

Data-driven learning in vocabulary instruction for South Korean learners of English

Aprendizado movido por dados no ensino de vocabulário para aprendizes sul-coreanos de inglês

Wesley Acorinti  

acorinti@arizona.edu

Northern Arizona University, Arizona, Estados Unidos.

Abstract

Corpus-based tools and techniques not only facilitate the description of learners' language but can also be used in educational settings to design targeted classroom activities. This approach is known as data-driven learning (DDL). Access to corpus data enables learners to observe and analyze patterns in real language use, addressing their specific needs and fostering learning autonomy. While several studies have examined the effectiveness of DDL in teaching various language skills, few have investigated its impact highlighting the implications for specific cultural groups. Adopting a cumulative knowledge building perspective, this study systematically synthesizes and builds upon previous empirical research to advance our understanding of the effectiveness of DDL for South Korean learners of English. The purpose of this paper is to (1) survey different DDL activities piloted in vocabulary instruction across various English language teaching contexts in Korea; (2) determine the effectiveness of DDL for vocabulary instruction for this demographic; and (3) explore Korean learners' attitudes towards DDL. To do so, empirical studies were systematically identified using the Korea Citation Index with the keywords "data-driven learning," "corpus-based," "vocabulary," "lexis," "Korea," and "Korean." The results indicate that DDL is generally welcomed by and effective for students in this demographic. By building on cumulative findings, this study provides an empirical foundation for curriculum development, classroom practices, and teacher training. Tailoring DDL activities to the Korean context can maximize their effectiveness in addressing learners' unique linguistic challenges.

Keywords: data-driven learning; vocabulary instruction; South Korean learners; English language learning.

Resumo

Ferramentas e técnicas baseadas em corpus não apenas facilitam a descrição da língua de aprendizes, mas também podem ser utilizadas para o desenvolvimento de atividades de sala de aula. Essa abordagem é conhecida como aprendizado movido por dados (AMD). O acesso a dados de corpus permite que os aprendizes observem e analisem padrões do uso real da língua, atendendo às suas necessidades específicas e promovendo a autonomia no aprendizado. Embora diversos estudos tenham examinado a eficácia do AMD no ensino de diferentes habilidades linguísticas, poucos investigaram seu impacto destacando as



10.23925/2318-7115.2026v47i1e72858



FLUXO DA SUBMISSÃO:

Submissão do trabalho: 03/08/2025

Aprovação do trabalho: 24/04/2026

Publicação do trabalho: 11/08/2026

AVALIADO POR:

Maria Eugenia Batista (Unifesp)

Neuda Alves do Lago (UFG)

Etelvo Ramos Filho (Ifes)

EDITADO POR:

Luciana Kool Modesto-Sarra (PUC-SP)

COMO CITAR:

ACORINTI, Wesley. Aprendizado movido por dados no ensino de vocabulário para aprendizes sul-coreanos de inglês. The Specialist, [S. l.], v. 47, n. 1, p. 200–225, 2026. DOI: 10.23925/2318-7115.2026v47i1e72858.



implicações para grupos culturais específicos. Adotando uma perspectiva de construção cumulativa do conhecimento, este estudo sistematicamente sintetiza e se apoia em pesquisas empíricas para avançar a compreensão sobre a eficácia do AMD para aprendizes sul-coreanos de inglês. O objetivo deste artigo é (1) mapear diferentes atividades de AMD implementadas no ensino de vocabulário em diversos contextos de ensino de inglês na Coreia; (2) determinar a eficácia do AMD para o ensino de vocabulário nesse grupo demográfico; e (3) explorar as atitudes dos aprendizes coreanos em relação ao AMD. Para isso, estudos empíricos foram identificados utilizando o Korea Citation Index com as palavras-chave “data-driven learning”, “corpus-based”, “vocabulary”, “lexis”, “Korea” e “Korean”. Os resultados indicam que o AMD é, em geral, bem recebido e eficaz para estudantes desse perfil. Ao se basear em achados cumulativos, este estudo fornece uma base empírica para o desenvolvimento curricular, práticas de sala de aula e formação de professores. Adaptar as atividades de AMD ao contexto coreano pode maximizar sua eficácia ao abordar os desafios linguísticos específicos dos aprendizes.

Palavras-chave: aprendizado movido por dados; ensino de vocabulário; aprendizes sul-coreanos; aprendizagem de língua inglesa.

1. Introduction

Corpus linguistics plays a relevant role in addressing challenges in second language education. In addition to facilitating the investigation of learners' language use (e.g., Yoon; Choi, 2015), corpus linguistics tools and techniques can be used to create classroom activities for students—an approach known as data-driven learning (DDL; Gilquin; Granger, 2022). In DDL activities, the language learner assumes the role of a linguistic researcher, who “has to form preliminary hypotheses on the basis of intuition or scanty evidence” from concordance lines, while also rejecting, refining, and integrating these hypotheses about language use within a suitable model of description (Johns, 1988, p. 14). This approach is grounded in the understanding that language is a probabilistic system, and speakers engage with “a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” (Sinclair, 1991, p. 110).

One of the most widely reported advantages of DDL is its ability to provide learners with vast quantities of input from collections of authentic language use, i.e., corpora (Boulton; Tyne, 2013; Hwang; Cho, 2022; J. Kim, 2024; Kang, 2019). This form of input is particularly valuable in light of the noticing hypothesis in the field of second language acquisition (SLA), which posits that the acquisition of language features is enhanced when learners' attention is consciously drawn to them (Flowerdew, 2015; Schmidt, 1990). Noticing is an essential step before learners can understand and use a linguistic feature (Gass et al., 2020).

Usage-based approaches in SLA further illuminate the pedagogical potential of DDL (Römer, 2024). In developing constructions, understood as cognitive associations between linguistic forms and their functions, learners tend to rely heavily on the frequency and meaningfulness of input. Increased exposure to recurrent forms supports the gradual development of networks connecting forms, functions, and the contexts in which they occur (Ellis; Wulff, 2020).

Another frequently reported advantage of DDL is learner autonomy, along with a shift in the teacher's role away from being the main focus in the classroom (Boulton; Tyne, 2013; Kim; Chun, 2008; Yun; Han, 2015). Autonomy has been argued to be closely related to motivation, which in turn has been shown to affect success in language learning (Dörnyei; Csizér, 1998; Dörnyei; Ryan, 2015). From the perspective of self-determination theory (SDT), autonomy is considered a fundamental psychological need that supports the development of intrinsic motivation (Ryan; Deci, 2023).

However, autonomy has also been characterized as a culturally specific concept rooted primarily in Western educational traditions. Hall (2018, p. 175) notes that “autonomy is a culturally specific Western concept,” while Asian learners are often described as placing greater value on “‘collectivism,’ conformity and ‘respect from authority,’ from which the idea of individuality is excluded” (Benson *et al.*, 2003, p. 24). These observations suggest that findings from previous research on learner autonomy, largely conducted in Western contexts, may not readily generalize to East Asian educational settings.

These cultural considerations become particularly salient in the South Korean context, where sociocultural learning norms and language-specific challenges shape learners’ experiences with English teaching. English is often considered one of the most difficult languages for native speakers of Korean to master (O'Donnell, 2006), and previous research suggests that failure to overcome this challenge can lead to negative consequences, including limited access to higher education and employment opportunities in South Korea (Choi, 2008). Beyond institutional outcomes, limited English proficiency may also restrict participation in everyday activities, such as appreciating humor or navigating consumer contexts that require some degree of English comprehension (J. S. Lee, 2020).

