

A Systematic Synthesis of Research on Popular Corpus Tools in English Language Learning and Teaching

İbrahim Halil Topal¹, Gazi University, College of Foreign Languages,
ibrahimtopal@gazi.edu.tr

Recommended citation: Topal, İ. H. (2026). A Systematic Synthesis of Research on Popular Corpus Tools in English Language Learning and Teaching. *Journal of Language Research (JLR)*, 10(2), 222-243.
DOI: <https://doi.org/10.51726/jlr.1793357>

Abstract: It is well-established that second/foreign language learning requires exposure to meaningful input. Thanks to advancing technologies, one way to achieve this in language teaching contexts is via online corpus tools. Given the potential pedagogical gains these tools offer for language skills (e.g., vocabulary, grammar, pronunciation, and writing), this study aims to synthesise the findings of research on five popular corpus tools – Sketch Engine, SkELL, PlayPhraseMe, Fraise.it, and CorpusMate. Following the PRISMA guidelines for systematic reviews, this study included five research papers from the Web of Science and Scopus databases until 2024. Applying qualitative and quantitative content analyses, the study found that corpus consultation promotes language skills and learner autonomy, supports contextually appropriate language use, and improves sensitivity to authentic language structures. Despite the promising outcomes, the scant number of studies analysed with some methodological issues and contextual constraints undermines the generalisability of the results. The study underscores the urgent need for more longitudinal, comparative, and large-scale research on using these corpus tools to harness them fully in language education. It also proposes that when systematically implemented, corpus tools may serve as powerful resources for language learning and teaching.

Keywords: corpus, English language learning, data-driven learning, systematic review, language corpora

INTRODUCTION

The proliferation of technology tools in foreign language education (FLE) contexts in recent years has added new dimensions to teaching-learning processes. In particular, corpus linguistics (CL) and data-driven learning (DDL) approaches offer alternative learning opportunities by accessing authentic language data. Corpus-based tools allow learners to explore how the target language is used in natural contexts and recognise language patterns. Such tools offer variety and innovation in FLE. One such tool is Sketch Engine, providing researchers and teachers with comprehensive textual analyses. Several studies have reported that it contributes to academic writing and vocabulary teaching (Kunilovskaya & Koviagina, 2017; Yang, 2025). Its student-friendly version, Sketch Engine for Language Learning (SkELL), enables learners to access contextual examples and concordance tables thanks to its simplified interface (Topal, 2025a). Accordingly, it may be preferred for selecting words and learning collocations among students.

Representing a different approach, PlayPhraseMe stands out as a multimedia corpus comprising short video clips from movies and series. Students are given the opportunity to be exposed to the target language features both auditorily and visually, practice listening, vocabulary, and pronunciation, and develop pragmatic awareness (Topal, 2025b). Fraise.it, which focuses on supporting writing skills, helps students see how words and expressions are used in authentic sentences. Finally, the newest tool, CorpusMate, is a user-friendly concordance tool supporting DDL in language education (Crosthwaite & Baisa, 2024).

¹ ORCID: [0000-0003-4220-3706](https://orcid.org/0000-0003-4220-3706)

Submitted: 29.09.2025

Accepted: 12.05.2026



Despite numerous studies on these tools with different learner groups and various language skills, they usually focus on a specific tool or context, thus requiring a comprehensive and holistic comparison. Accordingly, this study aims to systematically examine the use of these five prominent tools in the literature in learning and teaching English, reveal documentary and methodological characteristics, and present the major findings with pedagogical implications. Thus, it provides a clearer framework for teachers and researchers regarding the pedagogical potential of corpus-based tools.

Against this backdrop, the current study addressed the following research questions:

1. What are the documentary characteristics of the analysed publications?
2. What are the methodological properties of the analysed publications?
3. What skills in English language learning have the analysed corpus tools targeted?
4. What are the major findings presented in the analysed studies, and what do they imply for English language education?

THEORETICAL FRAMEWORK

Corpus Linguistics and Data-Driven Learning

The present study can primarily be grounded in CL – a research method using large-scale collections of texts (corpora) compiled to systematically study natural language use (McEnery & Hardie, 2012). Pioneering studies such as Sinclair (1991) and Biber et al. (1998) emphasized basing linguistic analysis on authentic language data rather than intuition. This approach has opened new language learning and teaching horizons by discovering linguistic patterns, word frequencies, and collocations. Indeed, CL provides a strong basis for preparing teaching materials and providing authentic language input to students in foreign language education.

The study can also be rooted in DDL, an approach proposed by Johns (1990) and an important paradigm combining CL and pedagogy. DDL gives students direct access to language knowledge, by which they move away from being passive recipients into “language detectives” (Moon & Oh, 2018). Students can use corpus features (e.g., wordlists and collocation examples) to discover linguistic structures and make learning more meaningful (Boulton, 2010; Römer, 2009). In addition, DDL is an effective method, especially in academic writing, vocabulary teaching, and developing grammatical awareness (Huang, 2014).

The study’s theoretical basis lies at the junction of CL and DDL. The tools examined (i.e., Sketch Engine, SkELL, PlayPhrase.me, Fraise.it, and CorpusMate) enable different uses of corpus data in language teaching. Therefore, this article examines corpus-based tools in English language teaching within CL principles and DDL pedagogy.

Corpora and Foreign Language Education

Corpus-driven approaches may contribute considerably to materials development through authentic language use in FLE. Unlike traditional textbooks, which usually include limited and artificial language input, corpora allow students to encounter language samples from real communication environments (McEnery & Xiao, 2011). This helps students develop a natural awareness of the target language’s word frequencies, formulaic expressions, collocations, and discourse patterns (Hunston, 2002; Reppen, 2011).

Using corpora in FLE has pedagogical value for teachers and students. Through corpora, teachers can create up-to-date and original teaching appropriate to the scope of the curriculum, and students experience a comprehensive learning experience in the target language through concordance lines and vocabulary lists. Corpora, especially with the DDL approach, allow students to develop their grammar and vocabulary knowledge independently (Johns, 1990; Boulton, 2010).



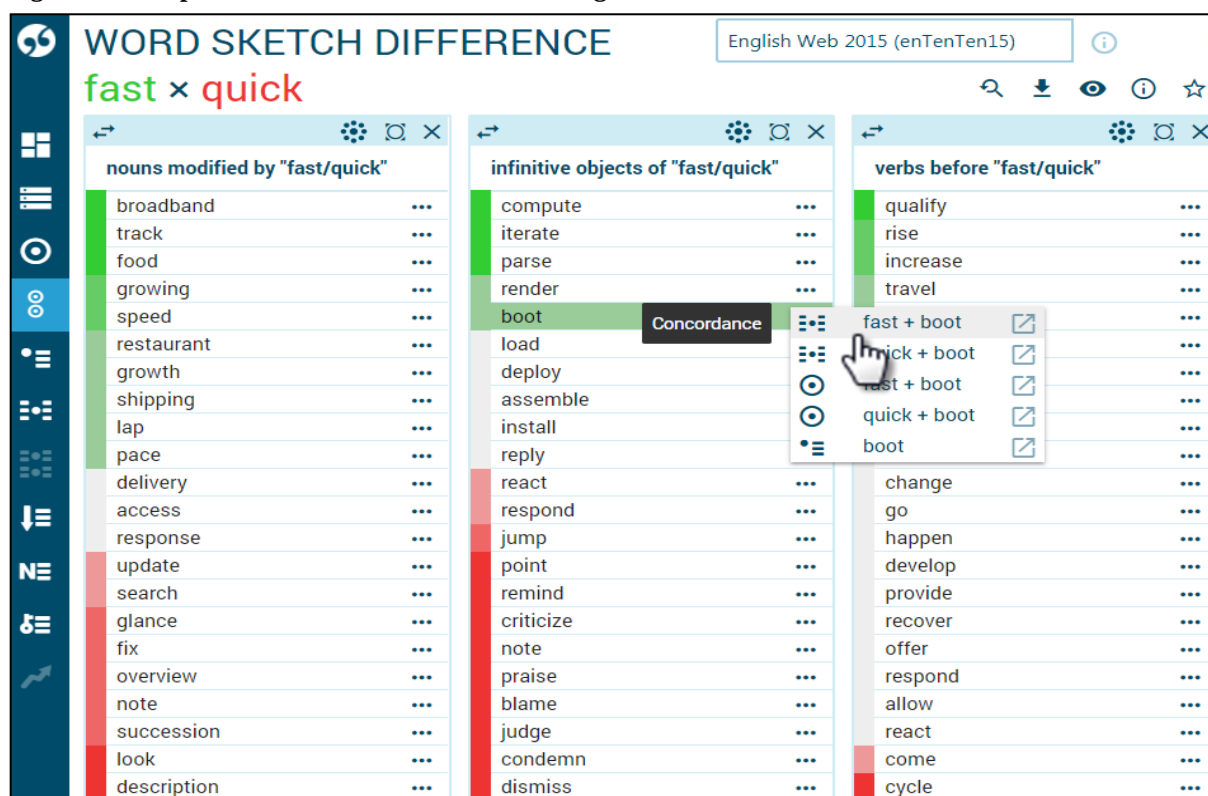
In recent years, the increasing accessibility of corpus tools in technology-supported language teaching has opened new horizons in foreign language education. Online tools such as Sketch Engine, SkELL, PlayPhraseMe, Fraise.it, and CorpusMate bring data-driven pedagogy to classroom practice and contribute to developing different language skills (Kilgarriff et al., 2015; Topal, 2025a, 2025b, 2025c). In this context, the corpus is positioned as an innovative resource and method not only in linguistic research but also in all aspects of foreign language education.

