

Structural and Semantic Taxonomy of English Phraseological Units: A Theoretical Perspective

¹ Hasan Alisoy

Accepted: 09.30.2025

Published: 10.18.2025

<https://doi.org/10.69760/portuni.0108005>

Abstract:

English phraseological units (multi-word expressions such as idioms, collocations, proverbs, etc.) can be classified by integrating their syntactic structure and degree of semantic compositionality. This theoretical article proposes a structured taxonomy that cross-classifies phraseological units by formal syntactic pattern (e.g. verb–object combinations, adjective–noun collocations, binomial phrases, clausal proverbs) and by semantic transparency (ranging from fully non-compositional idioms to partially compositional idioms, conventional collocations, and pragmatically fixed formulas or *pragmatemes*). Drawing on phraseological research and Meaning–Text Theory, we distinguish categories like **full idioms** (holistic, non-literal meanings), **semi-idioms** (partially analyzable metaphorical expressions), **collocations** (compositional yet convention-bound word combinations), and **pragmatemes** (fixed expressions tied to specific pragmatic contexts). Illustrative examples are provided for each category alongside clear classification criteria. The taxonomy accounts for structural diversity (from fixed noun phrases to whole proverbs) while highlighting the continuum of compositionality. The discussion addresses implications for linguistic analysis, cross-linguistic comparison (with brief reference to Azerbaijani equivalents), and theoretical modeling of the lexicon. This integrative classification aims to enhance our understanding of phraseological units, benefiting lexicology, semantics, syntax, and phraseology research by offering a comprehensive framework for describing multi-word expressions in English.

Keywords:

Phraseological units; Idioms; Collocations; Pragmatemes; Semantic compositionality; Syntactic structure; Phraseology taxonomy

1. Introduction

Phraseological units – variously termed *idioms*, *fixed expressions*, or *multimword expressions* – are a pervasive and richly varied component of the English lexicon. They encompass expressions like **idioms** (e.g. *kick the bucket* meaning “to die”), **proverbs** (e.g. *The early bird catches the worm*), **binomial phrases** (e.g. *bread and butter* meaning “livelihood”), **collocations** (e.g. *commit a crime*, *heavy rain*), and formulaic utterances like greetings and slogans. What unites these disparate expressions is their partial or complete **fixedness** and their resistance to literal, word-by-word interpretation. In other words, a

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

phraseological unit is a **non-free combination** of words whose overall meaning or usage cannot be fully predicted from its components in a *completely productive* way. At one extreme, we find highly opaque idioms such as *kicked the bucket*, whose meaning “died” bears little relation to the meanings of *kick* or *bucket*. At the other extreme, there are transparent but formulaic combinations such as *in the wrong place at the wrong time* (meaning simply that someone was unfortunate in timing and location) – here the phrase is semantically clear, yet it is conventionally fixed in form (one would not normally say “*in a bad place at a bad time*” to convey the same idea). Both types are considered phraseological units, contrasting with *free combinations* where words are selected freely and interpreted literally. This wide spectrum of fixed expressions makes phraseology an inherently “fuzzy” domain, requiring careful classification.

Understanding the different **types of phraseological units** is crucial for linguistics scholars in lexicology, semantics, syntax, and related fields. Each subtype of phraseological unit has distinct properties: some function as single lexical-semantic units with idiosyncratic meanings (posing challenges for semantics), some exhibit unique syntactic behaviors or structural constraints (of interest to syntax), and many carry cultural or pragmatic significance beyond their literal content (relevant to pragmatics and sociolinguistics). A robust classification helps linguists describe these phenomena systematically and supports applications in lexicography (e.g. idiom dictionaries, collocation manuals), language teaching, and cross-linguistic phraseology studies.

Traditional approaches to phraseology have often focused on either semantic opacity (idiomaticity) or syntactic form, but rarely both in an integrated way. This article proposes a **structural–semantic taxonomy** of English phraseological units that combines these two perspectives: we categorize expressions by (a) their syntactic structure or formal pattern, and (b) their degree of semantic compositionality (transparency of meaning). By cross-classifying structure and semantics, we aim to capture, for example, how a verb–object idiom that is non-compositional (*spill the beans*) differs from a verb–object collocation that is compositional (*make a decision*), or how a full-sentence proverb with figurative meaning (*A bird in the hand is worth two in the bush*) differs from a full-sentence pragmatic formula (*How do you do?* as a fixed polite greeting). The proposed taxonomy builds on insights from phraseological research including the classical European tradition (e.g., the work of Charles Bally and V. V. Vinogradov) and modern linguistic theories like **Meaning–Text Theory** (Mel’čuk) and others, in order to provide a comprehensive classification scheme.

The remainder of this article is organized as follows. **Section 2** reviews key theoretical background on phraseological units, including definitions and prior classifications by semantic and syntactic criteria. **Section 3** presents the proposed classification model, detailing the primary structural categories of English phraseological units and the hierarchy of semantic compositionality (from full idioms to pragmatically constrained formulas), and shows how these dimensions intersect. **Section 4** discusses the implications of this taxonomy: it addresses how the classification can aid phraseological analysis, considers cross-linguistic comparisons (with brief examples from Azerbaijani), and reflects on how theoretical frameworks like Meaning–Text Theory can model these units. **Section 5** concludes the article, summarizing the taxonomy’s contributions and suggesting avenues for further research in phraseology and linguistics.

2. Theoretical Background

Phraseology as a field has developed a rich terminology and various classification systems to describe multi-word lexical items. The term **phraseological unit** (often used interchangeably with *phraseeme*, *idiomatic expression*, or *fixed expression*) generally denotes “a stable combination of words with a fully or partially figurative meaning”. The concept was first clearly formulated by the Swiss linguist **Charles Bally** in the early 20th century, distinguishing fixed, non-compositional phrases from free word combinations [6]. Bally’s insights laid the groundwork for viewing certain multi-word combinations as integral units of meaning in the lexicon, not generated anew by syntax each time.

Subsequently, in the Russian linguistic tradition, **V. V. Vinogradov** developed one of the first systematic classifications of phraseological units by degree of semantic motivation. Vinogradov (1947) identified four main types (in translation): **phraseological adhesions** (completely opaque idioms, also called *fusions*), **phraseological unities** (partially motivated idioms), **phraseological combinations** (restricted collocations where one word’s use is constrained by convention), and **phraseological expressions** (fully compositional but institutionalized phrases). A *phraseological fusion* is a frozen idiom whose meaning cannot be derived from its parts at all (e.g., *to show the white feather* meaning “to display cowardice” in English – nothing in the literal image transparently hints at cowardice). A *phraseological unity*, in Vinogradov’s terms, has a figurative meaning that is somewhat suggestive of the literal meanings of its components, though still an idiom (for example, Russian *бить баклуши* “to be idle,” literally “to beat sticks,” retains a hint of an action but the overall meaning is idiomatic). *Phraseological combinations* are partly fixed collocations where components have their usual meanings, but certain words are conventionally paired (e.g., English *pay a visit* – the verb *pay* has a typical meaning but collocates with *visit* in a way that **make a visit* is less standard for the same idea). Finally, *phraseological expressions* are standard sayings or clichés that are fixed in form but semantically transparent (for instance, *Better late than never* – a proverb that means exactly what the words say, yet is used as a fixed expression). Vinogradov’s classification, though originally applied to Russian, has influenced how linguists think about idiomaticity in many languages, including English.