In South Korea, English education at various levels predominantly follows a teacher-centered instructional model characterized by rote memorization of grammar and vocabulary, often for translation purposes (Oak; Martin, 2003). These practices are largely shaped by exam-driven instructional priorities, as high-stakes English tests continue to exert strong washback effects on classroom teaching and learning (Moodie; Nam, 2016). Such conditions stand in direct contrast to the learner-centered, exploratory nature of DDL, leaving limited room for autonomous learning or engagement with authentic input (D. J. Lee, 2013; O'Donnell, 2006).

Over the past decades, research on DDL has generated a substantial body of literature documenting both its affordances and challenges across a range of educational contexts (e.g., Boulton; Tyne, 2013). As the field matures, there is increasing recognition of the need to synthesize these findings to build cumulative knowledge capable of informing more effective and context-sensitive pedagogical practices. Given the challenges characterizing the South Korean educational context, it is therefore essential to examine how DDL may address the specific needs of Korean learners.

While several studies have investigated the effectiveness of DDL for teaching different language skills (e.g., Holmberg, 2024), relatively few have focused on its impact within specific cultural and institutional contexts. To provide an empirically grounded perspective that may inform curriculum developers, teachers, and teacher trainers in South Korea, this paper addresses the following three research questions, with a specific focus on vocabulary instruction for Korean learners of English:

- (1) What types of DDL activities were piloted across various English language teaching contexts in South Korea?;
- (2) How effective is DDL for learners in this demographic?; and
- (3) What are the learners' attitudes towards DDL?

To address these questions, the present study adopts a cumulative knowledge building perspective that prioritizes the systematic synthesis and integration of prior empirical findings. Accordingly, a systematic review of existing studies was conducted based on research identified through a comprehensive search of the Korea Citation Index. By situating the analysis within the broader body of existing literature, this review seeks not only to summarize previous work but

also to extend understanding of the effectiveness of DDL for vocabulary instruction among Korean learners of English.

The present study emphasizes vocabulary instruction due to its critical role in communication. Limitations in vocabulary knowledge can directly hinder learners' ability to express themselves, and learners often recognize that a strong vocabulary is essential for improving their language skills (Nation, 2001). In this study, vocabulary knowledge is broadly defined to include knowledge of single lexical items, collocations, meanings, and meaning associations (Nation, 1990; Scrivener, 2011).

Vocabulary instruction tends to focus primarily on receptive skills (i.e., recognizing words for reading and listening) in South Korea, while the development of productive skills is often neglected (Kim; Chun, 2008). Therefore, further research in vocabulary instruction in this context can help identify effective strategies for enhancing Korean learners' communication abilities in real-world situations. Vocabulary seems to be particularly well suited to DDL because corpus exploration allows learners to observe recurrent patterns, collocational behavior, and contextualized meaning associations (Brezina; Gablasova, 2018; Gablasova *et al.*, 2017), which are dimensions of lexical knowledge that are often underdeveloped in traditional instruction.

In contrast to previous reviews that have addressed DDL from broad or global perspectives (e.g., Boulton; Tyne, 2013; Holmberg, 2024), the goal of the present study is to systematically synthesize empirical research specifically on DDL-based vocabulary instruction for Korean learners of English. By focusing exclusively on this population, the study aims to provide a more nuanced understanding of how demographic and contextual factors interact with the effectiveness of the DDL approach. Considering the dominant traditions of teacher-centered and exam-driven instruction in South Korea, it is hypothesized that DDL may be less effective in this context than has been reported in other educational settings.

While the present research centers on South Korea, the findings and implications might resonate with other English as a foreign language environments characterized by teacher-centered, exam-driven institutional traditions, such as China, Japan, and Vietnam. Thus, the present review is likely to inform curriculum development and to have a broader practical application for language educators working not only in South Korea but also in similar educational systems.

2. Method

2.1. Cumulative knowledge building and systematic review of the literature

Building knowledge from a cumulative perspective goes beyond simply listing what has or has not been done in previous studies. Instead, it emphasizes the importance of synthesizing empirical evidence to identify generalizable patterns, persistent challenges, and unresolved questions within a given research topic (Larsson; Biber, 2025). In the context of corpus linguistics, there has traditionally been a strong tendency for researchers to identify a gap (i.e., something that has not been precisely studied before) in the literature as a primary motivation for new research. However, given the decades-long development of corpus approaches to language study, a robust body of literature has in fact already been established across a wide range of topics, and many times, researchers tend to overlook it. In this scenario, Larsson and Biber (2025) suggest shifting the focus from exclusively what has been *done* to also consider what has been *learned* from previous research. Rather than focusing solely on gaps or novelty, cumulative knowledge building motivates new research by integrating and critically evaluating existing findings, providing a solid foundation for further inquiry and hypothesis formation (Larsson *et al.*, 2024).

Systematic reviews are one of several ways to contribute to this research perspective (Larsson; Biber, 2025). This method aims to produce knowledge by synthesizing the results of empirical studies to address new research questions. Accordingly, the present review was designed to cumulatively build on previous empirical studies concerning DDL in vocabulary instruction for South Korean learners of English. Following the recommendations of recent methodological literature (Chong; Plonsky, 2024; Li; Wang, 2018; Macaro, 2020), this paper intends to ensure transparent, replicable procedures for study selection and synthesis. Attention was given to identifying not only the topics covered in prior studies, but also the specific findings, areas of consensus, and persistent ambiguities in order to draw conclusions that can contribute to the advancement of both theoretical understanding and pedagogical practice regarding DDL.

2.2. Sampling

To address the three proposed questions based on the findings of previous empirical research, a systematic search was conducted in the Korea Citation Index, a database managed by the National Research Foundation of Korea. The Korea Citation Index was selected because it is a comprehensive repository of peer-reviewed academic publications produced in Korea, allowing for the retrieval of contextually relevant research that may not be indexed in international databases. It is worth noting that this methodological choice may constitute a limitation of this paper, as studies published or indexed outside this database were not included in the synthesis pool, thus narrowing the scope of the findings.

The search for primary studies was conducted on September 23, 2024. The search terms included “data-driven learning,” “corpus-based,” “vocabulary,” “lexis,” “Korea,” and “Korean.” A complex query was employed to retrieve research articles in a single search, using Boolean operators (AND and OR) and a wildcard (*). The full search string executed (including operators and wildcard use) was as follows: **"corpus-based"OR"data-driven learning" AND "vocabulary"OR"lexis" korea***. The search string was entered into the main search bar on the homepage of the database. No exclusion criteria were applied. This initial search produced 64 results.

During the screening stage, the abstracts of the 64 studies were examined. For a study to be included in the synthesis pool, it had to meet both of the following inclusion criteria: (1) the study reported results from an empirical experiment examining the effectiveness of data-driven learning for teaching vocabulary to Korean learners of English; and (2) the study was written in English, given the author’s limited proficiency in Korean and the role of English as the lingua franca of academia (see Hyland, 2009), which facilitates transparency in reporting the findings presented in this paper. Studies that did not meet one of these two criteria were excluded from the scope of the present review. After applying these criteria, 10 studies remained in the synthesis pool. The flow diagram in Figure 1 illustrates the sampling procedure.