Popular Corpus Tools in Language Education

Sketch Engine

The study utilises five corpus tools selected according to popularity, pedagogical utility, and research interest. The first one is Sketch Engine (<https://www.sketchengine.eu/>), a tool to explore how language works. Its algorithms process extensive corpora containing billions of words to reveal patterns of typical, rare, or emerging language use, and allow for advanced text analysis (Kilgarriff et al., 2014). The tool hosts over one trillion words across 800 ready-to-use corpora in more than 100 languages, with individual corpora reaching up to 80 billion words to ensure highly representative language samples (<https://www.sketchengine.eu/>). With its broad features, Sketch Engine is employed by linguists, lexicographers, translators, students, and educators worldwide. It is also a preferred tool for global publishers, universities, translation agencies, and national language institutions.

Figure 1. A snapshot of a word sketch on Sketch Engine



In language education contexts, Sketch Engine can be utilised for collocation discovery, word sketches, thesaurus in context, concordance lines, corpus building, frequency lists, and trends over time. In particular, language users may find words that naturally co-occur, observe a word's typical grammatical behaviour, distinguish near-synonyms, examine how specific words are used in authentic context, build specialised corpora by uploading texts, generate high-frequency vocabulary lists, and track lexical trends through diachronic corpora. Best for research and deep corpus analysis, Sketch Engine can be suitable for teachers and advanced language users.

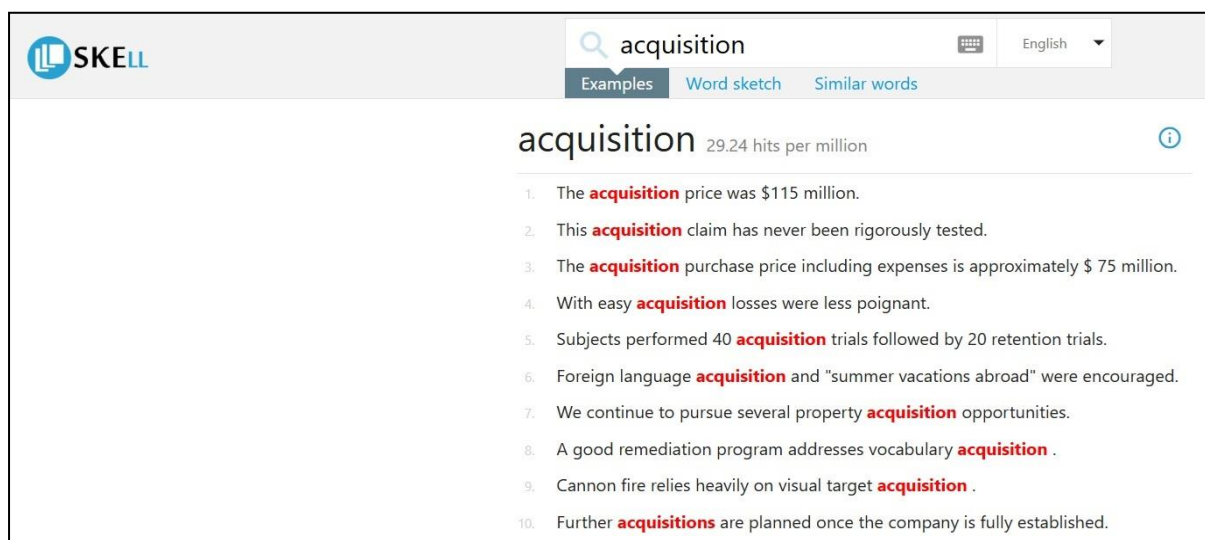
Sketch Engine for Language Learning (SkELL)

SkELL (<https://skell.sketchengine.eu/>) is a user-friendly resource for students and teachers of English to explore how words and phrases are used by real speakers (Topal, 2025a). It is free and



requires no registration. Examples, collocations, and synonyms are generated automatically through advanced algorithms and large-scale text analysis, with no manual intervention. Unlike Sketch Engine, SkELL allows students to access concordance tables and example sentences without having to deal with complex interfaces. Studies show that using SkELL increases students' vocabulary and strengthens their context awareness (Kilgariff et al., 2015; Troy, 2022).

Figure 2. A snapshot of a word search on SkELL

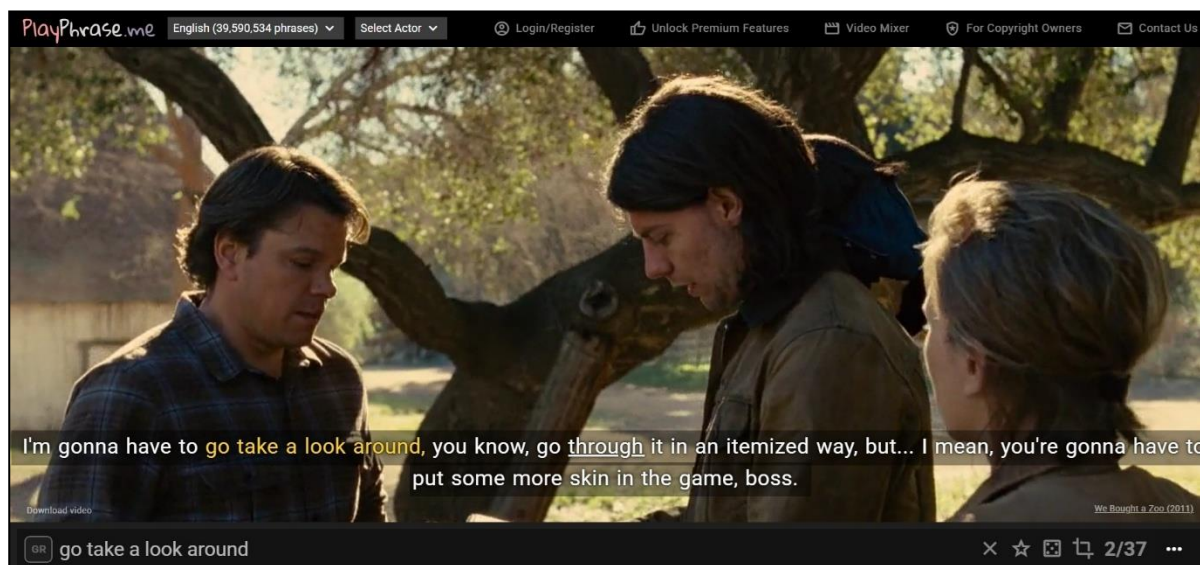


This simplified Sketch Engine interface, designed for learners (i.e., SkELL), may allow users to conduct example-sentence searches, provide a word-combination widget, access collocations for writing, promote independent learning, and foster classroom collocation exercises. Since it has almost the same qualities as the Sketch Engine, the most noticeable difference between SkELL and Sketch Engine is that the former does not necessitate expert knowledge of the corpus. Promoting vocabulary and collocations, SkELL can be used by B1-C2 learners and teachers.

PlayPhraseMe

PlayPhraseMe (<https://www.playphrase.me/>) is an online platform and a mobile application that allows users to watch video clips from movies and TV series. The tool enables search across six languages (English, German, French, Italian, Japanese, and Spanish) and filter by actors (Topal, 2025b). Unlike traditional corpus tools with a basic search engine, PlayPhraseMe serves as a multimedia corpus created from movie and TV series lines. That means, students watch short video clips of a particular word or expression and experience the use of the target language audio-visually. Research shows that PlayPhraseMe assists in comparing pronunciation varieties and highlighting the pragmatic aspects of the target language (Kontsioti & Karasimos, 2025; Veiga Norlander, 2022).



