ISSN: 2716-9189 E-ISSN: 2676-1750
Journal of Studies in Language, Culture, and Society (JSLCS). (08) 03, 2025. (PP. 37-48)
Received: 12/01/2025 Accepted: 29/05/2025 Published: 30/07/2025

ENHANCING ESP LANGUAGE LEARNING WITH CORPUS TOOLS AND DATA-DRIVEN APPROACHES

Houda Boumediene[1]
[1] **AILE Laboratory, Amar Telidji University of Laghouat, Algeria**
h.boumediene@lagh-univ.dz

Abstract: This article discusses how corpus tools and techniques can be integrated into the ESP classroom to promote learners' specialized language proficiency. Guided by the principles of DDL, this study demonstrates corpus-based methods that enable learners to develop a sophisticated understanding of professional and academic language use. The study, based on both existing corpora and custom-built datasets, manages to show some extra-mural purposes that could guide learners at different levels of proficiency to master the domain-specific varieties of languages. The methodology adopted in this study is both quantitative and qualitative in nature; the research work was supported by corpus analysis tools such as COCA and Sketch Engine in order to analyze authentic linguistic data. The results highlight the recurring lexical patterns and structures employed in the specialized discourse, focusing on compound nouns relevance to the field of pedagogy and psychology. The statistical methods are based on frequency analysis and concordance observation while the qualitative techniques involve synthesis, classification, and semantic analysis. All these are combined in a way that ESP instruction comprehensively examines corpus-driven insights. The findings show that the data-driven approach outperforms intuition-based instruction regarding the delivery of precise, contextualized insights into patterns of language. With this approach, learners can work directly with authentic texts and linguistic data; thus, developing in them a critical approach and deeper understanding of specialized information. The study further points out that educators and learners have to be specifically trained in corporations, that corpora have to be aligned with course goals, and that technologies must become more user-friendly for wider diffusion. It finally argues that corpus techniques should be systematically included in ESP curricula to enhance both language acquisition and critical thinking in specialized fields. This will help bridge both theoretical and practical knowledge and prepare the learner for effective communication in their respective professional domains.

***Keywords**:* Corpus Analysis, Corpus Tools, Critical Engagement, Data-Driven Learning (DDL), Datasets, ESP courses

How to cite the article :
Boumediene, H. (2025). Enhancing ESP Language Learning with Corpus Tools and Data-Driven Approaches. *Journal of Studies in Language, Culture, and Society (JSLCS)8*(3), 37-48.

[1] **Corresponding author :** Houda Boumediene , **ORCID** https://orcid.org/0000-0002-3264-7867

1. Introduction

Such a rise in the economy of the whole world and in interdisciplinary relationships within academic disciplines puts growingly higher demands on the quality of training and educational activities intended for future professionals from different walks of life. Yet, these factors influence the preparation of future teachers and social pedagogues, changing the core expectations surrounding their competency and knowledge (Boumediene, 2025, Baker et al. 2006). Now, modern educators and social pedagogues are obliged to include new theoretical and methodological principles in their professional development. Students also might not be able to perceive the need for such emerging trends at times (Idri 2025, Leleka & Moskalenko, 2018). One of the important tasks of a modern educator is explaining one's philosophies and approaches. In this connection, many scholars stress interdisciplinary, integrative, and content-based aspects concerning the developing need of interdisciplinarity. (Leláková,2018)

A modern professionally equipped teacher or a social pedagogue should possess elaborate skills both in mother tongue and foreign languages, especially in mastering the terminological vocabulary of pedagogy and psychology. This kind of terminological competence will enhance an ability to understand and process foreign language critical information while fostering professional growth (Idri et al.,2025, Hundarenko, 2020)

These considerations give reason to emphasize the establishment of interdisciplinary connections in teaching a specialized second language. In other words, the integration of knowledge from different scientific fields has turned into one of the most effective approaches to solving the problems of contemporary education (O'Keeffe et al., 2007). According to Karpenko (2017), Bouguebs & Djouima (2025) languages are complex and dynamic systems, and their properties are predetermined by linguistic and extra-linguistic factors that naturally allow for or even promote an interdisciplinary approach, which is an urgent need in searching for novel educational solutions.

