

Language Teaching Research Quarterly

2025, Vol. 47, 175–195



A Systematic Review of Corpus-Based Methodologies in Textbook Analyses and Evaluation

Laleh Khojasteh¹, Jayakaran Mukundan^{2*}

¹Department of English Language, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

²Taylors University, Lakeside Campus, Malaysia

Received 28 January 2025

Accepted 15 March 2025

Abstract

This systematic review explores the evolution and impact of corpus-based methodologies in textbook analysis within the field of language education between 2010 and 2024. Utilizing the PRISMA framework, 29 studies were identified and analyzed, highlighting the transition from basic frequency analyses to sophisticated examinations of lexical, semantic, and syntactic patterns. The findings indicate that advanced methodologies, particularly those employing specialized corpora and tools like AntConc and COCA, significantly contribute to improving the accuracy and relevance of textbook content. The review also reveals substantial regional variations, with countries like Malaysia, Iran, and South Korea leading in corpus-based research, driven by local educational needs and linguistic contexts. Logistic regression analysis further demonstrates that higher levels of methodological sophistication increase the likelihood of textbook revisions, underscoring the importance of rigorous research practices. The implications for pedagogy are profound, suggesting that integrating corpus-based findings can enhance the alignment of educational materials with authentic language use, thereby improving learner outcomes. This review offers insights for educators, curriculum developers, and researchers on how to effectively apply corpus-based methodologies to enhance language teaching materials and practices.

Keywords: *Textbook Analysis, Corpus, Corpus-Based Methodologies, Systematic Review, Educational Materials*

How to cite this article (APA 7th Edition):

Khojasteh, L., & Mukundan, J. (2025). A systematic review of corpus-based methodologies in textbook analyses and evaluation. *Language Teaching Research Quarterly*, 47, 175-195. <https://doi.org/10.32038/ltrq.2025.47.10>

Introduction

In recent years, language education has increasingly embraced empirical research to inform teaching practices, ensuring that instructional materials align with authentic language use (Khany & Kamalvand, 2022). One of the most impactful developments in this area has been

* Corresponding author.

E-mail address: jayakaran.mukundan@taylors.edu.my

<https://doi.org/10.32038/ltrq.2025.47.10>

the application of corpus-based methodologies to textbook analysis (Chan, 2021). Corpus linguistics provides researchers with robust tools for analyzing large datasets of authentic language, helping to identify gaps between the language presented in textbooks and the way it is actually used in real-world communication (Biber et al., 2002; Flowerdew, 2012). By leveraging these insights, educators can refine educational materials to better reflect the linguistic needs of learners (O’Keeffe, 2007).

Despite these advancements, existing textbooks still exhibit significant linguistic discrepancies (Liu et al., 2024). Early corpus-based studies primarily focused on surface-level frequency analyses, revealing mismatches in grammatical structures and vocabulary coverage (Khojasteh & Mukundan, 2011; Roslim et al., 2021). However, more recent research has adopted advanced computational techniques to examine lexical, semantic, and syntactic patterns, uncovering structural inefficiencies that may hinder effective language acquisition (Alavi & Rajabpoor, 2015; Gedik & Kolsal, 2022). These findings raise critical questions about whether current textbooks adequately prepare learners for real-world communication.

The educational implications of these discrepancies are particularly relevant in English as a Second Language (ESL) and English as a Foreign Language (EFL) contexts. If textbooks do not align with authentic language use, students may struggle with comprehension, fluency, and pragmatic competence in real-world settings (Biber et al., 2010; Flowerdew, 2015). While corpus-based methodologies offer solutions, their impact on textbook revisions and pedagogical practices remains inconsistent across different regions and educational systems.

This systematic review aims to synthesize corpus-based textbook analysis studies conducted between 2010 and 2024, identifying key methodological developments, regional variations, and their implications for curriculum design and instructional practices. By critically evaluating how these methodologies contribute to textbook revisions and language teaching effectiveness, this review seeks to provide evidence-based recommendations for educators, curriculum developers, and researchers to enhance the pedagogical relevance of corpus-driven insights.

Integration of Learning Theories in Corpus-Based Textbook Analysis in ESL/EFL Context
In corpus-based textbook analysis, multiple learning theories converge to create a framework that ensures textbooks align with authentic language use while being pedagogically effective. The Input Hypothesis (Krashen, 1982) highlights the necessity of providing learners with comprehensible input slightly beyond their current proficiency level ($i+1$). However, exposure alone is insufficient; the Noticing Hypothesis (Schmidt, 1990) posits that learners must actively recognize linguistic patterns to internalize them. To make this process engaging and effective, Tomlinson’s Principles of Materials Development emphasize the importance of rich, meaningful, and contextually appropriate input. Lastly, Communicative Language Teaching (CLT) ensures that textbook content not only presents linguistic structures but also fosters real life communication and interaction. Together, these theories form a comprehensive framework for evaluating and improving textbook design using corpus-based methodologies.

Aligning Theories to Corpus-Based Textbook Analysis

These learning theories are not isolated concepts but rather interconnected principles that collectively strengthen corpus-based textbook analysis. Table 1 outlines how each theory contributes to enhancing textbook evaluation and design.

Table 1*Integration of Learning Theories into Corpus-Based Textbook Analysis*

Theory	Description	Application in Corpus-Based Textbook Analysis
Input Hypothesis	Language acquisition via comprehensible input (i+1).	Identifies frequency and distribution of high-frequency words for logical difficulty progression.
Noticing Hypothesis	Learners must notice linguistic features for acquisition.	Highlights gaps between textbook and authentic usage; targets critical linguistic features.
Principles of Materials Development (Tomlinson)	Focus on rich, engaging, and meaningful input.	Ensures alignment with authentic, frequency-based lexical and grammatical examples.
Communicative Language Teaching (CLT)	Interaction and authentic communication.	Supports realistic role-plays and scenarios using real-life language data.

These theoretical foundations provide a structured approach to analyzing textbooks quantitatively (through corpus linguistics) and qualitatively (through learning principles), ensuring that instructional materials are both evidence-based and pedagogically sound.

The Input Hypothesis (Krashen, 1982) posits that language acquisition occurs when learners are exposed to comprehensible input slightly beyond their proficiency level (i+1). This underscores the need for materials that challenge learners while remaining accessible. Corpus-based methodologies operationalize this principle by enabling educators to analyze lexical and grammatical patterns in textbooks, ensuring that they align with the natural progression of linguistic difficulty.

Corpus-based analyses allow educators to identify and prioritize high-frequency vocabulary, such as those listed in the General Service List (GSL) (West, 1953) and the Oxford 3000. Studies indicate that textbooks that incorporate high-frequency words enhance learners' comprehension and engagement (Nation & Webb, 2011). However, research by Alavi and Rajabpoor (2015), Kim and Oh (2020), and Khojasteh and Mukundan (2011) suggests that many ESL/EFL textbooks omit critical high-frequency words or overemphasize less relevant vocabulary, disrupting the i+1 principle. By using corpus tools, educators can ensure that textbook vocabulary aligns with authentic language use and learner needs.

Corpus-based methods also evaluate lexical richness and diversity through token-type analysis (Kim & Oh, 2020). Studies by Laufer and Nation (1995), Meara and Fitzpatrick (2000), and (Fitzpatrick & Clenton, 2010) emphasize the importance of balancing lexical diversity with comprehensibility. For example, textbooks with an excessively high ratio of unique words (types) to total words (tokens) may overwhelm learners, violating Krashen's comprehensible input principle. Additionally, lexical progression—the increasing complexity of vocabulary across textbook levels—is crucial for supporting long-term acquisition. Research by Francis and Kucera (1982), Gilmore (2007), and Meara and Fitzpatrick (2000) demonstrates that textbooks often lack systematic progression, making it difficult for learners to transition to higher proficiency levels. Corpus tools can map lexical progression, ensuring that each textbook level appropriately advances linguistic difficulty, aligning with Krashen's i+1 principle.