Figure 3. A snapshot of a phrase search on PlayPhraseMe

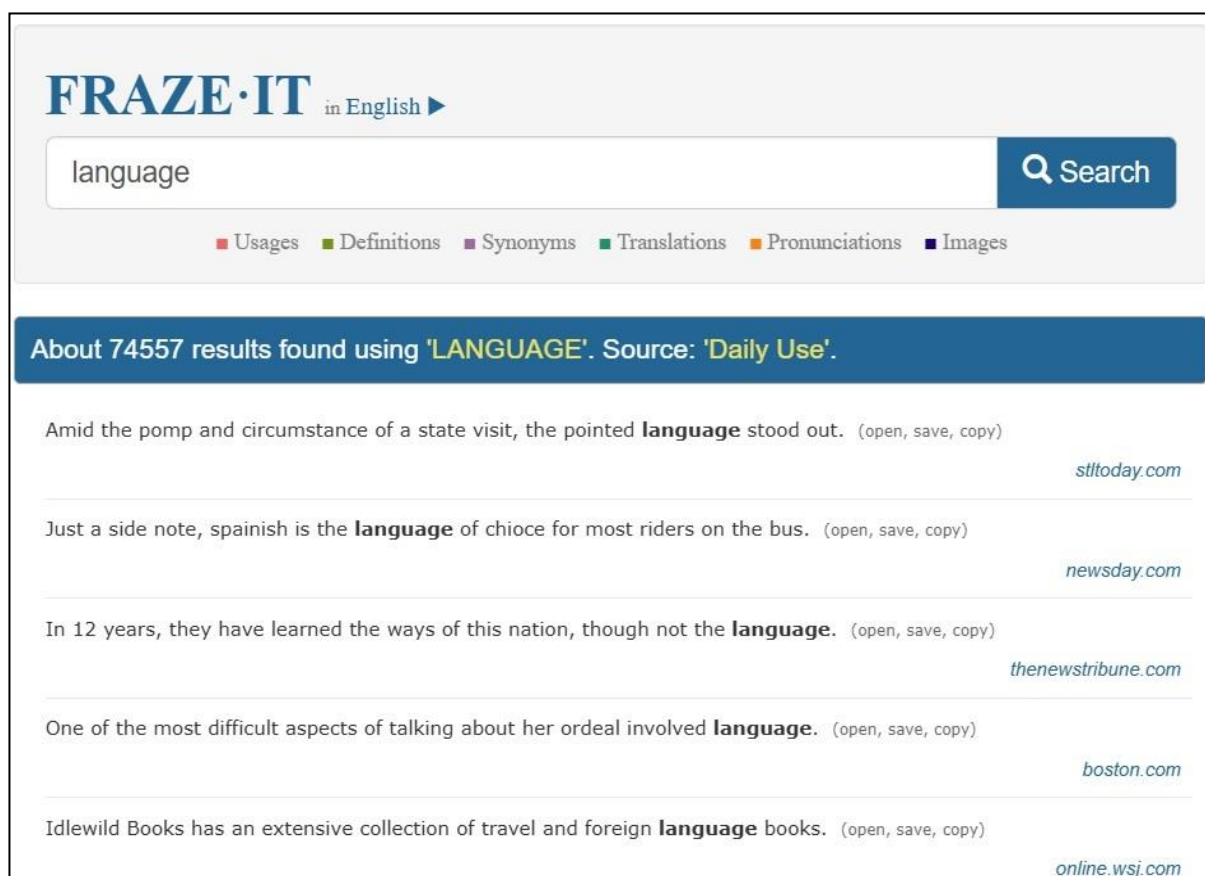
In language learning and teaching contexts, PlayPhraseMe provides users with contextualised pronunciation examples, informal register exposure, idioms and chunks, listening discrimination, cultural and pragmatic examples, vocabulary in emotional contexts, and video quiz opportunities. More specifically, learners can discover how specific words or phrases are pronounced, hear natural speech, search for exact scenes where certain lexical items are used, play clips at varying speed rates for dictation activities, observe how specific cultural/pragmatic expressions are used in actual context, and retain vocabulary through memorable and relatable scenes. Teachers, on the other hand, can design video quizzes on vocabulary, grammar, and pronunciation. Well-suited for listening, pronunciation, and pragmatics, this tool can be benefitted by language users between A2-C2 levels and teachers.



Fraze.it

Fraze.it (<https://frazе.it/>) is an online concordancer used especially for writing and vocabulary (Topal, 2025c). With 100 million sentence examples, the tool provides service in English, French, German, Italian, Portuguese, and Spanish. Filtering options include source (e.g., daily use and famous), form (e.g., affirmative and negative), tense (e.g., simple present and past continuous), rule (i.e., starting or ending with), zone (i.e., US or UK English), and context (e.g., business and education). Studies have found that using Fraze.it helps students choose more natural and accurate structures in their written production (Çakır & Özer, 2020).

Figure 4. A snapshot of a word search on Fraze.it



In language education settings, Fraze.it allows users to conduct phrase-in-context searches, distinguish similar structures, read authentic input, practise collocation, and verify phrases in natural contexts. Specifically, language learners can see thousands of examples of how specific words/phrases are sorted by frequency, compare two similar structures (e.g., *pensare a* vs. *pensare di*) with real examples, expose themselves to curated literary and journalistic sentences serving as high-quality reading materials at sentence level, and search for words (e.g., *grande*) to see what nouns they most typically modify in real prose. On the other hand, teachers can design assignments by extracting authentic sentence sets to build translation or gap-fill exercises. Being conducive to Italian phrase-in-context in particular, the tool can be more suitable for language users of A2-C2 levels, along with teachers.

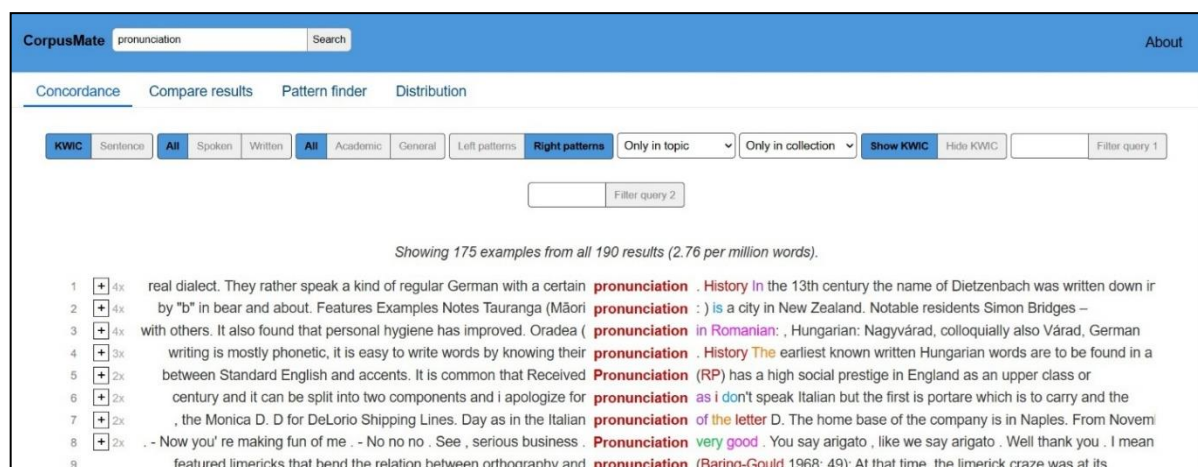
CorpusMate

The last corpus tool is CorpusMate (<https://corpusmate.com/>), a simple concordancer that allows users to perform queries across various topics and list results by concordance, compare results, pattern finder, and distribution. The platform also enables to refine the results by key word in context (KWIC) and sentence, spoken and written forms, academic and general discourse, and only in topic and collections. Further filtering options involve showing or hiding KWIC, sorting or not sorting lengths, and lengths (1-10 or 10-20). Despite limited studies in the literature, current research shows



that CorpusMate is a promising tool that positively affects prospective teachers' attitudes towards using corpora and helps students work more comfortably with concordance tables (Crosthwaite & Baisa, 2024; Idrizi, 2024).

Figure 5. A snapshot of a word search on Fraise.it



CorpusMate is a user-friendly corpus interface, often used with learner or pedagogical corpora. Teachers can upload student writing samples to identify common error patterns across the class, profile error frequency and design a lesson accordingly, build inductive grammar tasks through concordancing, test learners' lexical and grammatical knowledge by generating gap-fills directly from concordance lines, and compare registers. Learners, on the other hand, can build their own mini-corpora and investigate vocabulary and grammar patterns independently. Concerned particularly with classroom corpus work and learner corpora, this tool is mostly appropriate for advanced language learners and teachers.

Previous Studies and Reviews on Corpora and DDL in Language Education

In their meta-analysis of 64 studies on corpus use in language learning, Boulton and Cobb (2017) found that DDL approaches result in significant overall effects for both control/experimental group comparisons and pre-/posttest designs. In their review of CL and corpus-based research, Al-Hamzi et al. (2020) reported a notable trend toward integrating corpora in language teaching. In another study, Frigal et al. (2020) indicated that corpus-based instruction (CBI) and DDL foster language skills, particularly writing performance, vocabulary usage, collocation competence, and grammatical awareness.

Ma and Mei (2021) highlighted that corpus-based activities and tools can enhance vocabulary acquisition considerably by providing learners with authentic, contextualized language input. Looking at research from 2011 to 2015, Pérez-Paredes (2022) observed that much of the focus was on using concordancing tools and collocations to help college students develop their writing skills. Similarly, Gui et al. (2022) found that studies often aim to improve English writing quality, vocabulary knowledge, collocation accuracy, and learner autonomy through corpus-based approaches.

Lusta et al. (2023), in their systematic review, concluded that language corpora present significant pedagogical potential for the language classroom, particularly when accompanied by supportive tasks, guidance, extra assistance, and collaborative learning. They also noted that low learner proficiency and insufficient integration can limit the impact of corpus-based methods, and suggested more systematic strategies are needed to engage learners effectively. Emir and Saral (2025) supported the value of corpus applications as DDL tools for language improvement, but emphasized



the importance of developing teachers' corpus literacy, promoting collaborative learning, and offering comprehensive training for effective classroom integration.

