Journal of Critical Applied Linguistics Studies

ORIGINAL RESEARCH
E-ISSN: 2981-1678
Volume 3, Issue 1, pp 187-208

Gonbad Kavous University

Beyond Non-Compositionality: A Critical Framework for Idiom Recognition and Definition

Linguistics Society of Iran

Elham Farahani[1]

ABSTRACT

Idioms are traditionally defined as expressions whose meanings cannot be derived from their constituent words. This definition, however, is limited in that it focuses primarily on non-compositionality, overlooks fixed expressions characterized by structural oddity or partial semantic transparency, inadequately distinguishes idioms from other multi-word expressions, and fails to capture the gradient nature of idiomaticity. In response, this paper critically re-examines conventional definitions and proposes a more operational characterization of idioms as *institutionalized expressions of two words or more, but fewer than a sentence, whose individual words cannot be replaced with other similar words, and that are either non-compositional or odd in structure, or both*. Using a conceptual and critical review approach, the study develops a practical two-stage identification framework built around five criteria, namely multi-wordiness, institutionalization, restrictive flexibility, non-compositionality, and structural oddity. The framework provides a systematic procedure for distinguishing idioms from collocations, metaphors, and other multi-word expressions, thereby enhancing conceptual clarity in phraseological research and language pedagogy. Rather than treating idiomaticity as a binary property, the study conceptualizes it as a continuum shaped by multiple interacting criteria, reflecting the view that linguistic categories are not inherently fixed but are sensitive to context, use, and convention. This perspective offers researchers and language teachers a principled and practically applicable tool for identifying and analyzing idiomatic expressions.

Article History

Received:
2025-8-28
Revised:
2025-9-30
Accepted:
2025-10-30
Published:
2026-1-1

Keywords:
idioms,
restrictive flexibility,
non-compositionality,
structural oddity,
multi-wordiness,
institutionalization

[1] Assistant Professor of TEFL, Department of English Language and Literature, Faculty of Letters and Languages, Arak University, Arak, Iran; ORCID ID: https://orcid.org/0009-0001-9068-755X; *Email:* e-farahani@araku.ac.ir

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

1. Introduction

Idioms constitute an indispensable component of natural language. They are remarkably numerous. Brenner (2011) estimates that American English alone contains more than 10,000 idiomatic expressions, some of which have remained in use for over two millennia. Beyond their sheer quantity, idioms contribute to communicative efficiency, allowing speakers to convey complex or culturally loaded meanings concisely. As Liontas (2017) notes, idiomatic competence, the ability to understand and produce idiomatic expressions accurately and with minimal cognitive effort across varied contexts, represents a hallmark of advanced language proficiency.

In the context of language learning and teaching, idioms occupy a particularly significant role. Research has consistently demonstrated that exposure to and acquisition of idiomatic expressions correlates with gains in fluency (Cain et al., 2009; Lim et al., 2009; Maisa & Karunakaran, 2013; Teodorescu, 2015). This is partly attributable to the formulaic nature of idioms because they are stored and retrieved as relatively fixed chunks; they reduce the processing load on speakers during real-time production (Wray & Perkins, 2000). Discourse that incorporates idiomatic language also tends to be perceived as more native-like and authentic (Adel & Erman, 2012; Liontas, 2015). Liontas (2017) captures this developmental trajectory through the concept of idiomatization, the gradual process by which learners internalize the idiomatic stratum of a language through sustained exposure and practice.

Despite their centrality to language use and language learning, idioms remain among the most contested categories in linguistics. The standard definition, that an idiom is a multi-word expression whose meaning cannot be compositionally derived from the meanings of its parts, is both widely repeated and widely criticized. It accounts well for prototypical cases such as *kick the bucket* or *pull someone's leg*, but fails to accommodate the full range of expressions that speakers and lexicographers routinely treat as idiomatic, including expressions with partially transparent meanings, those with restricted syntactic behavior, and those whose idiomaticity resides primarily in their conventionalized form rather than in semantic opacity (Moon, 1998; Nunberg et al., 1994). The result is a definitional landscape in which the boundaries between idioms, collocations, proverbs, and other formulaic expressions remain unclear.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

This conceptual ambiguity carries concrete consequences. Definitions determine what is included in or excluded from analysis, and those boundaries in turn shape pedagogical practice, dictionary compilation, corpus annotation, language assessment, and the design of instructional materials. A narrowly compositional definition, for instance, systematically excludes large classes of formulaic language from idiom instruction, with direct implications for what learners are taught and tested on. Developing principled, transparent criteria for idiom identification is therefore not merely a theoretical exercise but an applied linguistic necessity.

The present study addresses this gap by proposing an operational definition of idioms and a criterion-based framework for distinguishing idioms from related multi-word expressions. The following research question guides the study:

RQ: How can idioms be operationally defined and systematically distinguished from other types of multi-word expressions?

2. Literature Review

2.1 Definitions of Idioms

The word *idiom* has a Greek root stemming from a lexeme meaning "own, private, and peculiar" (Simpson & Weiner, 1989, p. 624). In modern English, the term is used in two main senses: first, as a characteristic mode of expression in music, art, or theatre; second, as a group of words exhibiting phraseological peculiarities specific to a given language (Panou, 2014).

