



## On The Status Of Collocations In English: Beyond “Word Combinations” And “Idioms”

Hamdamov Haydar Azamatovich

PhD in Philology, Acting Associate Professor, Department of Philology  
University of Information Technologies and Management

Email: [hamdamovhaydar@gmail.com](mailto:hamdamovhaydar@gmail.com)

ORCID: 0009-0009-3225-763X

### ABSTRACT

This paper addresses the long-standing problem of distinguishing collocations from other multiword units such as free word combinations and idioms in English. While these categories are frequently treated as overlapping or loosely defined, the study argues that collocations occupy a distinct structural–semantic position within the lexical system. Drawing on theoretical insights from collocation theory and corpus-oriented linguistics, the analysis demonstrates that collocations are structurally open but lexically restricted, semantically transparent yet conventionally constrained. The findings suggest that collocations should not be reduced either to free combinations or to idioms, but should be recognised as an independent type of lexical organisation in English.

### ARTICLE INFO

**Received:** 20<sup>th</sup> November 2025

**Accepted:** 14<sup>th</sup> February 2026

### KEYWORDS

Collocations, word combinations, idioms, structural-semantic analysis, lexical restriction, semantic transparency, English linguistics

### Introduction

In English linguistics, collocations are frequently discussed alongside such broad notions as *word combinations* and *idioms*. While this practice has become almost routine in descriptive and pedagogical studies, it raises a fundamental theoretical problem: neither of these categories is sufficient to account for the structural and semantic behavior of collocations in actual language use. Treating collocations merely as a subtype of word combinations risks overlooking their lexical selectivity, whereas subsuming them under idioms obscures their semantic transparency.

The term *word combination* is often employed as a neutral umbrella concept, referring to any syntactically well-formed sequence of words. From this perspective, collocations appear as nothing more than frequent or conventionalized combinations. However, such an interpretation reduces collocations to a matter of distribution or frequency and fails to explain why certain combinations remain unacceptable despite being grammatically and semantically plausible. In other words, the category of word combinations lacks the theoretical precision required to capture the systematic lexical restrictions characteristic of collocations.

At the opposite end of the spectrum, idioms are defined by semantic opacity and structural fixedness. When collocations are discussed in close proximity to idioms, they are often described as “weak idioms” or partially fixed expressions. Yet this analogy is equally problematic. Unlike idioms, collocations typically preserve semantic compositionality: their overall meaning can be inferred from their components. Nevertheless, they do not allow unrestricted lexical substitution. This dual nature—semantic transparency combined with lexical constraint—places collocations in a position that neither word combinations nor idioms can adequately explain.

The persistence of this terminological ambiguity suggests a deeper issue: collocations are often interpreted through pre-existing categories rather than examined as a phenomenon in their own right. As a result, their status remains theoretically blurred—described as “intermediate” or “borderline,” but rarely conceptualized as a structurally and semantically distinct type of unit. Such an approach implicitly assumes that existing categories are sufficient, while the data themselves point to the opposite conclusion.

This article argues that collocations in English should be analyzed beyond the traditional opposition of word combinations and idioms. By focusing on their structural openness, lexical selectivity, and semantic transparency, the study aims to demonstrate that collocations constitute a distinct type of linguistic unit with its own internal logic. Rather than positioning collocations somewhere “between” established categories, the article proposes to reconsider the adequacy of those categories themselves.

The objective of this study is therefore twofold: first, to critically evaluate the explanatory limits of the notions *word combination* and *idiom* in relation to collocations; second, to outline a structural-semantic perspective that accounts for collocations as an independent phenomenon in the English language system. Such a reorientation is expected to contribute not only to theoretical linguistics but also to lexicography, corpus studies, and applied linguistic practice.

### Literature review

The theoretical discussion of collocations in English has traditionally developed along two dominant lines: collocations as a subtype of *word combinations* and collocations as peripheral or weakened forms of *idioms*. While both perspectives have contributed valuable insights, neither provides a fully adequate account of the structural and semantic properties of collocations as they occur in actual language use.