2. Literature Review

To meet the interdisciplinary needs of this research, we draw upon a corpus linguistic theoretical and methodological framework. Corpus linguistics, in general terms, refers to large collections of language data, or corpora, which are analyzed using computational tools to uncover linguistic patterns and insights. It was originally developed by Baker et al. (2006) and Frankenberg-Garcia et al. (2011) and, in the last couple of decades, corpus linguistics has become one of the most rapidly growing methodologies for analyzing various aspects of language. In return, the empirical nature of the vocabulary has made it very influential in vocabulary studies in that it provides the data to be analyzed directly, and changes methodological approaches taken; thus, far in this area.

As a sub-discipline that deals with discourse-related research, corpus linguistics takes representative samples of language, corpora, for its research. (Rafajlovičová, 2018) It allows grasping in detail how these linguistic systems are at work within contexts, and in this way offers evidence-based tools for deeper insights into real-world language use. In this direction, Lacková (2021) highlights the contribution of corpus linguistics to ESP teaching, especially regarding pedagogy and psychology. It evidences how corpora allow the study of specialized terms and their uses in professional settings.

With the integration of corpus-based methods, corpus resources have also been extended to language teaching and learning. Scholars have argued with increasing interest for the integration of corpus linguistics into a language curriculum at different levels. However, Gavioli (2005) and Granger (2002) identify certain issues while integrating corpus-based approaches; Carmen et al. (2010), Liu et al. (2017), McEnery et al. (2012), & O'Keeffe et al. (2007) discuss various challenges in research. These involve addressing the complexity of subjects and their relationship to pedagogical objectives. Although corpora are a very powerful tool, Granger (2009) warns against excessive reliance on them and asserts that they

should support natural communication and traditional teaching techniques rather than replace them.

Researchers investigate the properties of lexical units in discourse in terms of semantic dimensions for teaching purposes. Corpus concordances and collocations make apparent how lexical semantics is shaped by context; therefore, it questions whether terms have any inherent meaning or if they gain it from usage. From a psycholinguistic point of view, corpora disclose not only lexical contexts, but also how language users associate words with surrounding lexical units. Moreover, linguistic borrowing often reshapes the lexical structure of a language, borrowing words from one language and adapting them into another. As Lacková (2021) confirms, the application of corpus data in ESP teaching gives learners an opportunity to observe how linguistic structure expresses professional discourse, and therefore, this gives them even more field-specific navigation skills.

In the current study, in selecting items for corpus analysis, the basis is relative frequency; however, this is a varying measure across corpora. According to Mahlberg (2005), frequency lists shed light upon textual structures, but do not carry meaning in themselves. As Stubbs (2001) emphasizes, the results of computation, like collocations and frequencies, describe formal features but require interpretation in order to understand textual meaning. This is in tandem with Lacková's (2021) observation that interpreting corpus data needs sensitive alignment into the pedagogical objectives relevant to their applicability in ESP instruction.

Another central concept within corpus research, to inform teaching is that of the lexical syllabus, which enumerates recurrent patterns in order of descending frequency. Since the early 1990s, the lexical syllabus has combined basic lexical structures with grammatical features; thus, providing a usage-based approach to language teaching. This corresponds to the aims of professionally oriented teaching of compound nouns and phraseological units, which presupposes semantic units being contextualized within larger linguistic frameworks. Halliday et al. (2004) integrate such frameworks to show how corpus-driven insights into the nature of language can meet both linguistic and interdisciplinary needs; and hence, contributing to a more holistic approach to language education.

While there has certainly been a growing amount of research into how corpus tools can be integrated into ESP instruction, the present study offers a unique focus on the actual implementation of both ready corpora and self-compiled databases. Also, while many previous studies have targeted general or technical language in their respective domains, the current study will place greater emphasis on domain-specific varieties of language within pedagogy and psychology. Above all, it introduces a methodological framework embedding the most advanced tools, COCA and Sketch Engine, in the combined use of quantitative and qualitative approaches that allow for a sound analysis of recurring lexical patterns concerning their contextual relevance within specialized discourse. Addressing some specific challenges, like aligning corpus-driven techniques with course objectives and bridging theory with practice, the present study now provides a fresh look at how corpus linguistics can be systematically integrated into ESP curricula to enhance both professional communication and critical thinking skills.

3. Research Objectives

The present study focuses primarily on the development of expert-oriented lexical units in pedagogy and psychology, being a tool for equipping future teachers and social pedagogues with the linguistic means of communication. Both language teachers and subject specialists should work on the elaboration of the linguistic and communicative components of professionally oriented language.

