Original research article / Originalus mokslinis straipsnis
Vol. 1 No. 4 (2025): Birželis

A Taxonomic Approach to Structural and Semantic Dimensions in English Phraseology

[1] Hasan Alisoy

Accepted: 05.21.2025
Published: 05.26.2025
https://doi.org/10.69760/portuni.0104007

Abstract: Phraseological units are conventionalised multi-word expressions whose overall meaning cannot be straightforwardly derived from their individual parts. They include idioms, collocations, proverbs and other fixed expressions that are ubiquitous in English. Such units play a crucial role in language fluency, cultural expression and cognitive processing. This article aims to classify English phraseological units along two primary dimensions: structure (the syntactic form of the expression) and semantics (the transparency of meaning). We adopt a descriptive, corpus-based methodology, examining examples from the British National Corpus and authoritative idiom dictionaries (e.g. Oxford Dictionary of English Idioms) to ground our analysis. Structural categories are identified (e.g. noun phrases vs verb phrases vs full sentences) as well as semantic types (fully transparent vs semi-transparent vs opaque idioms). The proposed typology is summarized in terms analogous to nominative vs predicative vs communicative units. We also discuss how certain expressions blur category boundaries (e.g. literal vs figurative senses in context). This classification has practical implications: it can guide lexicographers in organizing idiom dictionaries, inform language teachers in grouping formulaic language, and assist computational linguists in multiword expression detection and processing. Future work may involve corpus-driven statistical modelling of phraseological regularities and the development of enriched phraseological databases for NLP applications.

Keywords: *phraseology; idiomatic expressions; structural linguistics; semantic typology; lexical combinations; English linguistics; phraseological classification*

INTRODUCTION

Phraseological units (also known as phrasemes or multi-word expressions) are fixed combinations of words which function as single semantic units. Such units include idioms (e.g. *kick the bucket*, *break the ice*), collocations (*make a decision*, *fast food*), proverbs and sayings ("*The early bird catches the worm*"), fixed metaphors (*white elephant*, *heart of gold*), and routine formulae ("*ladies and gentlemen*", *once upon a time*'). Although definitions vary, scholars agree that phraseological units are at least two-word sequences that are relatively stable in form and carry an idiosyncratic meaning not predictable from their parts. For example, *crocodile tears* denotes insincere sorrow, a meaning not found in crocodile or tears separately. Early work in phraseology by Vinogradov (1950) and others in Russian linguistics laid the foundation by identifying gradations of semantic motivation (fully opaque vs semi-transparent) in these expressions. In the Western tradition, researchers like Gläser (1984, 1998) and Cowie (1998)

[1] **Alisoy, H.** Lecturer in English, Nakhchivan State University, Azerbaijan. Email: alisoyhasan@ndu.edu.az. ORCID: https://orcid.org/0009-0007-0247-476X

have similarly emphasized the lexicalized and figurative nature of idiomatic phrases, treating them as a distinct stratum of the lexicon. Cognitive and functional linguists also highlight the importance of phraseological units in discourse and thought: they are highly frequent in everyday language, carry cultural connotations, and are argued to be stored and processed holistically in the mind (see Wray 2002 and references therein).

The primary objective of this paper is to provide a structural–semantic classification of English phraseological units. We restrict our focus to *English* and avoid cross-linguistic comparisons. Structurally, phraseological units can range from simple nominal phrases (*paper tiger*) to full clauses or sentence-like sayings. Semantically, they can range from fully transparent (meaning clear from context) to completely opaque (idiomatic). By combining these dimensions, we aim to map out the major types of English idiomatic expressions. Such classification has practical benefits: it helps lexicographers decide how to organize idioms in dictionaries, assists teachers in grouping and teaching idiomatic language, and provides NLP practitioners with categories for algorithmic idiom detection.

LITERATURE REVIEW

Research on English phraseology spans descriptive, theoretical and applied perspectives. Cowie's edited volume *Phraseology: Theory, Analysis and Applications* (1998/2001) is a foundational collection that surveys many issues in the field. Sinclair (2004) emphasized the "idiom principle" in corpus linguistics – the idea that language use is strongly influenced by prefabricated phrases – and argued for corpus-based methods to identify them. Moon's monograph *Fixed Expressions and Idioms in English* (1998) adopts a corpus-informed approach to document how fixed phrases are used in context. Fernando's work *Idioms and Idiomaticity* (1996) examines idioms from a functional perspective, noting that both literal and non-literal interpretations can co-occur even in formulaic language. More recently, Dobrovol'skij & Piirainen's *Phraseology: An International Handbook of Contemporary Research* (2006) compiles cross-disciplinary studies, highlighting semantic, syntactic and pragmatic views of phraseological units.