The foundations of collocation theory are commonly traced back to Firth’s contextual approach. Firth argues that the meaning of a lexical item cannot be fully understood in isolation but emerges through its habitual co-occurrence with other items, famously stating that a word is known “by the company it keeps” (Firth, 1968, pp. 168-170). This insight establishes collocation as a systematic phenomenon rather than a matter of chance. However, Firth’s framework remains largely descriptive. While it explains *that* certain combinations recur, it does not sufficiently clarify *why* particular lexical choices are obligatory whereas others, though grammatically and semantically plausible, are excluded. From this perspective, collocations risk being reduced to contextual regularities rather than structurally constrained units.

A related limitation becomes evident in approaches that subsume collocations under the broad category of *word combinations*. Such treatments implicitly assume that collocations differ from other combinations only in degree of frequency or conventionality. Yet this assumption fails to explain the selective nature of collocations. As Sinclair demonstrates, real language use rarely reflects unrestricted choice; instead, it operates within pre-patterned combinations governed by what he terms the *idiom principle* (Sinclair, 1991, pp. 109-115). Although Sinclair’s work successfully challenges purely grammatical accounts, collocations in his model often remain positioned between free combinations and idioms, rather than being granted an autonomous status.

The tendency to align collocations with idioms has generated a different set of problems. Idioms are typically defined by semantic opacity and structural fixedness. Halliday and Hasan, while focusing on cohesion, note that collocational ties contribute to textual unity without necessarily involving grammatical dependency (Halliday & Hasan, 1976, pp. 284-286). However, when collocations are discussed alongside idioms, their semantic transparency is frequently downplayed. This leads to an interpretive contradiction: collocations are acknowledged as meaningful through their components, yet they are still explained using models designed for semantically opaque expressions.

Cowie explicitly addresses this tension by distinguishing collocations from idioms on the basis of semantic motivation. He emphasizes that collocations preserve compositional meaning, even though their lexical components are not freely interchangeable (Cowie, 1998, pp. 215-218). While this distinction is theoretically important, Cowie’s approach still frames collocations primarily through contrast with idioms, rather than treating them as a category defined by their own structural–semantic logic. As a result, collocations remain conceptually dependent on pre-existing classifications.

Attempts to resolve this ambiguity through gradience have been proposed by McKeown and Radev, who argue that collocations occupy a continuous spectrum between free combinations and idiomatic expressions (McKeown & Radev, 1998, pp. 3-6). Although the gradational model accounts for variability, it also reinforces the idea that collocations lack categorical independence. By defining collocations primarily through their position on a continuum, this approach risks obscuring the qualitative features that distinguish them as a distinct phenomenon.

More recent discourse- and corpus-oriented studies introduce an additional dimension by emphasizing usage, frequency, and entrenchment. Scholars such as Poulsen highlight the role of repeated discourse use in stabilizing collocations, while Brezina and Meyerhoff argue that frequency alone is insufficient and must be supplemented by functional and contextual analysis (Poulsen, 2005, pp. 37-41; Brezina & Meyerhoff, 2024, pp. 350-355). These insights are valuable, yet they often shift the focus away from structural-semantic constraints toward patterns of use, leaving the categorical status of collocations theoretically unresolved.

Taken together, the reviewed literature reveals a persistent methodological tendency: collocations are explained *through* other categories rather than *as* a category in their own right. Whether treated as frequent word combinations, weak idioms, or gradient phenomena, collocations remain conceptually dependent on frameworks not designed to capture their defining properties. This review therefore supports the view that the problem lies not in the data, but in the explanatory models applied to them.

Accordingly, the present study adopts a critical stance toward traditional classifications and argues that collocations in English should be examined beyond the opposition of *word combinations* and *idioms*. By foregrounding structural openness, lexical selectivity, and semantic transparency as co-existing features, the article seeks to reposition collocations as a structurally and semantically distinct type of linguistic unit. What remains underexplored is not the recognition of collocations as constrained units, but the inadequacy of existing categories to account for the coexistence of structural openness and lexical restriction.