While Input Hypothesis focuses on exposure, the Noticing Hypothesis (Schmidt, 1990) posits that learners must consciously notice linguistic patterns for acquisition to occur. This

perspective is particularly relevant in corpus-based textbook analysis, as it helps identify underrepresented or misrepresented linguistic features. Corpus methodologies highlight gaps between textbook content and authentic language use. Studies by Boulton and Cobb (2017) and Lee (2020) demonstrate how corpus analysis refines materials by addressing underrepresented collocations, lexical bundles, and syntactic structures. By highlighting key features, corpus-based methodologies reinforce Schmidt's claim that noticing facilitates language internalization.

While Noticing Hypothesis ensures linguistic salience, Tomlinson's (2011) framework emphasizes rich, meaningful, and engaging input to support deep learning and motivation. Corpus-based methods enhance textbook design by incorporating authentic texts from corpora (e.g., BNC, COCA), ensuring that materials reflect real-world usage. Studies by Alavi and Rajabpoor (2015) and Vyatkina (2020) show that corpus-informed materials increase engagement by integrating idiomatic expressions and collocations common in natural discourse. Additionally, Tomlinson's principle of cognitive engagement is supported by corpus-based tasks, such as reconstructing collocations, analyzing verb patterns, and exploring prepositional usage, which promote deep processing and retention (Amare, 2022).

While Tomlinson's framework ensures input is meaningful, Communicative Language Teaching (CLT) (Sato & Kleinsasser, 1999) emphasizes real-life interaction. Though corpus-based methodologies are not inherently interactional, they support CLT by providing authentic language input for designing communicative tasks (Syahid et al., 2024). For example, corpus analyses of workplace speech and customer interactions inform the creation of task-based learning activities that simulate real-world scenarios (Friginal et al., 2025; Kim & Oh, 2020). Thus, corpus-based textbook analysis bridges theory and practice by integrating CLT principles into data-driven materials that enhance both linguistic competence and communicative ability.

Despite the increasing use of corpus-based methodologies in language education, existing studies have primarily focused on lexical frequency analysis and grammatical structures without a comprehensive theoretical framework that integrates key language acquisition principles. Furthermore, there is a lack of systematic reviews that critically evaluate how these methodologies have evolved over time, their varying levels of sophistication, and their regional implications for textbook development and pedagogical practice. This review addresses this gap by synthesizing research on corpus-based textbook analysis through the lens of four interconnected learning theories: the Input Hypothesis, Noticing Hypothesis, Tomlinson's Principles of Materials Development, and Communicative Language Teaching (CLT). By bridging theoretical insights with empirical findings, this study offers a holistic perspective on how corpus methodologies enhance ESL/EFL textbook design and instructional effectiveness.

Research Questions

This study seeks to address the following key research questions, focusing on the role and impact of corpus-based methodologies in textbook analysis and pedagogical practices:

RQ1: How have corpus-based methodologies evolved over time, and what impact have they had on improving the accuracy and relevance of textbook content in language education?

RQ2: To what extent do different levels of methodological sophistication (basic, lexical/semantic, advanced) influence the scope and nature of recommendations for textbook

revisions, and how do these recommendations vary based on the rigor of the research methodology?

RQ3: How do corpus-based methodologies and their findings vary across different regions, and what are the implications for textbook development in diverse educational contexts?

RQ4: How can corpus-based findings be integrated into pedagogical practices to improve language teaching methodologies and outcomes?

Materials and Methods

Design

This systematic review adopts an integrative approach, synthesizing qualitative and quantitative findings from diverse studies to offer a holistic understanding of corpus-based textbook analysis. This method enables a comprehensive examination of the evolution and impact of corpus-based methodologies in language education (Whittemore & Knafl, 2005), highlighting trends and demonstrating how these methodologies have informed textbook revisions and enhanced pedagogical practices. The design of this review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework (Moher et al., 2009) to ensure a rigorous and transparent methodology in identifying, selecting, and analyzing relevant studies.

The study selection process, conducted in accordance with the PRISMA framework, is illustrated in Figure 1. This diagram provides a visual summary of the identification, screening, eligibility, and inclusion stages, ensuring methodological transparency and replicability.

Step 1: Identification

A comprehensive search was performed across seven electronic databases: ERIC, LLBA, Web of Science, Scopus, JSTOR, ProQuest, and Academic Search Complete. Boolean search strings were refined to include keywords such as "corpus linguistics," "textbook analysis," "educational materials," and "data-driven learning" to ensure a broad yet relevant collection of studies. The initial search retrieved 150 studies for potential inclusion.

Step 2: Screening

To refine the selection, duplicate records ($n = 52$) were removed. The remaining 98 studies were screened based on their titles and abstracts. Studies were excluded if they lacked relevance to corpus-based textbook analysis, did not focus on ESP/EFL contexts, or were conference papers and non-peer-reviewed articles.

Step 3: Eligibility

A full-text review of 52 studies was conducted to assess their methodological rigor and relevance. Studies were evaluated on clarity in corpus-based methodology, empirical rigor, and relevance to textbook analysis. After this review, 23 studies were excluded due to insufficient methodological detail or lack of focus on corpus-based methodologies.

