

Edina Rizvić-Eminović
Mersad Dervić

COLLOCATIONAL COMPETENCE IN L2 ENGLISH: ORIGIN, DEFINITIONS, EMPIRICAL RESEARCH AND PEDAGOGICAL IMPLICATIONS

Abstract

This review article examines the concept of collocational competence in learning English as a foreign language, considering its theoretical foundations, definitional development, empirical research, and pedagogical implications. Drawing on Firth's principle that word meaning is shaped through its typical co-occurrence patterns and Halliday's concept of lexicogrammar, collocational competence is situated within contemporary theories of phraseology. The paper synthesizes findings from corpus-based studies and classroom research, which demonstrate that collocational knowledge plays a crucial role in the development of idiomaticity, fluency, and overall communicative competence across diverse educational contexts. Existing research consistently shows that collocational competence develops more slowly than vocabulary size, and that even advanced learners experience difficulties in the productive use of conventionalized word combinations. Verb–noun collocations, low-frequency and semantically opaque combinations, as well as collocations that are incongruent with learners' first-language patterns, are particularly problematic. These findings confirm that collocational knowledge does not emerge spontaneously through exposure to language alone but requires targeted and systematic pedagogical intervention. The paper further highlights the importance of explicit, corpus-informed, and strategy-based instruction, which has proven significantly more effective than implicit learning. Successful approaches include working with authentic corpora, developing noticing and memory strategies, and integrating collocations into tasks focused on real-life language use. At the same time, key methodological and conceptual challenges are identified, including inconsistent definitions, a limited number of longitudinal studies, and insufficiently contextualized assessment models. The study concludes by emphasizing the need for integrated curricular frameworks and mixed-method assessment approaches that link collocational competence to communicative outcomes across different proficiency levels.

Keywords: collocational competence, phraseology, corpus linguistics

KOLOKACIJSKA KOMPETENCIJA U ENGLISKOM KAO STRANOM JEZIKU: PORIJEKLO, DEFINICIJE, EMPIRIJSKA ISTRAŽIVANJA I PEDAGOŠKE IMPLIKACIJE

Sažetak

Ovaj pregledni rad analizira pojam kolokacijske kompetencije u učenju engleskog kao stranog jezika, sagledavajući teorijsko utemeljenje samog pojma, njegove definicija, empirijska istraživanja i pedagoške implikacije. Polazeći od Firthovog principa da se značenje riječi oblikuje kroz njene tipične spojeve sa drugim riječima i Hallidayevog koncepta leksikogramatike, kolokacijska kompetencija se pozicionira unutar savremenih teorija frazeologije. U radu se sintetiziraju nalazi korpusnih studija, te studija provedenih u učionicama, koje pokazuju da kolokacijsko znanje ima ključnu ulogu u razvoju idiomatičnosti, tečnosti i ukupne komunikacijske kompetencije u različitim obrazovnim kontekstima. Rezultati dosadašnjih istraživanja dosljedno ukazuju na to da se kolokacijska kompetencija razvija sporije od samog vokabulara, te da i napredni učenici pokazuju poteškoće u produktivnoj upotrebi ustaljenih leksičkih spojeva. Posebno su problematične glagolsko-imeničke kolokacije, rijetke i semantički netransparentne kombinacije, kao i kolokacije koje nisu podudarne s obrascima maternjeg jezika. Takvi nalazi potvrđuju da kolokacijsko znanje ne nastaje spontano kroz izloženost jeziku, već zahtijeva ciljanu i sistematsku pedagošku intervenciju. U radu se nadalje ukazuje na značaj eksplicitne, korpusno utemeljene i strategijski vođene nastave, koja se pokazala znatno efikasnijom od implicitnog učenja. Uspješni pristupi uključuju rad s autentičnim korpusima, razvijanje strategija uočavanja i pamćenja, te integraciju kolokacija u zadatke usmjerene na stvarnu upotrebu jezika. Istovremeno se identificiraju ključni metodološki i konceptualni izazovi, među kojima su neujednačene definicije, ograničen broj longitudinalnih studija i nedovoljno kontekstualizirani modeli procjene. U zaključku se naglašava potreba za integriranim kurikularnim okvirima i kombiniranim metodama vrednovanja koje kolokacijsku kompetenciju povezuju s komunikativnim ishodima na različitim nivoima jezičke kompetencije.

Ključne riječi: kolokacijska kompetencija, frazeologija, korpusna lingvistika

Introduction

Research on second language (L2) vocabulary increasingly emphasizes collocational competence - learners' knowledge of which words typically co-occur and how those combinations function in use. Foundational work argued that meaning is shaped by patterns of co-occurrence (Firth, 1957) and by the lexicogrammatical organization of language (Halliday, 1961). Later corpus-driven accounts situated collocations within the broader scope of formulaic language, emphasizing that fluent speakers rely on prefabricated multiword patterns, the "idiom principle," rather than constructing every utterance from first principles (Sinclair, 1991; Wray, 2002). Within contemporary vocabulary theory, collocational knowledge is widely treated as a dimension of depth, as opposed to breadth or size, alongside knowledge of word parts, associations, and grammatical behaviour (Nation, 2001; Gyllstad, 2007; Schmitt, 2014). These converging perspectives motivate a focused review of collocational competence in L2 English with particular attention to recent classroom-based and corpus-informed work (Rizvić-Eminović, Dervić, & Radoš, 2025).