Huang and Ma (2025) found that DDL might be effective for young learners across age groups and corpus interaction types, with learners exhibiting positive attitudes. Li et al. (2025) implied improved English skills and learner autonomy in EFL classrooms, albeit with challenges. They, therefore, suggested comprehensive teacher training, simplified tasks, and personalized learning can enhance corpus implementation. Finally, Tong et al. (2025) suggested that developing corpus-based wordlists effectively integrates various objective criteria (e.g., word frequency, range, and dispersion).

Overall, the systematic reviews and meta-syntheses on corpora in language education varied in scope and focus. The lack of a systematic review of corpus tools for English language education motivated the current study.

METHODOLOGY

Research Design

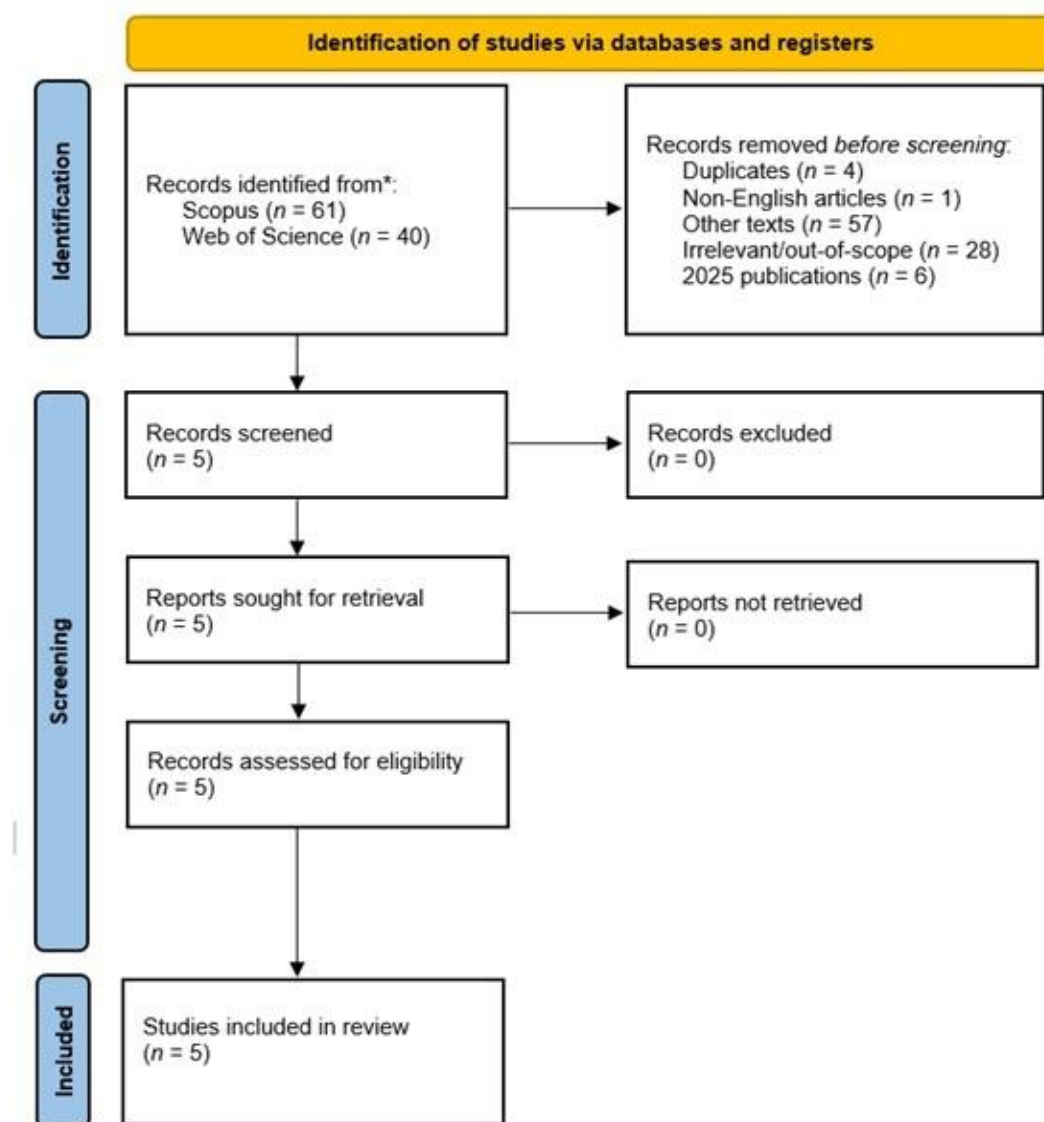
This systematic literature review (SLR) examined the research on the five popular corpus tools. SLR is a comprehensive investigation to identify all relevant studies addressing a specific research question and evaluate the validity of each study before drawing conclusions (Petticrew & Roberts, 2006). SLR is typically performed to (i) map the international evidence base, (ii) confirm existing practices, highlight variations, or reveal emerging approaches, (iii) inform future research directions, (iv) examine and reconcile conflicting findings, and (v) provide evidence-based guidance for decision-making (Munn et al., 2018, p. 2). These objectives also constituted the motivation for the present study. The updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021), which are widely recognized in academia for ensuring rigor, transparency, and comprehensiveness (Kolaski et al., 2023), were used in congruence with the SLR. A review protocol was designed to specify data sources, define the search strategy, formulate query strings, and set clear inclusion and exclusion criteria.

Databases

The study utilised two major databases: The Web of Science (WoS) and Scopus. The WoS is well-suited for conducting systematic reviews that demand top-notch, peer-reviewed research and for evaluating citations to gauge the impact of research in a particular field (Zhu & Liu, 2020). Equally, Scopus, with its access to high-quality, peer-reviewed research and advanced citation tracking (Baas et al., 2020), is essential in identifying leading studies and trends in an SLR, keeping you abreast of the latest developments.



Figure 6. The PRISMA flowchart for the SLR



Search Strategy

A systematic search in publication titles was conducted on Scopus and the WoS using the tools' names to obtain the most relevant and eligible studies. The Scopus query yielded an overall 61 results: Sketch Engine (n=48), SkELL (n=10), PlayPhraseMe (n=1), Frazee.it (n=1) and CorpusMate (n=1). The WoS search, on the other hand, revealed 40 documents: Sketch Engine (n=30), SkELL (n=8), PlayPhraseMe (n=0), Frazee.it (n=1) and CorpusMate (n=1). Without the refinements and inclusion/exclusion criteria applied, the total number of documents acquired were 101.

Inclusion and Exclusion Criteria

Studies were included if they were peer-reviewed articles in English on the five corpus tools published until 2024 in journals indexed in Scopus and/or the WoS. The temporal range was limited to 2024 due to the incompleteness of 2025. No limitation was imposed on the prior years of publication. Studies were excluded if published in languages other than English, were purely lexicographic or linguistic, had no pedagogical implications, or were published as different document types. Additionally, the relevant documents indexed in databases other than Scopus and the WoS were excluded. Table 1 summarises the inclusion/exclusion criteria established for this study.



Table 1. Inclusion and Exclusion Criteria

Criteria	Included	Excluded
Corpus tool	Sketch Engine, SkELL, PlayPhraseMe, Fraise.it, CorpusMate	Others (e.g., JustText, Spiderling, Chared, Onion, Unitok, wiki2corpus)
Document type	Article	Others (proceedings, reviews, book chapters, etc.)
Language of publication	English	Others (e.g., French, German, Spanish)
Year of publication	Until 2024	2025 and after
Focus/scope	English language learning/teaching	Others (e.g., corpus linguistics, computational linguistics, lexicography)
Database	Scopus and the WoS	Others (e.g., Google Scholar, JSTOR, ScienceDirect, PubMed, etc.)

Screening and Selection Process

Two independent researchers with a PhD in ELT screened the titles and abstracts (if necessary). Accordingly, 57 texts were removed since they were other document types (e.g., proceedings or reviews). Twenty-eight documents were excluded because of their irrelevance or lack of direct implications for English language education. Six articles were published in 2025; hence they are excluded. One article was published in another language (Spanish), and four documents overlapped in Scopus and the WoS and were removed from analysis. Following the inclusion/exclusion criteria, the Scopus query yielded five papers. After the refinement criteria, the WoS search generated four articles. Overall, both databases provided nine articles. Four of these articles were duplicates, resulting in five distinct articles.

Coding Scheme

For the first two questions, the research adapted a paper classification form (PCF), originally developed by Sozbilir et al. (2012) and adapted to English language teaching by Yağız et al. (2016) for documentary and methodological properties. The PCF included seven sections for documentary characteristics (i.e., title of publication, year of publication, authors' names, authors' affiliation, authors' country, keywords, and citation count) and five for methodological properties (i.e., research design/method, target audience, context, data collection tools, and data analysis methods). For the remaining two questions, the author used another spreadsheet with pre-determined codes (i.e., targeted skills and major findings).

Data Analysis

The study applied qualitative content analysis (QCA) to examine documentary and methodological aspects. QCA provides a systematic and context-sensitive approach to interpreting qualitative data (Selvi, 2020). In addition, thematic analysis (Braun & Clarke, 2006) was employed to identify the major findings (e.g., targeted skills, benefits/drawbacks of the corpus tools).