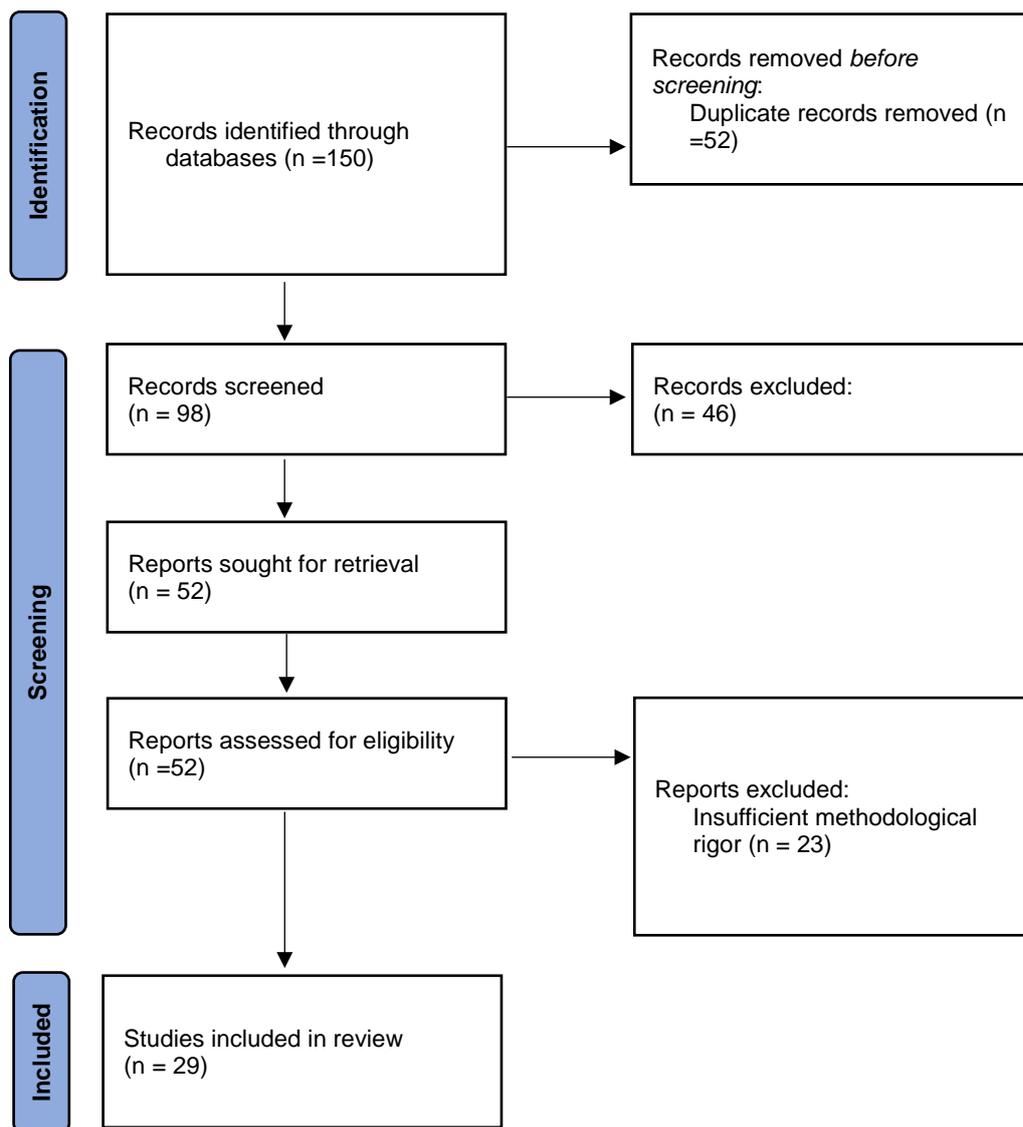
Step 4: Inclusion

A total of 29 studies met the inclusion criteria and were selected for detailed analysis. These studies provide insights into the evolution of corpus-based methodologies, their impact on ESP/EFL textbook revisions, and their broader pedagogical implications.

This structured selection process ensures that the review adheres to PRISMA guidelines, maintaining transparency and methodological consistency.

Figure 1

Identification of Studies via Databases and Registers



Findings

This section presents the findings of the systematic review, organized by the key research questions. Each research question is addressed with tables that provide only the relevant data needed to answer the specific question. This focused approach allows for easier interpretation of how corpus-based methodologies have evolved, how methodological rigor impacts textbook revisions, and how these methodologies vary across different regions.

The first research question aims to explore the progression of corpus-based methodologies from 2010 to 2024 and their contribution to improving the accuracy (alignment with authentic language use) and relevance (suitability for learner needs) of textbook content. Rather than treating each period in isolation, this section discusses the gradual shift from basic to advanced methodologies and their impact on textbook analysis. The studies are divided into two periods: early (2010–2014) and late (2015–2024), reflecting the methodological advancements over time.

During the early period (2010–2014), corpus-based studies focused on basic methodologies, such as frequency counts and collocational analyses, highlighting discrepancies in linguistic features like modal verbs and preposition usage (see Table 2).

Table 2*Evolution of Corpus-Based Methodologies (Early Period: 2010–2014)*

Study Authors	Year	Country	Methodological Details	Key Findings
Menon and Mukundan	2010	Malaysia	Basic corpus-based analysis of collocational patterns	Identified shared syntactic characteristics of collocations, but some combinations were arbitrarily blocked by usage.
Khojasteh & Mukundan	2011	Malaysia	Corpus-based comparison of modal verbs in textbooks vs. real language (BNC)	Found significant discrepancies in the frequency of modal verbs between textbooks and authentic usage.
Mukundan et al.	2012	Malaysia	Analysis of article usage and colligation patterns in textbooks	Detected inconsistencies in the use of articles in textbooks, leading to learner errors.
Zarifi & Mukundan	2012	Malaysia	Systematic analysis of phrasal verbs in textbooks vs. BNC	Highlighted overemphasis on certain phrasal verbs, with no systematic patterns.
Menon & Mukundan	2012	Malaysia	Corpus-based analysis of scientific collocations	Identified predominant NN and AN combinations, showing how textbooks were limited in lexical diversity.
Roslim et al.	2012	Malaysia	Corpus-based study comparing textbook preposition usage with BNC	Identified discrepancies in preposition usage, suggesting textbooks did not reflect real-world language use.
Mukundan et al.	2012	Malaysia	Corpus-based analysis of article distribution and colligation patterns	Highlighted inconsistent colligation patterns and overemphasis on certain patterns.
Alavi & Rajabpoor	2014	Iran	Cross-corpus comparison of idiomatic expressions in textbooks and authentic corpora	Found significant differences in the frequency and types of idioms between textbooks and real-world language corpora.
Molavi, & Hosseini	2014	Iran	Corpus-based analysis comparing textbook collocations with OANC	Found textbook collocations did not align with native speaker usage.
Zarifi, Mukundan & Rezvani	2014	Malaysia	Corpus-based study of phrasal verbs in textbooks	Found overemphasis on certain phrasal verbs and a lack of systematic patterns.

The key findings indicated that textbooks frequently failed to represent these linguistic features accurately. For instance, Khojasteh and Mukundan (2011) found that modal verbs were either underrepresented or inaccurately portrayed, limiting learners' ability to properly grasp modality. Similarly, Zarifi and Mukundan (2012) highlighted an overemphasis on certain phrasal verbs without systematic presentation, which could confuse learners and prevent them from fully grasping these critical linguistic elements in authentic communication.

Idiomatic expressions, another essential component of language learning, were also problematic. In a study by Molavi et al. (2014), idioms in Iranian advanced EFL textbooks were found to differ significantly in both frequency and type when compared to idioms in authentic corpora. Such inconsistencies suggest that learners relying solely on these textbooks may struggle with idiomatic fluency, potentially impacting their communicative competence.

While these early studies provided important insights into the accuracy gaps between textbook content and authentic language, they were largely descriptive in nature. Their primary contribution was identifying these discrepancies, rather than offering concrete solutions or revisions to address them. Consequently, while these studies laid the groundwork for methodological improvements, they did not yet propose explicit textbook revisions.

As it can be seen in Table 3, in the late period, corpus-based methodologies in textbook analysis became more sophisticated, shifting from purely descriptive approaches to more actionable and solution-oriented research. Advanced methodologies such as syntactic complexity analysis, cross-corpus comparisons, and multi-method approaches were widely employed. Researchers not only identified discrepancies between textbook content and authentic language but also made detailed recommendations for improving both the accuracy and relevance of textbooks.

One key trend during this period was the increased focus on ensuring that textbooks aligned with international standards. For example, Arellano (2018) found that Chilean ESL textbooks lacked alignment with CEFRL (Common European Framework of Reference for Languages) standards, which could result in learners being underprepared for international assessments. Another important area of focus was the representation of vocabulary and linguistic structures in textbooks. Norberg and Nordlund (2018), in their study of Swedish EFL textbooks, found notable variations in vocabulary selection, with many words appearing infrequently in real-world language use. This lack of lexical alignment raises concerns about how effectively these textbooks prepare learners for authentic communication.

Advanced corpus methodologies also shed light on issues such as collocational patterns and multiword units (MWUs). Kim and Oh (2020) observed that insufficient repetition of collocations in Korean EFL textbooks limited students' exposure to natural language patterns, which may negatively impact their fluency development. Similarly, Hoang and Crosthwaite (2024) identified discrepancies in the treatment of MWUs in Vietnamese textbooks, with significant differences between the way these units were presented in reading and listening materials. These findings highlight the need for more consistent integration of MWUs across textbook materials. Furthermore, several studies in this period focused on linguistic progression and syntactic complexity in textbooks. Lynn (2021) found that U.S. ESL/EFL reading textbooks did not gradually build linguistic complexity, which could hinder learners' ability to progress smoothly between proficiency levels. This study emphasized the importance of structured progression in textbook content to help learners transition from beginner to advanced language skills. In contrast to the early period, where findings were largely descriptive, the studies in this late period provided concrete, actionable recommendations for revising textbooks. The use of more advanced methodologies allowed researchers to offer specific solutions, such as aligning textbooks with international standards, ensuring adequate lexical coverage, and improving the presentation of complex linguistic features like collocations and MWUs. By integrating more sophisticated analytical techniques, these studies

not only identified deficiencies but also provided structured pathways for curriculum improvement.