Collocational competence is understood as learners' receptive and productive command of recurrent, conventionalized word-word combinations such as *make a decision*, *heavy rain*, or *commit a crime*, together with sensitivity to their frequency, semantic transparency, grammatical patterning, and register-specific distribution (Barfield & Gyllstad, 2009; Benson, Benson, & Ilson, 1986/2010; Henriksen, 2013). This definition integrates two traditions often distinguished in the literature, a frequency-based corpus or statistical perspective that identifies collocations via association measures, and a phraseological perspective that emphasizes conventionality and restricted co-occurrence (Henriksen, 2013). It also aligns collocations with the constructs of formulaic language and lexical competence (Dervić & Bećirović, 2020; Nation, 2001; Wray, 2002; Schmitt, 2014).

Empirical research links stronger collocational competence to more idiomatic L2 production, higher ratings of writing quality, and greater processing efficiency (Dervić & Spahić, 2018; Henriksen, 2013; Schmitt, 2014). Corpus linguistics has provided the methodological toolkit, including large reference corpora, association statistics, and concordancing, to operationalize collocational knowledge for research, pedagogy, and assessment (Sinclair, 1991; Nation, 2001; Henriksen, 2013). This has been complemented by local classroom studies of explicit writing instruction and corpus-based activities (Rizvić-Eminović, Dervić, & Bilić, 2025; Rizvić-Eminović & Neslanović, 2022). At the same time, the field continues to debate the details of definitions, i.e. what "counts" as a collocation, as well as the mechanisms by which receptive recognition transfers to productive, context-appropriate use (Henriksen, 2013; Schmitt, 2014).

This review synthesizes foundational theory and recent applied studies on collocational competence in L2 English with a focus on pedagogical implications for EFL and ESL contexts. It outlines conceptual developments, consolidates empirical evidence, and presents implications for curriculum, instruction, and assessment. The review is organized around the following questions:

1. How has collocational competence been defined across phraseological and corpus traditions, and how does it relate to communicative competence?
2. What is the importance of collocational competence for L2 learners across languages?
3. Which approaches and strategies can be used to develop collocational competence in L2 English?

Key historical and theoretical foundations

The concept of collocational competence is rooted in a long tradition of research on lexical relationships and phraseology. One of the earliest and most frequently cited statements about collocations is the one by Firth (1957:11), who remarked that “you shall know a word by the company it keeps.” This statement is important because it emphasizes that lexical meaning is not an inherent property of an individual. Instead, the meaning of words emerges through patterns of co-occurrence with other words. Such an insight shifted the focus of lexical studies from isolated words to recurrent word combinations, laying the foundation for what would later be termed collocations. Building on these ideas, Halliday (1961) advanced the concept of lexicogrammar, in which lexis and grammar are viewed as inseparable components of a single system. In his early functional linguistic work, Halliday proposed quantitative ways of defining collocation. Namely, collocations were seen as syntagmatic associations of lexical items, which were measured as probabilities within a certain word span. This concept was seen as a predecessor of the modern corpus-based approach.

The next major development came from Sinclair (1991), who contrasted the idiom principle with the open choice principle. According to the idiom principle, language users do not construct each sentence from scratch but draw heavily on pre-assembled multiword units such as collocations, formulaic expressions, and other phraseological patterns. This theoretical framing was crucial for conceptualising collocational competence, which was defined as the ability to recognise, process, and produce such expressions in real-time communication. Parallel to these theoretical advances, the rise of corpus linguistics and statistical approaches to lexis made it possible to identify and measure collocations empirically. Large electronic corpora enabled researchers to compute association measures such as mutual information (MI) and t-scores, to model distributional patterns and interpret the semantics of co-occurring words.

More recent research in phraseology and formulaic language situates collocations as a central component of lexical competence rather than as peripheral vocabulary. Wood (2015), Schmitt (2014), and Wray (2002) all highlight the role of formulaic sequences in language acquisition and use. Henriksen (2013) provides a comprehensive overview of L2 collocational research, emphasising its developmental and pedagogical dimensions. Within this framework, collocational competence is typically conceptualised as a dimension of depth of vocabulary knowledge, or knowledge of how words combine and function in context, rather than as a simple measure of vocabulary breadth. Additionally, collocational competence represents an important component of communicative competence and fluency, as it connects the knowledge of forms with their appropriate use across registers and genres. In regional contexts, overviews and practice-oriented work have mapped concrete pedagogical pathways for integrating collocations into English teaching (Rizvić-Eminović & Kasumagić-Kafedžić, 2024; Bureković, Rizvić-Eminović, & Pilav, 2023).

Collocational competence is widely treated as a depth construct, reflecting knowledge about how words behave and co-occur rather than a count of how many words learners know. In Nation’s (2001) framework, depth encompasses multiple facets – form, meaning, and use – including grammatical relations and typical companions and patterns of words. Empirically, depth adds unique explanatory power to proficiency outcomes. For example, in academic reading, depth measures that include collocational knowledge predict performance beyond vocabulary size alone (Qian, 2002). Reviews similarly conclude that depth and size are separable, with depth typically contributing additional variance to L2 performance (Schmitt, 2014).

From a phraseological perspective, collocations are a core subset of formulaic sequences, recurrent and conventionalised multiword units that speakers store and access holistically. Wray (2002) argues that a considerable proportion of everyday language is formulaic and appears to be represented as fixed or semi-fixed chunks. Collocational competence, on this view, reflects the ability to recognise and produce preferred, conventional combinations that underpin native-like selection and fluency, a phenomenon long noted as a central puzzle of linguistic theory (Pawley & Syder, 1983), and is consistent with corpus-based descriptions of usage (Sinclair, 1991).

In communicative-competence models, collocational competence aligns most directly with linguistic or grammatical competence, which encompasses knowledge of lexis and lexicogrammar, while also supporting discourse and sociolinguistic competence. Classic formulations (Canale & Swain, 1980) identify the lexical system as part of grammatical competence, and later pedagogical models by Celce-Murcia and colleagues make this linkage more explicit, some versions even including a formulaic or interactional component that acknowledges the importance of stored multiword patterns for real-time communication (Celce-Murcia, 2008). Collocational competence, therefore, functions as a bridge between linguistic knowledge and fluent, appropriate use of language across genres and registers.

