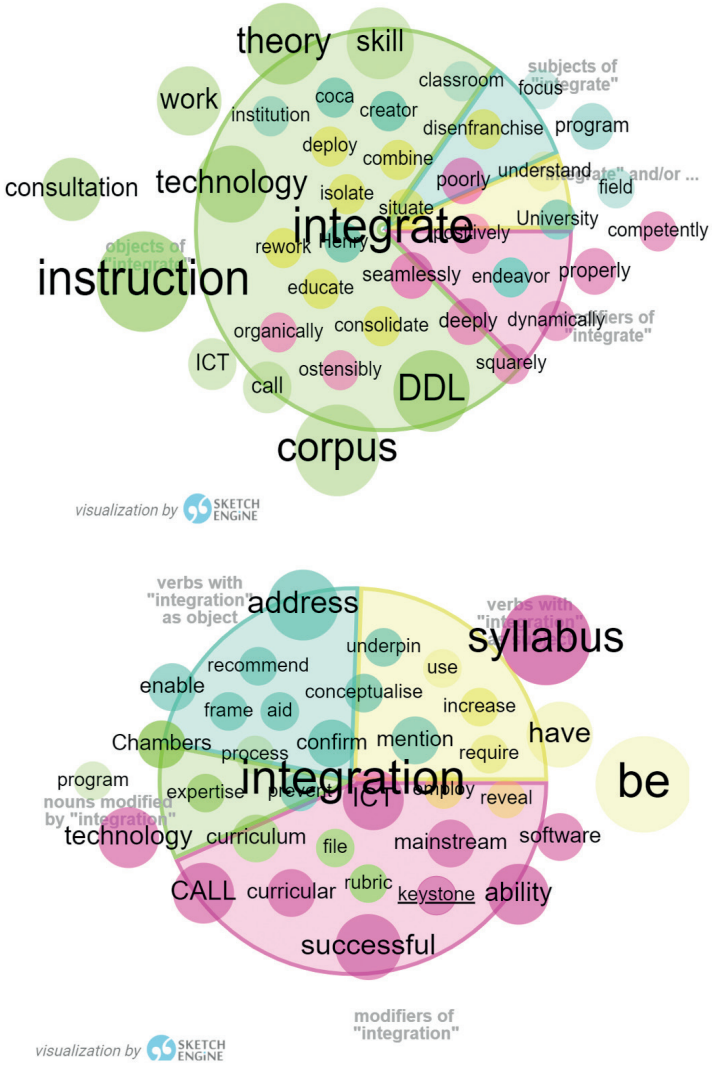

Figure 1. Visualisation of "integrate" and "integration"

Source: corpus (1,277,668 words) of DDL research articles.

refining teachers' pedagogical strategies. The underexplored terrain of matching specific linguistic resources to particular linguistic queries is echoed by Flowerdew (2009: 410) who notes that the question of which resource – be it a corpus, grammar, dictionary, or other – is most suited for specific linguistic inquiries has not been extensively examined in existing research.

5. A PROPOSAL FOR AN ESP TEACHER TRAINING MODEL

Expanding upon Flowerdew's (2009) insights on the optimal linguistic tools for specific inquiries, this paper explores how integrating DDL activities and aligning them with specific tools across the learning phases of the "six Is" framework can enhance the efficacy of these stages. This alignment may well delineate a clear progression of learning activities, ultimately boosting greater learner engagement and comprehension. As a result, the theoretical structured framework proposed by this study employs tools and corpus applications across the learning phases – *introduction, illustration, interaction, induction, intervention and integration* – making the argument for a more targeted and thoughtful application of resources and enhancing the pedagogical strategy.