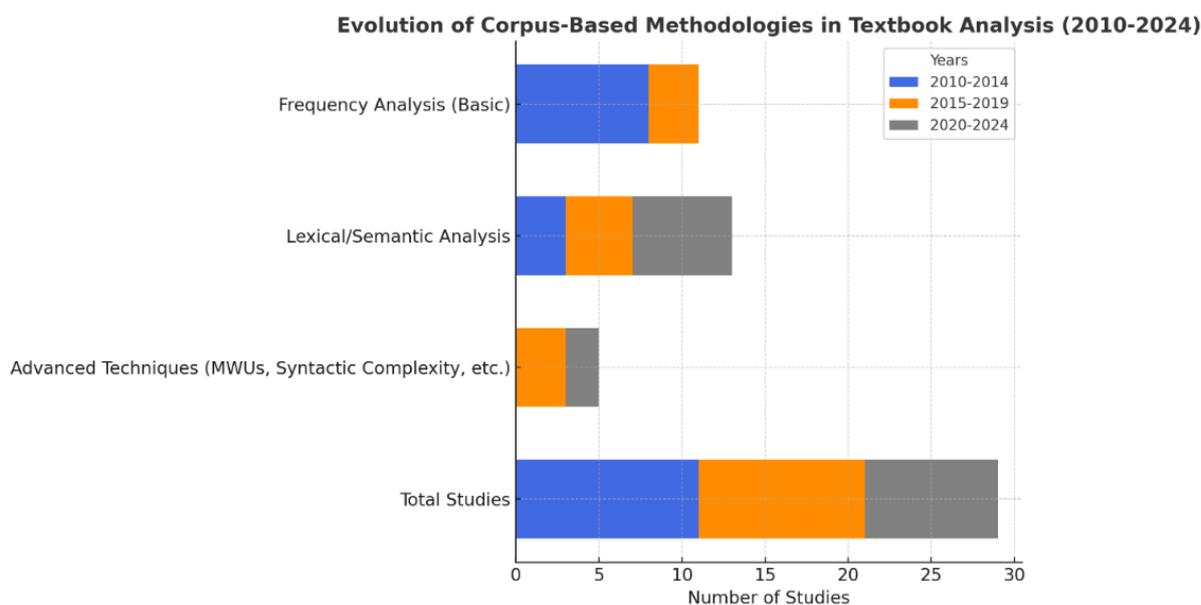
Table 3*Evolution of Corpus-Based Methodologies (Late Period: 2015–2024)*

Study Authors	Year	Country	Methodological Details	Key Findings
Aggeliki	2015	Greece	Corpus-based study comparing future tense usage in textbooks with real language (BNC)	Found discrepancies between "will" and "be going to" usage between textbooks and real language.
Arellano	2018	Chile	Corpus-based study comparing Chilean ESL textbooks with CEFRL	Found differences between Chilean textbooks and CEFRL in curriculum design and language assessment.
Norberg & Nordlund	2018	Sweden	Corpus-based study comparing vocabulary usage in textbooks with VP-Kids corpus	Found significant variation in vocabulary selection; many words infrequently used in real-world language.
Ainul Azmin Md Zamin et al.	2019	Malaysia	Advanced corpus-based content analysis of phrasal verbs in textbooks	Found that phrasal verbs in textbooks were over-repeated and lacked systematic presentation.
Zamin et al.	2019	Malaysia	Corpus-based analysis of phrasal verbs in textbooks using Longman Grammar corpus	Phrasal verbs were often selected based on authors' intuition rather than empirical data, leading to over-repetition.
Chanakarn Chareonkul	2020	Thailand	Corpus-based comparison of present perfect tense in textbooks vs. real corpora	Found significant differences in the frequency and meaning of present perfect tense between textbooks and real language.
Kim & Oh	2020	South Korea	Corpus-based analysis of collocational patterns in Korean EFL textbooks	Textbooks lacked sufficient repetition of key collocations, limiting students' exposure to natural language patterns.
Chalabian	2020	Canada	Comparative analysis of gerund/infinitive patterns in textbooks vs. COCA and TV Corpora	Identified inconsistent rules for gerund/infinitive use in textbooks, lacking authentic examples.
Lynn	2021	USA	Corpus analysis comparing ESL/EFL reading textbooks with university-level texts	Found that textbooks did not progressively build linguistic complexity, resulting in significant gaps.
Udaya	2021	India	Corpus-based comparison of vocabulary in textbooks vs. real corpora (BNC)	Significant gaps were identified in the vocabulary input, with insufficient real-world language use.
Ikonen	2021	Finland	Corpus-driven study analyzing idioms in EFL textbooks	Found that idioms in textbooks were not reflective of their frequency or usage in real-world language.
Hoang & Crosthwaite	2024	Vietnam	Corpus-based extraction of multiword units (MWUs) in reading and listening materials	Found significant differences in MWU occurrence and function between reading and listening materials in textbooks.
Khan	2024	Pakistan	Corpus-based comparison of vocabulary in Pakistani ESL textbooks vs. international standards	Found that Pakistani textbooks did not meet international standards for vocabulary range and complexity.
Farooq Khan	2024	Pakistan	Corpus-based analysis of Class X textbooks vs. international vocabulary standards	Found that textbooks provided insufficient vocabulary beyond basic levels.

As shown in Figure 2, the evolution of corpus-based methodologies in textbook analysis from 2010 to 2024 illustrates this progression in greater detail. The figure provides a breakdown of the 29 studies reviewed, categorized by their time period and the level of methodological sophistication employed. Early studies relied heavily on basic frequency analyses, whereas more recent research has incorporated advanced techniques, such as multiword unit extraction and syntactic complexity analysis. This visual representation reinforces the increasing complexity of research approaches and their growing impact on textbook development.

Figure 2

Evolution of Corpus-Based Methodologies in Textbook Analysis (2010-2024)



In addressing Research Question 2, we focused on studies that explicitly mentioned recommendations for textbook revisions based on the corpus-based methodologies employed. The recommendations included in this review are those explicitly stated by study authors and vary in scope depending on the sophistication of the methodology employed (Table 4).

Corpus-based studies that provided textbook revision recommendations demonstrated a clear progression in methodological sophistication, ranging from basic methodologies (e.g., frequency counts and surface-level grammatical analyses) to lexical/semantic methodologies (e.g., collocational and idiomatic expression analyses) and ultimately to advanced methodologies (e.g., syntactic complexity analysis and multiword unit extraction). Early studies that relied on basic methodologies primarily focused on identifying discrepancies between textbook content and authentic language use but lacked detailed implementation strategies. For example, Khojasteh & Mukundan (2011) suggested increasing the frequency of modal verbs in textbooks to better reflect real-world usage but did not outline a structured plan for revision. Similarly, Menon and Mukundan (2010) recommended aligning textbook collocations with real-world patterns but provided only general suggestions, lacking concrete instructional strategies.