Trustworthiness

This SLR was designed and reported in accordance with PRISMA guidelines. Multiple and reliable databases (Scopus and the WoS) were used to ensure the reliability of the literature review, and inclusion and exclusion criteria were predetermined. Based on Lincoln and Guba's (1985) trustworthiness components, a double-researcher review and consensus processes were conducted for credibility. Studies conducted in different contexts were reported in detail for transferability. Research steps were clearly documented to increase dependability, and all selections were justified and supported with tables or figures to ensure confirmability. These approaches aim to ensure transparency, reproducibility, and reliability of the findings.



FINDINGS

The findings are presented under relevant headings in line with the research questions.

Documentary Qualities

Table 2 displays the documentary characteristics of the analysed articles. Accordingly, the relevant studies ($n = 5$) indexed in the WoS and Scopus spanned a decade, albeit with individual papers published in specific years, thereby leaving wide gaps. This aligns with the findings of previous studies in CL, applied linguistics, and FLE reporting individual studies in separate years (Chambers, 2019; Park & Nam, 2017). The scarcity of publications on the popular corpus tools is one of the remarkable findings of the study, concurring with Römer's (2011) remarks about the insufficient implementation of corpora in language education contexts. When further analysed, two papers were published about Sketch Engine, one about SkELL, and the other about CorpusMate. This, again, coincided with the broader interest in Sketch Engine, notwithstanding the purely CL perspectives (e.g., Jakubíček et al., 2014; Pearce, 2008). No WoS- or Scopus-indexed articles were encountered until 2024. Yearly statistics show that the most publications were in 2019 ($n = 2$), followed by one in 2015, 2020, and 2024. This may be explained by the gradual interest in and recognition of the corpus tools (e.g., Sketch Engine and SkELL) among language teachers and researchers.

Table 2. The Documentary Characteristics of the Analysed Articles

Title of publication	Author(s)	Year of publication	Published in	Citation count*
Corpora and language learning with the sketch engine and SKELL	Kilgarriff, A., Marcowitz, F., Smith, S., Thomas, J.	2015	Revue Francaise De Linguistique Appliquee	39
How corpora can assist legal translation learners: The GENTT TransTools Corpora platform and Sketch Engine	Albi, A.B.	2019	Quaderns De Filologia Estudis Linguistics	5
Applying 'Sketch Engine for Language Learning' in the Japanese English classroom	Hirata, Y., Hirata, Y.	2019	Journal of Computing in Higher Education	9
Fostering intuitive competence in L2 for a better performance in EAP writing through Fraise.it in a Turkish context	Çakır, İ., Özer, M.	2020	Education and Information Technologies	2
A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate	Crosthwaite, P., Baisa, V.	2024	International Journal of Corpus Linguistics	4

* Citation counts are obtained from Scopus as of 2025 September.

Authorial descriptives reveals that multiple authorship was more common ($n = 4$) than the single one ($n = 1$). The variety of publication venues is another noteworthy finding, echoed previous research (Park & Nam, 2017; Salimi & Marami Hajikandi, 2025). The papers were included in journals from applied linguistics to CL and computers in higher education. Citation analysis indicates a predictable outcome, with the oldest publication (Kilgarriff et al., 2015) receiving the highest ($n=39$). However, the most recent one was not cited the least. This shows that article content and broad research interest may influence citational practices (Aksnes et al., 2019).



Further documentary characteristics involved analysis of authorial background (Table 2). In general, the authors showed institutional, national, and regional diversity, agreeing with previous research findings (Crosthwaite et al., 2023; Park & Nam, 2017; Zakaria & Aryadoust, 2024). Institutionally, Coventry, Masaryk, and Hokkai-Gakuen universities were represented by two authors. Nationally, the UK was the most prolific, followed respectively by two authors from Czechia, Japan, and Türkiye. Regional trends indicate the prevalence of European researchers ($n = 8$), succeeded by Asian ($n = 2$) and Australian ($n = 1$) ones.

Table 3. Authorial Background

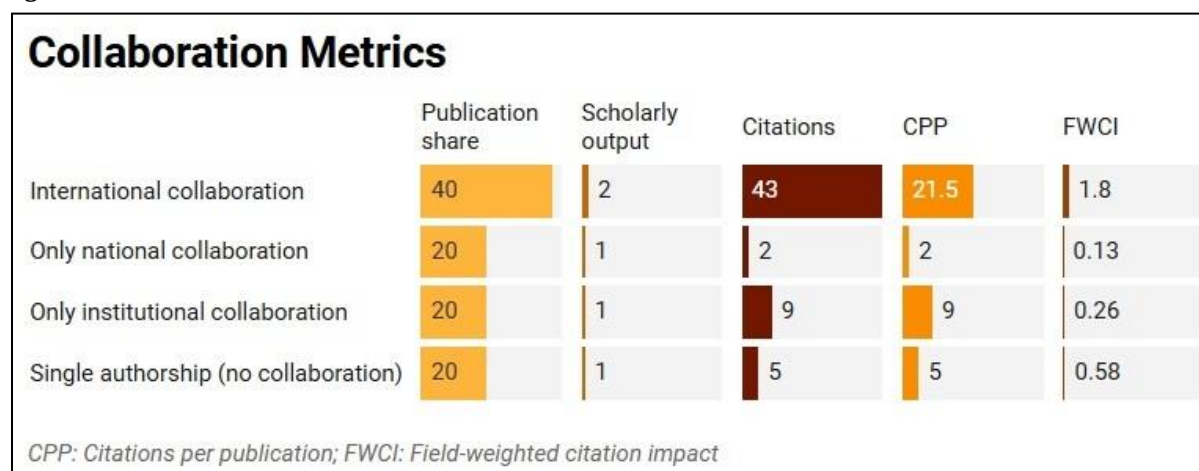
Author's name	Affiliation	h-index*	Country	Region
Adam Kilgarriff	Lexical Computing Ltd	39	UK	Europe
Fredrik Marcowitz	Coventry University	39	UK	Europe
Simon Smith	Coventry University	39	UK	Europe
James Thomas	Masaryk University	2	Czechia	Europe
Anabel Borja Albi	Universitat Jaume I	6	Spain	Europe
Yoko Hirata	Hokkai-Gakuen University	5	Japan	Asia
Yoshihiro Hirata	Hokkai-Gakuen University	3	Japan	Asia
İsmail Çakır	Ankara Yıldırım Beyazıt University	2	Türkiye	Europe
Mustafa Özer	Kayseri Abdullah Gül University	2	Türkiye	Europe
Peter Crosthwaite	University of Queensland	17	Australia	Australia
Vít Baisa	Masaryk University	10	Czechia	Europe

* H-indices were retrieved from Elsevier's SciVal.

A researcher has an h-index of n if they have n publications, each with at least n citations (Hirsch, 2005). H-index balances quantity (number of publications) with quality (number of citations). Accordingly, the authors of the Sketch Engine article were the most influential in terms of citational influence (h-index). They were followed by Peter Crosthwaite and Vít Baisa, who are also well-known CL researchers.

The collaboration practices between authors were also analysed within the documentary properties. As Figure 7 shows, international collaboration was more than other types of networks, concurring with earlier work (Park & Nam, 2017; Wang & Liang, 2024). Despite the low number of publications, only national, only institutional, and single-authorship (no collaboration) types were represented by a single paper. In terms of citational impact, international collaboration had the highest ($n=43$), followed by only institutional ($n=9$) and single authorship ($n=5$). This aligns with the previous scholarly research revealing similar results (Kwiek, 2021; Nguyen et al., 2017). Citations per publication followed a similar trend to the overall citations.

Figure 7. The collaboration network between authors



* The collaboration metrics were retrieved from Elsevier's SciVal.



Field-Weighted Citation Impact (FWCI) in SciVal shows how often an entity's publications are cited compared to the average citation rate of similar publications within the database (Scelles & Teixeira da Silva, 2025). Accordingly, FWCI scores of the analysed papers were congruent with the overall citations and CPP values.

Methodological Properties

The obtained studies were also analysed regarding methodological properties (Table 4). Kilgarriff et al. (2015) explicitly introduced the corpus tools (Sketch Engine and SkELL) to practitioners, used rich illustrative examples, and showed pedagogical applications, thus bridging CL with language pedagogy. However, the learner outcomes were not tested empirically, there was a lack of details on participants and data-driven classroom results, and pedagogical claims were not evidence-based. Therefore, future work could empirically test SKELL/Sketch Engine in classrooms with pre-/post learning measures and collect learner perceptions about the tools' efficacy.

The largely descriptive and illustrative approach taken in this study reflects a broader trend in early corpus tool research, where tool demonstration and pedagogical theorising often precede rigorous classroom validation. While such work is valuable for raising awareness and providing practical guidance, the lack of empirical testing limits the generalisability and evidence base for pedagogical claims. This highlights the need for the field to move toward more systematic impact evaluation, employing robust research designs that can substantiate the benefits and limitations of corpus tools in real educational settings.