Considering the above insights, initially, the *introduction* (Quinn 2015) part of a DDL training course could include an overview of corpus linguistics, its significance, and basic operations, alongside hands-off activities like paper-based sample activities (e.g. as in Figure 2, which can be provided as a printout) and demonstrations of online corpus searches, laying the groundwork for corpus referencing. At this stage, corpus searches could be performed with a user-friendly tool especially suitable for novice users. Further, in the *illustration* phase, tools such as concordancers can help demonstrate language patterns and usage, providing clear examples for learners, and can, therefore, serve to construct basic knowledge and functional understanding of corpus tools.

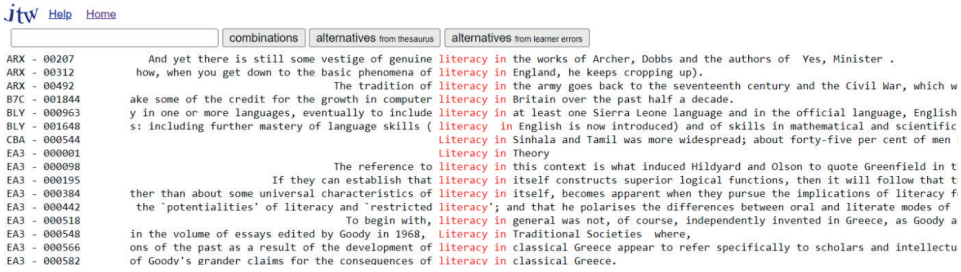

Figure 2. Concordance of the word "Literacy" through Just-the-word tool for Illustration Phase

Source: corpus (1,277,668 words) of DDL research articles.

Regarding the phase of *interaction*, corpus analysis tools may enable students to interactively engage directly with real language data, fostering active exploration and discovery, and engagement with authentic language scenarios, laying the groundwork for the later stages of corpus referencing and application in tasks such as essay revisions and enhancing basic knowledge through

interactive corpus analysis. At this stage, learners could also start exploring their own searches beyond the ones suggested (curriculum, pedagogy, assessment, interactive, inclusive, collaborative) in the sample activities in Table 1.

Table 1. DDL material initiating corpus analysis through Interaction Phase

WORKSHEET: Exploring Educational Terminology 1
Interaction Phase: Engaging with Corpus Data *Sample Activity 1:* Objective: Learn to formulate and refine corpus queries to investigate language use. Explore terms: curriculum, pedagogy, assessment, interactive, inclusive, collaborative 1. In pairs, select three of these educational terms and explore their concordances using a corpus tool. Each pair chooses one corpus tool for exploration (SkeLL, WebCorp, CorpusMate) – Which contexts or phrases are these terms most frequently associated with? – Are they part of specific educational theories or models? 2. Share and discuss your findings with other pairs, then present your findings in class. *Sample Activity 2*: Objective: Learn to formulate and refine corpus queries to investigate language use. In groups, perform queries in different tools to compare the usage of "formative assessment" vs. "summative assessment." Discuss the findings and their implications for teaching and learning.

Source: own study.

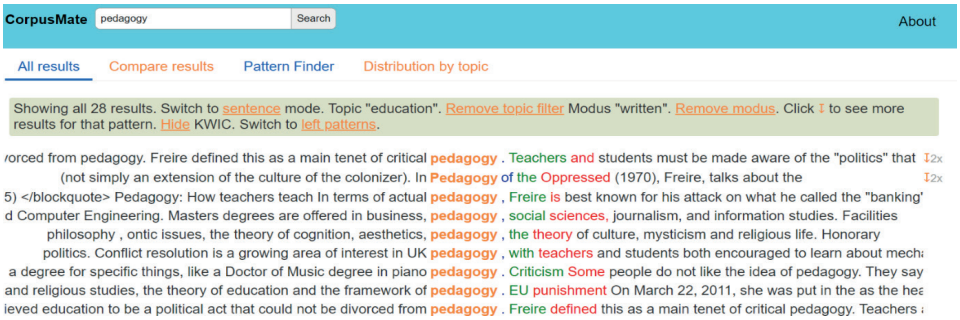

Figure 3. Concordance of the word "pedagogy"

Source: CorpusMate tool for Interaction Phase.