Western lexicographic and linguistic traditions have converged on similar distinctions. Many scholars differentiate **pure idioms**, **semi-idioms**, and **collocations** in English, drawing lines based on semantic compositionality (e.g., Cowie, 1985; Fernando, 1996; Moon, 1998). For example, Fernando (1996) defines *pure idioms* as completely non-compositional, *semi-idioms* as partially transparent (one component retains literal meaning, the other is figurative), and *literal idioms* (or *idiomatic collocations*) as those that are commonly used combinations with literal interpretation but restricted usage. Similarly, Nunberg, Sag, and Wasow (1994) distinguish idioms by their **analyzability** – whether parts of the idiom contribute identifiable sub-meanings. *Decomposable idioms* (like *spill the beans*, where *the beans* can be seen as “the secret” and *spill* as “reveal”) contrast with *non-decomposable idioms* (like *kick the bucket*, where no part corresponds to any aspect of “die”). These nuances correlate with how flexible an idiom is syntactically and how it might be paraphrased.

Another influential line of research comes from the **Meaning–Text Theory (MTT)**, particularly the work of Igor Mel’čuk and colleagues, who have elaborated a rigorous taxonomy of *phraseemes* (their term for phraseological units) within a lexicon-semantic framework. In Mel’čuk’s classification,

phraseological units are broadly divided by two intersecting criteria: (1) **semantic compositionality** – whether the whole expression’s meaning is the sum of meanings of its parts or not, and (2) **lexical vs. pragmatic fixation** – whether the restriction on word choice is due to lexical co-occurrence constraints or extralinguistic pragmatic convention. Crossing these dimensions yields three principal classes of lexical phrasemes: **idioms** (non-compositional phrasemes with arbitrary meaning), **collocations** (compositional phrasemes with one component constrained by lexical convention), and **clichés** (fully compositional, conventional expressions). Mel’čuk explicitly excludes the logically possible fourth combination (non-compositional *and* pragmatically constrained) as unattested – if an expression is non-compositional, its use is by definition a lexical matter rather than a purely situational convention. Under this view, *idioms* are defined strictly by non-compositionality: an idiom is “a phraseme whose meaning is not the predictable sum of the meanings of its components”. Idioms in MTT are further subclassified into **full idioms**, **semi-idioms**, and **quasi-idioms** based on whether *any* part of the expression’s meaning comes from a component word. For example, a full idiom’s meaning includes nothing of its parts (e.g., *go ballistic* “become very angry” – neither *go* nor *ballistic* contributes its usual meaning). A semi-idiom’s meaning includes the literal meaning of one component but not the other, plus an extra meaning element (often a metaphorical “pivot”) – for instance, *private eye* meaning “private investigator” includes *private* (suggesting “private” as in not official) but *eye* is metaphorical for “investigator”. A quasi-idiom (or “weak idiom”) includes the full meanings of all components, but the combination also carries an additional unpredictable meaning nuance – an example might be *red tape*, which literally involves “red” and “tape” but idiomatically means “bureaucratic paperwork” (here both *red* and *tape* have their normal meaning of a colored strip, yet historically the combination refers to a specific practice, acquiring an extra sense of official paperwork).

Collocations, in Mel’čuk’s framework and much lexicographic tradition, are **compositional** combinations of words that are **lexically constrained**, i.e. one chooses a particular word out of convention even though a literal synonym might seem logical. A classic example is how one talks about *making* a decision in English: *make a decision* is the normal collocation, whereas *do a decision* or *create a decision* are not used, despite *do* or *create* being literal synonyms in other contexts for the act of bringing about something. The meaning of *make a decision* is fully transparent (“perform the act of deciding”), but English convention dictates the verb “make” rather than alternatives. Such collocations typically involve a base word (the main semantic content) and a collocate (the word restricted by usage). For instance, in the phrase *heavy rain*, *rain* is the base and *heavy* is the selected collocate (one says *heavy rain* but not *strong rain*, whereas conversely one says *strong tea* not *heavy tea*, illustrating that *heavy* collocates with *rain* and *strong* with *tea* by convention). The collocate often carries a context-specific meaning that may not appear in its general dictionary definition but is contextually understood – e.g., *heavy* in *heavy rain* simply means “great in amount” (not literally heavy in weight). Yet, collocations are distinguishable from idioms because their overall meaning *does* equal the sum of parts (rain that is heavy = a large amount of rain). What makes them phraseological is the restricted word choice, often language-specific. Cross-linguistic comparisons demonstrate this: different languages choose different collocates for the concept “make a decision” (*prendre* “take” a decision in

French, *eine Entscheidung treffen* “meet” a decision in German, *karar vermek* “give a decision” in Turkish, etc.), reflecting that the selection of the support verb is arbitrary from a purely logical perspective.

The category of **clichés** or **formulas** refers to fixed, *fully compositional* phrases that are considered the standard way of expressing a given meaning in a language. They are often **pragmatically** restricted to certain communicative situations or conventionalized as social routines. Examples include *How are you?* as a greeting, *Happy birthday!* as a congratulatory formula, or *Yours sincerely* as a letter closing. These expressions are usually straightforward in meaning (no hidden idiomatic sense – *Happy birthday* literally conveys a wish for a happy birthday), and a person hearing them for the first time would understand the literal meaning. However, they count as phraseological units because they are **formulaic**: alternate phrasings are not customary (for instance, English speakers say *How are you?* rather than *How do you fare?* in everyday greeting, even though the latter is grammatically correct; and one says *Yours sincerely* rather than *Sincerely yours* or other variations in formal correspondence by convention). Within the realm of clichés, Mel’čuk and others further distinguish **pragmatemes** – phrases that are not only conventional but tied to a specific extralinguistic situation or communicative act. A pragmateme is a subtype of cliché “whose usage restrictions are imposed by the situation of utterance”. For example, *Will you marry me?* is a standard utterance used specifically for the speech act of making a marriage proposal. Its literal meaning is clear, and alternative wording (*Do you want to be my wife/husband?*) might be understandable but would be perceived as unusual – the pragmateme is the expected formula for that context. Other examples are fixed notices like *Keep off the grass* (found on lawn signs) or *Best before [date]* on food packaging. These could theoretically be phrased differently, but convention and situational appropriateness dictate a particular wording. Pragmatemes thus occupy one corner of the phraseological space: they are fully literal in meaning (hence compositional), yet fixed due to social or situational convention.

In summary, previous research provides us with two fundamental axes for classifying phraseological units: **structural form** (what syntactic shape the expression takes) and **semantic compositionality** (how the overall meaning relates to the meanings of the parts). Researchers like Alexander (1984, 1987) have enumerated the wide variety of structural types that idiomatic or fixed expressions can assume – including *binomials* (e.g. *wine and dine, wear and tear*), *compound-like noun phrases* (e.g. *red tape*), *phrasal verbs* (verb + particle combinations like *give up, look after*), *similes* (comparative idioms like *as cold as ice*), *full clauses and proverbs*, *stock phrases* (conventional sayings or quotations), and even *discourse markers* and *social formulas*[5]. Meanwhile, semantic-based classifications (Vinogradov, Mel’čuk, Fernando, etc.) draw lines between pure idioms, semi-idioms, collocations, and free combinations. What remains is to integrate these two perspectives into a single taxonomy that can describe, for instance, a “verb-object pure idiom” as opposed to an “adjective-noun collocation” or a “sentential pragmateme.” This integration is the goal of our proposed classification model.

3. A Structural–Semantic Classification Model for English Phraseological Units

In this section, we present a taxonomy of English phraseological units that combines **structural criteria** (the syntactic form of the expression) with **semantic criteria** (the degree and type of idiomaticity or compositionality). The classification is organized as a matrix of sorts: one dimension enumerates the major structural types of multi-word expressions, and the other dimension specifies

categories of semantic composition (from non-compositional to fully compositional but pragmatically fixed). We describe each structural category in turn, and within each, identify subcategories by semantic type. Illustrative examples are given for each intersection of structure and semantic type.