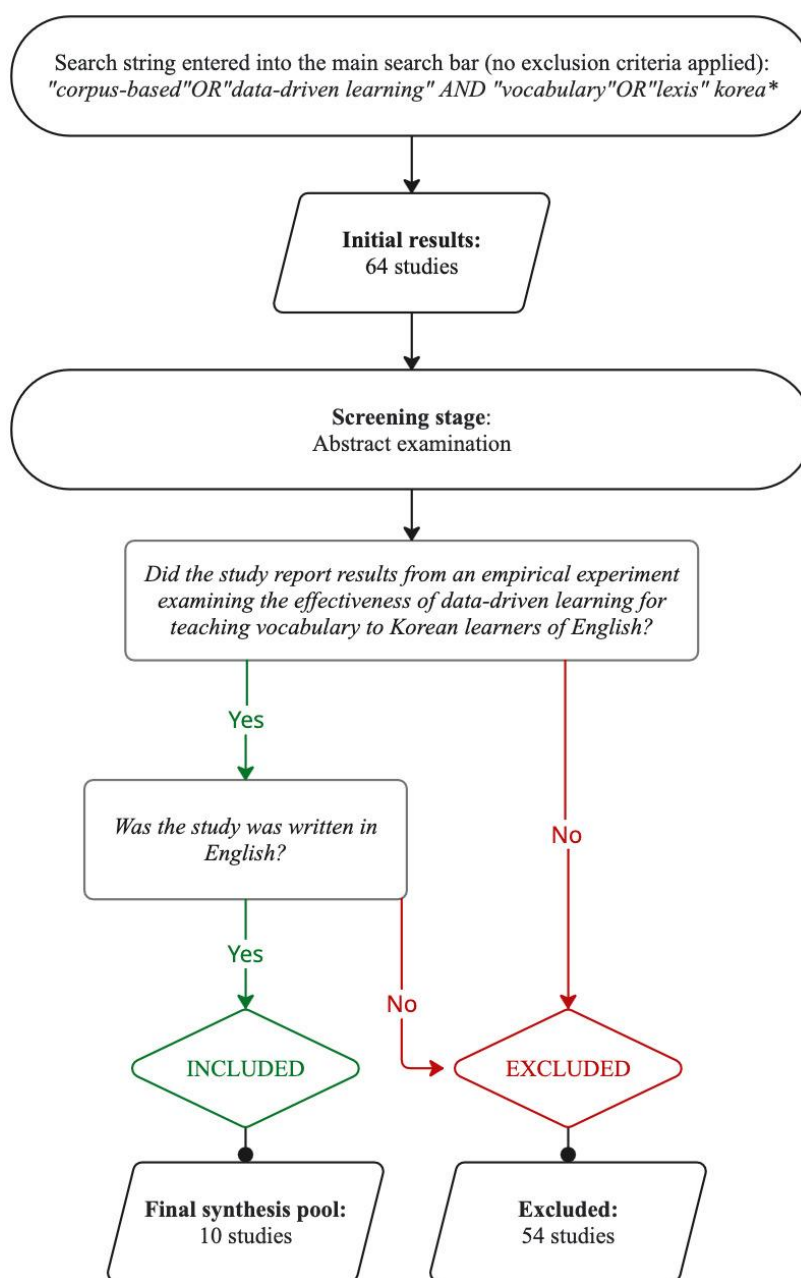
Figure 1. Flow diagram of the sampling procedure.

Table 1 presents the 10 selected studies, the educational contexts in which they were conducted, the number of participants, and their proficiency levels.

Table 1. Summary of selected studies.

Study	Educational context	Number of participants	Proficiency level of the participants
Chun (2016)	University	7	Intermediate
E. Kim (2015)	University	121	Pre-intermediate – Advanced
Hwang and Cho (2022)	Middle school	50	Not reported
J. Kim (2024)	University	55	Basic
Kang (2019)	University	24	Basic
Kim and Chun (2008)	University	48	Intermediate – Advanced
N.-Y. Kim (2018)	University	135	Intermediate
Schenck and Cho (2012)	In-service training program for English teachers	21	Advanced
Yang (2015)	University	115	Basic
Yun and Ham (2015)	University	41	Advanced

Despite the relatively small number of primary studies, which was anticipated given the specific focus of this research, as well as variations in participant numbers, proficiency levels, and the uneven distribution of studies across educational contexts, the present research provides meaningful insights into the implementation of DDL, its effectiveness, and students' attitudes towards it in South Korea. Some studies included only a small number of participants, and in a few cases, proficiency levels were not detailed, which should be considered when interpreting the findings. The criterion of including studies published only in English necessarily excludes studies published solely in Korean, which may limit the generalizability of the findings to English-language scholarship. Despite these limitations, the results contribute important perspectives that can inform further research and practical classroom application.

2.3. Data analysis

The data used to answer the research questions was manually extracted from each of the 10 studies. All studies were thoroughly read, and relevant information was recorded in a spreadsheet to help organize and compare findings. The spreadsheet listed number of participants, participants' proficiency level (e.g., intermediate), level of education (e.g., university), targeted vocabulary aspect (e.g., phrasal verbs), corpus tool(s) used (e.g., COCA), direct or indirect DDL (see 3.1), contrast teaching approach if any (e.g., dictionary-based vocabulary learning), main findings, statistical significance (yes/no), and learners' attitudes toward DDL. The findings were then integrated into a narrative synthesis, emphasizing patterns and relationships across studies.

3. Results and Discussion

3.1. Data-driven learning activities used in vocabulary instruction

The implementation of DDL in the classroom can take two distinct forms (Gabrielatos, 2005). In the *hard* version of DDL, often referred to as *direct* or *inductive* DDL, students themselves interact with the corpus: they learn how to use specialized tools, analyze patterns, and draw their own conclusions based on linguistic data. In other words, the hard version of DDL requires students to develop corpus literacy skills (see Mukherjee, 2006) that allow them to conduct queries using corpus tools independently. In contrast, the *soft* version of DDL, also known as the *indirect* or *deductive* DDL, involves the teacher conducting corpus queries and curating the resulting data to create learning materials for students. Out of the 10 studies surveyed, five piloted the hard version of DDL, while the remaining five adopted the soft version of the approach.

3.1.1. Indirect data-driven learning approach

In the studies that employed the soft version of DDL to evaluate its effectiveness, researchers demonstrated a variety of ways to design tasks using corpus resources to cater to different aspects of vocabulary knowledge in the classroom. J. Kim (2024) focused on teaching single lexical items (e.g., *first name*, *grandparents*), providing translation and gap-filling exercises

based on a set of five concordance lines for each target word. Chun (2016) targeted both single words (e.g., *count*, *fair*) and synonym pairs (e.g., *beat/defeat*, *choose/select*), while Hwang and Cho (2022, p. 1120) focused on polysemy, defined as “a lexical phenomenon where a single word has several meanings.” In both Chun (2016) and Hwang and Cho (2022), students received handouts to observe sets of 15 to 20 concordance lines for each target word and were asked to freely describe the patterns they identified. Yang (2015) focused on teaching the collocates of delexical verbs (e.g., *do*, *make*), requiring students to read 30 concordance lines and later identify collocates to create sentences. To teach phrasal verbs (e.g., *get into*), E. Kim (2015) selected sets of around 17 concordance lines and designed exercises where students had to choose the most appropriate verb, complete sentences, infer meanings, and replace words.