Moving on to the *induction* Phase, learners can utilize software that allows for detailed corpus querying, enabling them to independently make inferences about language rules and patterns. The induction phase "stands for making one's own, or the learning group's rule for a particular feature, a rule which will be refined and honed as more and more

data is encountered" (Carter & McCarthy 1995: 28), thus seamlessly moving to more refined vocabulary acquisition (Table 2), before culminating in the intervention and integration phase.

Table 2. DDL material-Deepening Linguistic Insight through Induction Phase

WORKSHEET: Exploring Educational Terminology 2
Induction Phase: Deepening Linguistic Insight – Objective: Investigate the collocational relationships of *student engagement* in educational discourse. – Approach: Examine the broader lexical context of *student engagement* by identifying words and phrases frequently associated with this term. – Sample Activity: Use Just-the-word to analyse and retrieve a detailed collocation report of *student engagement*. Identify verbs, adjectives, and related nouns that frequently co-occur with this term, and discuss the implications of these associations for strategies aimed at enhancing student participation.

Source: own study.

Next, the *intervention* phase, as suggested by Flowerdew (2009) is crucial in pedagogy, enabling educators to tailor their learning by addressing specific challenges or misunderstandings. Intervention thus, is able to facilitate personalized feedback and guidance, aiding learners in understanding complex language nuances. Through targeted interventions, educators can enhance the transition from theoretical learning to practical application, reinforcing students' mastery of language patterns and usage as in Table 3. As Flowerdew proposes (2009: 407), difficulties encountered by students while inducing phraseological tendencies can be remedied through "clues and prompts […] to mediate the inductive ↔ deductive continuum." As a result, the implementation of "a more delicate approach to corpus queries would help to reduce some of the difficulties associated with interpretation for students" (Flowerdew 2009: 407). As an example, in the suggested worksheet, I used the lexical verb *synthesize*. In the proposed worksheet, the lexical verb "synthesize" may be used as an example, illustrating its diverse application across lexical bundles with distinct register and genre variations across academic disciplines, as in the examples: "synthesize findings" (Data Analysis), "synthesize findings" (Research), "synthesize the main arguments" (Literature Review / Discussion), "synthesize a new compound" (Chemistry), "synthesize data" (Environmental Science), "synthesize accounts" (History), "teachers synthesize instructional strategies" (Pedagogy in Education), "synthesize observations" (Psychology), and finally, "synthesize market trends, consumer feedback, and competitive analysis" (Business).

Finally, in the *integration* phase, digital platforms facilitate the creation and sharing of corpus-based projects, enabling learners to apply their accumulated insights from earlier phases in new authentic communicative practices (see Table 4). Concluding, the proposed framework (Table 5) demonstrates the strategic incorporation of educational tools, directly supporting trainers and teachers, while indirectly benefiting learners by effectively bridging the gap between theoretical knowledge and practical application, thereby positioning the framework for future impact and further development.

Table 3. DDL material initiating corpus analysis through intervention phase

WORKSHEET: Exploring Educational Terminology 3
Intervention: Tailoring Pedagogical Practice – Objective: Customize teaching strategies by using specific corpus findings, facilitating personalized learning experiences that address the unique challenges of learners. – Approach: Use corpus-based activities for targeted feedback and guidance to resolve linguistic challenges and enhance language application skills. Tools: Corpus Mate, Compleat Lexical tutor, AntConc – Example Activity: Learners are presented with corpus examples of the lexical verb *synthesize* in various contexts. Analyse the examples to understand its use, then practice creating sentences or short paragraphs that correctly use *synthesize*, catering to their specific field of study. Review examples of 'differentiated instruction' from Flax and design an activity that incorporates these insights, aiming to meet diverse learner needs.

Source: own study.