Why collocational competence matters

Collocational competence underlies idiomaticity, fluency, and overall communicative adequacy in L2 use. Learners who master native-like combinations produce more natural and fluid speech and writing, avoiding awkward or overly general phrasing (Dervić & Bećirović, 2019). Durrant and Schmitt (2009) observed that non-native writers often rely heavily on extremely frequent collocations while underusing less common but strongly associated ones, which leads to output that is grammatical but non-idiomatic. In spoken language, Boers et al. (2014) found that learners who used a greater range of formulaic sequences were rated higher in oral proficiency, and experimental instruction that encouraged noticing of such sequences led to measurable improvements in fluency. With respect to writing assessment, Bestgen and Granger (2014) demonstrated that phraseological richness may be used as a predictor of overall text quality.

Research across languages indicates difficulties with collocations even among advanced learners. This is particularly the case with verb–noun combinations and collocations that lack direct L1 equivalents (Nesselhauf, 2005; Arnaut-Karović & Rizvić-Eminović, 2015). Durrant and Schmitt (2009) found that while learners' use of high-frequency collocations approached native norms, their use of low-frequency collocations remained significantly lower, which indicates limited depth of collocational competence. Eguchi (2023) reported that greater use of highly associated collocations correlates positively with proficiency measures across tasks, and Bestgen and Granger (2014) showed that texts lacking collocationals often receive lower quality ratings despite grammatical accuracy. Evidence from Bosnian university contexts links explicit instruction to improvements in writing performance and reductions in error-prone phrasing, which is reflected in the students' collocational competence (Rizvić-Eminović, Bureković, & Dervić, 2025). Further, studies of B2 students confirm significantly lower productive than receptive competence with respect to the use of grammatical collocations, with collocations integrating prepositions presenting the greatest difficulties for Bosnian EL learners (Rizvić-Eminović & Arnaut-Karović, 2015). The results improve as the students' proficiency level increases from B1 to C1; however, the difficulties remain unchanged (Arnaut-Karović & Rizvić-Eminović, 2015). These findings demonstrate that collocational development contributes not only to linguistic naturalness but also to measurable performance results.

Phraseological complexity of collocations, including their association strength, variety, and frequency, has become integrated into automated scoring systems and CEFR-aligned writing rubrics (Bestgen & Granger, 2014). Collocations are also register-sensitive and genre-specific: academic writing relies on a distinct set of high-association combinations, while conversational speech favours less fixed patterns. Growth in collocational competence thus enables learners to meet discourse expectations across genres and registers.

Approaches, methods, and strategies for improving collocational competence

Research on the teaching of collocations distinguishes primarily between explicit and implicit approaches. Explicit instruction, which directs learners' attention to collocations through explanation, contrastive examples, and data-driven teaching practice, has been shown to yield both short- and long-term gains in learning outcomes (Boers et al., 2014; El-Dakhs, 2015; Peters, 2016; Rizvić-Eminović & Hadžić, 2021). El-Dakhs (2015) demonstrated that learners who received targeted instruction on verb–noun collocations significantly outperformed those who encountered them incidentally in reading. Implicit instruction through exposure or extensive reading, while useful for developing receptive awareness, tends to have more limited effects on productive mastery (Peters, 2016). The lexical approach proposed by Lewis (1993) reconceptualised vocabulary learning around multiword chunks rather than individual words, arguing that collocations form the foundation of fluent language use. Task-based and genre-based activities, which embed collocations in meaningful contexts, are perceived as encouraging by learners and help them actively use such collocations in speech and writing (Barfield & Gyllstad, 2009; Chen & Flowerdew, 2018). Through collaborative writing or information-gap tasks, learners advance from noticing to accurate use of collocations within authentic discourse. Further, formulaic language teaching, as discussed by Wray (2002) and Wood (2015), emphasises awareness-raising and fluency development, thus placing focus on collocational learning as a part of broader phraseological competence rather than as isolated vocabulary study.

Studies show that learners benefit from strategies that enhance noticing, retention, and active use of collocations. Recording new collocations, comparing L1-L2 patterns, and using bilingual or collocation dictionaries have been shown to increase accuracy and recall (Fan, 2009). The use of corpora and concordancers further promotes learner autonomy, enabling students to investigate authentic usage, observe frequency patterns, and infer collocational norms (Rizvić-Eminović & Hadžić, 2021). A study with Croatian EL learners indicates that more proficient learners often employ deliberate strategies, such as grouping collocations semantically or noting syntactic frames, which correlate with higher collocational test performance (Tabak & Pavičić Takač, 2023). Classroom resources can also assist in developing collocational competence. Teachers can include authentic genres, such as song lyrics, to raise awareness of collocational patterns and translation incongruencies relevant to Bosnian learners (Rizvić-Eminović, Šukalić, & Tursanović, 2024).

The use of technological tools has significantly reshaped the way collocations are taught and learned. Through computer-assisted language learning and corpus-based methods, students today can explore authentic language data using online concordancers, collocation dictionaries, and large corpora such as COCA and the BNC (Boulton, 2010; Daskalovska, 2015). Integrating corpus exploration into classroom practice encourages learners to observe patterns and draw their own conclusions, thereby strengthening their awareness of typical word combinations. Automated feedback tools further support this process by drawing attention to inappropriate collocations and offering suitable alternatives. In this way, form-focused learning is closely connected to communicative use (Chen & Flowerdew, 2018). In blended and online contexts,

interactive digital platforms can enhance learner independence and motivation, although their effectiveness depends on adequate guidance and the users' digital competence (Arkhipova & Vlavatskaya, 2021).