This should underpin the need to include interdisciplinary phenomena in professionally oriented language teaching with the purpose of enhancing professional communication competence for future teachers and social workers. A core issue is how the use of a holistic approach in knowledge and skill development is applied to the process of teaching which supports comprehensive development of an individual.

Another objective of this research is to establish the relationship between the skills and knowledge gained by students using corpus-driven exercises and their general communicative competence in the English language. These exercises have been a challenge to the instructors in the task design, exercises, and activities that inspire students to achieve proficiency in English while responding to professional needs.

4. Methodology

The strategic combination of methods, techniques, and tools that would be crucial to realize the objectives of the study is based on aggregative needs for data collection, contributing to comprehensive analysis, and hence, a mixed-method research design was adopted wherein both quantitative-approaches, like frequency analysis of concordances and statistical calculations-and qualitative approaches-such as observation, comparison, synthesis, and generalization-could be combined in their procedural aspect. This would afford a balanced investigation into the problem at hand, which will be further elaborated procedurally in the sections to follow.

This study draws on COCA, a large and richly detailed corpus of contemporary American English, to explore the common patterns in which specific lexical items tend to appear. Identifying such patterns is an essential step in understanding how language functions in context and how meaning is shaped within discourse (Schmitt, 2010). Recognizing recurring structures in language not only deepens our understanding of linguistic behavior but also contributes to developing more effective communication skills (Beaugrande & de Dressler, 1981).

The decision to use COCA and Sketch Engine stems from their depth, flexibility, and reliability. COCA's extensive coverage—spanning spoken and written texts from various fields—makes it particularly valuable for studying how words are used in real-life situations. Its user-friendly design and robust tools for analyzing collocations and frequency patterns support both teaching and research.

At the same time, Sketch Engine adds an important dimension by allowing users to access a wide range of corpora in multiple languages. One of its standout features, Word Sketch, provides insightful summaries of how words behave grammatically and semantically. It also supports custom corpus creation and detailed searches, making it ideal for more specialized linguistic investigations.

Compared to other tools like AntConc, WebCorp, or WordSmith Tools, which may have limitations in terms of features or corpus diversity, COCA and Sketch Engine offer a more comprehensive and versatile approach. Their selection was guided by the needs of this research: to examine how language works in context and to uncover the subtle patterns that shape meaning in professional and academic communication.

Given the broad range of psychological and pedagogical issues covered in the training of future teachers and social pedagogues, such as emotional intelligence, neuropsychology, and social psychology, preparatory work involved the identification of certain terms on which the focus of analysis would be concentrated (Ouarniki & Alhasani,2025 , Flowerdew, 2012). The attention was focused on multi-word lexical units that refer, first of all, to a number of core concepts: thus, amnesia (retrograde amnesia, source amnesia), memory (working memory, episodic memory), and recall (cued recall, free recall).The research relies on corpus linguistics methodologies facilitated by tools like Sketch Engine, which enable detailed exploration of linguistic information within discourse. These tools help reveal context-specific

linguistic behaviors, rather than abstract features. Methodological procedures, such as classification, synthesis, and comparison, are applied to group collocates based on shared attributes while descriptive methods are used to organize and present findings effectively.

Computational tools and corpora provide the basis for analyzing qualitative linguistic features. These tools allow the analysis of immediate discourse contexts in which lexical units occur, therefore uncovering meaningful connections between patterns. Further, only the statistically significant lemmas are selected for analysis to ensure robustness, with exclusions of terms below 0.01 occurrences per million words.

The lexical units for this study are selected and are deliberately kept limited to allow for a more focused and intense analysis of the same. The compound nouns selected have been organized in three frequency tiers, namely,

1. High-frequency terms : long-term memory, working memory, episodic memory, false memory, short-term memory.
2. Middle-frequency terms : cued recall, free recall, explicit memory, declarative memory, implicit memory, procedural memory, recovered memory, and anterograde amnesia.
3. Low-frequency terms : indirect memory, state-dependent memory, source amnesia, infantile amnesia, childhood amnesia, retrograde amnesia.

This kind of categorization in tiers enables a systematic inquiry into what all these terms do in discourse and their interdisciplinary relevance to pedagogy and psychology. The integration of quantitative and qualitative approaches allows going into minute details of professionally oriented lexical units with a focus on the importance of such units for enhancing professional communicative skills among future educators and social pedagogues.