Table 4. DDL material-applying insights in ESP teaching through integration phase

WORKSHEET: Exploring Educational Terminology 4
Integration: Applying Insights in ESP Teaching – Objective: Incorporate corpus insights into practical teaching materials. – Approach: Develop or refine educational materials using corpus analysis tools. – Sample Activity 1: Based on corpus findings from SketchEngine, AntConc or CorpusMate on the collocational behaviour of 'development', create a lesson plan that includes authentic examples and activities to clarify the concept, as used in the *Pedagogy in Education* discipline. – Sample activity 2: Disambiguating Collocational Behaviour of three synonymous verbs. Explore and clarify the distinct collocational patterns associated of the verbs "attain," "accomplish," and "achieve" as used in *Pedagogy in Education* discipline, through corpus analysis.

Source: own study.

Table 5. Suggested Training Framework – "six Is"

Learning Phase	Contributing scholar	Corpus Tools/ Practices	Instructional activities	Goals
Introduction	Quinn (2015)	No tools required	Readings/Power Point Presentation	Developing a basic understanding of corpora
Illustration	Carter and McCarthy (1995)	SkeLL, Just the word, Flax, CorpusMate, WebCorp	Illustration of paper-based, *hands-off*, activities Pattern hunting	Gaining insight into corpus Functionality and language patterns
Interaction	Carter and McCarthy (1995) Kennedy and Miceli (2017)	Just the word, Flax, CorpusMate, WebCorp	Collocation / colligation Interactive activities, concordance analysis	Engagement with real language Discovering usage patterns, Pattern hunting
Induction	Carter and McCarthy (1995) Kennedy and Miceli (2017)	Just-the-Word, Corpus Mate, Flax, WebCorp, Compleat Lexical Tutor, No Sketch Engine, Sketch Engine, LancsBox, AntConc	Engaging with corpora for direct language pattern analysis	Enhanced engagement and identification of word associations / Patterns, pattern hunting
Intervention	Flowerdew (2009) Kennedy and Miceli (2017)	Flax, WebCorp, Compleat Lexical Tutor, No Sketch Engine, Sketch Engine, LancsBox, AntConc	Facilitated discussion and critical analysis of corpus findings	Strengthened corpus analytical skills, Corpus literacy, Pattern refining
Integration	Ammari (2025)	Integration of learned tools	Application of corpus insights to language-related, real-world tasks	Application of corpus insights into language tasks

Source: own study.

6. CONCLUSION

Corpus usage has revolutionized ESP and EAP by providing authentic language data, assisting in vocabulary selection, facilitating discourse analysis, revealing collocational patterns, aiding register and genre awareness, and supporting a learner-centred approach. These advancements have had a great impact

on language teaching and learning in specialized professional and academic contexts, enabling instructors to align with real-world demands.

This paper has suggested a theoretically grounded, phase-oriented DDL framework designed to address the practical and pedagogical barriers preventing ESP teachers from effectively incorporating corpus-based resources into their teaching. By detailing a progressive model aligned with the "6 Is" outlined framework, the study presents a structured developmental pathway for teachers to build the necessary corpus-related competencies in a systematic and scaffolded manner to ensure effective DDL integration into teaching practice. Through carefully sequenced instructional activities, this framework facilitates sustainable and accessible corpus-based ESP pedagogy, equipping teachers with the tools needed to navigate technical and methodological challenges.

Positioned at the intersection of ESP and DDL, the framework underscores the role of targeted, customized teacher training that mirrors the needs-based approach of ESP itself. By aligning the learning phases – from introduction to integration – with corpus tools and strategies, the framework fosters progressive mastery of corpus methodologies, ultimately empowering teachers to transition towards autonomy in corpus-based instruction. The addition of the final *integration* phase, in particular, aims to bridge the gap between corpus theory and classroom application, supporting the normalization of corpus-based methodologies in ESP instruction. The sample activities serve as illustrative examples, demonstrating potential strategies for incorporating DDL skills into teaching. They guide the progressive adoption of DDL techniques, enabling ESP educators to achieve autonomy, integration, and normalization and effectively transfer these skills to their students.