In three (Chun, 2016; Hwang; Cho, 2022; Yang, 2015) out of the five studies mentioned above, the data used to develop the pedagogical materials came from the Corpus of Contemporary American English (COCA; see Davies, 2009). Yang (2015) incorporated data from the British National Corpus (BNC; see Leech, 1992), LexTutor (see Cobb, 2015), and a corpus compiled by the researcher and analyzed using WordSmith (see Scott, 2001), which included “12 copies of textbooks widely used in schools” (Yang, 2015, p. 75). J. Kim (2024) exclusively used the corpora available on LexTutor, while E. Kim (2015) combined data from LexTutor and the BNC.

3.1.2. Direct data-driven learning approach

While the five studies reviewed above employed the soft version of DDL, five others adopted the hard version to evaluate its effectiveness in vocabulary instruction. Yun and Ham (2015), N.-Y. Kim (2018) and Schenck and Cho (2012) used direct DDL activities to focus on single words. In Yun and Ham (2015), students were tasked with querying corpora to complete several steps: selecting 10 useful example sentences, analyzing the word's meaning and parts of speech, verifying their guesses using a dictionary, identifying synonyms and antonyms, noting specific word features (such as affix, collocation, and usage), and creating their own example sentence. In contrast, students in N.-Y. Kim's (2018) study had a less structured task: they were simply asked to query the corpus, examine the input, and infer the meaning of the target word. In Schenck and

Cho (2012), students had to peer-edit texts written by their colleagues by systematically checking the usage patterns of each noun in the texts using a corpus.

Kim and Chun (2008, p. 223) do not provide details or sample activities used with learners to foster lexis awareness, but they note that both “practice in utilizing the functions of concordancers” and “web-based corpus exercises” were conducted. In Kang's (2019) study, students searched for collocates in a corpus and had to attend to part-of-speech tags in order to complete target collocation lists.

Interestingly, four (Kang, 2019; N.-Y. Kim, 2018; Schenck; Cho, 2012; Yun; Ham, 2015) out of the five studies that used the hard version of DDL relied on the COCA. In addition to COCA, Yun and Ham (2015) also used the BNC. Kim and Chun (2008) employed the BNC as well as the VLC Web Concordancer, the UGRU UAEU Concordance, and the Virtual DDL Library¹.

The studies surveyed demonstrate the diverse applications of DDL in vocabulary instruction, encompassing both the hard and soft versions. Notably, the corpus tools used in these studies (i.e., COCA², BNC³, LexTutor⁴, and WordSmith⁵) are all freely accessible, making DDL approaches available to a wide population. These studies also highlight the versatility of DDL activities in addressing various aspects of vocabulary knowledge, including the meaning, form, and structure components of lexical items. The examples included in this subsection may be of particular interest to researchers who are interested in how to go about assessing the effectiveness of DDL in different contexts, as well as to language educators who intend to integrate corpora into their classrooms. In addition, they serve as a foundation for contextualizing the other parts of the subsequent analyses of the results from the studies in the synthesis pool.

3.2. Effectiveness of data-driven learning for Korean learners of English

J. Kim (2024) and Yang (2015) compared the soft version of DDL with traditional approaches to vocabulary learning in South Korea. The instructional target of the first was individual lexical items (e.g., *information*, *social security number*), while the latter focused on the

¹ The links for the latter three tools are currently broken. As a result, further references to these resources could not be traced.

² <https://www.english-corpora.org/coca/>

³ <https://www.english-corpora.org/bnc/>

⁴ <https://www.lextutor.ca/>

⁵ <https://lexically.net/wordsmith/version5/index.html>

collocates of delexical verbs (e.g., *take*, *have*). The traditional methods included “instruction with teacher talk and explanation” of vocabulary items (Yang, 2015, p. 73) and “writing meanings in Korean and filling English words in the blanks” (J. Kim, 2024, p. 881). Both studies found that participants in the DDL groups showed a statistically significant improvement in vocabulary proficiency as measured by test scores. However, the two studies also found that students in the traditional approach groups improved their posttest scores as well.

In J. Kim’s (2024) study, the mean score⁶ on the vocabulary posttest of the DDL group increased, whereas the traditional instruction group showed a smaller gain. Both groups demonstrated statistically significant differences in the *t*-test⁷ results comparing their pretest and posttest mean scores. Relatedly, other studies have explored whether different ways of implementing DDL lead to differential learning outcomes. In addition to comparing corpus-based learning with traditional methods, Yang (2015) divided the students working with the DDL approach into two groups: one group collaborated with peers on tasks, while the other group worked independently. Interestingly, Yang’s (2015) findings showed that DDL had a more positive impact on the retention of targeted collocations compared to traditional methods, especially when students collaborated on corpus-based tasks, as opposed to working individually.

E. Kim (2015), Hwang and Cho (2022), Kim and Chun (2008), and Schenck and Cho (2012) primarily compared the outcomes of DDL to dictionary-based vocabulary learning. Generally, both approaches showed improvements in vocabulary knowledge, with DDL yielding more consistent gains. Focusing on polysemy, Hwang and Cho (2022) reported that the DDL group showed a larger increase in scores on an implicit vocabulary knowledge test from the pretest to the posttest compared to the group relying on dictionaries and online translators. Similarly, Kim and Chun (2008), which focused on the development of lexis awareness, found that the DDL group showed a significant improvement in posttest scores compared to the non-DDL group. This finding is consistent with E. Kim (2015), which targeted the learning of phrasal verbs (e.g., *get into*).

In contrast, Schenck and Cho (2012, p. 173) found no statistically significant differences between the DDL group and the group using “traditional methods of proofreading and the

⁶ The *mean score* indicates the central tendency of values in a dataset. It is calculated by summing all individual scores and dividing the total by the number of observations.

⁷ A *t*-test is an inferential statistical technique used to determine if the difference between two means is statistically significant. *t*-tests and other hypothesis tests applied to mean comparisons were commonly used in the studies included in the synthesis pool. A clear introduction to *t*-statistics can be found in Howell (2013, Chapter 7).

dictionary” to peer-edit texts. This suggests that while DDL is effective in many contexts, its impact may vary depending on the type of activity used. The researchers observed that participants in the DDL group from their study were unable to “substantively apply observations from the corpus to the editing process, despite systematic guidance and extensive prior experience in the English field,” and struggled to identify semantic aspects and common collocates of the target vocabulary (Schenck; Cho, 2012, p. 174).

In addition to comparing dictionary-based vocabulary instruction with DDL tasks, N.-Y. Kim (2018) introduced a third instructional condition: video-based vocabulary learning, designed to help learners infer word meanings from input. Similarly, focusing on collocation learning, Kang (2019) included a comparison group that relied on Internet searches and news articles from sources such as *The Korea Herald* and CNN. The first study found statistically significant improvements within groups (N.-Y. Kim, 2018). Specifically, the corpus-based group and the video-based group both showed significant increases from pretest to posttest, although there were no significant differences between the groups. The second study reported improvements in both the DDL group and the group using alternative sources but found no statistically significant difference between the two groups on posttest scores (Kang, 2019).