5. Results

As a starting point for our research, in Table 1 we provide the concordance frequency of the multi-word lexical units comprising the nouns amnesia, memory, and recall within the COCA corpus. The data in this table provides a detailed enumeration of the number of occurrences in which these collocations actually occurred, bringing to light the relative uses of each compound noun. This is an important point from which to start with a view to assessing how these lexical units function and/or interact in particular discourse contexts.

Table1.

Number of Occurrences of Investigated Compound Nouns in COCA

Compound Noun	Amount of Occurrences	Frequency per Million (f per mil.)
cued recall	140	0.01 per M
free recall	400	0.02 per M
declarative memory	460	0.03 per M
episodic memory	1,890	0.12 per M
explicit memory	420	0.03 per M
direct memory	1,010	0.07 per M
false memory	980	0.06 per M
implicit memory	590	0.04 per M
indirect memory	25	less than 0.01 per M
long-term memory	3,750	0.25 per M
procedural memory	380	0.02 per M
recovered memory	230	0.01 per M
semantic memory	770	0.05 per M
short-term memory	5,320	0.35 per M
state-dependent memory	15	less than 0.01 per M
working memory	14,500	0.93 per M
source amnesia	12	less than 0.01 per M
anterograde amnesia	200	0.01 per M
infantile amnesia	55	less than 0.01 per M
childhood amnesia	48	less than 0.01 per M
retrograde amnesia	210	less than 0.01 per M

The analysis of the COCA corpus provides valuable insights into the discourse significance of professionally oriented compound nouns, particularly those containing *amnesia*, *memory*, and *recall*. These terms appear in a variety of structural and semantic patterns, underscoring their importance in professional and educational contexts. By systematically analyzing their usage, it becomes possible to identify meaningful patterns while excluding irrelevant occurrences. This classification allows for a deeper understanding of their linguistic behaviors and practical applications.

These compound nouns display predictable patterns both in their use with determiner-possessive pronouns and definite articles-and with various modifiers. The modifiers include verbal phrases-evoke, erase, refresh-and adjectives such as traumatic, false, spatial. In COCA adjectives modifying such compound nouns can be semantic role classified as follows. For example, some adjectives denote the state of mind: dissociative, psychogenic, sensory; others denote the moment of time: lasting, recent, and historical. Further groups are represented by

subjective attitude descriptors like cherished, unforgettable, and bitter, while one more group shows terms which cannot fit the above categories, for example, institutional, nostalgic, or cultural. Such categorization will shed more light on the complex meaning and usage of these terms in discourse.

Semantic proliferation of these lexical units is evidenced also by their frequent collocational use with verbs. The verbs trigger, rekindle, and jog enrich the semantic features of these complex nouns, often with metaphorical overtones. Thus, for example, verbs like rekindle describe not only cognitive processes but also imply emotional aspects, which is why they will be of special relevance in the future, for social pedagogues and educators. Often, such terms are combined with prepositional items like of, with, in, and on, relating them to external conditions and further illustrating their discoursal flexibility.

Another aspect of the grammatical function of articles in modifying these compound nouns is the attention to the type of article used-in this case, the definite article. In the COCA corpus, definite articles appear in more than 1,483 concordances that fulfill two main functions: cataphoric and anaphoric references. In cataphoric reference, the article introduces information that shall be explained later in the discourse, as in "The episodic memory was clinically targeted by the advertisements." In anaphoric reference, the article links the term to ideas mentioned previously, as in "The signals were transferred to the long-term memory for further analysis." Indefinite articles, appearing in 41 concordances, are majorly used in descriptive purposes like "A cued recall process validates the character of memory issues." The proof of such patterns underscores the aspect of grammatical determination for these terms in effective communication.

The findings obtained within the framework of the COCA corpus are highly important for teaching. Given analysis underlines the possibility of professionally oriented compound nouns being successfully integrated into the interdisciplinary teaching strategies while working with future teachers and social pedagogues. Some creative and communicative exercises, for example vocabulary building activity, grammatical practice, and interactive tasks provide some very practical ways to enhance the students' communicative competence. Matching terms with their definitions, filling in gaps in sentences, or completion concordances allow students to work with the fine lexical and grammatical detail of these terms. Interactive exercises like crosswords and word searches further promote active learning and retention.

