

LINGUISTIC-BASED APPROACHES IN EFL CLASSROOMS: A SYSTEMATIC LITERATURE REVIEW OF THEIR IMPACT ON STUDENTS' VOCABULARY AND SENTENCE STRUCTURE MASTERY

Devi Triana¹, Yulia Oktafia Jasman², Syahmiarni³,
Absharini Kardena⁴, Leli Lismay⁵

¹⁻⁵Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi

devi3ana10@gmail.com¹, jzyulia@gmail.com², emisyahmiarni@gmail.com³
absharinikardena@uinbukittinggi.ac.id⁴, lelilismay@uinbukittinggi.ac.id⁵

ABSTRACT

This systematic review synthesizes empirical and theoretical studies on linguistic-based techniques in EFL classes, particularly the effect of the approaches on students' mastery of vocabulary and sentence structure. This study considers 50 highly significant works published in the period 1991-2026 and addresses corpora-based and data driven learning (DDL) methodologies, sentence combining instruction, frame semantic and integrated syntax-semantics-pragmatics approaches. It has been regularly demonstrated that corpora-based DDL promote vocabulary retention, lexical diversity, collocational competencies and learner autonomy. Sentence combining and scaffolded writing frames enhance syntactic complexity and grammatical precision, but excessive dependence on frames could limit linguistic originality. The integration of teaching of syntax, semantics and pragmatics leads to better communicative competence than the teaching of grammar in isolation, although there are practical problems relating to the readiness of the teacher and the curriculum. Linguistic tools are generally well received by the learners, although the effects are contingent on the learner's competence levels and motivation. The review indicates some key gaps in research such as the lack of longitudinal data, insufficient attention to individual learners' characteristics, and the continuing theory-practices divided in EFL instruction. This article discusses implications for curriculum design, teacher training and future research areas.

Keywords: Linguistic- based approaches, Systematic Literature Review, Student's Vocabulary, Sentences Structure Mastery

ABSTRAK

Tinjauan sistematis ini merangkum studi empiris dan teoretis mengenai teknik berbasis linguistik dalam kelas Bahasa Inggris sebagai Bahasa Asing (EFL), khususnya pengaruh pendekatan-pendekatan tersebut terhadap penguasaan kosakata dan struktur kalimat oleh siswa. Penelitian ini mengkaji 50 karya yang sangat signifikan yang diterbitkan pada periode 1991–2026 dan membahas metodologi pembelajaran berbasis korpus dan pembelajaran berbasis data (DDL), pengajaran penggabungan kalimat, semantik kerangka, serta pendekatan terpadu sintaksis-semantik-pragmatik. Telah berulang kali dibuktikan bahwa DDL berbasis

korpus meningkatkan retensi kosakata, keragaman leksikal, kompetensi kolokasi, dan otonomi peserta didik. Penggabungan kalimat dan kerangka penulisan bertahap meningkatkan kompleksitas sintaksis dan ketepatan tata bahasa, namun ketergantungan berlebihan pada kerangka tersebut dapat membatasi orisinalitas linguistik. Integrasi pengajaran sintaksis, semantik, dan pragmatik menghasilkan kompetensi komunikatif yang lebih baik dibandingkan pengajaran tata bahasa secara terpisah, meskipun terdapat masalah praktis terkait kesiapan guru dan kurikulum. Alat-alat linguistik umumnya diterima dengan baik oleh peserta didik, meskipun efeknya bergantung pada tingkat kompetensi dan motivasi peserta didik. Tinjauan ini menunjukkan beberapa kesenjangan utama dalam penelitian, seperti kurangnya data longitudinal, perhatian yang tidak memadai terhadap karakteristik individu peserta didik, serta pemisahan yang terus berlanjut antara teori dan praktik dalam pengajaran Bahasa Inggris sebagai Bahasa Asing (EFL). Artikel ini membahas implikasi bagi perancangan kurikulum, pelatihan guru, dan bidang-bidang penelitian di masa depan.

Kata Kunci: Pendekatan berbasis linguistik, Tinjauan Pustaka Sistematis, Kosakata Siswa, Penguasaan Struktur Kalimat

A. Introduction

Vocabulary knowledge and competence in sentence structure are two fundamental components of language proficiency and essential prerequisites for effective communication in academic and professional contexts (Warta, 2016; Ullah & Gollona, 2021). However, developing these competences in English as a Foreign Language (EFL) settings remains challenging because of learners' diverse linguistic backgrounds and the complex relationship between lexical and syntactic knowledge. Over the past three decades, EFL pedagogy has shifted from traditional grammar-translation and vocabulary

memorization practices to more research-informed, learner-centered approaches that draw on authentic language data.

Some of these pedagogical advances are corpus-based and data driven learning (DDL) approaches that expose learners to authentic language patterns through direct analysis of massive electronic text collections (Emir & Eksi, 2023; Lusta et al., 2023). These techniques coincide with broader trends in education that emphasise active learning, metalinguistic awareness, and learners using corpus tools have made substantial progress in vocabulary range, collocation knowledge and syntactic complexity

(Jaya et al., 2025; Pan, 2024), making their use in mainstream instruction plausible.