Table 4. The Methodological Properties of the Analysed Studies

Study	Research Method / Design	Target Audience	Context	Data Collection Tools	Data Analysis Methods
Kilgarriff et al. (2015)	Descriptive / conceptual tool-introduction with illustrative examples (not experimental)	Language learners, teachers, syllabus designers	General language learning and syllabus design contexts	Corpus outputs (concordances, collocations, learner corpora examples); illustrative tasks	Descriptive corpus analysis (frequency, collocation, overuse/underuse) + qualitative pedagogical discussion
Albi (2019)	Applied case study of corpus use in translation pedagogy	Legal translation learners (university students)	Legal translation classroom (Spain)	Learner translation tasks, corpus queries, classroom examples	Qualitative comparative analysis of translations; corpus inspection; descriptive commentary
Hirata & Hirata (2019)	Empirical classroom study; evaluative, mainly survey-based (mixed-methods orientation)	Japanese university EFL students	Japanese higher-education English classes	Student surveys (Likert + open-ended), classroom corpus queries	Descriptive statistics of survey items; qualitative thematic coding of open responses
Çakır & Özer (2020)	Quasi-experimental, mixed-methods (pre/post test + survey)	Turkish EAP university students	Turkish university academic writing course	Pre- and post-tests, delayed test, online survey with open-ended items	Inferential statistics (t-tests on scores); qualitative coding of survey responses; some corpus-informed analysis
Crosthwaite & Baisa (2024) –	Design-based/development research with pilot user feedback	Secondary-school learners and teachers (Australia)	Secondary education, disciplinary DDL integration	User feedback (questionnaires, focus groups, observation), system logs	Qualitative thematic analysis of feedback; design rationale synthesis; descriptive usability reporting



Albi (2019) addressed a specialised niche (legal translation) often overlooked in DDL research (Biel, 2017) and proposed concrete examples of corpus-supported translation practice. Nevertheless, case study orientation limits generalisability (Thomas, 2011). Also, the article did not clearly state sample size, participant demographics, and instruments. In addition, there was no robust measurement of learner performance gains. Prospective research might include pre-post translation quality assessment to test the impact and gather systematic data about learner perceptions and attitudes.

This methodological profile is indicative of a tension in the field between the value of in-depth, context-specific case studies and the need for broader generalisability. While the detailed practice-based insights are valuable for specialized domains like legal translation, greater methodological transparency and systematic measurement would help translate local findings into scalable best practices. This reflects a broader call in corpus-based pedagogy for mixed-methods research that balances rich qualitative data with quantitative validity

Employing a mixed-methods approach, Hirata and Hirata (2019) conducted an empirical classroom-based evaluation with real EFL students and captured student attitudes and perceptions. Still, their study relied heavily on self-reported perceptions rather than direct language outcomes. It was limited to one institutional context with small-scale sampling and had no control group to compare tool-based vs. non-tool-based learning. Future studies could add learning performance tasks (writing/speaking improvements) as outcome measures, use larger and more diverse samples across institutions, or track the long-term adoption of corpus tools (e.g., SkELL) in formal and informal contexts.

The mixed-methods design adopted here demonstrates progress toward methodological rigor, yet reliance on self-report data and limited sampling highlight persistent challenges in capturing the true impact of corpus tools. This finding underscores the importance of triangulating data sources and scaling research to include comparative and longitudinal elements. Enhanced methodological sophistication will be essential for building a robust evidence base for corpus-based pedagogies.

Çakır and Özer's (2020) study had the strongest methodological design (quasi-experimental) among the five. They used quantitative performance data and qualitative student feedback and demonstrated real classroom integration of the corpus tool (Fraise.it). Nonetheless, with no randomization, their study was limited by a small sample and a single context (one Turkish university), resulting in moderate internal validity (Patino & Ferreira, 2018). Plus, the survey relied on self-reports, which may be biased (Rosenman et al., 2011). Prospective researchers may replicate this study in other EAP contexts and employ longitudinal methods to track writing development.

The quasi-experimental approach marks a methodological advance in corpus tool research, offering a stronger basis for causal inference regarding tool effectiveness. However, limitations in sampling and randomization highlight the continued need for rigorous experimental controls and replication. Expanding methodological scope to include larger, more diverse participant pools and multi-site studies will help clarify the external validity of findings and support the adoption of corpus tools across educational contexts

Finally, Crosthwaite and Baisa's (2024) design-based research responded directly to usability gaps in DDL and focused us on secondary education, broadening DDL applications beyond universities. It was a tool introduction paper rather than a learning outcome study. Therefore, there was limited evidence of how much the tool actually improved performance. Follow-up studies might accordingly assess CorpusMate in classroom interventions with pre/post measures, compare usability/effectiveness against Sketch Engine/SkELL, or collect teacher adoption and integration data over full academic terms.

This focus on design-based research and tool usability reflects a welcome expansion of corpus pedagogy into underexplored educational levels and practical classroom concerns. However, the lack of outcome-focused evaluation underscores a recurring gap in the literature: the need to move from



tool development and feasibility studies toward evidence of actual learning gains. Methodologically, future research should prioritize integrating usability testing with systematic assessment of educational impact

Together, these papers advance corpus pedagogy by covering tools, contexts, and learner groups. They highlight both higher education (e.g., EAP, translation, and Japanese EFL) and secondary education (CorpusMate). Empirical robustness is generally low, except for Çakır and Özer (2020). Most are descriptive, case-based, or perception-oriented without rigorous outcome testing. Therefore, the field requires more mixed-methods studies to assess performance outcomes, measure longitudinal perceptions, and compare CBI with other methods.

Taken collectively, the methodological characteristics of the included studies illustrate both the field's progress and its ongoing challenges. While diversity in research design and participant context enriches the evidence base, the predominance of descriptive and small-scale investigations limits the strength of conclusions about corpus tool effectiveness. To move the field forward, a sustained shift toward methodologically robust, comparative, and longitudinal research is needed, enabling clearer guidance for educators and policymakers.

In line with the third research question, the study included an analysis of the targeted skills in the five articles published on the popular corpus tools. Results are presented in Table 5. Overall, the studies utilised the tools for vocabulary, grammar, corpus literacy, genre awareness, academic writing, and translation.

Table 5. The Targeted Skills in the Analysed Studies

Study	Targeted Skills
Kilgarriff et al. (2015) <i>Sketch Engine & SkELL</i>	• Vocabulary acquisition (collocations, word use) • Grammar awareness (patterns) • Corpus literacy (reading concordances)
Albi (2019) <i>Sketch Engine</i>	• Translation competence (legal contexts) • Genre awareness (legal text conventions) • Terminology & phraseology development
Hirata & Hirata (2019) <i>SkELL</i>	• Vocabulary learning • Collocations • Sentence construction • Learner autonomy in resource use
Çakır & Özer (2020) <i>Fraze.it</i>	• Academic writing (sentence variety, cohesion, accuracy) • Vocabulary use in academic contexts • Intuitive competence (“feel” for correct usage)
Crosthwaite & Baisa (2024) <i>CorpusMate</i>	• Disciplinary vocabulary learning • Reading comprehension of academic discourse • Corpus consultation skills (for novice users)

Vocabulary acquisition and collocations were targeted in the analysed studies (e.g., Kilgarriff et al., 2015; Hirata & Hirata, 2019; Çakır & Özer, 2020; Crosthwaite & Baisa, 2024). Nation (2013) states that vocabulary knowledge may predict overall language proficiency. Accordingly, corpus tools help shift vocabulary learning from memorisation to pattern noticing in authentic contexts, aligning with Lewis’s (1993) lexical approach.

The focus on vocabulary and collocations also underscores the potential for corpus tools to bridge the gap between classroom learning and real-world language use. By immersing learners in authentic linguistic input, these tools facilitate deeper understanding of contextual meaning and foster



more nuanced communicative competence. This interpretive insight highlights the importance of integrating corpus resources into vocabulary pedagogy for more robust language development.

Grammar and sentence construction were emphasised in Kilgarriff et al. (2015) and Hirata and Hirata (2019). Leaving aside discussions on rule-focused or usage-based grammar teaching, corpora might provide inductive grammar learning opportunities and encourage learners to discover patterns themselves (Biber et al., 1998).

The emphasis on inductive grammar learning through corpora suggests a shift from traditional prescriptive methods to more discovery-oriented approaches. This reflective turn enables learners to engage more actively with language structure, promoting critical thinking and deeper grammatical awareness. Such pedagogical shifts may enhance long-term retention and flexible use of grammar in diverse contexts.

Translation competence and genre awareness were also cited in Albi (2019). Considering the genre-based approaches in the translation pedagogy (Göpferich et al., 2009), specialised corpora may play a significant role in developing legal and technical translation competence (Bowker & Pearson, 2002).

Beyond improving translation accuracy, the integration of specialised corpora supports the development of genre-specific knowledge and professional digital literacy. This interpretive perspective points to the evolving requirements of the translation industry, where familiarity with corpus tools is increasingly seen as a core competency. Embedding such tools in translation pedagogy may therefore better prepare learners for real-world demands

Çakır and Özer's (2020) study mentioned academic writing and intuitive competence. Learners must master disciplinary discourse conventions in academic writing (Hyland, 2009). Research reports a lack of intuitive competence among students, which corpora can support (Cera & Sinclair, 2024). Their study confirms that DDL empowers learners to refine writing style (Chen & Flowerdew, 2018).