3.1 Structural Types of Phraseological Units

Structurally, English phraseological units can be found in virtually any syntactic category or construction type that allows multi-word combinations. The following are the primary structural types considered in our taxonomy:

- **(a) Verb Phrase Units (Verb + Object/Complement Idioms and Collocations):** A very large portion of idioms in English are verb-centered constructions, often involving a verb and object (or other complement). For example, *spill the beans* (verb + object) is an idiom meaning “reveal a secret,” *take off* (verb + particle) as in *take off [intransitive]* meaning “to suddenly succeed” (an idiomatic phrasal verb), or *look forward to [X]* meaning “anticipate [X] eagerly” (a prepositional verb phrase). Collocations in this category include cases like *commit a crime* (verb + object collocation, meaning “perpetrate a crime” – fully literal but conventionally *commit* is used) or *apply for a job* (verb + prepositional phrase – one *applies for* something, not **request for a job* in standard usage). Verb phrase units can function as a predicate in a sentence. Syntactically, some are fixed (e.g. *shoot the breeze* – “chat idly” – cannot be passivized or altered easily), while others allow some variation (*spill the beans* can become *the beans were spilled by X* in passive, albeit this might sound less common). We will further categorize these by semantic type in Section 3.2, but structurally they share the feature of being built around a verb and its complements.
- **(b) Noun Phrase Units (Nominal Idioms and Collocations):** These are multi-word noun phrases that act as a single nominal concept. For example, *a piece of cake* is a noun phrase idiom meaning “something very easy” (its surface structure is “indefinite article + noun + of + noun”). Another example: *red tape* meaning “bureaucratic procedures” is a noun phrase (Adj + N) with an idiomatic meaning (historically referring to red ribbon used to tie official documents). A semi-idiomatic noun phrase might be *a fresh pair of eyes* (meaning “an outsider’s perspective” – partially transparent metaphor). Collocational noun phrases are extremely common as well: e.g. *strong tea* (Adj + N collocation, meaning tea that is brewed concentrated – *strong* is a conventional choice; one wouldn’t say **powerful tea* even though the meaning “strong” is literal). Another is *high voltage* (Adj + N collocation in technical contexts – one says *high voltage* rather than **strong voltage*). Some noun phrase units include proper nouns or names (e.g. *John Bull* as a personification of England is an idiom-like name, or *the Big Apple* for New York City), though we will focus on general common-noun phrases here. Structurally, these can often take plural or possessive forms if the syntax allows (e.g. *skeletons in the closet* pluralizes an idiom to mean “hidden secrets”). Many nominal phraseologisms are also **compound-like** (in fact, English compounds often border on phraseology when they aren’t entirely single words – e.g. *brain drain*, *ivory tower*, *fat cat* in the sense of a rich person, etc., are written as two words but act as compound nouns with idiomatic meaning).

- **(c) Adjective + Noun Collocations and Idioms:** This is a subtype of noun phrase but worth noting separately because many collocations take the form of a modifier + noun. For instance, *cold war* (an international tension without direct conflict) is an idiom as a whole NP, but structurally Adj+N. Collocational examples abound: *heavy rain*, *rich vocabulary*, *social justice* (some nouns strongly prefer certain adjectives in idiomatic usage). While these could be lumped with general noun phrases, lexicologists often specifically discuss adjective–noun collocations as a distinct phenomenon since the noun is seen as the base and the adjective as a collocate that conveys a particular nuance.
- **(d) Binomial Expressions (X and Y):** Binomials are fixed pairings of words usually joined by *and* (or sometimes *or*) that function as a single conceptual unit^[5]. English has many of these: *bread and butter* (literally a common food pairing, but idiomatically “basic means of support or livelihood”), *pros and cons* (“advantages and disadvantages”), *wear and tear* (“damage that naturally and inevitably occurs as a result of normal use”), *fish or cut bait* (an American idiom meaning “do one thing or the other, but stop hesitating”), *more or less* (a fixed phrase meaning “approximately”). Binomials may exhibit irreversible order (one cannot swap the order without it sounding odd, e.g. *butter and bread* is not the usual way to say *bread and butter* meaning livelihood). Semantically, some binomials are idiomatic (*wine and dine* as a verb meaning “entertain lavishly with food and drink”), while others are simply conventional pairings (e.g. *husband and wife* is a fixed order for the literal pair). There are also rhythmic or alliterative binomials (*doom and gloom*, *riff-raff*) which often carry idiomatic meaning as a whole. We classify binomials separately because they often function as a single unit (noun phrase, adjective phrase like *deaf and dumb*, or even verb phrase as in the example *fish or cut bait*).
- **(e) Comparative Idioms (Similes):** A special subtype of phraseological unit in English are formulaic comparisons, often of the form “as X as Y” or “like a Y.” For example, *as blind as a bat* (meaning “very blind”), *as cool as a cucumber* (“calm under pressure”), *to fight like cats and dogs* (“to fight fiercely or quarrel constantly”). These simile idioms typically have a fixed pattern and a figurative standard of comparison. They are semantically semi-transparent: one can often see the metaphor (cats and dogs are proverbial enemies, hence the figurative meaning of fighting fiercely), but they are fixed expressions and one cannot freely substitute other animals or items without losing the idiomatic meaning. Comparative idioms in English form a well-known list of colorful expressions (e.g. *as stubborn as a mule*, *like a bull in a china shop*, etc.).
- **(f) Full Sentences: Proverbs and Sayings:** Many phraseological units are entire sentences (or clauses) that encapsulate a piece of wisdom, a folk saying, or a formulaic remark. **Proverbs** are a prime example: *The early bird catches the worm* (meaning initiative or prompt action leads to success), *You can't have your cake and eat it too*, *When it rains, it pours*. These are usually complete sentences, often metaphorical or proverbial in nature, and used as ready-made utterances. Their meanings are sometimes transparent metaphors but often require cultural or contextual knowledge. Another kind of sentential phraseological unit is the **maxim or catchphrase**, e.g. *Boys will be boys* (excusing typical male behavior), or famous quotes that have become clichés

(*Rome wasn't built in a day*). Structurally, these are clauses that can stand alone. Many proverbs are syntactically well-formed sentences, though some have archaic or elliptical grammar (e.g. *the more, the merrier* is a fragmentary clause). Because they function as complete utterances, their usage is often tied to communicative context (some proverbs are used to admonish, others to console, etc.). Semantically, proverbs often carry an idiomatic or metaphorical meaning distinct from the literal reading, but a few are almost literal truisms.

- **(g) Pragmatic Formulas (Illocutionary Units):** These include **speech-act formulas** and social routines, many of which are also full sentences or clauses, but distinguished by their pragmatic function. For instance, *How do you do?* (a formal greeting), *Long time no see!* (an informal greeting after absence), *I beg your pardon* (formulaic apology or polite excuse), *No smoking* (prohibition sign), *Thank you*, *Sorry to bother you, but...* These are often not idiomatic in the semantic sense (their literal meaning aligns with their function, or sometimes they have slightly fossilized syntax), but they are fixed ways of accomplishing communicative acts. In our classification, these pragmatic formulas – especially those that are restricted to certain contexts – will fall under the category of **pragmatemes** or clichés, depending on their properties (discussed in 3.2). They are worth noting structurally because some are single clauses (imperatives like “Enjoy your meal”), some are phrase fragments (e.g. *Best regards*), and some allow slots for insertion (e.g. *My name is ___* in introductions is a formulaic frame).

It is important to note that these structural categories can overlap or nest. For example, a proverb (sentential) might contain a smaller idiom within it, or a binomial may appear inside a longer idiomatic phrase. However, for classification purposes, we treat the *largest relevant unit* as the phraseological unit if it is commonly considered a fixed expression in its entirety.