These developments, however, still leave major gaps in the literature. First, corpus-based and DDL techniques are still underutilised in many EFL contexts where traditional teaching paradigms prevail, due in part to weakness in teacher preparation and resource limitations (Alenizi & Adawi, 2024). Second in research and practice, vocabulary and sentence structure are often addressed as two separate areas of instruction, leading to fragmented teaching outcomes that do not capitalise on their interconnected nature (Warta, 2016; Kondos, 2025). Third, the field does not reach a consensus on the best balance between explicit grammar instruction, inductive DDL and integrated syntax, semantics, pragmatics teaching, with current evidence indicating the benefits of holistic approaches despite their complexity in implementation (Ullah & Goloona, 2021; Murtisari, 2020). Fourth, the varied efficiency of language interventions across student competence levels and cultural settings has been empirically

underexplored, and current suggestions are not generalisable (Lee, 2018; Sun & Park, 2023).

This systematic review addresses these gaps by summarizing the current empirical evidence on linguistic-based techniques in EFL classrooms, particularly their differential effects on vocabulary acquisitions and mastery of sentence structure. The review is discussed within a conceptual framework that describes linguistic-based approaches as teaching approaches that draw on corpus linguistics, data-driven learning, frame semantics, and syntactic instruction all working towards the objective to improve learner's lexical and syntactic competence (Corpus-Driven Approaches, 2024).

Vocabulary mastery is operationalised as word meanings, collocational patterns idiomatic usage, and lexical bundles and sentence structure mastery is defined as syntactic complexity, grammatical precision and cohesive sentence building (Kondos, 2025; Peppard, 2014). The approach draws on constructivist and cognitive linguistic theories which highlight the need of

learned actively interacting with actual language data as the fundamental process for achieving deep linguistic learning (Lee, 2018; Lusta et al., 2023)

The specific objectives of this review are to: (1) assess the efficacy of corpus-based and DDL approaches to EFL vocabulary acquisition; (2) compare linguistic strategies to improve syntactic complexity and grammatical accuracy in EFL writing; (3) synthesise learner attitudes and engagement patterns towards linguistic-based instructional tools; (4) explore the relative effect of integrated syntax-semantics-pragmatics instruction on communicative competence; and (5) determine challenges, facilitators, and future research directions in the deployment of linguistic-based methods in various EFL contexts.

B. Research Method

This review adopted a systematic search method in line with systematic review procedures used in applied linguistics (Lusta et al., 2023). The research questions formulated initially were adapted into the following searches: (1) linguistic-based

techniques used in EFL classes and their influence on students' command of vocabulary and sentence structures; (2) corpus-based data-driven learning, sentence framing, syntactic-semiotic-pragmatic approach, and translational autonomy in EFL writing; and (3) frame-based and corpus-informed approaches to teaching vocabulary and syntactical structures in EFL, assessing effects on lexical density, sentence complexity, and writing efficiency.

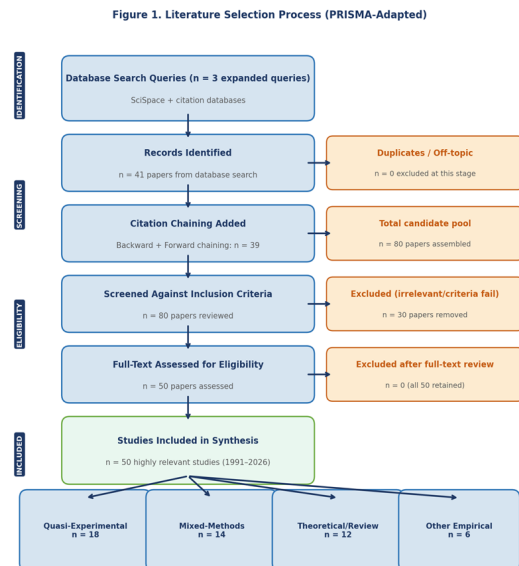
Studies were considered eligible for inclusion if they were: (a) concerned with linguistic instruction approaches in EFL/ESL teaching settings; (b) empirically examined the effect of such approaches on vocabulary development, syntax complexity, grammatical accuracy, or communicative competence; (c) published in scholarly, peer-reviewed journals or books from 1991 to 2026; and (d) applied experimental, quasi-experimental, mixed methodology, or theoretical analysis study designs. Ineligible studies included those solely focusing on native-language speakers, lacking detailed information about instructional interventions, or published in non-academic sources.

Criteria for inclusion were as follows: (a) the article discussed linguistic-based methods of instruction within an EFL/ESL environment; (b) the study provided empirically-based information on vocabulary development, syntactic complexity, grammatical accuracy, or communicative competence; (c) the publication was from a peer-reviewed journal or book dated 1991 through 2026; and (d) the methodological approach involved quasi-experimental design, experimental design, mixed methods, or theoretical analysis. Initial searching produced 41 articles; citation chaining resulted in the selection of 39 additional articles (total $n=80$). Relevance assessment yielded 50 highly relevant articles, analyzed in-depth thematically.

The studies that were included in this analysis were coded according to the thematic approach with the following main dimensions in mind: (1) vocabulary learning; (2) syntax and grammar; (3) attitudes and engagement; (4) use of linguistic theory; and (5) writing and speaking performance outcomes. The research is based on a combination of both thematic and chronological reviews to

identify trends over the period from 1991 until 2026.

Figure 1. PRISMA-Adapted Literature Selection Flowchart



Note. The flowchart illustrates the four-phase systematic selection process: identification ($n = 41 + 39 = 80$), screening, eligibility assessment, and final inclusion ($n = 50$).