Structural classifications. Linguists commonly classify idioms by their syntactic form. For example, Arnold (1973, cited in Fernando 1996) categorised idioms by parts of speech: noun phrases (*cat's paw* "dupe"), verb phrases (*take advantage*), adjective phrases (*high and mighty*), adverb phrases (*once in a blue moon*), and prepositional phrases (*in hot water*). Similarly, Klasinc (1985) and Al-Hassnawi (1989) distinguished idioms by grammatical type. In English, the bulk of phraseological units are noun phrases and verb phrases. Sinclair (2004) notes that many idioms function as heads of noun phrases (*white elephant*, *gag order*) or as predicates (*spill the beans*, *to pull one's leg*). Some are adjectival/adverbial similes (e.g. *as cold as ice*, *once in a blue moon*). At the extreme end, entire sentences (proverbs and sayings) count as fixed expressions (e.g. "A stitch in time saves nine", "What will Mrs. Grundy say?"). These structural distinctions often correlate with usage: noun-phrase idioms (sometimes called "nominative" in Russian tradition) typically *denote objects or concepts*, whereas verb-phrase idioms describe *actions or events*.

Semantic classifications. A widely-used criterion is transparency vs idiomaticity. Vinogradov's classic taxonomy (cited in Cowie 1998) divides phrasemes into *fusions* (completely opaque idioms), *unities* (partially motivated), and *combinations* (semi-transparent). In this scheme, fusions are fixed figurative expressions whose meaning is unrelated to their words (e.g. *let the cat out of the bag* = reveal a

secret). Unities are partially motivated by metaphor or metonymy (e.g. *break the ice* 'begin a conversation'; here *ice* stands metaphorically for social tension). Combinations are more loosely bound and are close to collocations: one component has a figurative sense and the other remains literal (e.g. *meet the demand*, where *meet* means satisfy a requirement figuratively). Fernando (1996) similarly notes a continuum: fully transparent collocations at one end, fully opaque idioms at the other, with many cases in between. In practice, scholars often use simpler terms: opaque idioms (non-compositional, like *kick the bucket* 'die'), semi-opaque/idiomatic (e.g. *cut corners* 'do poorly'; the metaphor of cutting is somewhat motivated), and transparent idiomatic combinations (e.g. *strong tea*, whose meaning is close to literal).

Overlap and continuum. Modern work emphasizes that fixed expressions form a spectrum rather than discrete boxes. For instance, Fernando (1996) and Wray (2002) argue that idioms and collocations overlap: compositionality alone does not separate them, since some idioms may be partially analyzable and some collocations may carry idiosyncratic nuance. Howarth (1998) similarly proposes a continuum from free combinations to pure idioms. The key distinguishing feature is conventionalisation and fixedness: idioms tend to have less internal variability. Sinclair (2004) and Moon (1998) also note that context and register affect phraseological status (e.g. proverbs are context-bound, slang fixed phrases occur in spoken registers). Vinogradov's distinctions of fusion/unity/combination remain influential, but researchers acknowledge fuzzy boundaries.

Communicative vs referential functions. Some linguists (especially in Russian tradition) classify idioms by communicative function. For example, Koonin's system (outlined by Masimova 2018) distinguishes four classes: (1) nominative units (word-groups denoting entities or qualities, e.g. *a bull in a China shop* for a clumsy person), (2) nominative-communicative (verb phrases that become full sentences in passive, e.g. *to break the ice* for "to initiate friendly interaction"), (3) interjectional fixed expressions conveying emotion (*By George!* "indeed"), and (4) communicative units (proverbs and sayings functioning as complete utterances, e.g. "*Too many cooks spoil the broth*"). While not part of mainstream Western taxonomy, this functional view underscores that idioms can act either as nominal labels or as pragmatic utterances.

In summary, existing literature provides multiple overlapping frameworks for categorising English phraseological units. Our work synthesizes these by focusing on two main axes – form (syntactic shape) and meaning (degree of semantic compositionality) – while illustrating each category with examples.