3.2 Semantic Categories by Compositionality

Orthogonal to the structural types, we classify phraseological units by **semantic compositionality** – essentially, how literal or non-literal the expression is, and what constrains its usage. We define four broad semantic categories: **(i) non-compositional idioms**, **(ii) partial idioms**, **(iii) collocations (conventionalized compositional combinations)**, and **(iv) pragmatemes (pragmatically fixed formulas)**. These correspond roughly to points along a spectrum from full idiomaticity to literalness, as well as distinctions in *why* the phrase is fixed (lexical-semantic idiosyncrasy vs. usage convention vs. situational necessity).

- **(i) Non-Compositional Idioms (Fully Idiomatic Expressions):** These are expressions whose meaning cannot be deduced from the meanings of the individual words – in other words, they are **semantically opaque** or “holistic.” The classic defining feature of an idiom is exactly this non-compositionality. Examples in English abound: *kick the bucket* (“die”), *spill the beans* (“reveal a secret”), *break the ice* (“initiate social interaction in a tense situation”), *on cloud nine* (“extremely happy”), *red herring* (“misleading clue or piece of information”), etc. In such idioms, either none of the words carry their literal meanings, or any literal connection is very indirect. Idioms often arise from metaphors or historical references that are not transparent to modern speakers. For instance, *kick the bucket*'s origin is obscure (folklore suggests it comes

from a method of suicide or slaughter, but that is not known to most speakers), so one hears it as a single lexical chunk meaning “to die.” Syntactically, many fully idiomatic expressions are fixed or allow only limited morphological variation (e.g., one can say *kicked the bucket* in past tense, but not **kicked a bucket* in reference to dying, nor **knock the bucket*). These idioms behave like lexical items – they have to be learned as units. According to Mel’čuk’s typology, these would be *full idioms* if none of the component’s meanings are present in the whole [18]. For example, *go bananas* (“go crazy”) is a full idiom: *go* doesn’t contribute its normal motion meaning, and *bananas* doesn’t contribute its literal fruit meaning (the expression has nothing to do with actual bananas).

Some fully idiomatic expressions can be longer than a phrase – e.g. proverbial idioms like *the coast is clear* (“no danger of being observed”) or *to beat around the bush* (“to avoid the main topic”). These still fit the non-compositional criterion. It’s worth noting that non-compositional idioms can sometimes be “dissected” by native speakers after the fact (through folk etymology or finding some metaphorical connection), but the key is that one *cannot predict* the meaning through normal decoding. For instance, *to let the cat out of the bag* meaning “to reveal a secret” might conjure a mental image related to surprise or chaos of a cat escaping, but there is nothing inherent that would let a learner know it means to divulge a secret – it must be learned as a whole.

Non-compositional idioms can occur in any structural form: verb phrases (most common), noun phrases (e.g. *a red herring*), adjectives (e.g. *brand new* – historically idiomatic meaning “completely new”), binomials (e.g. *by and large* meaning “on the whole”), or whole sentences (many proverbs). In our integrated classification, when we label something a “non-compositional idiom,” we will also note its structural form (e.g. “VP idiom,” “NP idiom,” “clausal idiom”) as needed.

- **(ii) Partially Compositional Idioms (Semi-Idioms and Metaphorical Units):** Between pure idioms and free combinations lies a large gray area of expressions that carry **figurative or specialized meanings** but still hint at their literal components. These are often called **semi-idioms** (Mel’čuk), **metaphorical idioms**, or just **figurative expressions**. In such cases, one part of the expression may retain its literal meaning while the other part is used metaphorically, or the expression as a whole is a metaphor that is somewhat transparent. For example, *to lose one’s touch* (meaning “to lose one’s ability or knack”) has a literal component *lose* (as in to lose something) and a figurative component *touch* (metaphorically meaning skill or expertise). The phrase *foot the bill* (meaning “to pay the bill”) is another case: *bill* is literal, but *foot* is a verb that originally meant “to add up the foot (total) of an account,” which is not obvious today – so it’s partly idiomatic (foot = pay). Another example: *draw the line [at]* (“set a limit”) – one can see a metaphor of literally drawing a boundary line, so it’s not entirely opaque, but it’s a fixed expression for establishing a limit.

Some idioms of this type are often termed **metonymic idioms** or **synecdoche idioms** – for instance, *lend a hand* meaning “help” (here *hand* stands for assistance, which is a part-for-whole metonymy). The phrase is somewhat transparent (because helping often literally involves one’s hands), yet it is a fixed idiom. Mel’čuk’s definition of *semi-idiom* requires that the meaning includes the full meaning of one component (not as the semantic pivot) and not of the other, plus an extra meaning element. Using that framework: *lose one’s head* (“to panic, lose self-control”) might be analyzed as including *lose* in its usual sense (to be deprived of) and *head* not in its usual “body part” sense, plus an extra meaning of “sanity/self-control” (the head representing the mind). Thus it’s a semi-idiom with *lose* literal and *head* idiomatic (meaning mind). Many idioms of emotion or state fit this pattern: *give someone the cold shoulder* (meaning “deliberately ignore someone”) includes *shoulder* in some metaphorical way (part of turning away?) but *cold* is metaphorical for “unfriendly” – one can parse it partially but not fully.

In our taxonomy, partially compositional idioms may still be considered “idioms” broadly, but it is useful to note their intermediate status because they often have different syntactic flexibility and different behavior in language. They tend to be more analyzable and sometimes allow wordplay or variation (for instance, people might joke “He lost his head – figuratively, not literally” acknowledging the idiom). This category also includes what some call **clichéd metaphors** or **idiomatic comparisons** like *X is in the driver’s seat* (metaphorically meaning X is in control). The image is understandable, but it’s a fixed way to express that particular meaning.

- **(iii) Conventional Collocations (Fully Literal but Fixed Combinations):** Here we have expressions that are **compositional in meaning** – the meaning is just the sum of the parts – yet they are **not freely chosen**, as there is a conventional pairing or habitual co-occurrence. This includes what are sometimes called **phraseological combinations** or **restricted collocations**. As discussed earlier, a typical collocation consists of a **base** word that is the central concept, and a **collocate** that is chosen by convention to express a particular nuance or to fit a usage pattern. The key characteristic is that an equivalent meaning could in theory be conveyed by other words, but actual usage favors a particular lexical combination. For example, English speakers say *make a mistake* (not *do a mistake*), *tell a lie* (not *say a lie*), *strong tea* (not *powerful tea*), *heavy rain* (not *strong rain*), *big mistake* (rather than *large mistake* in many contexts, even though large is a synonym of big). These are all perfectly literal – a mistake that is big is literally a big mistake – but collocational preferences impose a fixed choice.

Collocations also cover cases of words that habitually go together due to tradition or idiomatic preference even if synonyms could work: e.g. one *launches a campaign* but *wages a war*, one *holds a conversation* but *has a chat*. Often there is no strict rule other than lexical convention. The idea “cause someone trouble” is typically phrased as *give someone trouble* or *make trouble for someone*, but not **cause trouble to someone* (even though cause would be literally apt). These patterns are stored in speakers’ minds as part of knowing how to use the words idiomatically.

Collocations can be specific to certain registers or domains. For instance, in academic English, one speaks of *conducting research* (not *doing research* in formal writing) and *providing an explanation* (rather than *giving* in some contexts), etc. These are matters of usage rather than semantics per se. Because collocations are semantically transparent, a person can often understand them on first exposure, but producing them correctly is part of mastering the phraseological norms of the language (an area of interest in second language acquisition research, cf. Wray 2002; Nesselhauf 2005).

It should be noted that collocation is a gradient concept – virtually any word can collocate with any other given the right context, but phraseologists focus on **recurrent, relatively fixed collocations** that have an arbitrary element. The boundary between what is a “fixed collocation” and what is just a frequently occurring free combination can be fuzzy, and usage frequency plays a role. For our purposes, collocations are considered phraseological units if they show a notable element of lexical restriction (e.g., a native speaker will consistently use one particular word in that context where alternatives are theoretically possible, often to the point that alternatives sound odd or non-idiomatic).