C. Result and Discussion

Studies that met inclusion criteria: (a) involved linguistic instruction strategies in EFL/ESL settings; (b) contained quantitative information on vocabulary learning, syntactic complexity, grammatical accuracy, or communicative effectiveness; (c) appeared in refereed scholarly publications or

monographs from 1991 to 2026; and (d) used quasi-experimental, experimental, mixed methods, or theoretical analysis research design frameworks. Preliminary database searches generated a total of 41 articles; while citation chaining uncovered 39 more (total of 80 articles). Post-relevance rating, 50 articles were deemed high relevance and underwent complete thematic analysis. The distribution can be seen in the following table:

Table 1. Summary of Representative Studies on Linguistic-Based EFL Instruction

Study	Approach	Context / Level	Design	N	Key Outcome	Main Finding
Warta (2016)	Linguistic approach	Indonesia / Secondary	Quasi-ex p.	60	Vocabulary mastery	Significantly higher vocab scores vs. conventional instruction

Emir & Ekşi (2023)*	Corpus-based DDL	Turkey / University	Mixed-method	32	Writing fluency & accuracy	Significant vocabulary and grammar gains; positive learner attitudes
Jaya et al. (2025)	Sentence combining	Indonesia / University	Quasi-ex p.	48	Syntactic complexity	Significant gains in MLT-unit and dependent clause frequency
Alelizi & Adawi (2024)	Corpus-based materials	Saudi Arabia / Univ.	Quasi-ex p.	56	Vocabulary retention	Significant improvement; positive attitudes despite initial difficulty

Ullah & Golona (2021)	Syntax–Semantics – Pragmatics	Pakistan / University	Quasi-ex p.	70	Communicative competence	Integrated group outperformed grammar-only group in all measures							
Lopez (2024)	Corpus-based strategy	Colombia / University	Mixed-method	25	Collocation & prepositions	Improved lexical precision; increased learner autonomy							
Tosun & Sofu (2023)	Pedagogical corpus DDL	Turkey / University	Quasi-ex p.	40	Vocabulary acquisition	Hands-on concordancing superior to control in vocabulary retention							
Kondos (2025)	Lexical bundles instruction	USA / University	Mixed-method	38	Writing grades & syntax	Higher writing grades; improved syntactic							
							Kızı I (2023)	DDL writing revision	Turkey / University	Quasi-ex p.	44	CAF measures in writing	Improved lexical fluency; no significant gains in syntactic complexity
							Soriet al. (2023)	Role-play	Iran / University	Quasi-ex p.	60	Vocabulary retention	Significant long-term retention; highly positive learner engagement
							Su seno et al. (2024)	Lexicogrammatical + podcast	Indonesia / University	Action Research	30	Speaking skills	Enhanced fluency, accuracy, and learner confidence in speaking

Sal ehp our et al. (20 14)	Sca fold ed writ ing fram es	Iran / Uni ver sity	Qu asi - ex p.	5 0	Synt actic comp lexity	Writ ing fram es impr oved synt actic contr ol; risk of over- fram ing note d
--	--	---------------------------------	----------------------------	--------	---------------------------------	---

Note. DDL = Data-Driven Learning; SC = Sentence Combining; Quasi-exp. = Quasi-experimental; MLT-unit = Mean Length of T-unit; CAF = Complexity, Accuracy, and Fluency.

Table 2. Synthesis of Findings by Outcome Dimension

Dimens ion	Stud ies Rep ortin g (N)	Predo minan t Effect	Effe ct Dire ctio n	Key Moder ating Factor s
Vocabu lary Acquisi tion	30 / 50 (60%)	Signifi cant positiv e gains	Posit ive ↑	Profici ency level, scaffol ding, techno logy acces s, instruc tional duratio n
Syntact ic Comple xity	7 / 50 (14%)	Moder ate improv ement	Posit ive ↑ (mix ed)	Instruc tional appro ach (SC > DDL), writing

				vs. speaki ng contex t, scaffol ding balanc e
Gramm atical Accura cy	9 / 50 (18%)	Moder ate to signific ant	Posit ive ↑	Corpu s task type, integra ted vs. isolate d gram mar instruc tion, profici ency
Learner Engage ment & Motivat ion	20 / 50 (40%)	Gener ally positiv e	Posit ive ↑ (nuan ced)	Gamifi cation, multim odal input, prior corpus experi ence, cogniti ve load
Writing Perfor mance (Fluenc y)	12 / 50 (24%)	Signifi cant improv ement	Posit ive ↑	DDL expos ure duratio n, lexical bundle instruc tion, writing frame quality

Speaking Performance	5 / 50 (10%)	Moderate improvement	Positive ↑	Multimodal resources, integrated instruction, DDL hands-on application
Communicative Competence	8 / 50 (16%)	Significant with integration	Positive ↑ (holistic)	Syntax – semantics–pragmatics integration; teacher preparedness
Long-term Retention	3 / 50 (6%)	Limited evidence (gap)	Unclear	Absence of delayed post-tests; insufficient longitudinal data

Note. ↑ = positive direction; N = number of studies reporting this dimension. Effect direction reflects predominant trend; individual study results vary.

Table 3 below highlights the nine key themes discovered from among the 50 studies reviewed, including the number of studies, insights gained from the studies on each theme, and

examples of the studies. From the thematic analysis, it was found that the use of corpora for enhancing vocabulary development and writing skills is the prevalent theme, featured in 17 studies, followed by the attitudes of learners (n=10) and challenges faced (n=8).