Testing practices in the area of collocational competence remain challenging, largely because collocations vary in transparency, productivity, and the strength of association between their elements. Researchers generally distinguish between *receptive* tests, which measure learners' ability to recognise correct combinations, and *productive* tests, which assess their capacity to use collocations appropriately in context through cloze, recall, or translation tasks (Gyllstad, 2007). In addition to traditional testing formats, corpus-based analyses of learner writing have become valuable assessment tools, as they can quantify performance through measures such as collocational diversity, frequency, and accuracy (Bestgen & Granger, 2014; Paquot, 2018). Henriksen (2013) argues that a balanced approach, which combines controlled test formats with naturalistic data from essays or spoken transcripts, offers a more complete view of both learners' explicit knowledge and their ability to use collocations in communication. As language proficiency frameworks continue to develop, collocational competence is now increasingly regarded as a central and measurable dimension of lexical and communicative proficiency (Dervić & Bećirović, 2020; Schmitt, 2014).

Research findings consistently favour explicit instruction over incidental exposure when it comes to learning collocations. Boers et al. (2014) showed that treating collocations as unified lexical items leads to better retention than traditional matching or translation exercises. Similarly, Daskalovska (2015) demonstrated that data-driven learning, where learners use concordancers to observe authentic usage patterns, results in more durable learning outcomes than relying solely on textbook examples. Peters (2016) further observed that collocations which are infrequent or semantically incongruent demand explicit, repeated practice to be fully acquired.

Studies on individual learner differences indicate that students often recognise collocations without being able to use them accurately in production. This gap is largely influenced by first language transfer, the amount and quality of exposure, and the learning strategies employed (Gyllstad, 2007; Henriksen, 2013). Psycholinguistic research further shows that even advanced learners process congruent collocations more quickly and accurately than incongruent ones (Yamashita & Jiang, 2010; Wolter, 2013). Learners who demonstrate greater strategic awareness, such as effective use of dictionaries and conscious reflection on language, tend to achieve higher levels of collocational proficiency (Boers et al., 2014; Tabak & Pavičić Takač, 2023). Collectively, these findings highlight the importance of instructional approaches that incorporate strategy training, corpus consultation, and explicit feedback to narrow the gap between recognition and productive use.

Learner corpus research consistently confirms that collocational competence increases with proficiency, though more gradually than vocabulary breadth. Bestgen and Granger (2014), using the CollGram tool to analyse L2 essays, found that mean mutual information (MI) scores of bigrams correlated positively with writing quality, with higher-rated texts containing fewer unattested combinations and greater use of strongly associated pairs. They also observed that as learners progressed, their reliance on high-frequency but less native-like combinations declined. Similarly, Durrant and Schmitt (2009) reported that while advanced L2 writers approximate native norms on very frequent collocations (high *t*-score), they underuse low-frequency, high-association (high-MI) collocations, indicating a preference for safe, overlearned combinations and avoidance of more exclusive or idiomatic pairings. Again, these findings suggest that instruction should move beyond frequent word combinations to explicitly

develop learners' awareness and productive use of less common, strongly associated collocations through corpus-informed, strategy-based, and feedback-supported activities.

Error patterns concentrate around verb–noun collocations, such as *do a mistake* instead of *make a mistake* or *take a decision* instead of *make a decision* (Nesselhauf, 2005). Learners also face difficulties with adjective–noun and phrasal-verb–noun combinations, particularly when the L1 lacks a direct equivalent or when the collocation is semantically opaque (Peters, 2016). Durrant and Schmitt (2009) attributed this to the resistance of tightly bound, low-frequency collocations to acquisition. Cross-linguistic research further supports a strong congruency effect: collocations congruent with L1 patterns are acquired and processed more easily than incongruent ones, with speed and accuracy moderated by frequency and proficiency (Yamashita & Jiang, 2010; Wolter, 2013; Jiang et al., 2022). These findings imply that teaching should prioritise contrastive and corpus-informed approaches that raise learners' awareness of L1–L2 differences, provide focused practice with low-frequency and semantically opaque collocations, and integrate corrective feedback and repetition to promote accurate verb–noun and adjective–noun usage.

Instructional comparisons across classrooms consistently show that explicit, targeted instruction outperforms incidental exposure. Boers et al. (2014) found that teaching verb–noun collocations as intact lexical chunks produced better short-term and long-term recall than matching or decontextualised exercises. Corpus-based, data-driven learning approaches yield similar results. Daskalovska (2015) demonstrated that learners who explored concordance lines to identify collocational patterns performed significantly better on post-tests than those relying on textbook examples. Peters (2016) emphasised that low-frequency and incongruent collocations carry a higher learning burden and thus require sequenced guidance, spaced repetition, and explicit focus.

Across individuals, a persistent receptive–productive gap remains evident: learners can often recognise collocations they cannot yet produce (Gyllstad, 2007). This asymmetry is shaped by L1 influence, input frequency, motivation, and learning strategies. Psycholinguistic studies confirm faster, more accurate processing of congruent collocations even among advanced learners, suggesting that L1 interference continues to constrain acquisition (Dervić & Spahić, 2018; Yamashita & Jiang, 2010; Wolter, 2013). Learners who record collocations systematically, consult corpora or collocation dictionaries, and apply metacognitive planning tend to perform better on collocation tests (Tabak & Pavičić Takač, 2023). Because of the high difficulty of low-frequency, incongruent items, researchers increasingly recommend integrated interventions that combine explicit instruction, corpus-based discovery, and spaced retrieval practice.