Table 4*Recommended Revisions across Different Methodological Approaches*

Study Authors	Year	Methodology	Key Findings	Recommendations
Menon & Mukundan	2010	Basic Frequency Analysis	Identified patterns in scientific collocations, but textbooks showed arbitrary usage	Recommended broad alignment of textbook collocations with real-world patterns, but no specific strategy suggested.
Khojasteh & Mukundan	2011	Frequency Analysis	Underrepresentation of modal verbs in textbooks compared to corpora	General suggestion to increase modal verb representation to reflect authentic usage patterns.
Mukundan et al.	2012	Article Frequency Analysis	Inconsistent article usage ("a," "an," "the")	General recommendation for textbook authors to correct article usage inconsistencies, but no clear revision strategy provided.
Zarifi & Mukundan	2012	Lexical/Semantic Analysis	Overemphasis on certain phrasal verbs, leading to unnatural usage in textbooks	Recommended more systematic presentation of phrasal verbs, aligning with real-world usage.
Alavi & Rajabpoor	2014	Lexical/Semantic Analysis	Significant differences in idiom frequency/type compared to authentic corpora	Suggested targeted inclusion of idiomatic expressions that mirror real-world language use in textbooks.
Molavi et al.	2014	Lexical/Semantic Analysis	Collocations in textbooks did not align with authentic language corpora	Proposed targeted revisions to include more natural collocations to enhance textbook relevance and accuracy.
Kim & Oh	2020	Syntactic Complexity Analysis	Insufficient repetition of key collocations, limiting fluency development	Recommended significant revisions to increase repetition of collocations to improve learners' exposure and fluency.
Lynn	2021	Lexical/Syntactic Complexity	Gaps in linguistic progression across textbooks, causing inconsistency in language acquisition	Suggested restructuring textbooks to include a more consistent and progressive increase in linguistic complexity.
Hoang & Crosthwaite	2024	MWU (Multiword Unit) Analysis	Discrepancies in MWU usage between reading/listening textbooks, limiting MWU acquisition	Proposed comprehensive revisions to better align textbook MWU coverage with authentic language usage in both reading and listening.

As the field progressed, studies incorporating lexical/semantic methodologies offered more targeted recommendations by addressing specific linguistic features such as phrasal verbs and collocations. Zarifi & Mukundan (2012) emphasized the need for a systematic presentation of phrasal verbs, ensuring that textbooks align with authentic language use. Similarly, Alavi & Rajabpoor (2014) recommended integrating idiomatic expressions that reflect natural discourse, demonstrating a shift from broad recommendations to more detailed, actionable insights.

The most comprehensive and practical recommendations emerged from studies employing advanced methodologies. These studies not only identified discrepancies but also proposed

structured strategies to address linguistic progression, multiword unit representation, and textbook alignment with international language standards. Kim & Oh (2020) highlighted the insufficient repetition of collocations in Korean EFL textbooks and proposed systematic revisions to enhance learners' fluency development. Likewise, Lynn (2021) recommended a complete restructuring of textbooks to address gaps in syntactic complexity, ensuring a more gradual and structured increase in linguistic difficulty. By leveraging syntactic complexity analysis and MWU extraction, these studies contributed to a more pedagogically sound approach to textbook revision.

To further investigate the relationship between methodological sophistication and the likelihood of textbook revisions, we conducted a logistic regression analysis. The results indicated that studies employing lexical/semantic methodologies were 2.23 times more likely to lead to textbook revisions compared to those using basic methodologies, while studies employing advanced methodologies had an odds ratio of 4.48. These findings reinforce the strong link between methodological rigor and the specificity of recommendations for textbook revisions, emphasizing the value of increasingly sophisticated corpus-based approaches in language education.

In this model, the dependent variable was whether a textbook revision occurred (1 = revision occurred, 0 = no revision), and the independent variable was the type of methodology employed, divided into the three levels: basic, lexical/semantic, and advanced. The basic methodology served as the reference category in the model. Table 5 presents the coefficients, standard errors, odds ratios, and p-values for the impact of different methodologies on the likelihood of textbook revisions.

Table 5

Logistic Regression Analysis of Methodology Sophistication on Textbook Revisions

Variable	Coefficient (β)	Standard Error	Odds Ratio	p-value
Intercept	-1.2	0.5	-	-
Lexical/Semantic Methodology	0.8	0.4	2.23	0.04
Advanced Methodology	1.5	0.6	4.48	0.01

The baseline log odds of a textbook revision for studies employing basic methodologies is -1.2, suggesting that the likelihood of revision is relatively low when basic methodologies are used. For studies using lexical/semantic methodologies, the coefficient of 0.8 indicates that the log odds of a revision increase by 0.8 compared to basic methodologies. The corresponding odds ratio of 2.23 suggests that textbooks are approximately 2.23 times more likely to be revised when lexical/semantic methodologies are employed.

Similarly, the coefficient for advanced methodologies is 1.5, showing a substantial increase in the log odds of a textbook revision compared to basic methodologies. The odds ratio of 4.48 implies that textbooks are 4.48 times more likely to be revised when advanced methodologies are used.

These results indicate that the sophistication of the research methodology significantly predicts the likelihood of textbook revisions. Both lexical/semantic and advanced methodologies are associated with higher odds of revisions compared to basic methodologies, underscoring the impact of methodological rigor on educational content updates.

The findings from the logistic regression analysis further support the trends observed in the narrative interpretations and tables. Studies employing basic methodologies tend to make general recommendations for textbook revisions, which correspond to the relatively low likelihood of such recommendations leading to actual revisions, as reflected in the lower odds ratio. In contrast, studies using lexical/semantic and advanced methodologies make more targeted and comprehensive recommendations, which significantly increase the likelihood of textbook revisions, as evidenced by the higher odds ratios in the regression analysis.

Together, these findings demonstrate a clear relationship between the sophistication of research methodologies and the scope, specificity, and likelihood of textbook revisions. As methodologies become more sophisticated, the depth of recommendations increases, leading to a higher probability of textbook updates, as highlighted by both the narrative interpretations and the regression analysis.

In response to research question 3, we present the *geographical distribution of studies* and the *methodologies employed* across different regions over time (Table 6). This analysis reveals distinct regional patterns in the adoption of *basic*, *lexical/semantic*, and *advanced methodologies*, which have had varying impacts on textbook development.

Table 6
Regional Variations in Corpus-Based Studies and Methodologies

Region	Years Active	Methodologies Used	Key Focus Areas	Implications for Textbook Development
Malaysia	2010–2014	Basic Frequency Analysis	Article usage, modal verbs, collocations	General recommendations to align textbooks with authentic usage patterns.
Iran	2010–2014	Basic Frequency Analysis	Idiomatic expressions, frequency analysis	Broad recommendations for improving accuracy of idiom usage.
Greece	2015–2019	Lexical/Semantic Analysis	Future tense usage, modal verbs	Targeted revisions suggested to improve grammatical accuracy in EFL textbooks.
Chile	2015–2019	Lexical/Semantic Analysis	Vocabulary selection, collocations	Recommendations for textbook revisions to align vocabulary with real-world usage.
South Korea	2020–2024	Advanced Methodologies	Collocation and syntactic complexity analysis	Comprehensive recommendations for improving fluency and linguistic progression.
Thailand	2020–2024	Collocation Analysis	Frequency analysis using large corpora	Specific recommendations for aligning textbooks with real-world collocation patterns.
Turkey	2020–2024	Syntactic Complexity Analysis	Multiword units, advanced corpus tools	Detailed revisions to improve textbook progression and authenticity.

In Malaysia and Iran, during the early 2010s, researchers relied on basic methodologies to address fundamental issues related to article usage, modal verbs, and idiomatic expressions in textbooks. These studies provided broad recommendations, generally advocating for a closer alignment of textbooks with authentic language patterns but lacking detailed strategies for implementation.