These teaching methodologies enrich not only the linguistic proficiency of students but also motivate them to go deeper into the subject beyond the class. On incorporating data-driven exercises in the curriculum, the latter gain profound knowledge of professionally oriented lexical units, acquiring critical professional skills simultaneously. Interdisciplinary approaches like these specify the significance of corpus-driven insights applied to educational practices while fostering expertise in linguistic and professional fields.

The analysis of the COCA corpus points out the complexity and adaptability of the field for pedagogy and psychology. These findings underpin their crucial role in professional communication, including the possibility of using it in an interdisciplinary way. By integrating these findings into teaching, educators can contribute to developing both linguistic competence and professional skills for future teachers and social workers by bridging theoretical knowledge with its practical application.

6. Discussion

Data-driven exercises and real examples of language provide a more dynamic and realistic understanding of the language than traditional reference books. Such methodology of teaching terminology enables the acquisition of durable linguistic skills both in first mother tongue and in English. The data-driven sentences are selected on principles of exemplarity,

relevance, and typicality, which guarantee the meaningfulness of the examples and their adequacy in relation to the objectives of the lesson and the communicative intent.

For example, in medicine, the learning could analyze terms such as cardiac arrest and emergency response, bringing into focus modifiers like sudden or fatal. In business, it could review collocations in phrases such as profit margins or risk analysis. Such exercises would enable the students to connect very abstract terminology to concrete usage and building engagement and better retention as well.

An example is that in a lesson on memory terms, the sentence "The episodic memory enables the individual to remember certain events within his past" shows how a term is used in real life within a psychological framework. Similarly, a sentence like "Short-term memory is often affected by sleep deprivation" indicates the usage of the term in a scientific manner. These examples explain what the terms are and how they function in a real-life discourse.

By selecting sentences which are representative of typical usage patterns, students learn the language in context, therefore helping them understand both the terms and how they can be used. For example, if one were to consider a pedagogical term such as "Project-based learning prompts an active role from the students"; this places the term into a normal teaching situation to which the learner can more easily relate, therefore understanding the language. This approach ensures that the language examples support the comprehension of both content and the communicative aims of the lesson.

The results of the COCA analysis identify compound nouns such as episodic memory and working memory as important in developing the professional lexicon. Many of these terms have predictable modifiers, traumatic, false, short-term, that further identify their specialized use in the psychology and pedagogy discourses. This is in line with the current scholarship emphasizing authentic data-driven learning in ESP, e.g., Granger (2009), and further illustrates the ways these terms are used in discourse.

Exposed to real linguistic data, the learner can have a more sufficient understanding of the professional expressions and apply them correctly. For example, concordances of episodic memory can exemplify its semantic shading, and the exercises presented on this material could involve gap-filling, classification, and even role-play. Such findings confirm that more significance needs to be given to systematization and the use of data rather than intuition-based instruction.

However, some barriers remain: a lack of experience in the use of corpus tools by teachers and some technical aspects of platforms like COCA make the overall usage not very widespread. Moreover, less common phrases, such as state-dependent memory, may not have sufficient examples to get strong insights from. Notwithstanding these limitations, this study underlines the revolutionary potential of corpus-driven techniques in ESP teaching by highlighting how such methods should be included in curricula that support both linguistic and professional development.

7. Conclusion, Limitations and Recommendations

The lexicosemantic and morphological features of the selected multi-word lexical units including amnesia, memory, and recall in the COCA correspond to the laws of modern grammar of English. At the same time, they reflect the practical communicative needs of the language users conditioned by the psychological and sociological contexts. It follows from the discussion that the same word-formation procedures employed in the Germanic languages for general lexical items are applied to professional terminology in sciences such as psychology and pedagogy, specialist terms are created.

One of the significant outcomes that have emerged from the study is that these compound nouns can be flexible enough to suit contexts or situations in which they exist. This possibility depends on three major issues:

1. The speakers of English may not even discern every shade of meaning.
2. The desirability of using words depends upon the extralinguistic realities-the psychological and sociological context in which they were applied.
3. The semantic inter-relationship among these compound nouns is seen to be versatile in communication.

Moreover, the analysis highlights the lemmatization of words-such as the noun memory acting in an adjectival role when used in collocations like memory cells, memory capacity, and memory formation. COCA helps students draw these contextual differences for a better understanding of whether memory is functioning as a noun or adjective in context.