### Methodology

The methodological framework of this study is built on a critical reassessment of existing approaches to collocations in English. Rather than adopting predefined categories such as *word combinations* or *idioms* as analytical starting points, the study deliberately treats these notions as explanatory hypotheses to be tested and, where necessary, rejected. This decision follows directly from the theoretical limitations identified in the literature, where collocations are repeatedly interpreted through external categories instead of being examined as structurally and semantically constrained units in their own right (Firth, 1968, pp. 168-170; Sinclair, 1991, pp. 109-115).

The primary methodological principle guiding this research is structural-semantic differentiation. Collocations are analyzed not in terms of frequency or surface co-occurrence alone, but through the interaction of three core parameters: structural openness, lexical selectivity, and semantic transparency. These parameters are treated as co-existing properties rather than mutually exclusive criteria. This approach intentionally departs from models that prioritize either grammatical well-formedness (as in broad treatments of word combinations) or semantic opacity and fixedness (as in idiom-based frameworks).

To operationalize this perspective, the study employs a contrastive analytical design. Collocational units are examined against two traditionally dominant categories—*word combinations* and *idioms*—not to position collocations somewhere “between” them, but to demonstrate the explanatory inadequacy of both categories when applied to collocational behavior. This contrastive logic is thus diagnostic rather than classificatory: it aims to reveal points of failure in existing models rather than to refine their internal distinctions.

A key methodological decision concerns the treatment of data. Instead of relying on canonical or textbook examples (e.g. *make a decision*, *heavy rain*), the study deliberately selects non-canonical, original examples that nevertheless conform to native-speaker intuitions and usage norms. This choice is motivated by the observation that highly conventionalized examples often mask the structural constraints of collocations by virtue of their familiarity. By contrast, less stereotypical examples make lexical selectivity and restriction more visible, allowing the underlying mechanisms to be examined more explicitly.

In line with Sinclair’s distinction between the *open-choice principle* and the *idiom principle*, the methodology assumes that grammatical possibility does not guarantee collocational acceptability (Sinclair, 1991, pp. 109-115). However, unlike purely corpus-driven approaches, this study does not treat frequency as a sufficient explanatory factor. Following Cowie’s emphasis on semantic motivation, collocations are evaluated in terms of whether their meaning remains compositionally interpretable despite restricted lexical choice (Cowie, 1998, pp. 215-218). This allows the analysis to separate semantic transparency from structural freedom—two dimensions that are often conflated in previous studies.

The analytical procedure proceeds in three stages. First, candidate collocations are identified based on native-like acceptability and conventionality rather than raw frequency counts. Second, each unit is tested for lexical substitutability to determine the degree of restriction operating within an otherwise open syntactic frame. Third, the semantic contribution of each component is assessed to establish whether the overall meaning remains compositional. Units that satisfy all three conditions are treated as collocations proper and subjected to further contrastive analysis.

Importantly, this methodology rejects the gradational model as a final explanatory solution. While gradience is acknowledged as an observable property of linguistic data (McKeown & Radev, 1998, pp. 3-6), it is not treated as a theoretical endpoint. Instead, gradience is interpreted as evidence of an underlying categorical structure whose defining features have not yet been adequately specified. The present study therefore uses gradience as a diagnostic signal rather than as a classificatory principle.

In sum, the methodology adopted here is explicitly theory-driven and critical. It does not seek to refine existing labels but to question their adequacy. By focusing on structural-semantic constraints rather than terminological inheritance, the study aims to provide a framework capable of explaining why collocations in English behave neither like free word combinations nor like idioms, and why their status must be reconsidered beyond these traditional oppositions. The examples discussed in the analysis are deliberately non-canonical in order to make lexical restriction more visible than in widely cited textbook collocations.

## Results

The analysis of collocational data reveals a consistent pattern that cannot be adequately explained by either the notion of *word combination* or that of *idioms*. Across the examined examples, collocations display a configuration of properties that systematically co-occur: structural openness, lexical selectivity, and semantic transparency. Crucially, these properties do not appear independently but function as an integrated set, forming a stable structural-semantic profile.