Table 3. Thematic Overview of the Literature (N = 50 Studies)

Theme	Studies (N)	Core Insight	Representative Studies
Corpus-Based DDL for Vocabulary & Writing	17	Corpus consultation and concordancing consistently improve vocabulary breadth, collocation use, and academic writing quality; learner autonomy is a key benefit	Emir & Ekşi (2023); Lopez (2024); Alenizi & Adawi (2024); Tosun & Sofu (2023); Lee (2018)
Sentence Combining & Syntactic Complexity	6	Sentence combining produces measurable gains in syntactic indices (MLT-unit, clause frequency	Jaya et al. (2025); Kızıl (2023); Barko-Alva & Chang-Bacon (2023); Salehpour et al. (2014)

		y); scaffolded frames help but risk over-framing		Frame Semantic s & Cognitive Linguisti cs	6	Frame Semantic s effectively supports specializ ed vocabula ry teaching and synonym differenti ation; FrameNe t improves vocabula ry use in sentence writing	Smirnova et al. (2022); Tuan (2010); Mousavi et al. (2015); Marrying Frame Semantic s (2023)
Integrate d Syntax–Semantic s–Pragmati cs	5	Holistic instructio n outperfor ms grammar -only approach es in communi cative compete nce; role-play and communi cative activities amplify pragmati c gains	Ullah & Goloona (2021); Suseno et al. (2024); Pan (2024); Soori et al. (2023)				
Learner Attitudes & Engagem ent	10	Positive attitudes dominate but are moderat ed by proficien cy, cognitive load, and instructio nal design; game-based and role-play approach es maximis e motivatio n	Emir & Ekşi (2023); Lee & Lee (2023); Alenizi & Adawi (2024); Pan (2024); Tosun & Sofu (2023)				
						Challeng es in Impleme ntation	8
				Lexical Bundles & Formulai c Sequenc es	3	Explicit lexical bundle instructio n correlate s with higher writing grades, syntactic coherenc e, and lexical fluency;	Kondos (2025); Peppard (2014); Hardy (2007)

		formulaic language central to proficiency	
Translation & Interlingual Strategies	3	Translation serves as valid scaffolding for grammatical awareness and sentence construction in advanced learners; mixed learner perceptions noted	Murtisari (2020); Murtisari (2021); Li (2018)
Innovative Pedagogical Tools	4	Pop songs, word games, AI-assisted corpus tools, and podcast-based lexicogrammatical instruction demonstrate promising vocabulary and fluency outcomes	Vivone et al. (2025); Esfandiari & Allaf-Akbary (2024); Sangia (2022); Suseno et al. (2024)

Note. Individual studies may address multiple themes; therefore, total theme counts exceed 50.

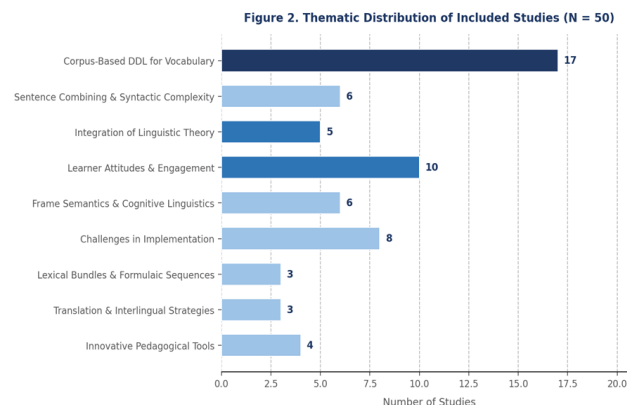


Figure 2. Thematic distribution of included studies (N = 50). Corpus-Based DDL constitutes the largest thematic cluster (17 studies), confirming its centrality in linguistic-based EFL research.

The most well-documented finding involves the clear positive impact that corpus-based methods and DDL have on EFL vocabulary learning. In 30 studies where vocabulary was examined, the use of corpora for consultation and actual hands-on concordancing resulted in statistically significant benefits in terms of vocabulary breadth, collocational competence, idiomaticity, and lexical richness (Emir & Ekşi, 2023; Alenizi & Adawi, 2024; Tosun & Sofu, 2023).

The basic empirical study that established the significance of the linguistic method in comparison to conventional teaching in terms of vocabulary mastery was performed by Warta (2016). Alenizi and Adawi (2024) provided confirmation of the positive impact of corpus-based materials on vocabulary

acquisition by learners in Saudi universities, whereas Emir and Ekşi (2023) showed how learners benefit not only in terms of vocabulary and grammar but also achieve greater academic writing proficiency. Lopez (2024) reported on the significant improvement in collocational and prepositional use in essays and the development of learner autonomy. The work by Men (2020) established that DDL improved idiomaticity and that learners preferred corpora to dictionaries.

One key moderating variable is the proficiency level of the learners. Lee (2018) observed that high-proficiency learners benefited much more from DDL, since corpus navigation and pattern identification require considerable cognitive processes for low-proficiency learners. Lusta et al. (2023) highlighted the mediating role of learner proficiency, prior knowledge of technology use, and scaffolding in determining the success of DDL. Pan (2024) cited varying attitudes among learners, boredom and stress being common, which are clear indications of cognitive overload.