Yun and Ham (2015, p. 38) compared the learning effectiveness of DDL tasks to word map tasks, in which participants needed to “write the target word and a sentence from a text; write out a dictionary definition; write a synonym and antonym; write down other forms of the word; make the student's own sentence; and create a mnemonic.” The findings of this study showed that both approaches led to significant vocabulary improvements, with the corpus group outperforming the word map group in overall vocabulary gains measured by test scores. While the word map group performed slightly better on the receptive knowledge posttest, the corpus group showed a more expressive improvement in productive knowledge, with a significant difference between the groups. These results suggest that DDL tasks are more effective for promoting productive vocabulary use, while both task types are similarly effective for receptive knowledge. This tendency is consistent with the findings of Chun (2016, p. 105), which, although not having a comparison group, observed that after participating in DDL activities, “students' recognition [of target vocabulary] increased and shifted from form or meaning to application knowledge.” Yun and Ham (2015, p. 43) suggest that this occurs because corpus-based tasks require “deeper cognitive processing,” particularly a higher “involvement index” and level of

evaluation of language use. This happens when learners devote cognitive effort to observing linguistic features, patterns, and regularities in authentic samples.

In sum, the studies reviewed here provide positive evidence for the effectiveness of DDL in improving vocabulary proficiency among Korean learners of English, despite being an approach that diverges from traditional educational practices in the Korean context (e.g., teacher-centered model, rote memorization). Across various settings, DDL consistently resulted in significant improvements in vocabulary knowledge. Interestingly, traditional and alternative methods (i.e., dictionary-based learning and word map tasks) also demonstrated to be effective.

Unfortunately, most of the studies included in the synthesis pool reported only whether or not the differences they found were statistically significant, without quantifying their magnitude using effect size measures (e.g., Cohen's *d*). Without such measures, it is difficult to make nuanced comparisons across studies or to discuss findings in terms of greater or lesser effectiveness of DDL with appropriate statistical support. It is important to be clear that a *statistically significant difference* does not necessarily indicate a practically meaningful difference. The term simply means that “the difference, even if trivial, is not likely due to chance” (Howell, 2013, p. 102). As a result, statistical significance alone does not allow for conclusions about the size or practical relevance of observed differences in the studies surveyed.

Since educational stakeholders may need further evidence to justify investing valuable classroom time and resources in a new approach, it can be hypothesized that the distinctiveness of DDL lies not only in its demonstrated effectiveness in vocabulary learning as reflected in test scores, but also in students' attitudes toward the approach. Given that eight (Chun, 2016; J. Kim, 2024; Kang, 2019; Kim; Chun, 2008; N.-Y. Kim, 2018; Schenck; Cho, 2012; Yang, 2015; Yun; Ham, 2015) out of the 10 studies surveyed provide data on students' beliefs, the next section will address the third research question of this study: What are Korean learners' attitudes toward DDL?

3.3. Korean learners' attitudes towards DDL

Learners often have preferences about how language instruction should be delivered, and these “beliefs about the kind of instruction that is best can influence their satisfaction and success” (Lightbown; Spada, 2021, p. 168). Hence, understanding students' perspectives is

valuable for preventing classroom difficulties that may arise from mismatches between teachers' and students' beliefs about effective learning approaches (Hall, 2018). Furthermore, learner beliefs and motivation are intrinsically linked, and “examining learner beliefs may help teachers understand why learners behave as they do and lead to a clearer understanding of classroom life” (Hall, 2018, p. 138). The eight studies reviewed in this paper that examined students' perceptions toward DDL accomplished that by administering questionnaire surveys and conducting interviews.

Chun (2016), J. Kim (2024), Kang (2019), Kim and Chun (2008), N.-Y. Kim (2018), Schenck and Cho (2012), Yang (2015), and Yun and Ham (2015) report overall positive attitudes towards the DDL approach in the questionnaire surveys administered to their participants. Yun and Ham (2015) and Kim and Chun (2008) found statistically significant evidence of positive perceptions of DDL. In the first study, researchers found greater satisfaction with corpus-based tasks compared to word map tasks, although both groups considered the tasks “less useful in learning the spellings, pronunciations, and affixes of words than in learning other aspects of vocabulary” (Yun; Ham, 2015, p. 49). In the second study, it was found that after DDL instruction, students were “more confident” and “more satisfied” with their answers and methods for deriving answers on collocation tests (Kim; Chun, 2008, p. 228).

Similarly, Yang (2015), J. Kim (2024) and Schenck and Cho (2012) demonstrated positive student attitudes toward DDL. Yang (2015) found that students who collaborated in corpus-based learning exhibited more positive attitudes, particularly in groups with higher proficiency levels. Lower-level students also showed positive attitudes toward the approach, but mainly when they worked on DDL tasks individually rather than in groups. This observation aligns with the literature on Korean learners' characteristics, which notes that Korean students are especially “afraid of making mistakes and being ridiculed in front of their classmates” (Cho, 2004, p. 35).

In addition to boosting students' confidence in their productive skills, J. Kim (2024, p. 886) found that students considered DDL-based lessons interesting, fun, helpful, and “easy enough to follow.” However, Schenck and Cho's (2012) survey results showed that while students acknowledged the effectiveness of using corpus tools, they struggled to use them effectively to achieve concrete outcomes.

As with other approaches to language teaching, DDL also has its drawbacks. N.-Y. Kim's (2018) findings show that students' motivation and interest were highest in the group that used

video-based vocabulary learning, followed by the corpus-based group, and lowest in the dictionary-based group. Although not detailed by the researcher, the reason for the DDL approach coming after video-based vocabulary learning might be that the latter approach allows students to hear words pronounced in natural contexts and to use visual cues to aid understanding—features not available in COCA, used by participants in the corpus-based group.

During an interview, one of the seven participants in Chun's (2016) study expressed dissatisfaction with the DDL approach, perceiving it as too time-consuming and ineffective. This perception of DDL as time-intensive aligns with previous literature, with both students and teachers reporting similar concerns in different educational settings (Boulton; Tyne, 2013; D. J. Lee, 2013). The challenge often stems from the training required for effective corpus use, which demands time from both teachers and learners. Notably, Chun's (2016) study lasted only three weeks, which may have made it overwhelming for one student to become familiar with indirect DDL activities, influencing their perception of the approach. Possible ways to address the issue include increasing the focus on corpus use in teacher training programs and encouraging publishers and curriculum developers to integrate corpora into educational materials and syllabi (McCarthy, 2008), which would reduce the demands placed on teachers.

Basic proficiency level participants in Kang's (2019, p. 73) study were overall "satisfied with the concordancing feature like a quick and easy search engine online." However, they also found the DDL-based assignments difficult and described them as "too overwhelming and complicated to find target collocations" (Kang, 2019, p. 67). This challenge was also previously noticed in the literature (Boulton; Tyne, 2013; D. J. Lee, 2013). Interestingly, while the reliance of DDL on authentic language is a key strength, it may also be its Achilles' heel, as authentic snippets of the English language can be more difficult to understand and thus require more scaffolded support from teachers. Yet, as Sinclair (1997, p. 31) emphasizes, "language cannot be invented; it can only be captured," and failing to present authentic language in the classroom risks impeding effective communication outside of it.