- **(iv) Pragmatemes (Pragmatically Fixed Formulas):** The last category comprises expressions that are *fully compositional and understandable* in literal terms, but are **fixed by convention of use in specific contexts or by social pragmatics**. These include the **routine formulas** and **clichés** of language – things we say in particular situations because convention so dictates. For example, *Happy New Year!* is literally just a happy wish, but it’s only used (sincerely) around New Year’s and is a fixed form (we don’t say *Joyous New Year* ordinarily). *Rest in peace* is a conventional expression used in funerary contexts or condolences; its literal words are simple, but no alternative phrasing (*Repose in tranquility*, *May you rest peacefully*, etc.) carries the same conventional force. Traffic signs and public notices are another example: *No Smoking*, *Keep Out*, *Handle with care*, *Fragile – do not drop*, etc., are standardized in phrasing. In correspondence or formal situations, we have expressions like *Thank you for your attention*, *All the best*, *To whom it may concern* (in letters). These examples illustrate that pragmatemes are tied to **situational contexts**: letter-writing, holidays, warnings, requests, etc.

Pragmatemes often have the form of complete sentences or imperative phrases. Some can include fill-in slots (e.g. *Dear [Name]*, in a letter – where only the name changes, or *Welcome to [Place]* in which the place name is inserted). In Meaning–Text Theory terms, pragmatemes are pragmatically constrained clichés. We treat them as a separate semantic category because what “fixes” them is not an idiosyncratic meaning (they are not idioms in the sense of having unpredictable meaning), but rather a **pragmatic function or conventional usage**. They highlight the fact that not all phraseological units are about idiomatic meaning – some are about doing things with words in a culturally expected way (for instance, English speakers say *Bless you* after someone sneezes – the words are ordinary but the usage is culturally fixed).

Pragmatemes have been less studied than idioms and collocations, but they are gaining attention as important building blocks of communication (especially in cross-cultural

pragmatics). In cross-linguistic perspective, pragmatemes often do not translate literally; e.g., the English telephone answering phrase “Hello?” corresponds to Russian “Алло?” and French “Allô?” – similar form, whereas something like *fingers crossed* (said to express hope) in Azerbaijani might have no direct literal equivalent and instead one might say a different conventional phrase or perform a gesture.

It is worth emphasizing that these four categories are not always cleanly separated – they blur into each other. We can imagine a continuum: at one end, a free combination like *read the book* (fully literal, not fixed at all – not a phraseme by any account). Moving a step into phraseology, we get collocations like *read aloud* or *read voraciously* (nearly free, but certain adverbs collocate typically). Further in, we have more fixed collocations like *make a decision*. Then semi-idioms like *make waves* (“cause trouble/attention”) – here *make* is literal but *waves* is metaphorical for disturbance. Then strong idioms like *make ends meet* (“manage financially”) – here *make* is not exactly “create” but “cause to meet,” and *ends* is metaphorical (ends of budget), partially understandable but largely idiomatic. Finally, full idioms like *make a killing* (“earn a lot of money easily”) – one can grasp the violent metaphor but it’s a specific idiom. Pragmatemes sort of occupy a different axis (since one could have a pragmatically fixed idiom, though Mel’čuk argues a non-compositional pragmateme is effectively just an idiom used in one context). For simplicity, in our taxonomy we treat pragmatemes as those expressions whose primary defining feature is context-specific usage, typically compositional in semantics.

3.3 Integrating Structure and Semantics: The Taxonomy Matrix

By combining the structural types from 3.1 with the semantic categories from 3.2, we arrive at an integrated classification. Essentially, any given phraseological unit can be described by a structural label and a semantic label. Table 1 (conceptually) presents a matrix of examples:

- **Verb Phrase – Non-Compositional Idiom:** e.g. *spill the beans* (verb + object, meaning “reveal a secret”). Structurally a VP; semantically a full idiom (non-compositional: *beans* are not literally involved in meaning “secret”). Another example: *kick the bucket* (V+NP idiom, meaning “die”)[3].
- **Verb Phrase – Partial Idiom:** e.g. *hold fire* (verb + object, meaning “wait, stop for now” as in *hold your fire* – literally *fire* implies gunfire, but meaning is “stop action”; *hold* is somewhat literal “hold back”). Also *carry weight* (as in *His opinion carries weight* – meaning “has influence,” where *weight* is metaphorical for importance but *carry* is used figuratively).
- **Verb Phrase – Collocation:** e.g. *draw a conclusion* (choose *draw* instead of say *make*; fully literal “draw” meaning “pull” is a bit metaphorical but standard usage). *run a business* (meaning “manage a business” – compositional, but one *runs* a business, doesn’t *drive* or *lead* a business in idiomatic English).
- **Verb Phrase – Pragmateme:** e.g. *God bless you* (after a sneeze – a fixed VP used as a response, pragmatically constrained to that situation). Or *Excuse me* (imperative form used for getting

attention or apologizing briefly – literally a verb phrase but functioning as a conventional politeness formula).

- **Noun Phrase – Non-Compositional Idiom:** e.g. *red tape* (Adj+N, meaning “bureaucracy” – no literal connection to colored tape). *a hot potato* (meaning “a controversial issue” – not a literal potato).
- **Noun Phrase – Partial Idiom:** e.g. *a silver bullet* (meaning “a simple miraculous solution” – based on metaphor of mythical silver bullet as cure-all; *bullet* is literal, *silver* metaphorical for special). *the icing on the cake* (meaning “an additional benefit” – one can see the cake metaphor, but it’s an idiom; components have literal referents yet refer to a figurative situation).
- **Noun Phrase – Collocation:** e.g. *a strong argument* (Adj+N, meaning an argument that is convincing – *strong* is a conventional adjective for argument strength). *a stroke of luck* (N+Prep+N, fully literal meaning a bit of luck, but “stroke” is a fixed collocate for luck, we don’t say **a bit of luck*). Also technical collocations like *mother tongue* (meaning native language, literal words but fixed pairing).
- **Noun Phrase – Pragmateme:** This is less common, since pragmatemes are often clausal, but could include something like *Your obedient servant* (as a closing in a formal letter – an NP used formulaically in historical usage). Or set titles/honorifics: *Your Majesty* is a phraseological NP used only in addressing a monarch (pragmatically restricted to that context).
- **Binomial – Non-Compositional Idiom:** e.g. *fun and games* (usually in negative or sarcastic use: “It’s not all fun and games,” meaning it’s not all entertaining – idiomatic meaning beyond just literal fun or games). *cloak and dagger* (Adj and N used as compound adjective meaning “related to espionage/secretive,” from the imagery of old spy fiction – nonliteral).
- **Binomial – Partial/Collocational:** Many binomials are somewhat literal pairs but become a fixed saying: *knife and fork* (literal tools, but always said in that order; a collocation in a sense), *high and dry* (“stranded”; originally literal meaning without resources, now idiom meaning without help – maybe partial idiom), *salt and pepper* (literal objects but fixed order when together).
- **Binomial – Pragmateme:** Hard to think of a binomial pragmateme, as binomials usually aren’t situational utterances. Perhaps something like standard pairs in ceremonies (*husband and wife* as pronounced in weddings – though that’s more literal collocation than situational formula).
- **Comparative Simile – Idiomatic:** e.g. *as cold as ice* (literal comparison but fixed form; in some cases these are more cliché than idiom), *fight like cats and dogs* (idiomatic simile for fighting fiercely – partial idiom because the image is transparent that cats and dogs fight, yet it’s a fixed phrase).