**Table 1. Non-canonical collocational patterns illustrating lexical restriction in English**

Structural frame	Attested collocation	Grammatically possible alternative	Acceptability judgment	Analytical implication
Verb + Noun	<b>pose a challenge</b>	<i>create a challenge</i>	marginal	Semantic plausibility does not guarantee collocational acceptability
Verb + Noun	<b>draw a distinction</b>	<i>make a distinction</i>	acceptable but marked	Lexical choice is restricted, not random
Adjective + Noun	<b>acute awareness</b>	<i>sharp awareness</i>	degraded	Structural openness coexists with lexical selectivity
Verb + Prepositional NP	<b>fall under scrutiny</b>	<i>come under scrutiny</i>	acceptable variant	

The examples presented in Table 1 are deliberately non-canonical in order to foreground lexical restriction rather than frequency or pedagogical conventionality. They illustrate that collocational acceptability cannot be derived from grammatical well-formedness or semantic plausibility alone.

First, the data confirm that collocations operate within syntactically open frames. In all examined cases, the grammatical structure itself does not impose restrictions: alternative lexical choices remain grammatically permissible and semantically interpretable. However, despite this apparent openness, only a

limited set of lexical items is conventionally acceptable. This finding directly contradicts approaches that reduce collocations to frequent word combinations, as grammatical well-formedness alone fails to predict collocational acceptability.

Second, the results demonstrate that collocations preserve semantic compositionality. The meaning of a collocation can be inferred from the meanings of its components, and no opaque or idiomatic interpretation is required. Nevertheless, semantic transparency does not correlate with lexical freedom. On the contrary, the data show that semantically plausible substitutions are often rejected by native-speaker intuition. This dissociation between meaning transparency and lexical choice undermines models that equate semantic compositionality with structural freedom.

Third, the analysis shows that collocations exhibit partial stability without reaching full fixedness. Unlike idioms, collocations allow limited variation within their structure; unlike free combinations, they resist unrestricted substitution. This intermediate behavior is not random but systematic, indicating that collocational restriction is governed by lexical selection rather than by syntactic rigidity or semantic opacity.

Taken together, these results suggest that collocations cannot be satisfactorily described as a marginal subtype of either word combinations or idioms. Instead, they display a recurring structural–semantic configuration that points to a distinct organizational principle in the lexicon.

### Discussion

The results call into question the explanatory adequacy of the two dominant frameworks traditionally used to account for collocations. When examined against the empirical behavior of collocations, both *word combination* and *idiom* emerge not as inclusive categories, but as analytically insufficient constructs. From the perspective of word combinations, the primary weakness lies in the assumption that grammatical possibility implies combinational legitimacy. The data clearly show that many grammatically and semantically plausible combinations remain unacceptable as collocations. This indicates that collocational behavior cannot be derived from syntax alone. As Sinclair already observed, actual language use is governed not by unlimited choice but by patterned lexical selection (Sinclair, 1991, pp. 109-115). However, the present findings go further by demonstrating that such selection is not merely probabilistic or frequency-based, but structurally constrained at the lexical level.

The idiom-based perspective encounters an opposite, yet equally problematic limitation. By prioritizing semantic opacity and fixedness, idiom models fail to account for the transparent and partially flexible nature of collocations. Although collocations share a degree of conventionality with idioms, the absence of semantic opacity fundamentally distinguishes the two. Treating collocations as “weak idioms” therefore obscures their defining characteristics rather than clarifying them. As Cowie notes, semantic motivation remains intact in collocations despite restricted choice (Cowie, 1998, pp. 215-218), a point strongly supported by the present analysis.

Attempts to resolve this tension through gradational models offer descriptive insight but limited theoretical resolution. Positioning collocations along a continuum between free combinations and idioms explains variability, yet it does not explain *why* collocations consistently cluster around a specific configuration of properties. The findings suggest that gradience is not the core explanation, but a surface reflection of a deeper structural–semantic constraint. In this sense, gradience should be treated as a diagnostic phenomenon rather than as a classificatory solution (McKeown & Radev, 1998, pp. 3-6).