Finally, this paper contributes to the growing body of literature on corpus linguistics by proposing a framework that, while promising, requires empirical validation to confirm its efficacy. Fully realizing the potential of corpus-based instruction, however, requires collaborative efforts among instructors, researchers, and learners, ensuring institutional support and quality assurance mechanisms (Belcher 2006). Such collaboration embeds quality assurance within institutional frameworks and bridges the gaps identified in the literature. Future research directions could, therefore, focus on assessing the effectiveness of this DDL framework within ESP contexts, and explore its impact on DDL teacher development and student learning outcomes in applied contexts.

Funding acknowledgement

The author does not declare any financial support for writing this paper.

REFERENCES

Aston, G. (2001). *Learning with corpora.* Houston, TX: Athelstan.

Bax, S. (2003). CALL – past, present and future. *System*, *31*(1), 13–28. https://doi.org/10.1016/S0346-251X(02)00071-4

Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *1*(2), 1–15.

Belcher, D.D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly, 40*(1), 133–156. https://doi.org/10.2307/40264514

Benesch, S. (2001). *Critical English for academic purposes: Theory, politics, and practice* (1st ed.). New York: Routledge.

Bennett, G.R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor: University of Michigan Press.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In: J. M. Sinclair (ed.), *How to use corpora in language teaching* (pp. 15–36). Amsterdam: John Benjamins.

Biesta, G. / Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults*, *39*(2), 132–149.

Bloch, J. (2012). Technology and ESP. In: B. Paltridge / S. Starfield (eds.), *The handbook of English for specific purposes* (pp. 385–401). Malden: Wiley-Blackwell. https://doi.org/10.1002/9781118339855.ch20

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 81–106.

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In: A. Boulton / S. Carter-Thomas / E. Rowley-Jolivet (eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–291). Amsterdam: John Benjamins.

Boulton, A. / Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348–393. https://onlinelibrary.wiley.com/doi/10.1111/lang.12224

Breyer, Y.A. (2011). *Corpora in language teaching and learning: Potential, evaluation, challenges*. Frankfurt: Peter Lang.

Cadman, K. (2002). English for academic possibilities: The research proposal as a contested site in postgraduate genre pedagogy. *Journal of English for Academic Purposes*, *1*(2), 85–104.

Callies, M. (2016). Towards corpus literacy in language teacher education. In: *Proceedings of the 12th Teaching and Language Corpora Conference (TaLC)* (pp. 1–2). Giessen: Justus Liebig University.

Carter, R. / McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158. https://academic.oup.com/applij/article/16/2/141/211140

Casanave, C.P. (2002). *Writing games: Multicultural case studies of academic literacy practices in higher education.* Mahwah: Laurence Erlbaum.

Chambers, A. (2019). Towards the corpus revolution? Bridging the research – practice gap. *Language Teaching, 52*(4), 460–475. https://doi.org/10.1017/S0261444819000089

Chambers, A. / O'Sullivan, I. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, *16*(1), 158–172. https://doi.org/10.1017/S0958344004001211

Charles, M. (2022). Corpora and autonomous language learning. In: R. Poole (ed.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 406–419). New York: Routledge. https://doi.org/10.4324/9780367824471-29

Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In: D. Wood (ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 107–125). London: Continuum.