The generally positive attitudes, increased confidence, and satisfaction reported by Korean learners engaging with DDL activities can be interpreted through the lens of educational psychology. The finding that DDL activities were perceived as interesting, helpful, and satisfying aligns with the emphasis of SDT on the importance of supporting learners' needs for competence,

autonomy, and relatedness (i.e., basic psychological needs), all of which promote motivation, which tends to lead to successful learning (Ryan; Deci, 2023). Variations in attitudes by proficiency and group format seem to indicate the influence of both individual and contextual factors in shaping motivation, as recognized by SDT. Thus, it can be argued that the positive attitudinal effects of DDL observed in these studies suggest that its broader adoption could contribute not only to vocabulary development but also to more sustained and self-determined engagement with language learning in Korea.

In conclusion, while overall attitudes toward the DDL approach among Korean learners are positive, various factors influence students' perceptions and experiences. Drawing on the empirical results summarized above, it is clear that, while DDL has demonstrated effectiveness in vocabulary learning, its distinctiveness may lie not only in improved test scores but also in the positive attitudes it fosters among students. Challenges remain, particularly regarding the time-intensive nature of DDL and the need to integrate authentic input in the classroom. Together, these findings seem to support the case for investing in the broader adoption and integration of DDL into language teaching practices in South Korea.

4. Conclusion

This paper examined the use of the DDL approach in vocabulary instruction for Korean learners of English through a systematic review of 10 empirical studies conducted at the middle school, university, and in-service teacher training levels in South Korea. The research addressed three key questions to review the state of the art of DDL research in Korean contexts regarding (1) the types of DDL activities piloted in this context; (2) the effectiveness of DDL for learners of this demographic; and (3) Korean learners' attitudes toward the approach.

First, the findings indicate the versatility of both the hard and soft versions of DDL activities in vocabulary instruction for Korean learners. In the hard version, students have direct access to corpus tools and can be tasked with activities such as grasping the meanings of target words from concordance lines (N.-Y. Kim, 2018) or peer-editing texts by conducting their own corpus queries (Schenck; Cho, 2012). In the soft version, students are not provided with direct access to corpora. Instead, the instructor or materials designer curates' data from corpus tools for learners to engage with. Activities in the soft version may include translation (J. Kim, 2024) or gap-filling

exercises (E. Kim, 2015; J. Kim, 2024) using authentic language and frequency information drawn from corpora, which diverges from traditional classroom practices that rely on inauthentic language (D. J. Lee, 2013). The findings for the first research question have direct pedagogical implications, as language researchers and instructors can better understand the classroom potential of the DDL approach in terms of the types of activities to design and the tools to use.

Second, regarding the effectiveness of DDL, the body of research reviewed provides compelling evidence that DDL outperforms traditional vocabulary learning approaches, such as lecture-based teaching (J. Kim, 2024; Yang, 2015), and alternative strategies, such as dictionary-based learning (E. Kim, 2015; Hwang; Cho, 2022; Kim; Chun, 2008) in the Korean context. This result is remarkable because the DDL approach contradicts current educational practices in Korea (e.g., teacher-centered methods, reliance on unauthentic examples). Interestingly, while the traditional and alternative approaches also showed varying degrees of effectiveness, DDL provided a constantly marked advantage, especially in enhancing students' productive knowledge of vocabulary (Chun, 2016; Yun; Ham, 2015), which is a relevant aspect of language learning that was observed to be neglected in the overall Korean context (Kim; Chun, 2008). The findings related to pedagogical implications of the second research question provide educational stakeholders with empirical evidence to support the adoption of DDL in their teaching practices.

Third, the positive attitudes of Korean learners toward DDL (J. Kim, 2024; Schenck; Cho, 2012; Yang, 2015) further reinforce its potential as a valuable pedagogical approach and justify the time and resource investments required by English instructors in this context. Most learners in the studies that were reviewed reported increased confidence in their language skills and satisfaction with learning through DDL activities (Kim; Chun, 2008; Yun; Ham, 2015). However, some learners expressed concerns about the time-consuming nature of DDL tasks (Chun, 2016) or found the approach overwhelming (Kang, 2019). The implications of the third research question for language pedagogy strengthen the body of empirical evidence for different educational stakeholders in the Korean context to consider adopting DDL activities in the classroom while being mindful of their potential drawbacks.

In conclusion, this paper finds the use of the DDL approach beneficial for vocabulary instruction in the context of English language education in South Korea. While traditional and alternative pedagogical approaches can also be effective to varying degrees, DDL stands out due

to the positive perceptions of Korean learners. In this country where both students and instructors are rarely exposed to real language use to support language production (D. J. Lee, 2013), DDL plays an important role in providing learners with authentic input. Notably, DDL has also been shown to be effective in bridging the gap between an overemphasis on receptive vocabulary knowledge (Kim; Chun, 2008) and the development of productive knowledge in English education in Korea (Chun, 2016; Yun; Ham, 2015). Moreover, as O'Donnell (2006) notes, the use of technology in language learning can help address Korea's socio-cultural educational challenges, such as large class sizes and limited class time for English instruction. In this context, DDL can encourage student autonomy by enabling learners to make discoveries in language through corpus tools. To ensure sustainable integration of DDL, coordinated efforts between curriculum designers, teacher educators, and technology providers will be essential.

While DDL offers several advantages, it also presents challenges for educators, such as the lack of ready-made materials tailored to the needs of Korean learners and instructors, as well as the time and resources required to acquire technical skills and computing equipment. Future research on DDL should further take into account the unique characteristics of different societies and learner groups, as well as its potential to influence instructors' classroom practices. This would include, for example, developing and making available pedagogical materials tailored to the specific needs of Korean learners that can be implemented in the classroom.

The description of learner corpora containing English produced by South Korean students would provide a starting point for needs analyses aimed at identifying learners' specific linguistic issues and informing the development of instructional materials. The materials targeted to English education in South Korea could include data from context-specific corpora reflecting relevant vocabulary and language use for indirect DDL activities, alongside scaffolded corpus literacy (Mukherjee, 2006) development and user-friendly web interfaces for direct DDL tasks. Furthermore, the materials could incorporate options for both collaborative and individual tasks, providing flexibility to suit learners' proficiency levels and preferences. As Yang (2015) highlights, collaborative corpus-based tasks, particularly for higher-proficiency students, can enhance retention and foster positive attitudes towards learning. In contrast, individual corpus-based tasks may be more suitable for lower-proficiency students, who often prefer working independently to avoid the risk of making mistakes in group settings (Cho, 2004).

From a cumulative knowledge building perspective, future research should move beyond establishing the general effectiveness of DDL activities and instead focus on systematically examining their differential effects across learner populations, cultural contexts, and receptive versus productive skills. Further empirical research is also needed to compare the relative effectiveness of hard and soft versions of DDL across different proficiency levels, as well as to explore how these approaches interact with learner motivation. By investigating more targeted comparisons, future work can yield increasingly specific and generalizable findings, contributing to a more robust and comprehensive understanding of DDL in diverse educational contexts.