Reflecting on academic writing outcomes, corpus-based approaches empower students to internalise disciplinary conventions and develop more sophisticated stylistic awareness. The move from intuition-based writing to evidence-informed practice may facilitate greater clarity, precision, and critical engagement in written work, supporting academic success and professional communication

Ultimately, corpus literacy and learner autonomy were targeted in all five studies, but explicitly in Kilgarriff et al. (2015), Hirata and Hirata (2019), and Crosthwaite and Baisa (2024). Students need corpus literacy (Boulton, 2010) to use corpus tools effectively. Kilgarriff et al.'s (2015) study supports the argument that tool design matters. This means that simplified interfaces improve adoption in diverse educational contexts (Zhao, 1998).

Reflecting on these targeted skills, the shift towards developing corpus literacy and learner autonomy signals a broader transformation in language education. Rather than focusing solely on discrete language competencies, corpus-based instruction fosters analytical and independent learning habits, equipping students to navigate authentic language data and adapt to evolving communicative demands. This approach may cultivate lifelong learning skills crucial for both academic and professional success

Major Findings and Pedagogical Implications

In accordance with the last research question, the major findings of each study were tackled. Table 6 summarised the important findings of each study.



Table 6. The Major Findings of the Analysed Studies

Study	Major Findings
Kilgarriff et al. (2015) – <i>Sketch Engine & SKELL</i>	• Corpus tools (SKELL, Sketch Engine) make authentic usage patterns accessible • Support vocabulary/grammar awareness and teacher syllabus design
Borja Albi & Anabel (2019) – <i>Sketch Engine</i>	• Specialised legal corpora improve translation competence, genre awareness, and terminology accuracy • Bridge theory–practice gap in translator training
Hirata & Hirata (2019) – <i>SkELL</i>	• Learners found SkELL user-friendly • Improved vocabulary and sentence construction • Positive attitudes despite limited digital literacy
Çakır & Özer (2020) – <i>Fraze.it in Turkish EAP</i>	• Fraise.it use led to significant gains in academic writing quality, vocabulary accuracy, and learner autonomy • Boosted motivation and confidence
Crosthwaite & Baisa (2024) – <i>CorpusMate</i>	• User-friendly interface enabled secondary learners to engage with disciplinary vocabulary • Teachers valued tool for supporting academic literacy • Increased accessibility of corpus pedagogy

Kilgarriff et al.'s (2015) study indicated the availability of accessible and learner-friendly corpus tools (e.g., Sketch Engine and SkELL). They also concluded that corpus tools help learners discover authentic patterns of usage. Additionally, it showed that corpus consultation supports teachers in designing syllabi based on actual learner needs. Accordingly, teachers can integrate corpus examples into classroom instruction to raise learner awareness of authentic usage. Corpora can support DDL and encourage learner autonomy (He & Xie, 2024; Lewandowska, 2014).

These findings, while descriptive, point to a deeper pedagogical shift enabled by corpus tools. The movement from teacher-centered to learner-centered approaches is evident in the way corpus tools encourage exploration, hypothesis testing, and independent learning. This suggests a redefinition of the teacher's role from sole knowledge provider to facilitator of discovery, which could have broader implications for curriculum design and teacher training in FLE.

Albi's (2019) study showed that corpus tools provide reliable access to legal terminology and phraseology. It also indicated that learners can use specialised corpora to improve accuracy and appropriateness in translation. These findings suggest that corpus-based resources can become central in specialised ESP/translation teaching (Boulton, 2012). They also highlight the importance of digital literacy in translation studies (Jianwen & Hao, 2025).

The implications here extend to the evolving demands of the translation profession, where translators are increasingly expected to leverage digital resources. The integration of corpus tools in translator education not only enhances terminology management and genre awareness but also develops essential digital competencies. This alignment with workplace realities may improve graduate employability and professional readiness.

In Hirata and Hirata's (2019) study, students with low digital literacy found SkELL easy to use. They generally held positive attitudes toward the tool, especially regarding vocabulary learning. These findings suggest that even low-proficiency learners can benefit from simplified corpus interfaces (Li et al., 2023; Yang, 2025). They also imply that corpus tools can reduce reliance on bilingual dictionaries and foster critical noticing of patterns (Friginal et al., 2020).



This result invites reflection on equity and access in corpus-based pedagogy. That low-proficiency and digitally inexperienced learners can successfully engage with corpus tools challenges assumptions that DDL is best suited for advanced or tech-savvy students. This could encourage more inclusive implementation, provided sufficient support and scaffolding are in place

In Çakır and Özer's (2020) study, the experimental group using Fraise.it showed significant improvement in EAP writing compared to the control. Learners also reported greater autonomy and confidence in academic writing.

Beyond this positive outcome, the findings suggest that corpus tools do not merely enhance language skills in isolation but may also foster transferable academic competencies such as self-regulation, critical thinking, and reflective practice. These broader educational benefits argue for the integration of DDL approaches in EAP curricula and potentially across other domains

CONCLUSION AND SUGGESTIONS

Given the significant place corpora and corpus tools hold in language learning and teaching, this SLR qualitatively and quantitatively examined the content of research articles on Sketch Engine, SkELL, PlayPhraseMe, Fraise.it, and CorpusMate published in Scopus and WoS-indexed journals until 2024. The search yielded five articles after the inclusion/exclusion criteria. Collectively, the five studies indicate the considerable potential of these corpus tools for improving English language education across diverse contexts. A consistent finding is that corpus consultation fosters learner autonomy, raises awareness of authentic usage patterns, and supports more accurate and context-sensitive language production.

However, the studies also share certain limitations. Many were conducted with relatively small sample sizes or in localized contexts (e.g., Japanese EFL classrooms, Turkish EAP writing courses, specialised legal translation training), which restricts the generalisability of the findings. In addition, some studies relied on learner perceptions and self-reports rather than long-term performance measures, making it difficult to assess the sustained impact of corpus-based pedagogy. Finally, specific issues (e.g., varying levels of digital literacy, the need for explicit training in corpus consultation, and limited integration of corpus tools into mainstream curricula) remain challenges for broader implementation. Addressing these limitations through larger-scale, longitudinal, and more systematically comparative studies would help solidify the role of corpora in language learning and teacher education.

REFERENCES

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open*, 9(1), 2158244019829575. <https://doi.org/10.1177/2158244019829575>
- Al-Hamzi, A. M. S., Gougui, A., Amalia, Y. S., & Suhardijanto, T. (2020). Corpus linguistics and corpus-based research and its implication in applied linguistics: A systematic review. *Parole: Journal of Linguistics and Education*, 10(2), 176-181. <https://doi.org/10.14710/parole.v10i2.176-181>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biel, Ł. (2017). Enhancing the communicative dimension of legal translation: comparable corpora in the research-informed classroom. *The Interpreter and Translator Trainer*, 11(4), 316-336. <https://doi.org/10.1080/1750399X.2017.1359761>
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–291). John Benjamins.



- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. <https://doi.org/10.1111/lang.12224>
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. Routledge.
- Çakır, İ., & Özer, M. (2020). Fostering intuitive competence in L2 for a better performance in EAP writing through fraze. it in a Turkish context. *Education and Information Technologies*, 25(6), 5405-5426. <https://doi.org/10.1007/s10639-020-10225-0>
- Cera, R., & Sinclair, M. (2024). What kind of intuition for what kind of education: A scoping review. *Encyclopaideia*, 28(70), 53-70. <https://doi.org/10.6092/issn.1825-8670/19650>
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52(4), 460-475. <https://doi.org/10.1017/S0261444819000089>
- Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335-369. <https://doi.org/10.1075/ijcl.16130.che>
- Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. *International Journal of Corpus Linguistics*, 29(4), 595-610. <https://doi.org/10.1075/ijcl.23056.cro>
- Crosthwaite, P., Ningrum, S., & Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research in arts and humanities. *International Journal of Corpus Linguistics*, 28(3), 344-377. <https://doi.org/10.1075/ijcl.21072.cro>
- Emir, G., & Saral, N. Ç. (2025). Exploring emerging issues in corpus applications in ELT: A systematic review. *Participatory Educational Research*, 12(2), 1-19. <https://doi.org/10.17275/per.25.16.12.2>
- Friginal, E., Dye, P., & Nolen, M. (2020). Corpus-based approaches in language teaching: Outcomes, observations, and teacher perspectives. *Bogazici University Journal of Education*, 37(1), 43-68. <https://izlik.org/JA78MB44XR>
- Göpferich, S., Alves, F., & Mees, I. M. (Eds.). (2009). *New approaches in translation process research* (Vol. 39). Samfundslitteratur.
- Gui, X., Deng, L., & Jiang, G. (2022). The use of corpus in English writing: Scholarly development and implications. *Journal of Education and Educational Research*, 1(1), 64-71. <https://doi.org/10.54097/jeer.v1i1.2476>
- He, M., & Xie, Q. (2024). Empowering autonomy in language learning: the sustainable impact of data-driven learning on noun collocation acquisition. *Humanities and Social Sciences Communications*, 11(1), 1-14. <https://doi.org/10.1057/s41599-024-04038-6>
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Huang, S., & Ma, Q. (2025). A systematic review of data-driven learning research on language learning and teaching for pre-tertiary learners: Balancing qualitative and quantitative research. *Learning and Individual Differences*, 122, 102752. <https://doi.org/10.1016/j.lindif.2025.102752>
- Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, 26(2), 163-183. <https://doi.org/10.1017/S0958344014000020>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hyland, K. (2009). *Academic discourse*. Continuum.
- Idrizi, E. (2024). Computer-assisted DDL vs. Mobile-assisted DDL: Exploring EFL learners' preferences and experiences with two corpus interfaces. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 13(3), 11-24. <https://doi.org/10.58885/ijllis.v13i3.11.ei>
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2014). Finding terms in corpora for many languages with the sketch engine. In S. Wintner, M. Tadić & B. Babych (Eds.), *Proceedings of the demonstrations at the 14th conference of the European chapter of the Association for Computational Linguistics* (pp. 53–56). Association for Computational Linguistics. <https://doi.org/10.3115/v1/E14-2014>