Chong, C.S. (2016). Ten innovations that have changed English language teaching. *British Council: Voices Magazine*. https://www.britishcouncil.org/voices-magazine/ten-innovations-have-changed-english-language-teaching

Cobb, T. / Boulton, A. (2015). Classroom applications of corpus analysis. In: D. Biber / R. Reppen (eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478–497). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.025

Corino, E. / Onesti, C. (2019). Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education, 4*(7), 1–12. https://doi.org/10.3389/feduc.2019.00007

Coxhead, A. / Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, *16*(3), 129–147. https://doi.org/10.1016/j.jslw.2007.07.002

Crosthwaite, P. / Boulton, A. (in press). DDL is dead? Long live DDL! Expanding the boundaries of data-driven learning. In: H. Tyne / M. Bilger / L. Buscail / M. Leray / N. Curry / C. Pérez-Sabater (eds.), *Discovering language: Learning and affordance*. Peter Lang [in press].

Crosthwaite, P. / Luciana, L. / Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, *36*(7), 1392–1420.

De Chazal, E. (2014). *English for academic purposes: Oxford handbooks for language teachers*. Oxford / New York: Oxford University Press.

Dudley-Evans, T. / St. John, M.J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.

Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes, 7*(2), 77–86. https://doi.org/10.1016/j.jeap.2008.03.002

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, *14*(3), 393–417.

Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, *45*(4), 475–489.

Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam / Philadelphia: John Benjamins.

Gillespie, J. (2020). CALL research: Where are we now? *ReCALL, 32*(2), 127–144. https://doi.org/10.1017/S0958344020000051

Halliday, M.A.K. / Martin, J.R. (1993). *Writing science: Literacy and discursive power*. London / New York: Routledge.

Han, S. / Shin, J.A. (2017). Teaching Google search techniques in an L2 academic writing context. *Language Learning & Technology, 21*(3), 172–194. https://doi.org/10125/44626

Heather, J. / Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education, 20*(4), 415–440.

Higgins, J.J. (1988). *Language, learners and computers*. Harlow: Addison-Wesley.

Hutchinson, T. / Waters, A. (1987). *English for specific purposes*. Cambridge: Cambridge University Press.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes, 21*(4), 385–395. Amsterdam: Elsevier. https://doi.org/10.1016/S0889-4906(01)00028-X

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing* (Michigan classics ed.). Ann Arbor: University of Michigan Press.

Johns, T. (1991a). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In: T. Odlin (ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge: Cambridge University Press.

Johns, T. (1991b). Should you be persuaded: Two samples of data-driven learning materials. Classroom concordancing. *English Language Research Journal, 4*, 1–16.

Kennedy, C. / Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1–2), 91–114. https://doi.org/10.1080/09588221.2016.1264427

Kennedy, G. (2008). Phraseology and language pedagogy. In: F. Meunier / S. Granger (eds.), *Phraseology in foreign language learning and teaching* (pp. 21–41). Amsterdam / Philadelphia: John Benjamins.

Leńko-Szymańska, A. (2014a). *Data driven learning in teacher training: Tackling the challenge.* Paper presented at the 11th TALC Conference, Lancaster, UK. https://ucrel.lancs.ac.uk/talc2014/doc/talc2014-abstract-book.pdf

Leńko-Szymańska, A. (2014b). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL, 26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology, 21*(3), 217–241.

Long, M.H. (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.

McEnery, T. / Xiao, R. (2011). What corpora can offer in language teaching and learning. In: E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (pp. 364–380). New York / London: Routledge.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In: U. Connor / T. Upton (eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 239–250). Amsterdam: Rodopi. https://doi.org/10.1163/9789004333772_014

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269–286.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/scl.2

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1–2), 36–61. https://doi.org/10.1080/09588221.2019.1667832

Pho, P.D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, *10*(2), 231–250.

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *Elt Journal*, *69*(2), 165–177.

Richards, J.C. / Schmidt, R. (2002). *Longman dictionary of applied linguistics and language teaching*. Harlow: Longman.

Römer, U. / Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research, 2*(2), 99–127. https://doi.org/10.17239/jowr-2010.02.02.3

Rundell, M. / Granger, S. (2007). From corpora to confidence. *English Teaching Professional, 50*(1), 15–18.