Informações complementares:

a) Declaração de contribuição das autoras e dos autores:

Text produced by a single author.

b) Disponibilidade de dados de pesquisa e outros materiais:

The primary studies supporting the conclusions of this systematic review of the literature can be accessed through the information provided in the References section.

c) Declaração de conflito de interesse:

The author has declared no conflict of interest.

d) Avaliação por pares:

✓ **Avaliador 1:** Maria Eugenia Batista (correções obrigatórias)

The paper is well-written, presents and addresses very clear and objective research questions. There are minor writing issues highlighted in the file attached. As statistics is used to address the second research question, I do recommend using footnotes in order to explain the statistic terms. Differently from our way of organizing a paper having a specific section for the Theoretical Framework, the author intertwines the results with the pertaining theory which is perfectly fine. See comments on the attached file.

✓ **Avaliador 2:** Neuda Alves do Lago (rejeitar)

O manuscrito discute um tema atual ao reunir estudos sobre Data-Driven Learning no ensino de vocabulário para aprendizes sul-coreanos de inglês e aponta implicações pedagógicas pertinentes. O objetivo e as perguntas de pesquisa estão bem formulados. Contudo, para sustentar o enquadramento como revisão sistemática, o texto ainda carece de transparência metodológica: é necessário detalhar a estratégia de busca (termos, período, filtros), explicitar critérios de inclusão/exclusão, relatar o processo de triagem com justificativas (preferencialmente com fluxograma) e descrever como os achados foram extraídos e categorizados (categorias, exemplos e controle de vieses). Recomenda-se incluir algum procedimento, ainda que simples, de apreciação da qualidade dos estudos selecionados. A opção por restringir-se à base KCI deve ser melhor justificada e acompanhada de discussão das limitações. Sugere-se

reforçar a contextualização do ELT na Coreia e revisar linguagem e padronização das referências. Recomendação: aprovado com restrições.

✓ **Avaliador 3:** Etelvo Ramos Filho (correções obrigatórias)

This paper represents a valuable contribution to the teaching of English as a Foreign Language. The objective proposed in the article “Data-driven learning in vocabulary instruction for South Korean learners of English” is threefold: (1) survey different DDL activities piloted in vocabulary instruction across various English language teaching contexts in Korea; (2) determine the effectiveness of DDL for vocabulary instruction for this demographic; and (3) explore Korean learners' attitudes towards DDL. The three aspects of the objective are accomplished and presented to the reader in clear, academic language. It highlights the use of corpus linguistics in language teaching in the Korean context, which may be of interest to language educators concerned with integrating corpora into their classrooms. Furthermore, the article offers insights which may be applicable to other environments. As a result, its contribution to the field of Applied Linguistics, particularly the teaching of English as a Foreign Language is relevant. The strengths and limitations of the research are presented and discussed in an unbiased manner. Moreover, the overall organization of the article: Introduction, Method, Results and Discussion, Conclusion, and References facilitates readability. Its transparency makes the research replicable. The article complies with the guidelines established for publication by The ESpecialist. In summary, I recommend that the article be accepted for publication with minor language revisions, as suggested in the comments section.

References

- BENSON, P.; CHIK, A.; LIM, H-Y. Becoming autonomous in an Asian context: Autonomy as a sociocultural process. In: PALFREYMAN, D.; SMITH, R. (eds.) **Learner autonomy across cultures: Language education perspectives**. London: Palgrave Macmillan, 2003. p. 23–40.
- BOULTON, A.; TYNE, H. Corpus linguistics and data-driven learning: A critical overview. **Bulletin Suisse de Linguistique Appliquée**, v. 97, p. 97–118, 2013. Available at: <https://hal.science/hal-01850687>. Accessed: 1 Aug. 2025.
- BREZINA, V.; GABLASOVA, D. The corpus method. In: CULPEPER, J.; KERSWILL, P.; WODAK, R.; MCENERY, T.; KATAMBA, F. (eds.). **English language: Description, variation, and context**. 2. ed. London: Bloomsbury Publishing, 2018. p. 595–609.
- CHO, B.-E. Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. **The East Asian Learner**, v. 1, n. 2, p. 31–36, 2004.
- CHOI, I.-C. The impact of EFL testing on EFL education in Korea. **Language Testing**, v. 25, n. 1, p. 39–32, 2008. DOI: <https://doi.org/10.1177/0265532207083744>
- CHONG, S. W.; PLONSKY, L. A typology of secondary research in Applied Linguistics. **Applied Linguistics Review**, v. 15, n. 4, p. 1569–1594, 2024. DOI: <https://doi.org/10.1515/applirev-2022-0189>

CHUN, S. Word knowledge expansion using different types of target words in guided data-driven learning. **Modern English Education**, v. 17, n. 4, p. 95–118, 2016. DOI: <http://dx.doi.org/10.18095/meeso.2016.17.4.05>

COBB, T. **Compleat lexical tutor** [Website]. 2015. Available at: <http://www.lex tutor.ca/>. Accessed: 1 Aug. 2025.

DAVIES, M. The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. **International Journal of Corpus Linguistics**, v. 14, n. 2, p. 159–190, 2009. DOI: <https://doi.org/10.1075/ijcl.14.2.02dav>

DÖRNYEI, Z.; CSIZÉR, K. Ten commandments for motivating language learners: Results of an empirical study. **Language Teaching Research**, n. 2, v. 3, p. 203–29, 1998. DOI: <https://doi.org/10.1177/136216889800200303>

DÖRNYEI, Z.; RYAN, S. **The psychology of the language learner revisited**. New York: Routledge, 2015.

ELLIS, N. C.; WULFF, S. Usage-based approaches to L2 acquisition. In: VANPATTEN, B.; KEATING, G. D.; WULFF, S. (eds.). **Theories in second language acquisition: An introduction**. 3. ed. New York: Routledge, 2020. p. 63–82.

FLOWERDEW, L. Data-driven learning and language learning theories: Whither the twain shall meet. In: LEŃKO-SZYMAŃSKA, A.; BOULTON, A. (eds.). **Multiple affordances of language corpora for data-driven learning**. Amsterdam: John Benjamins, 2015. p. 15–36.

GABLASOVA, D.; BREZINA, V.; MCENERY, T. Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. **Language Learning**, v. 67, n. S1, p. 155–179, 2017. DOI: <https://doi.org/10.1111/lang.12225>

GABRIELATOS, C. Corpora and language teaching: Just a fling or wedding bells? **The Electronic Journal for English as a Second Language**, v. 8, n. 4, p. 1–34, 2005. Available at: <https://www.tesl-ej.org/ej32/a1.html>. Accessed: 1 Aug. 2025.

GASS, S.; BEHNEY, J.; PLONSKY, L. **Second language acquisition: An introductory course**. 5. ed. New York: Routledge, 2020.

GILQUIN, G.; GRANGER, S. Using data-driven learning in language teaching. In: O'KEEFFE, A.; MCCARTHY, M. J. (eds.). **The Routledge handbook of corpus linguistics**. 2. ed. Oxon: Routledge, 2002. p. 430–442.

HALL, G. **Exploring English language teaching: Language in action**. 2. ed. Oxon: Routledge, 2018.

HOLMBERG, A. J. Using data-driven learning for second language academic writing acquisition. **Journal of English Language and Pedagogy**, n. 3, v. 1, p. 1–11, 2024. DOI: <https://doi.org/10.58518/jelp.v3i1.1880>

HOWELL, D. C. **Statistical methods for psychology**. 7. ed. Belmont: Wadsworth, 2013.