- **Sentence/Proverb – Idiom:** e.g. *The straw that broke the camel's back* (meaning the final minor action that causes a big reaction – proverb, non-compositional unless one knows the metaphor). *You can't judge a book by its cover* (proverb, metaphorical but understandable – perhaps partial idiom).
- **Sentence – Collocational (literal proverb):** e.g. *Easy come, easy go* (meaning “quick gains can be lost quickly” – fairly literal and simple but a fixed proverbial expression). *Business is business* (literal truth said as a fixed phrase to justify something).
- **Sentence – Pragmateme:** Many formulas fall here: *May I help you?* (said by a clerk to a customer – literal but conventional in context), *I now pronounce you husband and wife* (literal but part of a marriage ritual), *Do you take X to be your lawfully wedded...* (marriage vow formula), etc. Also social scripts like greetings: *How are you?* – literal question but used as routine, where the expected answer *I'm fine, thanks* is also formulaic regardless of actual state, demonstrating pragmatic fixedness.

For conciseness, we may not enumerate all possible combinations, but the idea is each phraseological unit can be located in this classification. For instance, consider *raining cats and dogs* (meaning “raining very heavily”). Structurally, it's a gerund clause or participial expression (*it's raining X*). Semantically, it's a non-compositional idiom (the animals do not contribute to meaning except via an obscure metaphor). So in our taxonomy we'd label it a **clausal idiom (weather expression)**. Another example: *by heart* (meaning “from memory”). Structurally a prepositional phrase, semantically a full idiom (one cannot guess that “heart” means memory or understanding here). Meanwhile, *on the one hand... on the other hand...* (for contrasting points) is a formulaic construction – structurally two paired prepositional phrases acting as discourse markers, semantically mostly literal but it's an idiomatic way to structure an argument, thus more of a cliché.

To illustrate the classification more formally, let us take a few concrete examples and classify them:

- **Example 1:** *Spill the beans*. Structure: Verb + Object (verb phrase). Semantics: Non-compositional idiom (full idiom: ‘spill’ and ‘beans’ don't retain literal meaning “cause beans to fall out” – it means “reveal secret”). So category: **Verb Phrase – Full Idiom**.
- **Example 2:** *Make a decision*. Structure: Verb + Object (verb phrase). Semantics: Collocation (compositional meaning “take a decision” but conventional verb choice). Category: **Verb Phrase – Collocation**.
- **Example 3:** *Break the ice*. Structure: Verb + Object (verb phrase). Semantics: Partial idiom (metaphorical meaning “initiate friendliness,” *break* partly metaphorical, *ice* metaphor for stiffness in relations – one can see some image). It might be considered fairly opaque historically, but contemporary speakers often understand the metaphor of “ice” as social tension. So we label it **Verb Phrase – Semi-idiom**.

- **Example 4:** *Bread and butter*. Structure: Binomial noun phrase (N and N). Semantics: When used idiomatically to mean “livelihood” (as in *Teaching is my bread and butter*), it is a non-compositional idiom (one isn’t literally talking about food) – though metaphorically one’s income buys one’s bread and butter. Category: **Binomial – Idiom**.
- **Example 5:** *On cloud nine*. Structure: Prepositional phrase (idiomatic PP functioning as adjective or adverb meaning “elated”). Semantics: Non-compositional (the literal image of cloud nine doesn’t inherently mean very happy). **PP Idiom**.
- **Example 6:** *Under the weather*. Structure: PP (meaning “slightly unwell”). Semantics: Non-compositional idiom (though arguably metaphorical, one might imagine weather pressing down – but it’s not obvious). **PP Idiom**.
- **Example 7:** *The X that Y* (as in *the straw that broke the camel’s back*). Structure: complex NP clause (proverbial template). Semantics: idiomatic metaphor. **Clausal Idiom (Proverb)**.
- **Example 8:** *How do you do?* Structure: Full sentence (interrogative). Semantics: Pragmateme – a formulaic greeting not meant as a literal information-seeking question in context. **Sentence – Pragmateme**.
- **Example 9:** *No pain, no gain*. Structure: Clause (proverb, elliptical two-part saying). Semantics: Semi-idiom (literal words but succinct metaphor: without suffering work, you won’t achieve benefit – each word literal, overall meaning proverbial truth). It’s almost literal but we class it as a fixed proverbial expression (could be considered a cliché). **Sentence – Cliché (Proverb)**.

Each of these examples demonstrates how the two-dimensional classification works.

We can visually imagine the taxonomy as a table where rows are structural types (Verb Phrase, Noun Phrase, Adjective–Noun, Binomial, Prepositional Phrase, Clause, Formulaic utterance, etc.) and columns are semantic types (Full Idiom, Partial Idiom, Collocation, Pragmateme). Not every cell in the matrix will be populated by a common example (some combinations might be rare or logically not applicable), but most are. For instance, an **Adjective–Noun Pragmateme** might not have many examples aside from fixed titles or honorifics as noted (e.g. *Your Excellency* – an NP used only in protocol). On the other hand, **Clause Collocations** is not a typical notion – collocation is usually discussed for smaller structures, though one could argue that certain clausal constructions with literal meaning are conventional (like the paired *on the one hand... on the other hand* example which is like a collocational frame in discourse).

4. Discussion

The proposed structural–semantic taxonomy provides a comprehensive framework for understanding English phraseological units. In this section, we discuss several implications and applications of this classification: how it aids phraseological analysis, how it facilitates cross-linguistic and contrastive studies (with a brief look at English vs. Azerbaijani phraseology), and how it interfaces with theoretical

models such as Meaning–Text Theory and other lexicon models. We also consider the relevance for lexicography and computational linguistics.

4.1 Implications for Phraseological Analysis and Lexicology

For linguists and lexicographers, having a clear taxonomy helps in several ways. Firstly, it clarifies **terminology** and avoids talking past one another when referring to “idioms” or “collocations.” By specifying, for example, that we are dealing with an “Adj+N collocation (compositional, restricted)” versus a “Verb+N full idiom (non-compositional),” we make explicit both the syntactic behavior and the semantic nature. This is crucial because the **lexico-grammatical behavior** of phraseological units often correlates with their semantic type. Non-compositional idioms tend to be more **lexically frozen and syntactically inflexible** – they often resist modification, passivization, or variation in their components (e.g., one cannot normally say *the bucket was kicked by John* to mean John died, at least not without irony). Collocations, being semantically transparent, are usually **more flexible**: they might allow synonyms in some contexts or translations, but still have a preferred form (for instance, we can sometimes transform collocations passively or adjectivally: *a crime was committed*, *decision-making* derived from *make a decision*). Partial idioms lie in between, sometimes allowing certain variations (people play with semi-idioms in jokes or headlines by substituting components, capitalizing on the partial transparency). Knowing the idiomaticity level can guide what transformations or morphological changes an expression can undergo.

Another area is **lexical productivity vs. fixedness**. By definition, phraseological units are fixed to some degree, but our classification highlights which part of the phrase is fixed and which might be an open slot. Collocations usually have one main fixed element (the collocate) and one that is the base and can combine with other partners (e.g. *commit a crime*, *commit treason*, *commit an error*, here *commit* is the collocate that specifically goes with certain nouns). Idioms are often fixed in all parts, though some idioms allow limited lexical variations (like *kick the bucket* has a variant *kick off* as a slang for “die,” or *give up the ghost* as a variant of *give up the ghost* meaning die – well, those are different idioms entirely but share a component). Recognizing patterns in structure can help identify **families of idioms** or **idiomatic constructions**. For example, English has a family of “the X to end all X” (as in *the movie to end all movies*, meaning the ultimate movie) – structurally a noun phrase template, semantically a hyperbolic idiom. Our structural taxonomy would place that as a **frame** that generates idiomatic instances. A classification that is too coarse (just idiom vs collocation) might overlook such constructional patterns.