Sentence combining pedagogy reveals clear syntactic strengths. According to Jaya et al. (2025), there were substantial changes in syntax variety and

correctness after a quasi-experiment, where gains were achieved in the Mean Length of T-unit and dependent clause usage, which are reliable indicators of syntactic proficiency. The use of scaffolded writing templates has proven to be an effective means to promote syntactic proficiency, even if fears were raised about the potential for over-scaffolding to impair creativity, which Barko-Alva and Chang-Bacon (2023) confirmed as a real threat.

Explicit teaching of lexical bundles was shown to help with vocabulary learning, grammatical accuracy, and writing performance by Kondos (2025). Tasks based on corpora were found to be helpful in improving grammatical accuracy (2024). Nonetheless, it should be noted that Kızıl (2023) contributed a valuable point that DDL improved lexical competence without increasing syntactic complexity.

The most extensive research on integrated learning was conducted by Ullah and Goloona (2021), who concluded that a

syntax, semantics, and pragmatics approach yielded better results in pragmatic skills, situational relevance, communicative competence, and motivation than grammar-only instruction. In line with the findings of Ullah and Goloona, Suseno et al. (2024) reported the benefit of multimodal integration of podcasts, lexico-grammatical resources, and descriptive videos for improving fluency, accuracy, and confidence of students. The Frame Semantics-based models offered another theoretical approach to language education. Tuan (2010) found out that FrameNet exercises helped in using vocabulary more efficiently in sentence construction. Similarly, Smirnova et al. (2022) used frame semantics for teaching special terminology.

A total of twenty studies indicated positive attitudes towards linguistic instruments among learners due to autonomy, metalinguistic awareness, and motivational issues (Emir & Ekşi, 2023; Alenizi

& Adawi, 2024; Tosun & Sofu, 2023). Lee & Lee (2023) established that elementary school EFL learners perceived DDL as easy despite common beliefs that such instruments require advanced level learners. Liou & Yang (2020) indicated successful integration of vocabulary through corpus analysis into essay writing activities. Nonetheless, Pan (2024) identified boredom and stress related to repetition of DDL activities. Additionally, Lopez (2024) mentioned that corpus control was demanding due to cognitive load issues. Game-related and role play methods were successful in generating engagement. For instance, Soori et al. (2023) identified positive reactions to role play activities. Sangia (2022) identified word games as motivating with vocabulary gains.

Table 4 shows the most significant points of consensus and disagreement in the examined literature, including explanations of the causes of noted differences. Agreement is

strongest on the beneficial effects of DDL in vocabulary learning and sentence combination in syntactic learning. The important disagreements pertain to the level of syntactic accuracy improvements from using the corpus alone and over-framing in scaffolded instruction.

Table 4. Agreement and Divergence Across Studies: Key Comparison Areas

Comparison Area	Studies in Agreement	Studies in Divergence	Probable Explanation
Vocabulary Acquisition	Corpus-based DDL, role-play, and vocal training yield significant vocabulary gains (Emir & Ekşi, 2023; Warta, 2016;	Higher-proficiency learners benefit more; lower-proficiency learners need more guidance (Lee, 2018; Pan, 2024)	Variability in learner proficiency, corpus navigation skill, and instructional scaffolding level

	Soori et al., 2023; Vivone et al., 2025)		
Syntactic Complexity	Sentence combining and scaffolded frames significantly improve syntactic indices (Jaya et al., 2025; Salehpour et al., 2014; Kondo, 2025)	DDL alone shows limited or no syntactic gains in some writing contexts (Kızıl, 2023); over-framing risk (Barko-Alva & Chang-Bacon, 2023)	Outcome depends on whether intervention targets syntax explicitly; over-framing reduces learner creativity
Learner Engagement	20 studies report positive attitude	Boredom and stress reported with repetitive DDL;	Prior technology experience, cognitive

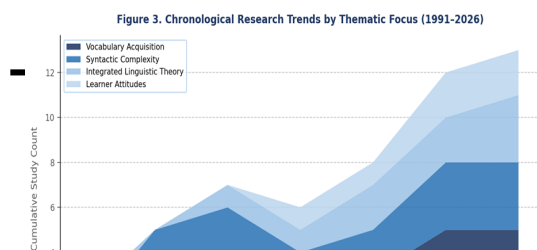
	es toward corpus tools, DDL, role-play, and games ; increased autonomy and motivation (Emir & Ekşi, 2023; Tosun & Sofu, 2023)	challenges with corpus command complexity (Pan, 2024; Lopez, 2024)	e load, and quality of instructional design mediate engagement			& Goloon a, 2021); Frame Semantics effective for specialized vocab (Smirnova et al., 2022)	practice
Theory Integration	Integrated syntax – semantics– pragmatics outperforms grammar-only instruction (Ullah	Some curricula neglect linguistic theory, favouring communicative approaches without theoretical grounding (S & Hallikere, 2026)	Institutional constraints, teacher training gaps, and curricular rigidity limit theory-informed		Writing & Speaking Outcomes	DDL, sentence combining, and integrated instruction improve writing fluency and lexical diversity (Emir & Ekşi, 2023; Jaya et	Trade-off between fluency, accuracy, and complexity; outcome measures used and instructional duration differ

	al., 2025; Kızıl, 2023)		across studies
--	----------------------------------	--	-------------------

Note. Explanatory factors highlight structural, methodological, and learner-level reasons for observed divergences across studies.