By the mid-2010s, regions such as Greece and Chile moved towards lexical/semantic methodologies, allowing for a more refined analysis of specific linguistic features, including

collocations and future tense usage. These studies produced more targeted recommendations, often suggesting precise grammatical and vocabulary revisions to improve textbook alignment with real-world language use.

More recently, regions such as South Korea, Thailand, and Turkey have employed advanced methodologies, including collocation analysis and syntactic complexity analysis, to assess textbook content. These methodologies have facilitated comprehensive textbook revisions, focusing on linguistic progression, fluency development, and alignment with international standards. Researchers in these regions have increasingly emphasized the need for textbooks to support progressive linguistic complexity, ensuring learners advance systematically through different proficiency levels.

The findings suggest that as more regions adopted advanced methodologies, the depth and specificity of textbook revisions also increased. Earlier studies primarily identified broad, foundational issues, while later research provided detailed, data-driven recommendations to improve textbook accuracy and pedagogical effectiveness.

Building on the findings from Research Questions 1 to 3, this section explores how corpus-based research can be applied to enhance pedagogical practices in language education. These insights serve as a foundation for understanding existing shortcomings in textbook content and provide actionable strategies for educators, curriculum developers, and language institutions.

Aligning Textbook Content with Authentic Language Use

Corpus-based analyses from Research Question 1 revealed substantial discrepancies between textbook content and real-world language usage. Studies such as Mukundan and Aziz (2010) and Khojasteh and Mukundan (2011) highlighted misalignment in the presentation of articles and modal verbs, particularly in Malaysian secondary school textbooks. These findings underscore the necessity of revising textbooks to better reflect authentic communication.

To integrate these insights, educators can: (a) Regularly update textbook content using corpus-based authentic language corpora (e.g., British National Corpus), ensuring that language patterns are accurately represented. (b) Develop supplementary materials where immediate textbook revisions are impractical. Teachers can incorporate real-world dialogues and reading passages based on corpus data to supplement existing materials.

Enhancing Vocabulary and Collocation Teaching

Studies from Research Question 2 emphasized the importance of incorporating high-frequency collocations and idiomatic expressions that are commonly found in authentic discourse but are often underrepresented in textbooks. For example, Alavi and Rajabpoor (2015) and Molavi et al. (2014) found that idioms (e.g., "hit the nail on the head") and collocations in Iranian EFL textbooks did not align with their frequency and usage in real-world corpora. This misalignment can hinder students' ability to use language naturally and effectively in real-life communication.

Effective pedagogical applications include: (a) Using corpus-based frequency lists to introduce collocations in context, rather than in isolation. Teachers can incorporate high-frequency collocations and idiomatic expressions drawn from corpus-based frequency lists into their teaching (Alavi & Rajabpoor, 2015). For example, instead of teaching students isolated words like "coffee" or "decision," teachers can introduce common collocations such as "strong

coffee" or "make a decision," which are far more useful in authentic communication. (b) Designing collocation-focused exercises, where students practice common word combinations in real-life scenarios, such as business meetings or casual conversations. For example, teachers might use news articles, conversations, or podcasts as sources for these exercises, encouraging students to identify and use collocations effectively.

Addressing Linguistic Features That Are Commonly Misrepresented

Corpus research has identified key linguistic features that are often inaccurately portrayed in textbooks. For instance, studies such as *Zarifi and Mukundan (2012)* and *Khojasteh and Shokrpour (2015)* found that textbooks overemphasized certain phrasal verbs while neglecting others, leading to imbalanced language exposure.

To address this issue: (a) Educators should supplement textbooks by focusing on underrepresented phrasal verbs and modal auxiliaries. For example, instead of just teaching basic phrasal verbs like "get up" or "give up," teachers could introduce more complex forms like "put up with" (to tolerate) or "come across" (to find by chance), which are often used in everyday communication. (b) Teachers can incorporate corpus-based exercises where students analyze authentic language use across different contexts. For instance, teachers might provide students with examples from a corpus to illustrate how certain phrasal verbs are used differently in formal versus informal contexts.

Improving Grammatical Competence through Corpus-Based Feedback

One of the key findings from Research Question 3 was that grammatical structures, such as tense markers and verb complementation patterns, are often simplified or distorted in textbooks (Chareonkul, 2020). Such distortions impede learners' ability to use grammar accurately in real-world contexts.

Potential solutions include: (a) Using authentic corpus-based examples to contrast textbook grammar rules with real-world usage. This involves comparing the simplified or artificial examples found in textbooks with authentic language from corpora, such as the British National Corpus (BNC) or Corpus of Contemporary American English (COCA). (b) Designing awareness-raising grammar exercises, where students identify and correct discrepancies between textbook examples and corpus-based data. Teachers can ask students to analyze authentic language examples from corpora and identify how they differ from the rules or examples provided in the textbook.

Developing Contextualized Learning Materials Using Corpus Data

Corpus-based research provides valuable insights into how linguistic features differ across spoken and written contexts. For example, Hoang and Crosthwaite (2024) found that multiword units (MWUs) were used differently in reading and listening contexts, highlighting the need for context-specific instructional strategies.

Practical implementations include: (a) Incorporating authentic materials such as podcasts, news articles, and conversation transcripts into classroom activities. These materials expose students to real-world language used in various contexts, helping them develop a more nuanced understanding of how language functions in different communicative situations (Hoang & Crosthwaite, 2024). (b) Designing contextualized exercises, where students explore

differences in spoken vs. written language use. Teachers can design activities that focus on the differences in how language features (such as MWUs or collocations) are used in spoken versus written contexts, helping students understand how to apply language appropriately in each scenario (Kim & Oh, 2020).

Aligning Assessment Materials with Real-World Language Complexity

Findings from Research Question 2 highlighted gaps between textbook content and high-stakes exams. Studies such as Gedik and Kolsal (2022) and Lynn (2021) found that many textbooks fail to prepare students for the lexical and syntactic demands of standardized tests.

To bridge this gap: (a) Test designers should incorporate corpus-based complexity analysis to ensure that assessments reflect real-world language demands (Gedik & Kolsal, 2022). For example, test designers can create assessment items that mirror authentic language tasks, such as interpreting real-life conversations, understanding formal emails, or summarizing news articles. (b) Educators can integrate authentic exam preparation materials, mirroring language tasks found in professional and academic settings. By analyzing authentic corpora, such as the British National Corpus or COCA, educators can ensure that exams include complex sentence structures and advanced vocabulary that reflect how language is used in professional, academic, and everyday communication.

Discussion

This review provides a comprehensive examination of the evolution of corpus-based methodologies in textbook analysis over the past decade and a half. By analyzing 29 studies published between 2010 and 2024, we observed a significant shift from basic frequency analyses of grammatical structures to more sophisticated, nuanced methodologies that incorporate lexical, semantic, and advanced linguistic features. These findings highlight the increasing complexity and refinement of corpus-based approaches, which have been pivotal in enhancing the accuracy and relevance of textbook content in language education. However, beyond these observations, a deeper examination is necessary to understand why these shifts occurred and how they have influenced educational materials and pedagogical practices.

The transition from basic to advanced methodologies mirrors broader trends in corpus linguistics and applied linguistics. Early studies primarily employed surface-level frequency counts to highlight disparities in textbook content (e.g., Khojasteh & Mukundan, 2011; Menon & Mukundan, 2010), but these approaches were inherently limited in scope, unable to capture the nuanced complexity of authentic language use. From the perspective of the Input Hypothesis (Krashen, 1982), these early corpus-based studies played an important role in ensuring that textbooks provided adequate comprehensible input (i+1) by identifying lexical and grammatical patterns that could scaffold learner progression. However, mere exposure to high-frequency forms is not sufficient for acquisition unless learners notice critical linguistic features (Schmidt, 1990). As corpus tools and analytical techniques improved, researchers expanded their focus to collocations, idioms, and syntactic complexity (Alavi & Rajabpoor, 2015; Molavi et al., 2014), leading to more actionable recommendations for textbook revisions. These advancements align with the Noticing Hypothesis, as corpus analyses have helped identify lexical and syntactic gaps in textbooks that learners might otherwise overlook, thereby guiding targeted interventions to enhance linguistic salience. This shift reflects a growing

recognition that language learning materials must align with real-world usage patterns rather than artificially constructed, outdated representations of language (Egbert et al., 2022).