- Jianwen, L., & Hao, Y. (2025). enhancing translation students' digital literacy: Using a web-based parallel corpus as an example. In S.K.K. Ng, T.S. Bai, N.P.K. Lo, A.W.B. Tso, & W.W.L. Chan (Eds.), *Transcending Boundaries in the digital age* (pp. 83-99). Springer Nature Singapore.
- Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34. https://corpus.bfsu.edu.cn/Tim_Johns_From_printout_to_handout.pdf
- Kilgariff, A., Marcowitz, F., Smith, S., & Thomas, J. (2015). Corpora and language learning with the Sketch Engine and SKELL. *Revue française de linguistique appliquée*, 1, 61-80. <https://doi.org/10.3917/rfla.201.0061>
- Kolaski, K., Romeiser Logan, L., & Ioannidis, J. P. (2023). Guidance to best tools and practices for systematic reviews. *Journal of Pediatric Rehabilitation Medicine*, 16(2), 241-273. <https://doi.org/10.3233/PRM-230019>
- Kontsioti, V., & Karasimos, A. (2025). Teaching collocations to teenagers through a fantasy-adventure movies/series subtitle corpus: A soft DDL approach. *Research Papers in Language Teaching and Learning*, 15(1), 117-134. <https://rpltl.eap.gr/images/2025/RPLTL15.pdf>
- Kunilovskaya, M., & Koviagina, M. (2017). Sketch engine: A toolbox for linguistic discovery. *Jazykovedny Casopis*, 68(3), 503-507. <https://doi.org/10.2478/jazcas-2018-0006>
- Kwiek, M. (2021). What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. *Studies in Higher Education*, 46(12), 2629-2649. <https://doi.org/10.1080/03075079.2020.1749254>
- Lewandowska, A. (2014). Using corpus-based classroom activities to enhance learner autonomy. *Konińskie Studia Językowe*, 2(3), 237-255. https://www.ksj.konin.edu.pl/wp-content/uploads/2015/05/KSJ-2_3_237-255.pdf
- Li, D., Noordin, N., Ismail, L., & Cao, D. (2025). A systematic review of corpus-based instruction in EFL classroom. *Heliyon*, 11(2), e42016. <https://doi.org/10.1016/j.heliyon.2025.e42016>
- Li, M., Zhang, X., & Reynolds, B. L. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: an ETS corpus study. *Applied Linguistics Review*, 14(4), 847-873. <https://doi.org/10.1515/applirev-2020-0129>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9(12), e22731. <https://doi.org/10.1016/j.heliyon.2023.e22731>
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177-190. <https://doi.org/10.1515/jccall-2021-2008>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 364-380). Routledge.
- Moon, S., & Oh, S. Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48-67. <https://doi.org/10.1017/S0958344017000246>
- Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18, 1-7. <https://doi.org/10.1186/s12874-018-0611-x>
- Nguyen, T. V., Ho-Le, T. P., & Le, U. V. (2017). International collaboration in scientific research in Vietnam: an analysis of patterns and impact. *Scientometrics*, 110(2), 1035-1051. <https://doi.org/10.1007/s11192-016-2201-1>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ...& Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal*, 372 (71). <https://doi.org/10.1136/bmj.n71>



- Park, H., & Nam, D. (2017). Corpus linguistics research trends from 1997 to 2016: A co-citation analysis. *Linguistic Research*, 34(3), 427-457. <https://doi.org/10.17250/khisli.34.3.201712.008>
- Patino, C. M., & Ferreira, J. C. (2018). Internal and external validity: can you apply research study results to your patients?. *Jornal Brasileiro de Pneumologia*, 44(03), 183-183. <https://doi.org/10.1590/S1806-37562018000000164>
- Pearce, M. (2008). Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora*, 3(1), 1-29. <https://doi.org/10.3366/E174950320800004X>
- Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning*, 35(1-2), 36-61. <https://doi.org/10.1080/09588221.2019.1667832>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell.
- Reppen, R. (2011). *Using corpora in the language classroom*. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 35-50). Cambridge University Press.
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* [Studies in Corpus Linguistics 33]. (pp. 83–98). John Benjamins.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. <https://doi.org/10.1017/S0267190511000055>
- Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, 2(4), 320-332. <https://doi.org/10.1504/IJBHR.2011.043414>
- Salimi, E. A., & Marami Hajikandi, A. (2025). Publishers and authorial voice in applied linguistics research articles: A corpus-based study. *Journal of Scholarly Publishing*, 56(1), 1-33. <https://doi.org/10.3138/jsp-2023-0002>
- Scelles, N., & Teixeira da Silva, J. A. (2025). Making the impact of publications within a field comparable by improving the field-weighted citation impact (FWCI): the case of sport management. *Scientometrics*, 130(3), 1571-1586. <https://doi.org/10.1007/s11192-025-05268-4>
- Selvi, A. F. (2020). Qualitative content analysis. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 440–452). Routledge.
- Sinclair, J. (1991). *Corpus concordance collocation*. Oxford University Press.
- Sozibilir, M., Kutu, H., & Yasar, M. D. (2012). Science education research in Turkey: A content analysis of selected features of published papers. In D. Jorde & J. Dillon (Eds.), *Science education research and practice in Europe: Retrospective and prospective* (pp. 341-374). Sense Publishers.
- Thomas, G. (2011). The case: Generalisation, theory and phronesis in case study. *Oxford Review of Education*, 37(1), 21-35. <https://doi.org/10.1080/03054985.2010.521622>
- Tong, J., Che Abdul Rahman, A. N., & Hamat, A. (2025). A systematic literature review on word selection criteria in corpus-based wordlists development. *Cogent Arts & Humanities*, 12(1), 2432131. <https://doi.org/10.1080/23311983.2024.2432131>
- Topal, I. H. (2025a). A critical review of SkELL (Sketch Engine for language learning)–V1. 11.8. *Technology in Language Teaching & Learning*, 7(1), 102690. <https://doi.org/10.29140/tltl.v7n1.102690>
- Topal, I. H. (2025b). PlayPhrase. Me: A multimedia corpus for foreign language education. *Technology in Language Teaching & Learning*, 7(1), 102405. <https://doi.org/10.29140/tltl.v7n1.102405>
- Topal, I. H. (2025c). Review of Fraise. it: a multilingual sentence concordancer. *Journal of China Computer-Assisted Language Learning*. <https://doi.org/10.1515/jccall-2025-0008>
- Troy, H. R. (2022). Learning vocabulary with SkELL: Developing a methodology with university students in Japan using action research. *Bulletin of Nagoya University of Foreign Studies*, 10, 163–183. <http://id.ndl.go.jp/bib/032107918>
- Veiga Norlander, C. (2022). *L2 vocabulary acquisition: An investigation into the effectiveness of playphrase.me as a tool for learning English vocabulary for Swedish level 9 students*



- (Unpublished Master's thesis). Mid Sweden University, Sweden. <https://miun.diva-portal.org/smash/get/diva2:1708176/FULLTEXT01.pdf>
- Wang, D., & Liang, L. (2024). Mapping academic collaboration patterns of scholarly publications in fifteen linguistics journals from 2014 to 2023 via social network analysis. *Journal of Scholarly Publishing*, 55(4), 524-558. <https://doi.org/10.3138/jsp-2023-0094>
- Yağız, O., Aydın, B., & Akdemir, A. S. (2016). ELT research in Turkey: A content analysis of selected features of published articles. *Journal of Language and Linguistic Studies*, 12(2), 117-134. <https://izlik.org/JA37EE54WA>
- Yang, S. (2025). Examining the effectiveness of Sketch Engine for verb-noun collocation learning among Chinese low-proficiency learners. *Arab World English Journal (AWEJ)*, 16(2), 233-261. <https://dx.doi.org/10.24093/awej/vol16no2.14>
- Zakaria, A., & Aryadoust, V. (2024). A scientometric analysis of applied linguistics research (1970–2022): Methodology and future directions. *Applied Linguistics Review*, 15(4), 1595-1619. <https://doi.org/10.1515/applirev-2022-0210>
- Zhao, Y. (1998). Design for adoption: The development of an integrated web-based education environment. *Journal of Research on Computing in Education*, 30(3), 307-328. <https://doi.org/10.1080/08886504.1998.10782229>
- Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321-335. <https://doi.org/10.1007/s11192-020-03387-8>