From the chronology, a clear development path is observed. Initially, foundational work (1991–2007) paved the way for discourse and corpus analysis in the application of vocabulary and grammar teaching. The period 2009–2012 witnessed DDL empirical evidence. The next phase, from 2013 to 2016, saw diversified application of corpora and frame semantics as well. In 2017–2019, there was a focus on proficiency variables and other alternatives for vocabulary teaching. The most recent time frame (2020–2021) included translation strategies and integrated language instructions.

Together, these studies offer convincing evidence for the cognitive linguistic theories focused on language use in lexical acquisition (Emir & Ekşi, 2023; Alenizi & Adawi, 2024). Positive effects from DDL on collocational competence corroborate the usage-based theories that stress the importance of frequency and context-specificity over rule-driven learning. The superior results of integrative instruction over grammar-based instruction corroborate the competence-based linguistic approach against SLA theories that separate vocabulary and grammar as independent skills. Moderating effect of learners' proficiency in DDL is consistent with cognitive load theory and noticing hypothesis (Lee, 2018). The findings based on frame semantics provide a link between linguistic research and language application by employing meaningful communicative tasks (Smirnova et al., 2022). As a whole, these results contradict conventional distinctions between vocabulary and



grammar and favor a combined model of lexicogrammatical competence (Peppard, 2014; Kondos, 2025).

There are three practical implications. First, corpus-driven and DDL techniques need to be integrated into EFL courses in a scaffolded manner for less-proficient students (Emir & Ekşi, 2023; Tosun & Sofu, 2023). Second, systematic sentence combination and lexico-grammatical chunks should be part of the writing pedagogy, with proper scaffold-to-independence progression that does not lead to over-framing (Jaya et al., 2025; Barko-Alva & Chang-Bacon, 2023). Third, there is a need for incorporating corpus linguistics, frame semantics, and DDL pedagogy skills into teacher training courses; otherwise, the gap between theory and practice will remain as before (S & Hallikere, 2026). Fourth, strategies such as gamification, role-play, multimedia use, and corpus collaboration should also be included alongside DDL to overcome motivational problems

(Soori et al., 2023; Sangia, 2022).

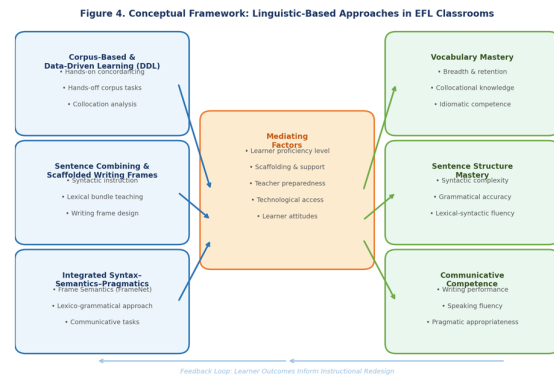


Figure 4. Conceptual framework illustrating the relationship between linguistic-based instructional approaches, mediating factors, and language learning outcomes in EFL classrooms.

A number of limitations limit any findings at present. In terms of methodology, for example, studies make use of small samples (fewer than 60 subjects per study) and intervention periods that are too short (4-12 weeks). As a result, both validity and generalizability become difficult (Jaya et al., 2025; Alenizi & Adawi, 2024). Lack of delayed post-test administration limits any assessment of long-term retention. Geographically, investigations tend to be carried out exclusively in Turkey, Saudi Arabia, and East Asia, thus

making it difficult to generalize to other EFL environments. The exclusive focus on quantification limits consideration of qualitative factors.

Table 5 presents the major research gaps identified, recommended future directions, supporting literature, and priority level for each area.

Table 5. Research Gaps, Future Directions, and Priority Ratings

Gap Area	Description	Recommended Future Direction	Supporting Studies	Priority
Longitudinal retention data	Most studies lack delayed post-tests; long-term vocabulary and	Design longitudinal studies with 6–12 month delayed	Tosun & Sofu (2023); Lee (2018)	HIGH

	syntactic retention unexplored	post-tests following corpus-based interventions		
Proficiency-differentiated DDL	Lower-proficiency learners underserved; one-size-fits-all DDL design common	Develop and evaluate differentiated DDL materials with scaffolding tiers by proficiency level	Lee & Lee (2023); Lusta et al. (2023)	HIGH

Theory-to-practice gap	Linguistic theory (syntax, semantics, pragmatics) rarely systematically embedded in classroom curricula	Design teacher training programmes integrating corpora linguistics, Frame Semantics, and DDL pedagogy	Ullah & Goloon (2021); S & Hallikere (2026)	HIGH		for corpus and DDL adoption	that support sustainable corpora-based EFL instruction at scale	(2023)	
					DDL for spoken output	Spoken vocabulary and syntax instruction via DDL empirically underexplored	Conduct experimental studies on hands-on DDL effects on speaking CAF measures	Pan (2024); Suseno et al. (2024)	MEDIUM
Teacher training models	Insufficient professional development infrastructure	Investigate PD programme designs	S & Hallikere (2026); Lust et al.	HIGH					

Cognitive & strategy factors	Working memory and DDL-strategy use as moderators largely ignored	Mixed-methods studies examining cognitive factors and metacognitive strategies in corpus use	Lee (2018)	MEDIUM		Asia, Africa underrepresented	veness across diverse EFL settings		
Geographic diversity	Majority of studies from Turkey, Saudi Arabia, East Asia; Southeast	Cross-cultural comparative studies of DDL effecti	Sun & Park (2023); Emir & Ekşi (2023)	MEDIUM	Fluency–accuracy – complexity trade-off	Improvements in fluency without corresponding complexity/accuracy gains documented	Design tasks explicitly targeting all three CAF dimensions simultaneously	Kızıl (2023); Corpus-Driven (2024)	MEDIUM

Note. Priority ratings (HIGH / MEDIUM) reflect the urgency and potential impact of addressing each gap based on current evidence and field-wide needs.