Moreover, the increasing emphasis on corpus-based methodologies can be linked to technological advancements in natural language processing (NLP) and the availability of larger, more diverse corpora (McEnery & Hardie, 2011). With more powerful computational tools, researchers are now better equipped to conduct fine-grained analyses of language structures, leading to richer insights into linguistic patterns and their role in language acquisition. Studies have also shown that corpus-based approaches directly impact language learning outcomes by ensuring that textbooks reflect authentic language usage, thereby fostering deeper learning and increased engagement (Li, 2024; Chan, 2020). These developments reinforce Tomlinson's (2011) Principles of Materials Development, which stress the importance of providing learners with rich, engaging, and meaningful input. Corpus-based textbook analyses ensure that linguistic input is not only authentic but also reflective of the natural frequency and co-occurrence of words and structures, making materials more effective for long-term retention and engagement. This has direct pedagogical implications, as findings from corpus research can inform curriculum development, teacher training, and instructional materials, ensuring that students are exposed to language that is both meaningful and authentic (Nation, 2013).

The findings of this review also underscore the critical role of methodological rigor in driving textbook revisions. Our logistic regression analysis revealed that more sophisticated methodologies significantly increase the likelihood of textbook revisions, demonstrating that studies employing advanced corpus techniques have a stronger impact on educational materials. This suggests that as research methods become more refined, they yield findings that are more persuasive and actionable for curriculum developers and textbook authors. This claim is supported by Arshad (2020), who found that corpus-based analyses significantly influenced textbook revisions in EFL contexts, leading to more authentic language representation and a greater focus on communicative competence. This is consistent with previous research emphasizing the importance of robust corpus methodologies in educational content updates (Callies, 2019). In this regard, Communicative Language Teaching (CLT) principles come into play, as corpus findings inform the development of task-based activities that foster real-world communication. Corpus-informed textbooks increasingly integrate data-driven dialogues, role-plays, and situational tasks, ensuring that learners develop both linguistic competence and communicative fluency (Friginal, 2024; Syahid et al., 2024).

However, this trend also raises concerns about accessibility and adoption. While advanced methodologies provide deeper insights, their implementation in textbook revisions is not always straightforward. Many educational institutions, particularly in under-resourced regions, lack access to the necessary corpus tools and expertise to incorporate these findings effectively. Research by Xodabande and Nazari (2022) shows that while teachers recognize the importance of corpus-driven materials, there are significant barriers to adoption, including limited professional development and institutional support. This highlights a gap between research and practice, necessitating further efforts to bridge the divide between corpus-based research and practical textbook development (Boulton & Cobb, 2017). Moreover, in regions where textbooks remain largely traditional, the potential of corpus methodologies to support i+1 input, promote noticing, and enhance communicative engagement is underutilized. The

effectiveness of corpus-driven materials is ultimately dependent on how well they are integrated into classroom practices, reinforcing the need for teacher training programs that align corpus findings with pedagogical theories.

Regional variations in the adoption of corpus-based methodologies were also evident, with significant contributions from countries like Malaysia, Iran, Greece, and Chile over different periods. These geographical trends suggest that educational policies, linguistic priorities, and institutional support play a crucial role in shaping how corpus research is applied. The regional focus of some studies indicates that localized educational challenges drive the emphasis on specific linguistic features. For example, early research in Malaysia and Iran primarily addressed fundamental grammatical issues in textbooks, while more recent studies in South Korea and Thailand have leveraged corpus methodologies to refine textbook content at a syntactic and collocational level. This progression further illustrates the interplay between the Noticing Hypothesis and Communicative Language Teaching. While earlier studies sought to correct overt linguistic inaccuracies, contemporary research is more focused on enhancing fluency and interaction by ensuring that textbooks reflect the lexical and structural patterns of real-world discourse. This progression further illustrates the interplay between the Noticing Hypothesis and Communicative Language Teaching. While earlier studies sought to correct overt linguistic inaccuracies, contemporary research is more focused on enhancing fluency and interaction by ensuring that textbooks reflect the lexical and structural patterns of real-world discourse (Kim & Oh, 2020; Boulton & Cobb, 2017). This supports previous findings that contextual factors influence corpus-based research applications (Boulton & Cobb, 2017; Callies, 2019).

Despite the clear benefits of corpus-based approaches, several challenges remain. One primary concern is the difficulty of integrating corpus findings into classroom practice, particularly in regions where traditional teaching methods still dominate. While corpus research offers valuable insights into language use, many educators lack the training or institutional support to effectively incorporate these findings into their teaching strategies (Xodabande & Nazari, 2022). This disconnect between research and classroom application necessitates targeted professional development programs that equip educators with the necessary skills to utilize corpus-based resources (Flowerdew, 2015). Moreover, teachers' attitudes toward corpus-based teaching can be influenced by institutional factors and pedagogical context, as demonstrated by Poole (2022), who found that even with positive attitudes toward technology, teachers were often constrained by the demands of traditional curricula.

Another limitation is the scope and generalizability of corpus analyses. While corpora provide extensive data on language patterns, they are often limited by the specific texts they include. For example, textbooks analyzed in corpus-based studies may not fully capture regional language variations or genre-specific linguistic norms. Li (2024) argues that many corpora primarily reflect native speaker usage, potentially underrepresenting the linguistic diversity found in global English contexts. Future research should expand corpus datasets to include a more diverse range of text types, registers, and dialects to enhance the applicability of findings (McEnery, Xiao, & Tono, 2006). As research evolves, corpora should strive to better represent the language needs of learners in specific sociocultural and regional settings,

a point emphasized by Arshad (2020) in their study of localized textbook revisions in EFL contexts.

Practical Implications and Future Research Directions

Given the increasing reliance on corpus-based methodologies in language education and textbook analysis, there are several practical implications that educators, policymakers, and curriculum developers should consider:

Enhanced teacher training

Institutions should implement professional development programs that equip teachers with the skills to interpret and apply corpus-based findings in classroom instruction. This would help bridge the gap between research insights and practical teaching applications.

Integration of corpus data into curriculum design

Textbook authors and curriculum developers should collaborate with corpus linguists to ensure that language materials reflect authentic usage patterns. This could involve the use of data-driven learning techniques, where students actively engage with corpora to explore language in context.

Leveraging AI for corpus-based textbook analysis

With advances in artificial intelligence (AI) and machine learning, corpus-based textbook analysis can be automated and enhanced to identify linguistic patterns more efficiently. AI-powered tools can analyze vast datasets, detect trends in language use, and even suggest automated textbook revisions that align with authentic language models. Future research should explore how AI-driven corpus analysis can facilitate real-time updates in textbook development, ensuring continuous alignment with evolving language trends.

Expanding research on corpus-based textbook revisions

Future studies should examine how corpus-informed revisions impact language learning outcomes, providing empirical evidence on the efficacy of these changes in classroom settings. More longitudinal studies are needed to assess whether these revisions lead to measurable improvements in language proficiency.