4.2 Cross-Linguistic Considerations: English and Azerbaijani Phraseology

Phraseological units are notoriously language-specific in their surface form, even if underlying meanings or metaphors are often shared across cultures. An advantage of a structured classification is that it provides a basis for **contrastive phraseology**: we can compare how different languages express similar idiomatic concepts, and whether they do so with similar structures or different ones.

Taking **Azerbaijani** as a point of contrast (a Turkic language spoken in Azerbaijan), we find both parallels and divergences with English in phraseology. For instance, English idioms that involve animals or vivid imagery often have counterparts in Azerbaijani, but not always with the same animals

or images. English says *When pigs fly* to mean “never” (an impossible time); Azerbaijani uses a different idiom – one common equivalent is *filan iş qiyamətə qalandı* literally “that thing will happen when the apocalypse comes,” effectively meaning “when hell freezes over” or “never.” Here the concept (impossible time) is shared, but the imagery differs (pigs flying vs. doomsday). Structurally, English *when pigs fly* is a clause idiom; the Azerbaijani equivalent is also a clause referencing Judgment Day (a full sentential proverb). Both are non-compositional idioms.

Another example: English *to kill two birds with one stone* (achieve two goals with one action) has an Azerbaijani counterpart *bir güllə ilə iki dovşan vurmaq* – “to hit two rabbits with one bullet.” The structure is identical (verb phrase idiom, with quantitative formula “kill X two Y with one Z”), but the specific animals and instruments differ (birds vs rabbits; stone vs bullet). This shows a cross-linguistic idiom mapping: same type of idiom (V+object idiom with instrumental phrase) but different lexical choices. A structured classification helps in aligning such idioms for translation or bilingual lexicography by separating the abstract form and the idiomatic meaning from the surface words.

Collocations often do not align between languages, as previously noted with the *make a decision* example across languages[26]. Azerbaijani might use a different verb or construction. Indeed, Azerbaijani for “make a decision” is *qərar vermək* (literally “to give a decision”) – analogous to Turkish and many languages using “give” or “take” instead of “make.” So in classification terms, both are Verb+Object collocations in their respective languages, but the collocate differs (*make* vs *vermək* “give”). Recognizing this as the same category (collocation) highlights that what differs is a lexical choice tied to each language’s norms, rather than an entirely different phraseological phenomenon.

Some phraseological categories are more language-specific. **Pragmatemes**, being tied to cultural context, often vary widely. For example, Azerbaijani has the expression *Gözün aydın!* (literally “bright be your eye!”) which is a pragmateme used to congratulate someone on receiving good news or recovering from illness (similar in usage to saying “Congratulations!” or “I’m happy for you” depending on context). English doesn’t have a literal equivalent (“May your eye be bright” has no meaning to an English speaker), so this is a pragmateme that doesn’t translate word-for-word. Instead, one would use a different expression appropriate to the context (e.g. *Good to hear!* or *That’s wonderful news!*). Conversely, English *Bless you!* (after a sneeze) might be rendered in Azerbaijani simply by *Çox yaşa!* (“Live long!” – literally “live much,” which is the local sneeze-response). Both fall under the category of pragmatic formula (a wish uttered after a sneeze), showing how languages fulfill the same social function with different fixed phrases.

By classifying phraseological units, we also can better document **what types of expressions each language has a lot of**. For example, English has many verbal idioms and similes, while some other languages might have more proverbial sayings or more honorific formulas. Azerbaijani (like Turkish) is rich in proverbial expressions and also has many idioms involving body parts (e.g. *başı sky idiom* meaning *someone is very busy* – literally “his head is in the sky” or similar). *If we see a structural gap – say, does Azerbaijani use binomials like English? It does, though perhaps differently* (e.g. *itim-it oğlu** can be an idiomatic insult meaning “scoundrel,” literally “dog and son of dog” – a binomial structure in an insult formula). Each language’s phraseology can be systematically studied using the taxonomy, and differences can be

categorized: some English collocations may be free combinations in Azerbaijani and vice versa, some English idioms might be expressed as single words in Azerbaijani, etc.

Such comparisons have practical relevance for **translation studies and language teaching**. Translators need to recognize phraseological units and find appropriate equivalents, which are often not literal. If an English idiom is a verb phrase idiom, the translator will look for an idiom of similar meaning in Azerbaijani, but the structure might change (maybe it's a proverb instead). Knowing the classification can help predict that “non-compositional idiom likely won't translate literally; search for an idiom or phrase with comparable function” whereas “collocation might translate word-by-word if the same concept exists, but check if a different collocate is needed.”

4.3 Relevance to Theoretical Modeling (Meaning–Text, Lexical Functions, etc.)

The classification presented resonates strongly with the Meaning–Text Theory approach to the lexicon. In MTT, **lexical functions** are used to systematically describe collocations and other multiword lexical relations (Mel'čuk et al., 1995). For example, the collocation *heavy rain* can be described by a lexical function like Magn(rain) = heavy (meaning the word *heavy* expresses a standard intensifier for *rain*). Similarly, idioms are treated as **lexical units** with a special representation. Our taxonomy's distinction between idioms, collocations, and pragmatemes aligns with the divisions used in MTT[15]. The structural aspect adds another layer that MTT does account for (in terms of syntactic representation in the Explanatory Combinatorial Dictionary, each idiom or collocation entry specifies its syntactic construction).

In computational linguistics, multiword expressions (MWEs) are recognized as a “pain in the neck” (to use the idiom) for parsing and generation. A taxonomy like ours can inform how to handle different types in natural language processing: e.g., idioms might be stored as units with an idiomatic meaning that overrides compositional analysis, collocations might be handled through statistical preferences in language models, and pragmatemes might be managed by templates or specialized modules for certain communicative contexts (like dialogue systems that use fixed polite formulas). Recent advances in AI language models actually benefit from understanding MWEs as units because it improves translation and idiomatic usage.

Another theoretical implication is for **syntax-semantics interface**: Phrases that are structurally regular but lexically fixed challenge traditional grammar, which assumes word-by-word generation. Construction Grammar, for instance, would be quite comfortable with idiomatic constructions (treating them as constructions with holistic meaning). Our structural categories could be seen as different construction types that can carry idiomatic meanings. A constructionist approach might catalog an “Idiomatic VP Construction” schema that covers many idioms, a “Binomial X and Y construction” that accounts for fixed binomials, etc. This ties in with work by Fillmore, Kay, and O'Connor (1988) on idiomatic constructions like *let alone* or *What's X doing Y?* (as in *What's this fly doing in my soup?* – a semi-productive idiom frame).

From the perspective of **semantics and cognition**, classifying by compositionality relates to how idioms are processed and stored. Psycholinguistic studies (e.g. Abel, 2003; Sprenger, Levelt & Kempen, 2006) suggest that highly familiar idioms are retrieved as wholes, while less familiar or more

transparent ones might be computed or at least semantically decomposed by speakers. Our category of partial idioms corresponds to idioms that tend to be more decomposable (Nunberg et al.'s analyzable idioms), which are often easier for people to infer and sometimes more flexible in use. Understanding these differences is key for anyone modeling mental lexicon or language acquisition: learners often struggle with pure idioms (having to just memorize them), whereas collocations require exposure and practice to notice which combinations sound “right.” The taxonomy therefore also has pedagogical implications: teaching materials can organize phraseological content from more transparent collocations to more opaque idioms.