D. Kesimpulan

This systematic literature review has reviewed 50 empirical and

theoretical studies about linguistic-based strategies used in EFL classroom teaching to reveal that there have been consistent positive results in relation to the development of students' vocabulary acquisition and sentence structures. There have been consistent positive benefits from using corpus-based learning approaches, DDL, and learner independence. Sentence combining and writing frame strategies contribute to syntactic sophistication and accuracy, but at the same time care must be taken to avoid over-framing. The best outcome is achieved through an integrated approach incorporating syntax, semantics, and pragmatics.

However, certain issues still remain unresolved, including the differential efficiency of methods depending on learners' language proficiency, the problem of translating theoretical knowledge into practical skills, and the lack of longitudinal studies proving the lasting effect of the method on language learning results. It is important to continue conducting empirical research in order to address these problems, invest in teacher training in applying the described method, and change curricula to

introduce corpus and integrated language approach systematically into the classroom. In this light, all the research findings support the need for a linguistically based, learner-oriented, context-sensitive, and effective EFL methodology.

DAFTAR PUSTAKA

- Alenizi, A., & Adawi, R. (2024). Investigating the effectiveness of using corpus-based developed materials in vocabulary learning for Saudi EFL students. *Forum for Linguistic Studies*, 6(3), 721–745.
<https://doi.org/10.30564/fls.v6i3.6551>
- Barko-Alva, K., & Chang-Bacon, C. K. (2023). Over-framing: Interrogating sentence frames as pedagogical support vs. language restriction. *Language Culture and Curriculum*, 1–17.
<https://doi.org/10.1080/07908318.2023.2218096>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

- Canale, M., & Swain, M. (1980). The theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/l.1.1>
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written English grammar and usage*. Cambridge University Press.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Heinle & Heinle.
- Corpus-driven approaches for advancing L2 grammar: An empirical investigation. (2024). <https://doi.org/10.5281/zenodo.14284068>
- Emir, G., & Ekşi, G. Y. (2023). Corpus used as a data-driven learning tool in L2 academic writing: Evidence from Turkish contexts. *TEFLIN Journal*, 34(2), 209–225. <https://doi.org/10.15639/teflinjournal.v34i2/209-225>
- Esfandiari, R., & Allaf-Akbary, O. (2024). Assessing interactional metadiscourse in EFL writing through intelligent data-driven learning: The Microsoft Copilot in the spotlight. *Language Testing in Asia*, 14(1). <https://doi.org/10.1186/s40468-024-00326-9>
- Hardy, D. E. (2007). Textual patterns: Key words and corpus analysis in language education. *Style*, 41(4).
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin.
- Jaya, S., Susyla, D., & Melati. (2025). Developing syntactic complexity and accuracy in EFL students' essays through sentence combining. *Indonesian Journal of Educational Development*, 6(2), 408–421. <https://doi.org/10.59672/ijed.v6i2.4802>
- Kızıl, A. Ş. (2023). Data-driven learning: English as a foreign language writing and complexity, accuracy and

- fluency measures. *Journal of Computer Assisted Learning*.
<https://doi.org/10.1111/jcal.12807>
- Kondos, S. (2025). Lexical bundles: A focused framework for enhancing vocabulary and syntax in English composition courses. *IAFOR Journal of Education*, 13(1), 171–193.
<https://doi.org/10.22492/ije.13.1.07>
- Lee, H. (2018). Exploring corpus use in second language vocabulary learning: Toward the establishment of a data-driven learning model [Doctoral dissertation].
- Lee, I., & Lee, D. J. (2023). Effect of data-driven learning (DDL) on 6th grade elementary school English sentence writing. *Foreign Languages Education*, 30(1), 33–55.
<https://doi.org/10.15334/fle.2023.30.1.33>
- Lee, J., & Yoon, K. (2019). Alternative vocabulary learning approaches in EFL setting: Bottom-up or top-down? *English Teaching*, 74(3), 141–160.
<https://doi.org/10.15858/ENGT.EA.74.3.201909.141>
- Li, X. (2018). Language enhancement for translation undergraduates: An evaluation of a comparable and parallel text analysis course. *Journal of Multilingual and Multicultural Development*, 39(6), 475–490.
<https://doi.org/10.1080/01434632.2018.1439948>
- Liou, H., & Yang, T. (2020). Data-driven learning at the English drafting stage.
<https://doi.org/10.4018/978-1-7998-2591-3.CH014>
- Lopez, L. A. (2024). Configuring collocations and prepositions in essay writing through a corpus-based strategy. *HOW*, 31(2).
<https://doi.org/10.19183/how.31.2.790>
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). Language corpus and data driven learning (DDL) in language classrooms: A systematic review. *Heliyon*.
<https://doi.org/10.1016/j.heliyon.2023.e22731>