Conclusion

In conclusion, the evolution of corpus-based methodologies has had a profound impact on the accuracy and relevance of textbook content in language education. As these methodologies continue to develop, they offer valuable insights for educators and curriculum developers seeking to align instructional materials with authentic language use. Future research should continue to explore these trends, with a particular focus on expanding the regional application of advanced corpus methodologies and further integrating these findings into pedagogical practices.

ORCID

 <https://orcid.org/0000-0002-6393-2759>

 <https://orcid.org/0000-0003-3673-6127>

Acknowledgements

Not applicable.

Funding

Not applicable.

Ethics Declarations

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

Open Access

This article is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which grants permission to use, share, adapt, distribute and reproduce in any medium or format provided that proper credit is given to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if any changes were made.

References

- Alavi, S., & Rajabpoor, A. (2015). Analyzing idioms and their frequency in three advanced ILI textbooks: A corpus-based study. *English Language Teaching*, 8(1), 170-179. <http://doi:10.5539/elt.v8n1p170>
- Amare, T. (2022). *Effects of corpus-based instruction on EFL students' academic writing skills, critical thinking skills, engagement, and perception*. Doctoral dissertation, Bahir Dar University. <http://ir.bdu.edu.et/handle/123456789/13874>
- Arellano, R. (2018). A corpus linguistics application in the analysis of textbooks as national teaching instruments of English as a Second Language in Chile. *Actualidades Investigativas en Educación*, 18(1), 376-395. <http://doi:10.15517/aie.v18i1.31807>
- Biber, D., Reppen, R., & Conrad, S. (2002). Developing linguistic literacy: Perspectives from corpus linguistics and multi-dimensional analysis. *Journal of Child Language*, 29(2), 449-488. <http://doi:10.1017/S0305000902235345>
- Biber, D., Reppen, R., & Friginal, E. (2010). *Research in corpus linguistics*. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed.). Oxford Handbooks. <https://doi.org/10.1093/oxfordhb/9780195384253.013.0038>
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. <http://doi.org/10.1111/lang.12224>
- Callies, M. (2019). Integrating corpus literacy into language teacher education. *Learner corpora and language teaching*, 245-263. <https://doi.org/10.1075/scl.92.12cal>
- Chan, H. (2021). A corpus-based approach to comparing English textbook pedagogies: The case of Hong Kong and the People's Republic of China. *Pedagogies: An International Journal*, 16(3), 256-285. <https://doi.org/10.1080/1554480X.2020.1763799>
- Chareonkul, C. (2020). A corpus-driven study of the English present perfect in general and textbook corpora. Chulalongkorn University, Theses and Dissertations (Chula ETD). 42. <https://digital.car.chula.ac.th/chulaetd/42>
- Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge University Press.
- Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27(4), 537-554. <http://doi.org/10.1177/0265532209354771>
- Flowerdew, J. (2012). *Discourse in English language education*. Routledge. <https://doi.org/10.4324/9780203080870>
- Flowerdew, L. (2015). Data-driven learning and language learning theories. *Multiple Affordances of Language Corpora for Data-Driven Learning*, 69, 15-36.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Friginal, E., Prado, M., & Roberts, J. (2025). *Advancements in Teaching and Assessment Practices for Global Aviation English*. Bloomsbury. <https://www.bloomsbury.com/us/teaching-and-assessment-in-global-aviation-english-9781350411821/>
- Gedik, T. A., & Kolsal, Y. S. (2022). A corpus-based analysis of high school English textbooks and English university entrance exams in Turkey. *Theory and Practice of Second Language Acquisition*, 8(1), 157-176. <https://doi.org/10.31261/TAPSLA.9152>

- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97-118. <https://doi.org/10.1017/S0261444807004144>
- Hoang, H., & Crosthwaite, P. (2024). A comparative analysis of multiword units in the reading and listening input of English textbooks. *System*, 121, Article 103224. <https://doi.org/10.1016/j.system.2024.103224>
- Khany, R., & Kamalvand, A. (2022). 100 years of research on English language learning/teaching materials: A systematic literature review. *TESL-EJ*, 25(4), n4.
- Khojasteh, L., & Mukundan, J. (2011). A pedagogic corpus analysis: Modal auxiliary verbs in Malaysian English textbooks. *Journal on English Language Teaching*, 1(2), 45-55.
- Khojasteh, L., & Shokrpour, N. (2015). The “permission/possibility/ability” modals in Malaysian English textbooks: A corpus-based analysis. *Khazar University Press*, 18(2). <http://dspace.khazar.org/handle/20.500.12323/3258>
- Kim, Y. S., & Oh, S.-Y. (2020). A corpus-based analysis of collocations in Korean middle and high school English textbooks. *Language Research*, 56(3), 437-461. <https://doi.org/10.30961/lr.2020.56.3.437>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: OUP.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Lee, H.-K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research*, 37(1). <https://doi.org/10.17250/khisli.37.1.202003.005>
- Liu, Y., May, S., & Zhang, L. J. (2024). The ideological underpinnings in university English textbooks in China: A critical discourse analysis and corpus linguistics approach. *Journal of Multilingual and Multicultural Development*, 45(10), 4448-4467. <https://doi.org/10.1080/01434632.2023.2173208>
- Lynn, E. M. (2021). *The Linguistic nature of ESL/EFL reading textbooks: A corpus analysis. dissertation*, Northern Arizona University.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30. [https://doi.org/10.1016/S0346-251X\(99\)00058-5](https://doi.org/10.1016/S0346-251X(99)00058-5)
- Menon, S., & Mukundan, J. (2010). Analysing collocational patterns of semi-technical words in science textbooks. *Pertanika Journal of Social Sciences and Humanities*, 18(2), 241-258.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Molavi, A., Koosha, M., & Hosseini, H. (2014). A comparative corpus-based analysis of lexical collocations used in EFL textbooks. *Latin American Journal of Content & Language Integrated Learning*, 7(1), 66-81. <https://doi.org/10.5294/4087>
- Nation, I., & Webb, S. (2011). Content-based instruction and vocabulary learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Volume II (pp. 631–644). New York: Routledge/Taylor & Francis.
- Norberg, C., & Nordlund, M. (2018). A corpus-based study of lexis in L2 English textbooks. *Journal of Language Teaching and Research*, 9(3), 463-473. <http://dx.doi.org/10.17507/jltr.0903.01>
- O’Keeffe, A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University.
- Roslim, N., Aziz, A., Abdullah, M. H. T., & Nimehchisalem, V. (2021). Corpus-informed prepositions: What to learn and what is suitable for learning. *Journal of Asia TEFL*, 18(3), 1003.
- Sato, K., & Kleinsasser, R. C. (1999). Communicative language teaching (CLT): Practical understandings. *The Modern Language Journal*, 83(4), 494-517. <https://doi.org/10.1111/0026-7902.00037>
- Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied Linguistics*, 11(2), 129-158.
- Syahid, A., Julianto, A., Maulana, A., Zidan, M. R., & Rahman, M. H. (2024). An analysis of the ESP Textbook 'English for specific purposes for meetings' based Tomlinson's Theory. *Inspirasi Dunia: Jurnal Riset Pendidikan dan Bahasa*, 3(2), 96-106. <https://doi.org/10.58192/insdun.v3i2.2080>
- Tomlinson, B. (2011). *Materials development in language teaching*. Cambridge University Press.
- Vyatkina, N. (2020). Corpus-informed pedagogy in a language course: Design, implementation, and evaluation. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching* (pp. 306–335). Hershey, PA: IGI Global.
- West, M. (1953). *A general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longman.
- Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Zarifi, A., & Mukundan, J. (2012). Phrasal verbs in Malaysian ESL textbooks. *English Language Teaching*, 5(5), 9-18. <https://doi.org/10.5539/elt.v5n5p9>