Language-specific research provides further support for the broader trends in collocational development. Studies with Chinese learners of English, for instance, show an overreliance on a limited range of high-frequency combinations and persistent verb–noun difficulties even at advanced levels (Wang, 2023; Du et al., 2022). When guided by appropriate scaffolding, data-driven tools such as SkELL and Sketch Engine have been shown to enhance collocational learning. Yue and Ruangjaroon (2025) found that for intermediate Chinese learners, collocational accuracy, associative strength, and diversity improved with proficiency, though unevenly across spoken and written modes. Research on Japanese learners similarly confirms strong congruency and frequency effects: collocations that overlap with L1 structures are processed more efficiently, while proficiency, immersion, and instructional input moderate performance (Chen & Flowerdew, 2018; Wolter, 2013). Even advanced learners exhibit processing patterns distinct from native speakers, as shown by Matsuno (2017), reinforcing the

view that chunk-based processing underlies collocational proficiency. Comparable evidence from studies on Spanish and Arabic learners points to limited collocational range, idiomaticity gaps, and persistent verb–noun errors driven by L1 transfer (González & Ramos, 2013; Gudmestad, 2022; Alangari, 2019; Alsakran, 2011). For Spanish learners, error corpora and automated feedback tools now support targeted correction (Rodríguez-Fernández & Casillas, 2018). Similar tendencies appear in learners of Chinese as a foreign language, where verb–noun collocations are particularly problematic and acquisition patterns follow frequency and transparency effects (Cai, 2017). Taken together, these findings suggest that certain “hard zones” in collocational learning are consistent across languages and learner populations, underscoring the need for explicit, corpus-informed pedagogical support (El-Dakhs, 2015; Shiha, 2024).

Within Bosnian, Croatian, and Serbian learner populations, comparable trends are observed. Among Bosnian students of English, research identifies a clear gap between receptive and productive knowledge, with verb–noun combinations posing persistent challenges (Begagić, 2014; Rizvić-Eminović & Arnaut-Karović, 2015). Analyses of B2-level essays show a predominance of lexical over grammatical collocations and frequent avoidance of idiomatic collocations (Rizvić-Eminović, Dervić, & Bilić, 2025). Additionally, studies explore the university learners’ collocation use by using different task types (Rizvić-Eminović, Bujak, & Bureković, 2018; Rizvić-Eminović & Bujak, 2017). Croatian studies reinforce the importance of congruency, demonstrating that learners produce congruent collocations more accurately and that proficiency-related gains are limited mainly to frequent, transparent combinations (Patekar, 2022; Lukač, 2016; Lovrić, 2020). In ESP and Business English contexts, learners also show gaps in verb–noun and noun–noun collocations, leading to calls for corpus-based syllabi (Miščin et al., 2024/2025). Although fewer studies focus specifically on Serbian learners, contrastive analyses indicate parallel difficulties. Work on English-Serbian collocations in tourism discourse reveals systematic differences in noun–noun patterning, pointing to transfer effects (Vojnović, 2021). Related research on dictionary use and strategy awareness highlights the importance of collocation-conscious learning behaviours (Knežević, 2021; Tabak & Pavičić Takač, 2023). Across Bosnian, Croatian, and Serbian contexts, learners struggle most with verb–noun combinations, incongruent collocations without direct L1 equivalents, and low-frequency or highly idiomatic pairs typical of academic and ESP registers. These findings strengthen the case for corpus-informed, explicit instruction complemented by strategy training and spaced retrieval to promote the transfer of receptive collocational knowledge into productive, context-appropriate use.

Challenges and gaps in research

Despite extensive research, several conceptual and practical challenges continue to shape how collocational competence is studied and taught. A major difficulty lies in defining what it actually means to *know* a collocation. The construct remains only partly delineated, with existing tests assessing mixed components such as form and meaning recognition or recall, often overlapping with broader measures of lexical depth. This limits the comparability of research findings and construct validity (Ding et al., 2024). Methodological variations, such as timing, association strength, and task type, further complicate interpretation.

A second long-standing issue concerns the transfer from receptive to productive use: learners frequently recognise collocations passively but fail to retrieve them accurately in communication. Tasks that integrate both receptive and productive processing tend to yield better results than exposure alone, although outcomes are influenced by frequency, association strength, and L1–L2 congruency (Zhang, 2017). Research findings also indicate that

collocational development progresses more slowly than vocabulary breadth, particularly for low-frequency, high-association combinations (Du et al., 2022), reinforcing the value of retrieval practice and contextualised use.

Phraseological development research also faces methodological gaps: most studies are short-term or cross-sectional, offering limited insight into longitudinal trajectories across skills or modalities. Although ongoing projects such as LONGDALE provide valuable learner-corpus data, large-scale, multi-site studies linking instructional practices to developmental outcomes remain rare (Bestgen & Granger, 2014; Vandeweerd, 2023). Scaling corpus-based pedagogy beyond controlled research settings poses additional challenges. While meta-analyses confirm the overall effectiveness of data-driven learning, generalisability remains limited. Namely, participants are typically university students, delayed post-tests are rare, and evidence of transfer beyond the immediate learning context is scarce (Boulton & Cobb, 2017; Boulton, 2010; Ding et al., 2024). Emerging work at pre-tertiary levels shows encouraging results but underscores the importance of teacher mediation and guidance (He, Hu, & Liang, 2024).

Teacher readiness and resource availability further constrain implementation. Many educators report limited corpus literacy and low confidence in integrating data-driven techniques into everyday instruction (Ma, 2023). Although case studies demonstrate that experienced teachers can successfully embed corpus use, doing so is time-intensive and dependent on institutional support (Ma, 2023). Ongoing professional development initiatives are addressing these needs, but more evidence is required on training models, classroom application, and learner impact (Bennett, 2023; Ma, 2023).

