



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

DEVELOPING LEXICAL AWARENESS THROUGH CORPUS ANALYSIS TOOLS

Barnoyeva Nilufar Yoqubovna

Associate Professor (PhD), Department of Foreign Languages

Tashkent University of Information Technologies

named after Muhammad al- Khwarizmi

barnoyevanilufar1984@gmail.com

Abstract

Lexical awareness is an essential component of communicative competence, enabling learners to recognize, interpret, and use vocabulary appropriately in different linguistic and professional contexts. In English language teaching, vocabulary instruction has traditionally emphasized word meanings, translation, and memorization. However, effective language use requires a broader understanding of lexical relationships, collocations, semantic associations, grammatical patterns, and contextual variation. Corpus analysis tools provide opportunities to develop this understanding by enabling learners to investigate authentic language data and observe how words function in naturally occurring texts. This article examines the pedagogical potential of corpus analysis tools for developing lexical awareness in English language teaching. Particular attention is paid to concordancers, frequency lists, collocation analysis, keyword analysis, and data-driven learning. The article also discusses the relevance of corpus-based vocabulary instruction for English for Specific Purposes (ESP), particularly in information and communication technology (ICT) education. Practical classroom activities are proposed to demonstrate how corpus tools can support contextual vocabulary learning, collocational competence, learner autonomy, and analytical thinking. The article argues that the systematic integration of corpus analysis into



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

language instruction can complement traditional vocabulary teaching by encouraging learners to explore authentic examples, formulate hypotheses, and develop a more comprehensive understanding of lexical use.

Keywords: Lexical awareness, corpus linguistics, corpus analysis tools, vocabulary acquisition, collocations, data-driven learning, concordancer, English language teaching, ESP, ICT students.

Introduction

Vocabulary is a fundamental component of foreign language competence because it enables learners to comprehend texts, express ideas, participate in communication, and acquire professional knowledge. In English language teaching, lexical competence involves more than memorizing individual words and their definitions. Learners must also understand the relationships between words, their grammatical behaviour, typical collocations, semantic associations, register, and contextual meanings. These dimensions collectively contribute to lexical awareness, which supports the accurate interpretation and appropriate use of vocabulary.

Traditional vocabulary instruction frequently relies on bilingual word lists, dictionary definitions, textbook exercises, and repeated memorization. Although these techniques can support the initial acquisition of vocabulary, they may not provide sufficient opportunities for learners to investigate how lexical items function in authentic communication. For example, students may know the meaning of the verb *make* but remain uncertain about its conventional combinations, such as *make a decision*, *make progress*, and *make an attempt*. Such difficulties demonstrate that vocabulary knowledge includes the ability to recognize and use conventional lexical patterns.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

Corpus linguistics offers a practical means of addressing these challenges. A corpus is a systematically compiled collection of spoken or written language that can be examined using digital tools. Corpus analysis enables learners to investigate word frequencies, examine concordance lines, compare contextual meanings, and identify recurrent collocations. Rather than depending exclusively on prescriptive explanations, students can explore examples of actual language use and draw conclusions from the evidence. Johns (1990) established an influential pedagogical approach through data-driven learning (DDL), which encourages learners to investigate language data and discover linguistic patterns. Sinclair (1991) also emphasized the importance of examining words through their relationships with surrounding linguistic elements, particularly collocation. These perspectives provide a theoretical foundation for incorporating corpus analysis into vocabulary instruction.

The relevance of this approach is supported by Boulton and Cobb (2017), whose meta-analysis identified positive overall effects of corpus-based data-driven learning across the studies examined. Nevertheless, the effectiveness of corpus-based instruction depends on the design of learning tasks, the suitability of the selected corpus, and the guidance provided by teachers. The purpose of this article is to examine how corpus analysis tools can contribute to the development of lexical awareness in English language teaching. It considers the principal functions of corpus tools, their pedagogical applications, their relevance to ESP and ICT students, and the challenges associated with their implementation.

Lexical awareness refers to learners' conscious understanding of vocabulary, including word meanings, relationships between lexical items, patterns of use, and contextual appropriateness. It enables students to recognize that knowing a word involves more than identifying its translation or dictionary definition.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

Several dimensions of lexical awareness are particularly relevant to foreign language learning.

Semantic awareness concerns understanding word meanings and recognizing how meanings change according to context. For instance, *cloud* refers to an atmospheric phenomenon in general English but has a specialized technological meaning in expressions such as *cloud computing* and *cloud storage*.

Collocational awareness involves recognizing words that frequently or conventionally occur together. Expressions such as *strong evidence*, *heavy traffic*, *conduct research*, and *install software* illustrate the importance of learning vocabulary in combinations rather than as isolated items.

Grammatical awareness involves understanding the grammatical patterns associated with particular words. Learners need to recognize, for example, that English commonly uses *depend on*, *interested in*, and *provide someone with something*.