- Ma, R. (2012). Vocabulary proficiency instruction for Chinese EFL learners. *Theory and Practice in Language Studies*, 2(6), 1199–1205.
<https://doi.org/10.4304/TPLS.2.6.1199-1205>
- Marrying frame semantics with task-based language teaching. (2023).
<https://doi.org/10.26262/istal.v24i0.9160>
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge University Press.
- Mei, L., Kun, Y., Yuhua, N., & Yiping, J. (2009). New way to lexical acquisition: Approach of corpus-based data driven learning. *Proceedings of IFCSTA*.
<https://doi.org/10.1109/IFCSTA.2009.248>
- Men, H. (2020). Data-driven learning in enhancing learners' language idiomaticity. *International Journal of Emerging Technologies in Learning*, 15(23), 27–41.
<https://doi.org/10.3991/IJET.V15I23.19023>
- Mousavi, S. H., Amouzadeh, M., & Rezaei, V. (2015). Application of frame semantics to teaching seeing and hearing vocabulary to Iranian EFL learners. *Research in Applied Linguistics*, 6(1), 98–117.
- Murtisari, E. T. (2020). Enhancing isolated grammar teaching through translation: Sentence level and beyond. *LLT Journal*, 23(1).
<https://doi.org/10.24071/LLT.V23I1.2416>
- Murtisari, E. T. (2021). Use of translation strategies in writing: Advanced EFL students. *LLT Journal*, 24(1).
<https://doi.org/10.24071/LLT.V24I1.2663>
- Pan, Z. (2024). EFL learners' use of data-driven learning and their attitudes in the improvement of English speaking skills. *International Journal of Learning, Teaching and Educational Research*, 23(8), 98–115.
<https://doi.org/10.26803/ijlter.23.8.6>

- Peppard, J. (2014). A lexicogrammatical approach to fluency. https://doi.org/10.1057/9781137449405_6
- Promoting independent sentence construction in ESL learners by minimizing overreliance on translation tools. (2025). <https://doi.org/10.5281/zenodo.15726718>
- S, N., & Hallikere, G. (2026). Integrating linguistic theory into contemporary teaching practices: From structural understanding to classroom application. *International Journal for Multidisciplinary Research*. <https://doi.org/10.36948/ijfmr.2026.v08i01.66167>
- Salehpour, S., HadidiTamjid, N., & Behnam, B. (2014). Scaffolding advanced writing through writing frames. *Applied Linguistics*, 7(14), 146–169.
- Sangia, R. A. (2022). Emerging word games to improve young learners' vocabulary skills in Ambonese Islamic school.
- Journal of English Language Teaching, 9(1), 72–72.
- Smirnova, O., Rackeviciene, S., & Mockiene, L. (2022). Frame semantics methodology for teaching terminology of specialised domains. *Journal of Teaching English for Specific and Academic Purposes*, 765–765. <https://doi.org/10.22190/jtesap.2104765s>
- Soori, A., Kafipour, R., Dakhil, T. A., Khojasteh, L., & Behrosi, B. (2023). The effect of role-play on vocabulary learning and retention in Iranian EFL learners. *Forum for Linguistic Studies*. <https://doi.org/10.59400/fls.v5i2.1835>
- Sun, W., & Park, E. (2023). EFL learners' collocation acquisition and learning in corpus-based instruction: A systematic review. *Sustainability*. <https://doi.org/10.3390/su151713242>
- Suseno, E., Purwati, O., & Anam, S. (2024). Using podcasts, lexicogrammatical resources, and

- videos with descriptive images to improve speaking skills. IJORER.
<https://doi.org/10.46245/ijorer.v5i1.424>
- Tosun, S., & Sofu, H. (2023). The effectiveness of data-driven vocabulary learning: Hands-on concordancing through a pedagogical corpus. *Journal of Language and Education*.
<https://doi.org/10.17323/jle.2023.12426>
- Tuan, L. T. (2010). Improving EFL learners' use of English vocabulary in sentence writing through FrameNet. *Studies in Literature and Language*, 1(2), 48–57.
- Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. Routledge.
- Ullah, I., & Goloona, A. (2021). The interplay between syntax, semantics, and pragmatics in language education: How teaching sentence structure, meaning, and context shape effective communication skills. *Global Language Review*.
[https://doi.org/10.31703/glr.2021\(vi-ii\).36](https://doi.org/10.31703/glr.2021(vi-ii).36)
- Vivone, A., Rumlich, D., & Lehmann-Wermser, A. (2025). Maximizing linguistic and cognitive potential: The role of vocal training with pop songs in teaching English as a foreign language.
<https://doi.org/10.1080/23735082.2025.2501043>
- Wang, Y., & Yang, J. (n.d.). A data-driven lexical approach to L2 lexis instruction.
<https://doi.org/10.3969/j.issn.1002-2643.2008.04.010>
- Warta, I. K. (2016). Improving students' vocabulary mastery using linguistic approach. *Jurnal Ilmu Pendidikan*, 14(3).
<https://doi.org/10.17977/JIP.V14I3.39>
- Yang, K., & Li, M. (2010). New approach to enhance composition writing: The application of corpus-based data driven learning. *Proceedings of CISE*.
<https://doi.org/10.1109/CISE.2010.5677124>