HWANG, I.; CHO, M. Increasing lexical awareness through data-driven learning: Polysemy in EFL pedagogy. **Korean Journal of English Language and Linguistics**, v. 22, p. 1116–1132, 2022. DOI: <https://doi.org/10.15738/kjell.22..202210.1116>

HYLAND, K. **Academic discourse: English in a global context**. London: Continuum, 2009.

JOHNS, T. Whence and whither classroom concordancing? In: BONGAERTS, T.; HAAN, P.; LOBBE, S.; WEKKER, H. (eds.). **Computer applications in language learning**. Dordrecht: Foris, 1988. p. 9–34.

KANG, N. Effectiveness of collocation learning through a web-concordancer in a Korean EFL university course. **Multimedia-Assisted Language Learning**, v. 22, n. 1, p. 41–84, 2019. DOI: <https://doi.org/10.15702/mall.2019.22.1.41>

KIM, E. Enhancing a corpus-based approach to teach English phrasal verbs to Korean learners. **The Journal of Studies in Language**, v. 31, n. 2, p. 295–312, 2015. DOI: <https://doi.org/10.18627/jslg.31.2.201508.295>

KIM, H.; CHUN, S. Fostering lexis awareness and autonomy by corpus-based data-driven learning. **English Teaching**, v. 63, n. 2, p. 213–235, 2008. DOI: <https://doi.org/10.15858/engtea.63.2.200806.213>

KIM, J. A methodology for teaching corpus-based English vocabulary to Korean college learners. **Journal of Learner-Centered Curriculum and Instruction**, v. 24, n. 1, p. 877–894, 2024. DOI: <https://doi.org/10.22251/jlcci.2024.24.1.877>

KIM, N.-Y. A study on mobile-assisted vocabulary learning with dictionary-based, corpus-based, and video-based tasks. **Modern English Education**, v. 19, n. 4, p. 27–38, 2018. DOI: <http://dx.doi.org/10.18095/meeso.2018.19.4.27>

LARSSON, T.; BIBER, D. Encouraging cumulative knowledge building as normal practice in (learner) corpus research. **International Journal of Learner Corpus Research**, v. 11, n. 2, p. 1–16, 2025. DOI: <https://doi.org/10.1075/ijlcr.24022.lar>

LARSSON, T.; BIBER, D.; HANCOCK, G. R. On the role of cumulative knowledge building and specific hypotheses: The case of grammatical complexity. **Corpora**, v. 19, n. 3, p. 263–284, 2024. DOI: <https://doi.org/10.3366/cor.2024.0314>

LEE, D. J. Data-driven learning (DDL) approach: Is it applicable to Korean EFL classrooms? **Foreign Languages Education**, n. 20, v. 3, p. 45–86, 2013.

LEE, J. S. English in Korea. In: BOLTON, K.; BOTHA, W.; KIRKPATRICK, A. (eds.). **The handbook of Asian Englishes**. Hoboken: John Wiley & Sons, 2020. p. 585–604.

LEECH, G. 100 million words of English: The British National Corpus (BNC). **Language Research**, v. 28, n. 1, p. 1–13, 1992. Available at: <https://s-space.snu.ac.kr/handle/10371/85926>. Accessed: 27 Mar. 2026.

LI, S.; WANG, H. Traditional literature review and research synthesis. In: PHAKITI, A.; DE COSTA, P.; PLONSKY, L.; STARFIELD, S. (eds.). **The Palgrave handbook of applied linguistics research methodology**. London: Palgrave Macmillan, 2018. p. 123–144.

LIGHTBOWN, P.; SPADA, N. **How languages are learned**. 5. ed. Oxford: Oxford University Press, 2021.

MACARO, E. Systematic reviews in applied linguistics. In: MCKINLEY, J.; ROSE, H. (eds.). **The Routledge handbook of research methods in applied linguistics**. Oxon: Routledge, 2020, p. 230–239.

MCCARTHY, M. Accessing and interpreting corpus information in the teacher education context. **Language Teaching**, v. 41, n. 4, p. 563–574, 2008. DOI: <https://doi.org/10.1017/S0261444808005247>

MOODIE I; NAM H-J. English language teaching research in South Korea: A review of recent studies (2009–2014). **Language Teaching**. v. 49, n. 1, p. 63–98, 2016. DOI: <https://10.1017/S026144481500035X>

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

NATION, I. S. P. **Learning vocabulary in another language**. Cambridge: Cambridge University Press, 2001.

NATION, I. S. P. **Teaching and learning vocabulary**. Boston: Heinle & Heinle, 1990.

O'DONNELL, T. J. Learning English as a foreign language in Korea: Does CALL have a place? **Asian EFL Journal**, v. 10, n. 1, p. 1–27, 2006. Available at: http://asian-efl-journal.com/pta_april06_TJO.pdf. Accessed: 1 Aug. 2025.

OAK, S.; MARTIN, V. (eds.). **Teaching English to Koreans**. Seoul: Hollym, 2003.

RÖMER, U. Usage-based approaches to second language acquisition vis-à-vis data-driven learning. **TESOL Quarterly**, v. 58, n. 3, p. 1235–1245, 2024. DOI: <https://doi.org/10.1002/tesq.3278>

RYAN, R. M.; DECI, E. L. Self-determination theory. In: MAGGINO, F. (ed.) **Encyclopedia of quality of life and well-being research**. 2 ed. Cham: Springer, 2022. p. 6229–6235.

SCHENCK, A.; CHO, Y. W. The efficacy of corpus-based pedagogical techniques for academic writing. **Multimedia-Assisted Language Learning**, v. 15, n. 2, p. 167–186, 2012. DOI: <https://doi.org/10.15702/mall.2012.15.2.167>

SCHMIDT, R. The role of consciousness in second language learning. **Applied Linguistics**, v. 11, n. 1, p. 17–46, 1990. DOI: <https://doi.org/10.1093/applin/11.2.129>

SCOTT, M. Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In: GHADESSY, M.; HENRY, A.; ROSEBERRY, R. L. (eds.). **Small corpus studies and ELT: Theory and practice**. Amsterdam: John Benjamins, 2001. p. 47–67.

SCRIVENER, J. **Learning teaching: The essential guide to English language teaching**. 3. ed. Oxford: Macmillan Education, 2011.

SINCLAIR, J. (1997). Corpus evidence in language description. In: KNOWLES, G.; MCENERY, T.; FLIGELSTONE, S.; WICHMAN, A. (eds.). **Teaching and language corpora**. London: Pearson. p. 27–39.

SINCLAIR, J. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

YANG, J. Effects of collaborative corpus-based learning on the acquisition and retention of delexical verb collocation. **The Journal of Studies in Language**, v. 31, n. 1, p. 67–94, 2015. DOI: <https://doi.org/10.18627/jslg.31.1.201505.67>

YOON, C.; CHOI, J. Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. **Modern English Education**, v. 16, n. 3, p. 47–69, 2015. Available at: <https://uci.kci.go.kr/resolver/G704-001860.2015.16.3.008>. Accessed: 1 Aug. 2025.

YUN, J. U.; HAN, J-T. Effects of vocabulary task types on EFL learners' vocabulary learning. **Foreign Languages Education**, v. 22, n. 3, p. 27–55, 2015. DOI: <http://dx.doi.org/10.15334/FLE.2015.22.3.27>