7.1.Limitations

The following are some of the limitations of using COCA for corpus-driven analysis despite its utility:

- Typology of Texts: COCA does not always differentiate between professionally oriented texts and those written for the general audience. This may introduce some noise into the analysis.
- Constraints in the Text: The morphological and lexical patterns in the corpus may not represent all the variations necessary for specialized language teaching.
- Corpus Representation: Although COCA is rich in data, it represents mainly American English, and that alone may not explain the different terminologies used in other varieties or even contexts of English.

The present study once again testifies to the usefulness of COCA in professional terminology teaching, since the authentic linguistic data given therein serve as a boost toward effective second-language learning.

7.2.Recommendations

These findings emphasize the possibility of exploiting the COCA corpus further in teaching professional terms, particularly ESP. To make this possible, there should be some attention paid to the following proposals, which are targeted at educators, institutions, researchers, and learners.

The educators should introduce COCA in the syllabus to give the students a chance to get acquainted with real examples of the professional use of the language. Emphasizing authentic settings, the educator can draw learners' attention to semantic relation patterns, word-formation mechanisms, and collocations relevant to the special purpose. In this respect, for example, the phrase episodic memory or working memory needs an in-depth analysis regarding the ways these terms function in speech. This should then be further enhanced with appropriate linguistic exercises of contextual analysis or gap-filling-type. For example, pointing out the contrastive difference between memory as a noun in 'long-term memory', and as an adjective in 'memory formation' develops critical thinking and increases lexical delicacy. Secondly, the teacher should also have the students practice independent exploration of COCA. This puts the student in an apt position of understanding patterns and meanings independently, which is another aspect of autonomous learning.

Institutional level: Training programs for educators are of essence. That means workshops and professional development that introduce teachers to corpus tools, such as COCA, and how to use the resources in the classroom. Institutions should also create easy-to-use materials for specific professional fields, such as psychology or pedagogy. That is, resources could include concordance-based worksheets, visualizations of frequency data, or pre-designed exercises to suit course goals. Such support enables the teacher to easily embed

corpus-based methods into teaching, even in cases where teachers had limited experience with such tools.

One promising direction of further research would be an expansion in the scope of corpus-based studies. A search for additional corpora, either general ones like the BNC, or domain-specific academic datasets, would allow for comparative investigations into the ways that terminology differs across varieties of English or professional genres. The study of COCA integration into multilingual learning environments opens a great number of possibilities for bridges between English and other languages. The research might be focused on how learners transfer corpus-based insights in English to their native or additional languages with the purpose of enhancing general linguistic competence. There is also a need for longitudinal studies that would measure the effectiveness of the instructions based on COCA over a period of time for students in being able to use professional terminology correctly in real-life settings.

Learners themselves may immensely benefit from the use of COCA as a useful tool in language acquisition. The analysis of frequent terms, modifiers, and contexts gives them valuable insight into how language works within the areas of their concern. Data-driven exercises, such as classification tasks or contextual analysis, provide them with numerous opportunities for developing their vocabulary and using specialized terms correctly in communication. With this, learners are considerably involved in their own learning; this leads to increased motivation and more involvement.

The recommendations put forward in this article further introduce the level of change that COCA might represent in ESP instruction. In this respect, pedagogic incorporation of corpus-driven techniques bridges the gap between theoretical knowledge and practical application, embedding within learners the necessary linguistic and professional competencies to function comfortably in global, interdisciplinary environments. This approach will need to be effectively taken up by institutions and educators alike, in addition to the learners themselves, to make the educational experience more dynamic, relevant, and effective.

References

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.

Beaugrande, R. A., & Dressler, W. U. (1981). *Introduction to text linguistics*. Longman.

Bouguebs, R., & Djouima, L. (2025). Adopting CAQDAS in Algerian EFL Research: Barriers, Challenges, and the Need for Professional Development. *Langues & Cultures*, *6*(1), 269-283.

Boumediene, H. (2025). The AI Evolution in Higher Education: Enhancing Teaching with ChatGPT. *Journal of Studies in Language, Culture, and Society (JSLCS*), 8(1), 15-29 https://asjp.cerist.dz/en/article/266930

Carmen, M., Campoy-Cubillo, M. C., Bellés-Fortuño, B., & Gea-Valor, M. L. (2010). *Corpus-based approaches to English language teaching*. Continuum.

Corpus of Contemporary American English (COCA). (n.d.). https://www.english-corpora.org/coca/

Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan.