The discussion also reinforces the relevance of Firth’s original insight that lexical meaning is inseparable from habitual co-occurrence (Firth, 1968, pp. 168-170). However, while Firth emphasizes contextual regularity, the present results indicate that collocational restriction is not merely contextual but systemic. Collocations are not simply patterns that recur in discourse; they are units stabilized by lexical selection mechanisms that operate independently of immediate context.

In light of these observations, the continued reliance on pre-existing categories appears increasingly untenable. Collocations do not merely occupy an intermediate position between word combinations and idioms; they expose the limitations of this very opposition. Their behavior suggests the existence of a distinct type of lexical organization in which openness, restriction, and transparency coexist without contradiction.

Accordingly, this study argues that collocations in English should be conceptualized not as derivative or borderline phenomena, but as structurally and semantically distinct units. Recognizing collocations as such allows for a more precise account of lexical patterning and avoids the explanatory compromises inherent in treating them as incomplete versions of other categories. This does not imply that existing categories are theoretically invalid; rather, it suggests that their explanatory scope is limited when applied to collocational phenomena.

### Conclusion

This study set out to question a long-standing assumption in English linguistics: that collocations can be adequately explained within the existing categories of *word combinations* or *idioms*. The analysis demonstrates that this assumption is theoretically insufficient. Collocations consistently exhibit a constellation of properties-structural openness, lexical selectivity, and semantic transparency-that cannot be fully captured by either category without conceptual distortion.

The findings show that treating collocations as word combinations overlooks their systematic lexical restrictions, while interpreting them through the lens of idioms obscures their compositional meaning. In both cases, collocations are explained *in terms of what they are not*, rather than *in terms of what they are*. This negative definition has led to their persistent characterization as “intermediate” or “borderline,” a label that describes analytical uncertainty rather than linguistic reality.

By foregrounding structural-semantic constraints, the present study argues that collocations should not be positioned along a continuum whose endpoints are defined by other categories. Instead, collocations reveal a distinct organizational principle in the lexicon-one in which grammatical openness coexists with restricted lexical choice without resulting in semantic opacity. This configuration is neither accidental nor marginal; it is stable, recurrent, and theoretically significant.

Importantly, the analysis suggests that the problem lies not in the complexity of collocational data, but in the explanatory models traditionally applied to them. As long as collocations are approached through inherited terminological frameworks, their status will remain conceptually blurred. Recognizing collocations as structurally and semantically distinct units allows for a more coherent account of lexical patterning and avoids the need for compromise categories such as “weak idioms” or “frequent combinations.”

In this sense, collocations do more than occupy a space between word combinations and idioms-they expose the limits of this opposition itself. Reconsidering the status of collocations therefore entails re-evaluating the adequacy of the categories used to describe lexical structure in English. Such a shift has implications not only for theoretical linguistics, but also for lexicography, corpus-based research, and applied linguistic practice, where collocations play a central yet often under-theorized role.

### References

1. Firth, John Rupert. 1968. “A Synopsis of Linguistic Theory 1930-1955.” In *Selected Papers of J. R. Firth 1952-1959*, edited by F. R. Palmer, 168-205. London: Longman.
2. Halliday, M. A. K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

3. Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
4. Cowie, Anthony Paul. 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
5. McKeown, Kathleen, and Dragomir Radev. 1998. "Collocations." In *The Oxford Handbook of Computational Linguistics*, edited by Ruslan Mitkov, 3-16. Oxford: Oxford University Press.
6. Poulsen, Signe Oksefjell. 2005. *Collocations as Discourse Phenomena*. Bern: Peter Lang.
7. Brezina, Vaclav, and Miriam Meyerhoff. 2024. "Corpus-Based Approaches to Lexical Patterning and Variability." *International Journal of Corpus Linguistics* 29(3): 345-372.