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In: J.M. Sinclair (ed.), *How to use corpora in language teaching* (pp. 271–299). Amsterdam: John Benjamins.

Stubbs, M. (2007). Notes on the history of corpus linguistics and empirical semantics. In: M. Hoey / M. Mahlberg / M. Stubbs / W. Teubert (eds.), *Text, discourse and corpora: Theory and analysis* (pp. 317–329). London / New York: Bloomsbury.

Swales, J. (1985). English as the international language of research. *RELC Journal, 16*(1), 1–7. https://doi.org/10.1177/003368828501600101

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/scl.6

Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In: A. Leńko-Szymańska / A. Boulton (eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37–62). Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/scl.69.03tri

Vyatkina, N. / Boulton, A. (2017). Corpora in language teaching and learning. *Language Learning & Technology, 21*(3), 66–89.

Wu, S. / Fitzgerald, A. / Franken, M. (2019). *Making use of and adapting MOOC text resources for language learning*. Paper presented at the 2019 International Conference of Artificial Intelligence and Technology-Enhanced Language Learning (AiTELL) together with the Post-Graduate Academic Forum, Shanghai, China. https://hdl.handle.net/10289/13201

Yoon, H. / Jo, J. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *ReCALL, 26*(1), 54–69. http://dx.doi.org/10125/44356

ONLINE MEDIA

AntConc, http://laurenceantony.net
Compleat Lexical Tutor, http://lextutor.ca
CorpusMate, http://corpusmate.com
English Corpora, http://www.english-corpora.org
LancsBox, http://lancsbox.lancs.ac.uk
Flax, http://flax.nzdl.org
FrazeIt, http://frazeit.it
Just-the-word, http://www.just-the-word.com
Ozdic, http://ozdic.com
SkeLL, http://skell.sketchengine.eu
Sketch Engine, http://sketchengine.eu
WebCorp, http://www.webcorp.org.uk/live/

Received: 8.04.2024; **revised:** 14.02.2025

Nauczanie języka angielskiego do celów zawodowych poprzez DDL: ustrukturyzowane ramy szkoleniowe

Abstrakt. Badanie ukazuje potrzebę włączenia językoznawstwa korpusowego do nauczania języka angielskiego do celów zawodowych, wskazując na duży, choć niedostatecznie wykorzystywany potencjał tego podejścia. Ograniczone zastosowanie korpusów często wynika z braku przygotowania nauczycieli, co podkreśla konieczność rozwoju odpowiednich kompetencji i dostosowania dydaktyki do zróżnicowanych potrzeb językowych uczących się. W artykule zaproponowano zintegrowany model kształcenia, oparty na koncepcji „3I" Cartera i McCarthy'ego (1995) i rozszerzonej wersji „4I" Flowerdewa (2009), rozbudowany do „6I", który równoważy podejście indukcyjne i dedukcyjne. Badanie pokazuje, jak specjalistyczne szkolenia umożliwiają skuteczne wdrażanie Data-Driven Learning (DDL) oraz integrację analiz korpusowych w procesie dydaktycznym. Przedstawione rozwiązania wspierają rozwój zawodowy nauczycieli języków specjalistycznych,

wzmacniając ich kompetencje w zakresie pracy z materiałami korpusowymi. Zaproponowany model kształcenia stanowi narzędzie ułatwiające efektywne wykorzystanie korpusów w nauczaniu oraz wpisuje się w bieżącą debatę dotyczącą normalizacji dydaktyki korpusowej i modernizacji szkoleń dla nauczycieli języków specjalistycznych.

Słowa kluczowe: językoznawstwo korpusowe, data-driven learning, szkolenie nauczycieli języków specjalistycznych, nauczanie indukcyjne i dedukcyjne, model szkoleniowy „6I".

MARIA AMMARI
Adam Mickiewicz University, Poznań, Poland
maria.ammari@amu.edu.pl
https://orcid.org/0000-0002-5386-2427