Frankenberg-Garcia, A., Flowerdew, L., & Aston, G. (2011). *New trends in corpora and language learning*. Continuum.

Gavioli, L. (2005). *Exploring corpora for ESP learning*. John Benjamins.

Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. John Benjamins.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–32). John Benjamins. https://doi.org/10.1075/scl.33.04gra

Halliday, M. A. K., Teubert, W., Yallop, C., & Čermáková, A. (2004). *Lexicology and corpus linguistics: An introduction*. Continuum.

Hundarenko, O. (2020). Students' perspectives on academic writing in European higher education (Based on 2019 Erasmus teaching experience in Slovak and Hungarian universities). *Revista Românească pentru Educație Multidimensională, 12*(4), 87–102. https://doi.org/10.18662/rrem/12.4/335

Idri, N. (2025, April 21–22). *Policy to practice: Implementing Algeria's Vision 2030 through EMI-ESP-AI convergence in higher education*. Paper presented at the National Hybrid Conference on Empowering ESP Practitioners for Enhanced EMI in Higher Education: Addressing Challenges, Seizing Opportunities, and Strategizing for Academic Excellence, École Normale Supérieure of Constantine.

Idri, N., Sarnou, H., & Schug, D. (2025). AI in education: (Dis)embodied interactions. *Journal of Studies in Language, Culture, and Society (JSLCS), 8*(1). https://hal.archives-ouvertes.fr/hal-05075932

Karpenko, Y. (2017). Multimedia computer software for the professional training of prospective specialists in foreign languages for preschool and primary education. *Information Technologies and Learning Tools, 57*(1), 50–55. https://doi.org/10.33407/itlt.v57i1.1513

Lacková, M. (2021). The use of corpora in the teaching of ESP (The sphere of pedagogy and psychology). *University of Žilina, Žilina, Slovakia.*

Leláková, E. (2018). *Learning gastronomy vocabulary through corpus*. EDIS: Publishing House of the University of Žilina.

Leleka, T., & Moskalenko, O. (2018). Psycholinguistic peculiarities of the assimilation of Angloamericanisms by the speakers of Ukrainian-English bilingualism. *Psycholinguistics: Philology, 24*(2), 144–162. https://doi.org/10.31470/2309-1797-2018-24-2-144-162

Liu, D., & Lei, L. (2017). *Using corpora for language learning and teaching*. TESOL Press.

Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach*. John Benjamins. https://doi.org/10.1075/scl.20

McEnery, T., Xiao, R., & Tono, Y. (2012). *Corpus-based language studies: An advanced resource book*. Routledge.

O'Keeffe, A., Carter, R., & McCarthy, M. (2007). *From corpus to classroom*. Cambridge University Press.

Ouarniki, O., & Alhasani, M. (2025). Listening Beyond Silence: Subject Matter's Echoes from the EMI Classroom. *Journal of Studies in Language, Culture, and Society (JSLCS) 8*(2), 171-181.https://asjp.cerist.dz/en/article/273152

Rafajlovičová, R. (2018). Structures of subordinate clauses in English oral and written discourse. *X Linguae – European Scientific Journal, 11*(2), 738–757. http://dx.doi.org/10.18355/XL.2018.11.02.58

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. http://dx.doi.org/10.1057/9780230293977

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Appendix(ces)

Appendix I

Table 1. Number of Occurrences of Investigated Compound Nouns in COCA *Main Results*

Compound Noun	Amount of Occurrences	Frequency per Million (f per mil.)
cued recall	140	0.01 per M
free recall	400	0.02 per M
declarative memory	460	0.03 per M
episodic memory	1,890	0.12 per M
explicit memory	420	0.03 per M
direct memory	1,010	0.07 per M
false memory	980	0.06 per M
implicit memory	590	0.04 per M
indirect memory	25	less than 0.01 per M
long-term memory	3,750	0.25 per M
procedural memory	380	0.02 per M
recovered memory	230	0.01 per M
semantic memory	770	0.05 per M
short-term memory	5,320	0.35 per M
state-dependent memory	15	less than 0.01 per M
working memory	14,500	0.93 per M
source amnesia	12	less than 0.01 per M
anterograde amnesia	200	0.01 per M
infantile amnesia	55	less than 0.01 per M
childhood amnesia	48	less than 0.01 per M
retrograde amnesia	210	less than 0.01 per M