Finally, assessment practices often remain detached from communicative context. Standardised collocation tests isolate lexical pairs, while corpus-based tests lack agreement on which metrics best capture communicative competence (Bestgen & Granger, 2014; Du et al., 2022). Current reviews advocate mixed-method assessment frameworks, which would combine controlled receptive and productive testing with corpus-informed, performance-based evaluation, all anchored in explicit control of item features (Ding et al., 2024). Developing comprehensive, context-sensitive assessment models that integrate language authenticity remains essential for advancing both the measurement and pedagogy of collocational competence across languages.

Implications for teaching, pedagogy, and assessment

A growing body of evidence suggests that collocations should be introduced from the earliest stages of language learning and revisited regularly as proficiency develops. Well-designed curricula embed collocational learning explicitly across levels rather than treating it as an optional extension of vocabulary study. Initial instruction should prioritise high-frequency verb–noun and adjective–noun combinations, particularly those common in the target registers of academic or professional communication (Peters, 2016; Paquot, 2018; Bestgen & Granger, 2014). As learners gain confidence with these core patterns, teaching can gradually incorporate less frequent but more tightly associated collocations that enhance idiomaticity and naturalness. Research consistently shows that such combinations are among the strongest indicators of writing and speaking quality, underscoring the need for intentional selection, systematic recycling, and focused practice instead of relying on incidental exposure (Bestgen & Granger, 2014; Durrant & Schmitt, 2009).

Corpus-informed pedagogy and data-driven learning (DDL) approaches are particularly effective when implemented in blended formats. At the course level, combining teacher-

selected corpus input, such as concordance lines and collocation lists, with guided discovery tasks helps learners notice patterns and transfer them to production. Meta-analytic evidence confirms that corpus use enhances learning outcomes, particularly when activities are scaffolded and tailored to learners' proficiency levels (Boulton & Cobb, 2017). Classroom-oriented syntheses recommend practical scaffolding techniques, including small-scale searches, pre-highlighted concordance lines, collaborative interpretation, and explicit follow-up production tasks (Cheng, 2022).

Strategy training is essential to help learners become autonomous in acquiring collocations. Effective routines include noticing activities (e.g. underlining nodes and collocates), the use of bilingual or collocation dictionaries, corpus consultation, and systematic spaced retrieval (e.g. 1–3–7 day review cycles). Productive practice should involve slot-and-frame prompts that encourage active use (e.g., *commit* ____, *raise serious* ____). Instruction should prioritise collocations according to known difficulty factors, such as congruency, word-class pairing, association strength, and semantic transparency, since incongruent verb–noun pairs impose a higher cognitive load and require more intensive support (Peters, 2016).

Materials development should integrate collocations directly into core instructional texts rather than treating them as supplementary. Reading and listening materials can include pre-extracted collocational bundles with marginal glosses or annotation cues modelled on the BBI format (e.g., “V + N: *pose a threat, wreak havoc*”). Short contrastive tables clarifying near-synonyms (*do/make/take* + noun) help refine learners' intuitive understanding of collocational preferences. Collocation dictionaries (Benson, Benson, & Ilson, 2010) and learner-friendly corpus tools (e.g., COCA, Sketch Engine) should be embedded in task sheets and classroom activities to promote hands-on engagement.

Assessment practices need to reflect both collocational knowledge and its application in communicative contexts. Multi-method testing offers the most comprehensive approach, combining receptive recognition tasks (e.g., odd-one-out, matching), controlled productive formats (e.g., gapped frames, constrained translation), and performance-based tasks (e.g., writing or speaking analysed with corpus-based metrics) (Gyllstad, 2007; Bestgen & Granger, 2014). Test items should be calibrated according to statistical measures such as *t*-score and MI, and include both congruent and incongruent pairs to capture varying difficulty levels (Gyllstad, 2007; Paquot, 2018).

Analytic rubrics and automated indices can enhance feedback by linking linguistic evidence to quality judgments. In writing assessment, metrics such as the proportion of high-MI bigrams, collocational diversity, and accuracy align well with CEFR proficiency bands and provide actionable feedback for learners (Bestgen & Granger, 2014; Paquot, 2018). Teachers should move beyond generic comments like “awkward phrasing” and instead provide contrastive feedback that presents the target collocation alongside the learner's version. Corpus snippets can illustrate authentic usage, while automated feedback systems can flag miscollocations and propose alternatives (Chen & Flowerdew, 2018). Scheduling delayed revision sessions, typically within 24 to 72 hours, has been shown to enhance retrieval and long-term retention.

Implementing collocation-rich pedagogy on a larger, programmatic scale depends fundamentally on teachers' corpus literacy and the institutional conditions that support it. Research indicates that corpus literacy is a strong predictor of teachers' confidence and willingness to integrate corpora into classroom practice (Ma, 2023). Professional development initiatives should therefore prioritise the development of search skills, pedagogical mediation, and task design, progressing systematically from initial familiarisation to advanced corpus-

based assessment and feedback. Institutional provision of corpus access and intuitive, user-friendly tools can further lower the barriers to regular classroom use (Boulton, 2010; Ma, 2023). In the long term, meaningful progress in learners' collocational competence will rely not only on high-quality materials and assessment instruments but also on sustained teacher training and infrastructural support that embed corpus-informed pedagogy as a routine part of language instruction.