METHODOLOGY

This study adopts a descriptive, corpus-informed approach. Data were drawn from the British National Corpus (BNC) to ensure examples reflect authentic English usage across genres, supplemented by examples cited in standard idiom dictionaries (e.g. *Oxford Dictionary of English Idioms*). We systematically surveyed the corpora for candidate phraseological units and manually verified their idiomaticity and stability with dictionary definitions. Each identified unit was classified first by structural type (nominal, verbal, adjectival/adverbial, or sentential) and second by semantic transparency (opaque, semi-transparent, transparent). We prioritize widely recognized English idioms and fixed expressions. Examples were selected to illustrate each category clearly; many of them (such as *crocodile tears* or *break the ice*) are well-attested in usage and dictionaries. The classification scheme was iteratively refined by checking consistency with sources like Cowie's dictionary and Moon's corpus

analysis. No experimental or quantitative procedures were used – the aim is a qualitative typology backed by examples.

RESULTS AND DISCUSSION

Structural Categories

English phraseological units exhibit diverse syntactic forms. We identify four main structural classes:

- **Nominal (Noun Phrase) idioms:** These are fixed expressions functioning syntactically as noun phrases (often serving as subjects or objects). Examples include *crocodile tears* ("insincere display of emotion") and *Pandora's box* ("a source of many troubles"). Other examples: *white elephant* ("useless possession"), *paper tiger*, *elephant in the room*. Such units denote objects, persons or abstract concepts. They typically appear with determiners or as bare plurals (e.g. *the elephant in the room*, *crocodile tears*). Many of Vinogradov's "phraseological fusions" are of this type.

- **Verbal (Predicative) idioms:** These are verb-centered idiomatic phrases, functioning as predicates (often with objects and sometimes prepositions). Examples: *to go to pot* ("to deteriorate"), *kick the bucket* ("die"), *spill the beans* ("reveal a secret"), *take one's hat off (to)*, *pull someone's leg*. They may be phrased actively or passively (*the beans were spilled*). Verbal idioms can involve auxiliary verbs or particles (*hang tight*, *give in*). In Koonin's terms, some of these become full sentences in passive voice ("the ice is broken" from *break the ice*), linking them to the nominative-communicative class.

- **Adjectival/Adverbial idioms:** These fixed expressions include adjectives or adverbs. For example, simile-based adjectival idioms like *as mad as a hatter* ("completely crazy"), *as cool as a cucumber* ("calm under pressure"), *as good as gold* ("well-behaved"). Adverbial idioms include *by and by* ("soon"), *to and fro* ("backwards and forwards"), *once in a blue moon* ("very rarely"), *tooth and nail* ("fiercely"). These often function adverbially in sentences. They typically follow the pattern *as [adjective] as [noun]* or are fixed adverbial phrases.

- **Sentential idioms (Proverbs/Sayings):** These are complete clauses or sentences with proverbial meaning. E.g. "*Queen Anne is dead!*" (an old rhetorical formula meaning "I'm telling a truth that might displease people") and "*What will Mrs. Grundy say?*" (asking about social reputation). Other examples: "*An apple a day keeps the doctor away*", "*It goes without saying*", "*Better safe than sorry*". They function as stand-alone statements or responses. Because these have full sentence form, they often carry general wisdom or social norms (the communicative class).

These structural categories can be summarised as in Table 1 below:

Structural Class	***Example(s)***	***Meaning***
Nominal Idiom	*crocodile tears* (n)	insincere tears; *Pandora's box*
Verbal Idiom	*to go to pot* (v)	deteriorate; *spill the beans* (idiomatic)
Adjectival/Adverbial Idiom	*as cool as a cucumber* (adj simile); *by and by* (adv)	metaphorical cal(mness), literal sense "soon"
Sentential Idiom	"*What will Mrs. Grundy say?*" (full sentence)	"What will people say?" (social norm)

Table 1. Examples of structural categories of phraseological units.

Within each class, idioms may show slight syntactic variability. For instance, noun-phrase idioms sometimes take plural or possessive forms (*"wheels of justice"* vs *"wheel of justice"*), and verb idioms may allow tense or aspect changes (*kicked the bucket*, *taking one's hat off*). However, many phraseological units resist internal substitutions or permutations (as noted by Cowie et al., 1993–1994) – e.g. *a bull in a china shop* cannot easily be altered.