Frequency awareness concerns recognizing which words and expressions occur frequently in particular contexts or genres. This knowledge can help learners prioritize vocabulary for academic reading and professional communication.

Pragmatic and register awareness involves understanding whether a word or expression is appropriate in a specific communicative situation. An expression common in informal conversation may be unsuitable for a formal academic report.

These dimensions are interconnected. Learners who recognize the meaning of a word but are unfamiliar with its collocations or grammatical patterns may experience difficulties in both comprehension and production. Corpus analysis tools can make these relationships more visible by providing authentic examples and allowing systematic comparisons. Different corpus tools facilitate different forms of lexical investigation. Their pedagogical value lies in enabling learners



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

to observe language patterns and connect vocabulary knowledge with actual language use.

A concordancer displays examples of a selected word or phrase in its surrounding context. One common format is Keywords in Context (KWIC), which presents the target item together with the words immediately preceding and following it. For example, a search for *develop* in an academic or technical corpus may reveal combinations such as *develop software*, *develop skills*, *develop a strategy*, and *develop an application*. By comparing these examples, learners can identify the nouns commonly associated with the verb and distinguish different meanings according to context.

Concordance analysis encourages students to investigate vocabulary rather than merely receive definitions. It can also help them identify grammatical patterns, semantic preferences, and differences in usage. Frequency lists display words according to how often they occur in a corpus. They can support the selection of high-priority vocabulary for reading and vocabulary instruction. Teachers may use frequency information to identify recurring words in academic or technical materials, while learners can investigate the relative importance of unfamiliar items. However, frequency should not be treated as the only criterion for vocabulary selection. A technical term may be relatively infrequent in general English but essential for understanding a specialized text.

Collocation analysis identifies words that tend to occur together. It is especially useful for developing awareness of conventional word combinations. For example, ICT students may investigate expressions such as *data security*, *network infrastructure*, *software development*, *protect sensitive information*, and *prevent unauthorized access*. Examining authentic examples helps learners understand which combinations are conventional and how their meanings are shaped by context.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

Keyword analysis identifies words that are unusually frequent in one text or corpus compared with a reference corpus. It can help learners recognize the characteristic vocabulary of a particular subject area. When studying cybersecurity, for example, students may investigate words such as *encryption*, *vulnerability*, *authentication*, and *threat*. They can then examine these words in context to determine how they contribute to the meaning and organization of specialized texts. Tools such as AntConc and Sketch Engine support several of these forms of analysis. Their functions can be used to develop guided classroom activities or independent investigations, depending on learners' proficiency and digital literacy.

Data-driven learning provides a pedagogical framework for using corpus evidence in language education. In this approach, learners examine authentic examples, identify patterns, formulate hypotheses, and evaluate their conclusions. The teacher acts as a facilitator who supports the interpretation of linguistic data and connects discoveries with learning objectives. DDL is particularly relevant to lexical awareness because it encourages students to explore the relationships between words and their contexts. Instead of learning a collocation as an isolated item, learners can examine multiple examples and identify the contexts in which it occurs.

A corpus-based vocabulary lesson can follow five stages.

Vocabulary selection. The teacher selects target words or expressions relevant to the lesson topic, students' proficiency levels, and their academic needs.

Corpus exploration. Students examine concordance lines, frequency information, or collocation results containing the selected vocabulary.

Pattern identification. Learners compare examples and identify recurring combinations, grammatical structures, or differences in meaning.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

Interpretation and verification. Students formulate explanations and verify their interpretations using dictionaries, reference materials, or additional corpus examples.

Application. Learners apply their findings in reading comprehension, sentence construction, summaries, discussions, or short written texts.

This sequence combines observation, analysis, and practical language use. It also promotes learner autonomy because students gradually develop the ability to investigate vocabulary independently. Boulton and Cobb (2017) reported positive overall effects for data-driven learning in their meta-analysis of 64 studies representing 88 samples. Their findings support the educational potential of corpus-based approaches, although they do not establish that every corpus activity produces the same outcomes or that all dimensions of lexical awareness improve equally. Consequently, DDL should be implemented through carefully designed tasks rather than unrestricted access to large quantities of corpus data. Beginners and intermediate learners may benefit from teacher-prepared examples before progressing to independent searches.

Collocational knowledge is an important dimension of lexical competence because many English expressions follow conventional patterns that cannot always be predicted from individual word meanings. Learners may understand each word separately but still produce combinations that sound unnatural or fail to recognize familiar expressions in reading.