References

1. Alangari, M., 2019. Investigating Arab EFL learners' collocational competence: Verb–noun and adjective–noun combinations. *Arab World English Journal*, 10(3), pp.345–364.
2. Alsakran, R.A., 2011. The productive and receptive knowledge of collocations by advanced Arabic-speaking EFL learners. *International Journal of English Linguistics*, 1(1), pp.107–122.
3. Arkhipova, E.V. & Vlavatskaya, M., 2021. Digital technologies in phraseological competence development. *Education and Information Technologies*, 26, pp.1127–1145.
4. Arnaut-Karović, K. & Rizvić-Eminović, E., 2015. Poznavanje gramatičkih kolokacija među studentima engleskog jezika na Univerzitetu u Zenici. *Zbornik radova Filozofskog fakulteta univerziteta u Zenici*, 13, pp.23–38.
5. Barfield, A. & Gyllstad, H., 2009. *Researching collocations in another language: Multiple interpretations*. Palgrave Macmillan.
6. Begagić, M., 2014. Collocational competence of Bosnian EFL learners. *Zbornik radova Filozofskog fakulteta Univerziteta u Zenici*, 12, pp.131–146.
7. Benson, M., Benson, E. & Ilson, R., 2010. *The BBI combinatory dictionary of English: Your guide to collocations and grammar*. 3rd ed. John Benjamins.
8. Bestgen, Y. & Granger, S., 2014. Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, pp.28–41.
9. Boers, F., Demecheleer, M., Coxhead, A. & Webb, S., 2014. Gauging the impact of exercises on verb–noun collocations. *Language Teaching Research*, 18(1), pp.54–74.
10. Boulton, A., 2010. Data-driven learning: On paper, in practice. *Applied Linguistics*, 31(3), pp.390–412.
11. Boulton, A. & Cobb, T., 2017. Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), pp.348–393.
12. Canale, M. & Swain, M., 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), pp.1–47.
13. Celce-Murcia, M., 2008. Rethinking the role of communicative competence in language teaching. In: E. Alcón Soler & P. Safont Jordà, eds. *Intercultural language use and language learning*. Springer, pp.41–57.

14. Chen, Q. & Flowerdew, L., 2018. Applying corpus tools in feedback on academic writing. *Journal of English for Academic Purposes*, 33, pp.1–11.
15. Cheng, W., 2022. *Data-driven learning revisited: Pedagogical scaffolding and learner engagement*. Routledge.
16. Cai, Y. 2017. *Second language acquisition of Chinese verb–noun collocations*. Doctoral dissertation, University of Massachusetts Amherst.
17. Daskalovska, N., 2015. Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), pp. 130–144.
18. Dervić, M. & Bećirović, S., 2020. Prerogative of the lexical approach in communicative language teaching. *European Journal of Education Studies*, 7(3), pp.1–13.
19. Dervić, M. & Spahić, N., 2018. An evaluation of five key aspects of language acquisition (LA) – A critical review of Krashen’s theory. *ILUM*, 16(16), pp.391–408.
20. Dervić, M. & Bećirović, S., 2019. Native and non-native EFL teachers dichotomy: Terminological, competitiveness and employment discrimination. *Journal of Language and Education*, 5(3), pp.114–127.
21. Ding, Y., Reynolds, B.L. & Lee, J., 2024. Validating the construct of collocational competence. *Language Testing*, 41(2), pp.189–210.
22. Du, X., Afzaal, M. & Al Fadda, H., 2022. Collocation use in EFL learners’ writing across multiple proficiency levels: A corpus-driven study. *Frontiers in Psychology*, 13.
23. Durrant, P. & Schmitt, N., 2009. To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47(2), pp.157–177.
24. El-Dakhs, D., 2015. The lexical collocational competence of Arab EFL learners. *English Language Teaching*, 8(6), pp.112–126.
25. Eguchi, M. & Kyle, K., 2023. L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. *Journal of Second Language Writing*, 60.
26. Fan, M., 2009. An exploratory study of collocational use by ESL students. *Asian EFL Journal*, 11(1), pp.7–28.
27. Firth, J.R., 1957. *Papers in linguistics, 1934–1951*. Oxford University Press.
28. González, A.O. & Ramos, J., 2013. A comparative study of collocations in a native corpus and a learner corpus of Spanish, *Procedia – Social and Behavioral Sciences*, 95, pp. 83–92.
29. Gudmestad, A., 2022. Insights from Spanish learner corpus research. *Research in Corpus Linguistics*, 10(2), pp. 34–51.
30. Gyllstad, H., 2007. *Testing English collocations: Developing receptive tests for advanced learners*. Lund University Press.
31. Halliday, M.A.K., 1961. Categories of the theory of grammar. *Word*, 17(3), pp.241–292.

32. He, M., Hu, S. & Liang, C., 2024. The sustainable impact of data-driven learning on noun collocations in pre-tertiary EFL. *Humanities & Social Sciences Communications*, 11.
33. Henriksen, B., 2013. Research on L2 learners' collocational competence and development: A progress report. In: C. Bardel, C. Lindqvist & B. Laufer, eds. *L2 vocabulary acquisition, knowledge and use*. Eurosla, pp.29–56.
34. Jiang, N., Zhang, Y. & Chen, L., 2022. Cross-linguistic congruency effects on L2 collocation processing. *Studies in Second Language Acquisition*, 44(1), pp.45–69.
35. Lewis, M., 1993. *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.
36. Lovrić, M., 2020. Collocational competence and translation congruency in Croatian EFL learners. *Suvremena lingvistika*, 46, pp.217–236.
37. Lukač, M., 2016. Congruency and frequency in Croatian learners' collocational development. *Jezikoslovlje*, 17(1–2), pp.45–65.
38. Ma, Q., 2023. Teacher corpus literacy and technology integration in EFL contexts. *System*, 113.
39. Matsuno, T., 2017. L2 collocation processing in Japanese learners of English. *Asian EFL Journal*, 19(4), pp.56–84.
40. Miščin, A., Hadžić, A. & Bureković, M., 2024/2025. Corpus-based instruction of ESP collocations in Croatian tertiary education. *ESP Today*, 13(1).
41. Nation, I.S.P., 2001. *Learning vocabulary in another language*. Cambridge University Press.
42. Nesselhauf, N., 2005. *Collocations in a learner corpus*. John Benjamins.
43. Patekar, J., 2022. Collocational development and congruency effects in Croatian EFL learners. *Strani jezici*, 51, pp.45–68.
44. Paquot, M., 2018. Phraseological complexity in academic writing. *International Journal of Applied Linguistics*, 28(1), pp.26–51.
45. Pawley, A. & Syder, F.H., 1983. Two puzzles for linguistic theory: Native-like selection and native-like fluency. In: J.C. Richards & R.W. Schmidt, eds. *Language and communication*. Longman, pp.191–225.
46. Peters, E., 2016. The learning burden of collocations: The role of interlexical factors. *Language Teaching Research*, 20(1), pp.113–138.
47. Qian, D.D., 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), pp. 513–536.
48. Rizvić-Eminović, E. & Arnaut-Karović, K., 2015. The use of collocations by B1, B2 and C1 level students of English as L2 at the University of Zenica. *Epiphany*, 8(2), pp.115–130.