Semantic Categories

Phraseological units also differ in how their overall meaning relates to the meanings of their parts. We adopt the traditional three-way semantic typology, illustrated below with examples:

- **Fully transparent (highly compositional) units:** The overall meaning is essentially deducible from the literal meanings of the words. These are at the borderline of idioms and strong collocations. For example, *at the drop of a hat* (meaning *immediately*) is partly transparent (a drop of a hat implies immediacy). A truly transparent example might be *red herring* ("distraction" – literally a smoked fish, but the origin is metonymic). Such units are relatively rare as true idioms, and some lists exclude them as collocations. In practice, we treat transparent units as the far end of a continuum (most collocations, few idioms).
- **Semi-transparent (motivated) idioms:** Here one or more components contributes to the figurative meaning by metaphor or metonymy. E.g. *break the ice* means 'initiate friendly interaction'; the literal idea of cracking ice is metaphorically extended to warming up a social situation. Likewise *add fuel to the fire* ('intensify hostility'), where the metaphor of fire stands for conflict. Phraseological unities in Vinogradov's sense fall here: the phrase is metaphorical but still anchored in its parts. The meaning can often be inferred with some cultural knowledge: *rock bottom* meaning *the lowest possible level* is partly transparent (a rock on the bottom). Transparent and semi-transparent idioms may allow learners to reason about their meaning.
- **Opaque (non-compositional) idioms:** The meaning cannot be deduced from the components. For example, *kick the bucket* ('to die') has no clear connection between the words *kick*/*bucket* and dying; *white elephant* ('useless costly possession') has no literal link to elephants. These correspond to Vinogradov's fusions and to non-compositional idioms in Cowie (1998). Such idioms are the most lexicalized and often culturally specific. Other examples: *dead ringer* (exact duplicate), *red tape* (bureaucratic procedure). These are prototypical idioms and are fully listed as such in idiom dictionaries.

These semantic types exist on a spectrum. As Fernando (1996) notes, both idioms and collocations include literal and figurative uses. In practice, one must often rely on usage evidence to judge transparency. For instance, *fly off the handle* can be understood metaphorically ("lose one's temper suddenly"), but a naive learner hearing it literally might guess wrongly. Context disambiguates: "He flew off the handle when he saw the mess" is clearly idiomatic. Some expressions (like *tooth and nail* meaning *fiercely*) are conventionally figurative but remain somewhat transparent via imagery (teeth and nails are related to fight).

Discussion of Ambiguity and Context

Many phraseological units exhibit **context-dependence**. An expression may appear idiomatic in one context and literal in another. For example, "*we blew the whistle on corruption*" is figurative (reveal

wrongdoing) whereas in sports it can be literal (use a whistle). Similarly, *turn tail* can be literal (an animal twisting) or idiomatic (*retreat in fear*). Metaphorical meanings can bleed into new phrases over time (e.g. "*launch pad*" – originally literal, now figurative for starting projects).

There are also **edge cases** between idioms and collocations. Some common verb-noun pairs (*commit a crime*, *make a decision*) are so frequent that one might call them "collocations" rather than idioms. They are fully compositional, so we do not include them as idiomatic. Yet other multiword units are ambiguous: for example, *cold shoulder* is idiomatic ("dismissively ignore") but also has a literal meaning (a shoulder that is physically cold). Only by convention do we know which meaning is intended in context.

Across registers, phraseological units vary. In colloquial speech one hears many idioms (*hang in there*, *hit the sack*), whereas in formal writing one finds more adverbial and nominal idioms (*by and large*, *coup de grâce*). The British National Corpus confirms that idioms occur in both spoken (e.g. *"I've had it up to here"*) and written texts (*"under the aegis of"*). Literary language abounds in metaphors turned idiomatic (Shakespeare's *"heart of gold"*). Journalistic English often uses vivid idioms for impact (*"cornered like a wild animal"*, *"on the ropes"*). All registers contain phraseological units, but their frequency and form can shift by genre.

Taken together, the structural-semantic classification helps clarify the internal structure of English phraseology. For instance, *crocodile tears* (noun idiom, opaque) is distinct from *cry crocodile tears* (verb idiom, somewhat compositional). Recognizing these categories helps learners and analysts: a textbook might group idioms by type (noun phrase vs verb phrase) and by transparency (explaining *break the ice* via its metaphor).

CONCLUSION

Phraseological units in English form a rich and varied class of expressions. We have proposed a two-dimensional classification: one axis is structural form (nominal vs verbal vs adjectival/adverbial vs sentential), and the other is semantic transparency (fully transparent collocations through semi-compositional idioms to opaque idioms). This typology captures most fixed expressions encountered in corpora and dictionaries. It highlights that **"idiomaticity" is a gradient property**: some units (like *break the ice* or *as good as gold*) are partially analyzable, while others (*kick the bucket*, *white elephant*) are wholly conventional.