4.4 Lexicographic and Analytical Utility

For dictionary makers and corpus linguists, the integrated taxonomy offers a guide for deciding **which expressions merit inclusion** in phraseological dictionaries or as fixed entries. A dictionary of idioms might focus on the non-compositional idioms primarily, whereas a collocations dictionary (like the Oxford Collocations Dictionary or Benson's BBI Dictionary) focuses on collocations. A general learner's dictionary might mark certain collocations or include common pragmatic phrases in usage notes. By recognizing pragmatemes, lexicographers can also ensure to list those essential formulas that learners might not generate by themselves (like the fact we say *Happy Birthday* but not *Congratulations on your birthday* as a direct wish).

In linguistic analysis, having terms for these categories allows precision. Instead of vaguely saying “this phrase is idiomatic,” one can say “this is a pragmatic formula” or “this is a collocational restriction.” For example, in analyzing a text, one might annotate all phraseological units and classify them: results could show a text is full of clichés and idioms (perhaps a very idiomatic informal text) or perhaps mostly collocations (academic text with many technical collocations but few vivid idioms). This can yield insights into style and register.

4.5 Limitations and Further Delimitations

While our taxonomy is broad, it still may not capture every nuance. Some expressions defy easy categorization. **Phrasal verbs** like *turn up*, *give in*, *get along* are a special case: are they idioms or just part of the verb's lexicon? Phrasal verbs can range from literal (*sit down*, *come in* – compositional) to highly idiomatic (*egg someone on* – meaning “urge/encourage,” where *egg on* has nothing to do with the egg). They form a continuum themselves. In our classification, a phrasal verb would just be a kind of verb phrase; we classify each case by meaning opacity. But some linguists treat phrasal verbs as a separate category of idiomaticity because the particle often contributes an aspectual or idiomatic sense that's not a full idiom but also not literal (e.g. *write up*, *clean up* where *up* adds a sense of completion). Those might be called **aspectual or aktionsart phrasemes** by some. We haven't broken those out separately, but an extended taxonomy might.

Also, our focus is synchronic (present-day usage). Historically, many idioms originated as transparent phrases (e.g. *spill the beans* might have once conjured spilling secrets metaphorically) but have since fossilized. Diachronic phraseology looks at how phrases change category over time (some collocations become idioms, idioms can sometimes become literal or lose idiomatic force – think of *awesome*, once

literally “awe-inspiring,” became a colloquial single word meaning “great,” not phrase but a metaphor turned literal in usage).

Additionally, **register and style** influence phraseology. Some idioms are informal/slang (*shoot the breeze*), others are formal (*to whom it may concern* formula). Classifying by semantic and structural type is one dimension; an orthogonal tagging for register would further enrich description.

5. Conclusion

Phraseological units are a vital and fascinating part of the English language, sitting at the intersection of lexicon, syntax, semantics, and pragmatics. In this article, we have presented a theoretical classification system that brings together structural and semantic perspectives on English phraseology. By distinguishing the structural patterns (from verb-object constructions and noun phrases to binomials and full sentences) and the degrees of semantic compositionality (from opaque idioms to transparent collocations and context-bound formulas), we offer a two-dimensional taxonomy that can categorize any given phraseological expression in English.

This structural–semantic taxonomy clarifies how different subtypes of phraseological units relate to each other. It highlights, for instance, that an idiom like *kick the bucket* and a collocation like *kick a habit* share a syntax (verb + object) but differ profoundly in semantic transparency, while *kick the bucket* and *bite the dust* share idiomatic meaning (both meaning “to die”) but differ in wording. The taxonomy thus aids in systematically comparing and contrasting phraseological units. It also underscores the continuum and gradience in phraseology: there are not always hard lines between idioms and collocations or between collocations and free combinations, but prototypical categories exist and are useful for description.

For linguistics scholars, especially in lexicology, semantics, and syntax, this classification provides a framework for analyzing phrases that behave like single units of meaning. It is directly relevant to tasks such as parsing idiomatic language, constructing dictionaries (where one must decide what multi-word entries to include), and understanding how idiomatic meaning is represented in the mental lexicon. The classification also has a cross-linguistic dimension – we briefly saw how Azerbaijani idioms and formulas compare to English ones, illustrating that while the structural–semantic types may be universal, each language populates them differently. This can inform contrastive linguistics and translation studies by providing a common reference for what kind of unit we are dealing with.

In theoretical modeling, our approach resonates with the Meaning–Text Theory’s categorization of phrasemes into idioms, collocations, and clichés, and expands on it by making syntactic form an explicit part of the classification. We also touched on how such taxonomies facilitate computational handling of multiword expressions and how they align with constructional approaches to grammar.

Finally, this work emphasizes that phraseological units are not peripheral oddities of language; rather, they are central and indispensable elements, contributing to the expressive richness of English. From everyday colloquial idioms to academic collocations and polite formulas, these units reflect both the creativity and the conventionality of language use. A structured understanding of them not only benefits theoretical analysis but also has practical applications in language education (helping learners

navigate idiomatic usage) and natural language processing (improving idiom recognition and translation).

In conclusion, the structural–semantic classification proposed here offers a comprehensive way to navigate the “jungle” of English phraseology. It allows us to map out the landscape of idioms and other fixed expressions with greater clarity – identifying what they mean, how they are built, and how they function – thereby contributing to a deeper understanding of the interplay between structure, meaning, and usage in the lexicon. With this foundation in place, future research can further refine each category (for instance, exploring subtypes of pragmatemes or degrees of idiom flexibility) and extend the comparative analysis to other languages, enriching the global study of phraseology.

References:

- Babayev, J. (2025). Algorithmic Autonomy or Dependence? A Mixed-Methods Study on AI Personalization and Self-Regulated Learning in Higher Education. *Journal of Azerbaijan Language and Education Studies*, 2(4), 32-39. <https://doi.org/10.69760/jales.2025004002>
- Babazade, Y. (2024). The Impact of Digital Tools on Vocabulary Development in Second Language Learning. *Journal of Azerbaijan Language and Education Studies*, 1(1), 35-41. <https://doi.org/10.69760/jales.2024.00103>
- Bally, C. (1909/1951). *Traité de stylistique française*. Berne: Francke. (Original work published 1909)
- Burger, H., Dobrovolskij, D., Kühn, P., & Norrick, N. (Eds.). (2007). *Phraseology: An international handbook of contemporary research* (Vol. 1–2). Berlin: Mouton de Gruyter.
- Coulmas, F. (Ed.). (1981). *Conversational Routine: Explorations in standardized communication situations and prepatterned speech*. The Hague: Mouton.
- Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge, UK: Cambridge University Press.
- Dobrovolskij, D., & Piirainen, E. (2005). *Figurative language: Cross-cultural and cross-linguistic perspectives*. Amsterdam: Elsevier.
- Fernando, C. (1996). *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis* (pp. 1–32). Oxford: Philological Society. (Quote: “You shall know a word by the company it keeps.”)
- Ismayilli, T. (2024). Navigating complexities in medical text translation: Challenges, strategies, and solutions. *Acta Globalis Humanitatis Et Linguarum*, 1(2), 170-176.
- Gläser, R. (1984). The translation aspect of phraseological units. *Papers and Studies in Contrastive Linguistics*, 18, 123–134.
- Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25(1), 38–61.
- Makkai, A. (1972). *Idiom structure in English*. The Hague: Mouton.

- Mel'čuk, I. (1995). Phrasemes in language and phraseology in linguistics. In M. Everaert, E.-J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 167–232). Hillsdale, NJ: Lawrence Erlbaum.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3(1), 31–56. [\[13\]](#)[\[15\]](#)
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, 70(3), 491–538.
- Sadiqzade, Z. (2025). Idiomatic Expressions and Their Impact on Lexical Competence. *Journal of Azerbaijan Language and Education Studies*, 2(1), 26–33. <https://doi.org/10.69760/jales.2025001002>
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *Substance and structure of language* (pp. 23–81). Berkeley, CA: University of California Press.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.