49. Rizvić-Eminović, E. & Hadžić, A., 2021. Using corpora in English language teaching: A brief overview and sample corpus-based activities. *Zbornik radova Filozofskog fakulteta*, 19, pp.27–35.
50. Rizvić-Eminović, E. & Kasumagić-Kafedžić, L., 2024. *Engleski jezik: učenje, podučavanje [English Language: Learning and Teaching]*. Zenica: Meligrafprint.
51. Rizvić-Eminović, E. & Neslanović, V., 2022. The effect of corpus-based activities on vocabulary acquisition of A1–A2 level seventh graders: A case study. *European Journal of English Language Teaching*, 7(4), pp.1–14.
52. Rizvić-Eminović, E. & Bujak, A., 2017. Gramatička kompetencija bruoša Univerziteta u Zenici [Grammatical competence of the University of Zenica freshmen]. In: *Zbornik radova VI međunarodnog naučno-stručnog skupa "Jezik – književnost – kultura"* [Proceedings of the 6th International Scientific Conference "Language – Literature – Culture"]. Bosnia and Herzegovina: University of Zenica, pp.4–12.
53. Rizvić-Eminović, E., Bujak, A. & Bureković, M., 2018. A study of trends in the level of English language grammatical competence at Zenica University. In: M. Mulalić, A. Mulalić, N. Obralić & E. Jelešković, eds. *Education, culture and identity: The future of humanities, education and creative industries. Book of proceedings*. Sarajevo: International University of Sarajevo, pp.197–207.
54. Rizvić-Eminović, E., Bureković, M. & Dervić, M., 2025. Patterns of language errors in the writing of Bosnian EFL students: A case study. *Studies in English Language and Education*, 12(2), pp.648–662.
55. Rizvić-Eminović, E., Dervić, M. & Bilić, A., 2023. The effect of explicit writing instruction on EFL students' writing performance: A case study. *Human Research in Rehabilitation*, 15(2), pp.411–422.
56. Rizvic-Eminovic, E., Dervic, M. & Rados, A., 2025. Collocational competence of B2 level EFL students: A case study. *Jezikoslovlje*, 26(1), pp.77–94.
57. Rizvić-Eminović, E., Šukalić, Đ. & Tursanović, A., 2024. Analysis of Collocations in English Song Lyrics and Their Bosnian Translation Equivalents: Potential for English Language Learning. *European Journal of Applied Linguistics Studies*, 7(2), pp.56–67.
58. Rodríguez-Fernández, I. & Casillas, A., 2018. Detecting collocational errors in Spanish learner corpora. *ReCALL*, 30(2), pp.181–200.
59. Schmitt, N., 2014. *Size and depth of vocabulary knowledge: What the research shows*. Routledge.
60. Shiha, P., 2024. Arab EFL learners' collocational competence, paper presented at ACE 2024, The IAFOR Research Archive.
61. Sinclair, J.M., 1991. *Corpus, concordance, collocation*. Oxford University Press.
62. Sonbul, S., 2023. Re-examining the construct validity of collocational tests. *Language Assessment Quarterly*, 20(2), pp.156–178.
63. Tabak, T. & Pavičić Takač, V., 2023. Strategy use and collocational knowledge among Croatian EFL learners. *Suvremena lingvistika*, 49(1), pp.101–122.

64. Vandeweerd, L., 2023. Longitudinal development of collocational competence in L2 English writing. *System*, 114, 103033.
65. Vojnović, S., 2021. English–Serbian noun–noun collocations in tourism discourse. *Philologia*, 19(1), pp.73–86.
66. Wang, Y., 2023. A comparative study on verb–noun collocation errors of Chinese EFL learners with different proficiency levels. *Journal of Literature and Art Studies*, 13(6), pp. 586–600.
67. Wolter, B., 2013. The effects of frequency and collocational congruency on L2 collocational processing. *Studies in Second Language Acquisition*, 35(3), pp.451–482.
68. Wood, D., 2015. *Foundations of formulaic language: Acquisition, processing and use*. Bloomsbury.
69. Wray, A., 2002. *Formulaic language and the lexicon*. Cambridge University Press.
70. Yamashita, J. & Jiang, N., 2010. L1 influence on the acquisition of L2 collocations. *TESOL Quarterly*, 44(4), pp.647–668.
71. Yue, Z. & Ruangjaron, S., 2025. Putting in the last piece: A comprehensive profiling of learners' collocational competence. *LEARN Journal*, 18(1), pp. 141–171.
72. Zhang, X., 2017. Effects of receptive–productive integration tasks on collocation learning: Roles of prior knowledge and feedback. *System*, 66, pp. 87–98.