Corpus analysis helps learners investigate these patterns by displaying multiple instances of words in authentic contexts. For example, students can compare the combinations *make a decision*, *reach a conclusion*, *conduct an experiment*, and *draw attention*. Such comparisons demonstrate that semantically related words do not necessarily share the same collocational patterns.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

Contextual analysis is equally important. A single word may have several meanings, and the surrounding words often provide clues to its interpretation. The term *security*, for example, may refer to personal safety, financial assets, or protection of computer systems. By examining concordance lines from relevant corpora, students can identify how the word is used in different contexts. Sinclair (1991) highlighted the importance of studying language through recurring patterns and the relationships between words. This perspective is particularly useful in vocabulary instruction because it encourages learners to develop knowledge of lexical combinations rather than treating vocabulary as a collection of independent units.

Corpus-based collocation activities may also contribute to productive competence. When learners become familiar with common combinations, they can use them more accurately in academic writing, oral presentations, and professional communication. Nevertheless, repeated exposure alone may not ensure acquisition; learners also need opportunities to retrieve, practise, and apply the expressions.

Corpus analysis tools are especially relevant to English for Specific Purposes because students need vocabulary connected with their academic disciplines and future professional activities. ICT students regularly encounter English in technical documentation, programming resources, cybersecurity reports, software manuals, and online professional materials. A corpus-based lesson can focus on the topic *Software Development and Cybersecurity*. The teacher selects several target expressions, including *develop software*, *detect vulnerabilities*, *protect sensitive data*, *install updates*, and *prevent unauthorized access*. Students first identify the main ideas of a short technical text. They then use a corpus interface or prepared concordance lines to investigate the target expressions. Their task is to identify common collocations, examine the grammatical



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

structures in which the expressions occur, and explain how the expressions contribute to the meaning of the text.

Corpus analysis tools offer several potential benefits for developing lexical awareness. First, they expose learners to authentic language and multiple examples of word use. Second, they make lexical relationships and collocations more visible. Third, they support inductive learning by encouraging learners to discover patterns. Fourth, they can promote learner autonomy by enabling students to investigate vocabulary questions independently. Corpus-based activities may also strengthen analytical thinking. Students compare examples, recognize similarities and differences, formulate hypotheses, and revise their interpretations when the evidence does not support their initial assumptions.

Ultimately, corpus analysis should complement rather than replace explicit instruction, dictionary use, extensive reading, and communicative practice. Its effectiveness depends on the integration of corpus evidence into coherent learning objectives and meaningful classroom tasks.

Lexical awareness is an essential component of English language competence because it involves understanding word meanings, collocations, grammatical patterns, frequency, and contextual appropriateness. Developing this awareness requires more than memorizing individual words; learners need opportunities to investigate how vocabulary functions in authentic communication.

Corpus analysis tools provide practical means of supporting this process. Concordancers enable learners to examine words in context, frequency lists help identify recurring vocabulary, collocation analysis reveals conventional word combinations, and keyword analysis highlights vocabulary characteristic of particular texts or disciplines. Through data-driven learning, these tools encourage learners to observe linguistic evidence, formulate hypotheses, and develop independent strategies for vocabulary investigation.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

The integration of corpus analysis is particularly valuable in ESP and ICT education, where students need to understand specialized terminology and use technical expressions accurately. Carefully designed corpus-based activities can connect vocabulary exploration with reading comprehension, academic writing, and professional communication.

Nevertheless, successful implementation requires appropriate corpus selection, accessible materials, teacher guidance, and opportunities for learners to apply their findings. Corpus tools are most effective when they complement established teaching methods and are integrated into purposeful communicative activities.

In conclusion, corpus-based vocabulary instruction offers a valuable approach to developing lexical awareness by connecting language learning with authentic evidence of language use. Further empirical research could examine the effects of specific corpus tools on learners' collocational knowledge, contextual vocabulary interpretation, retention, and independent language-learning strategies.

REFERENCES

1. Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // *Language Learning*. — 2017. — Vol. 67, No. 2. — P. 348–393. — DOI: <https://doi.org/10.1111/lang.12224>.
2. Hunston S. *Corpora in Applied Linguistics*. — Cambridge : Cambridge University Press, 2002.
3. Johns T. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning // *Perspectives on Pedagogical Grammar* / ed. by T. Odlin. — Cambridge : Cambridge University Press, 1994. — P. 293–313. — DOI: <https://doi.org/10.1017/CBO9781139524605.014>.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th October, 2026

Website: <https://econferencia.com>

-
4. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. — Cambridge : Cambridge University Press, 2012.
 5. Nation I. S. P. Learning Vocabulary in Another Language. — 2nd ed. — Cambridge : Cambridge University Press, 2013.
 6. O’Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: Language Use and Language Teaching. — Cambridge : Cambridge University Press, 2007.
 7. Schmitt N. Vocabulary in Language Teaching. — Cambridge : Cambridge University Press, 2000.
 8. Sinclair J. Corpus, Concordance, Collocation. — Oxford : Oxford University Press, 1991.