This framework has practical relevance. In *lexicography*, dictionaries such as Cowie et al. (1994) organise entries by part of speech; our structural categories align with that practice. Knowing that *slow and steady* is a fixed adverbial phrase, for instance, helps lexicographers list it under adverbs. In *language teaching*, instructors can cluster idioms by type (e.g. all adjective similes like *as hot as hell*) and by semantic opacity, focusing student attention appropriately. In *computational linguistics*, algorithms for multi-word expression detection can exploit such categories: a system might use a pattern for noun-phrase idioms (adjective + noun) differently than for verb idioms.

Future research directions include **corpus-driven statistical modelling** of phraseology. Large corpora can reveal the degree of fixedness and collocational strength of candidate units, which could refine our categories quantitatively. Machine learning approaches might predict idiomaticity scores for n-grams, testing the transparent–opaque continuum. Additionally, expanding phraseological

dictionaries with usage notes (e.g. register and collocates from corpus data) would aid both humans and NLP systems. Finally, investigating how phraseological use evolves in new media (social networks, for instance) could show how fixed expressions gain or lose transparency over time.

In conclusion, a combined structural-semantic classification provides a clear map of English phraseological units, from simple collocations to full proverb sentencess. It underscores that phraseology is a core part of the language: mastering it is essential for fluency, and understanding it enriches our linguistic analysis of meaning.

REFERENCES

Alisoy, H. (2023). Enhancing Understanding of English Phrasal Verbs in First-Year ELT Students Through Cognitive-Linguistic Methods.

Alisoy, H. (2023). Evolving vocabulary: Bridging colloquial and standard English in communication. *Norwegian Journal of development of the International Science No*, *120*, 89.

Alisoy, H. (2024). The magic of languages: Introducing a second language in early childhood education. *Retrieved from ResearchGate: https://doi. org/10.13140/RG*, *2*(11042.81603).

Alisoy, H., Mammadova, I., Asadova, B., Ismayilli, F., & Aliyeva, T. (2024). The future of language education: Integrating smartphones into the curriculum. *Edelweiss Applied Science and Technology*, *8*(6), 4539-4556.

Babazade, Y. (2025). Speech Acts and Hidden Meaning: A Journey into Pragmatics. *Acta Globalis Humanitatis Et Linguarum*, *2*(1), 221-228. https://doi.org/10.69760/aghel.02500128

Cowie, A. P. (Ed.). (2001). *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198294252.001.0001

Cowie, A. P., Mackin, R., & McCaig, I. R. (1993/1994). *Oxford Dictionary of English Idioms* (rev. ed.). Oxford: Oxford University Press.

Dobrovol'skij, D., & Piirainen, E. (Eds.). (2006). *Phraseology: An international handbook of contemporary research* (Vols. 1–2). Berlin: Mouton de Gruyter.

Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 125–143). Oxford: Oxford University Press.

Hajiyeva, B. (2024). Challenges in Understanding Idiomatic Expressions. *Acta Globalis Humanitatis Et Linguarum*, *1*(2), 67-73. https://doi.org/10.69760/aghel.01024066

Hajiyeva, B. (2025). Translating Idioms and Slang: Problems, Strategies, and Cultural Implications. *Acta Globalis Humanitatis Et Linguarum*, *2*(2), 284-293. https://doi.org/10.69760/aghel.025002123

Khudaverdiyeva, T. (2024). Enhancing Language Learning in Young Learners Through Azerbaijani Fairy Tales. *EuroGlobal Journal of Linguistics and Language Education*, *1*(1), 141-150. https://doi.org/10.69760/g6vv0g32

Moon, R. E. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.

Sadigzadeh, Z. Interplay of Linguistic Relativism and Universalism. *Akademik Tarih ve Düşünce Dergisi*, *11*(3), 1750-1758.

Sadiqzade, Z. (2025). The Linguistic Expression of Emotion: A Cross-Cultural Analysis. *EuroGlobal Journal of Linguistics and Language Education*, *2*(3), 42-54.

Sadiqzade, Z., & Alisoy, H. (2024). Level-Up Learning: Using Games to Teach English Across Student Levels. *EuroGlobal Journal of Linguistics and Language Education*, *1*(3), 181-194. https://doi.org/10.69760/egjlle.20240104

Sinclair, J. (2004). *Trust the Text: Language, corpus and discourse*. London: Routledge.

Vinogradov, V. V. (1950). *Theoretical foundations of phraseology*. Moscow: Academy of Sciences.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20*(1), 1–28.