Any serious account of idioms must engage with Sinclair's (1991) foundational distinction between the open-choice principle and the idiom principle. Under the open-choice principle, speakers construct utterances by selecting individual words from the available lexical stock, subject to grammatical constraints. Under the idiom principle, speakers draw instead on a large inventory of semi-preconstructed phrases, giving "the language user a choice of using several semi-preconstructed phrases rather than stringing together individual words" (p. 109). Sinclair's account was influential in establishing that prefabricated multi-word units are not marginal exceptions but central to ordinary language use. However, the idiom principle as originally formulated does not provide criteria for distinguishing idioms from other types of formulaic sequences, nor does it account for the fact that idiomaticity varies in degree. As subsequent research has demonstrated, the boundary between the open-choice and idiom poles is

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

not fixed but is sensitive to familiarity, frequency, and contextual factors (Moon, 1998; Wray & Perkins, 2000). This limitation makes Sinclair's framework a useful starting point but an insufficient basis for idiom definition and identification.

In light of these complexities, providing a definition that adequately captures all aspects of idioms has been one of the most persistent challenges in phraseological research. As Langlotz (2006) observes, "idioms are peculiar linguistic constructions that have raised many eyebrows in linguistics and often confuse newcomers to a language" (p. 1). Consequently, definitions in the literature range from the very broad to the highly restrictive. Hockett (1958) treats any expression whose meaning cannot be derived from its structural composition as an idiom, regardless of size. At the other extreme, Grant and Bauer (2004), applying strict criteria of non-compositionality and non-figurativeness, arrive at a list of only 140 expressions they consider pure idioms in English.

This lack of consensus is compounded by cross-linguistic and disciplinary factors. Miller (2011) notes that phraseology developed as a formal discipline in Eastern Europe before Western researchers engaged with it, and many of its technical terms resist straightforward translation into English. The result is that even major reference dictionaries diverge; the expression *to be at play*, for instance, is listed as an idiom in Merriam-Webster (n.d.) but not in other standard dictionaries.

Regarding classification, Makkai (1972) proposed an influential division between encoding idioms and decoding idioms. Encoding idioms are expressions whose meanings remain relatively transparent; his example is the prepositional use of *at* in *the girl drove at 90 mph*. Decoding idioms are expressions whose meanings are opaque and unpredictable, such as *to be under the weather*. Although Makkai's terminology has not been widely adopted, his distinction anticipates the later understanding of idiomaticity as a continuum, with expressions ranging from fully transparent to entirely opaque rather than falling into a simple binary of idiomatic versus non-idiomatic.

Fernando (1996) proposed a three-way scale: pure idioms (e.g., *pull someone's leg*), semi-literal idioms (e.g., *fat chance*), and literal idioms (e.g., *according to*, *in sum*). Grant and Bauer (2004) similarly distinguished core idioms (non-compositional and non-figurative), figuratives (non-compositional but figurative), and ONCE expressions (containing one non-compositional element). Despite its greater precision, the Grant and Bauer framework remains

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

vulnerable to subjectivity, since determining whether an expression is figurative or non-figurative depends on the analyst's judgment and is not independently verifiable.

2.2 Criteria for Idiom Identification

Gramley and Patzold (2004) treat meaning as the primary criterion for idiom identification and offer several diagnostic tests. First, they propose a test of recurrent semantic contrast: if substituting a synonym for one word in an expression changes the idiomatic meaning, the expression is idiomatic. In *kick the bucket*, replacing *kick* with *hit* and *bucket* with *pail* eliminates the meaning of dying. Second, they observe that most idioms carry both a literal and a figurative reading. Third, they note that a subset of idioms depicts situations that are physically impossible or implausible, such as *to fly off the handle* and *it is raining cats and dogs*, and that this incongruence with the real world itself signals idiomaticity.

Flavell and Flavell (2011) focus on grammatical restrictions as a complementary criterion. Certain idioms violate standard grammatical rules; for instance, *to come a cropper* features an intransitive verb followed by a direct object. Additionally, idioms resist synonym substitution (*end* in *in the end* cannot be replaced with *finish* or *stop*) and block standard syntactic transformations. A *hot dog* cannot be paraphrased as *the dog is hot* or *the heat of the dog*, whereas the non-idiomatic phrase *hot sun* permits both.

More recent research has challenged the assumption that idioms are processed uniformly as non-compositional wholes. Studies of analyzability, decomposability, and idiomatic processing indicate that language users frequently access the relationship between the literal meanings of component words and the idiomatic meaning of the whole during comprehension (Fotovatnia & Goudarzi, 2014). These findings reinforce the view that idiomaticity is better understood as a continuum than as a binary distinction between compositional and non-compositional expressions, a position that informs the framework proposed in the present study.

2.3. Idioms and Other Multi-word Expressions

Distinguishing idioms from other multi-word expressions is essential for any operational definition, and the boundary with collocations is particularly contested. A collocation is broadly understood as a natural, recurrent co-occurrence of words, such as *advanced knowledge* (not *high knowledge*) or *considerable experience* (not *big* or *wide experience*). As Bauer (2018) observes,

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

meaning shifts according to collocational context, since *dry* conveys different meanings in *a dry state*, *dry toast*, and *a dry lecture*. Gramley and Patzold (2004) specify that collocations consist of two lexical items, each making an individual semantic contribution, belonging to different word classes, and occurring with a restricted but variable range.

The key distinction between collocations and idioms, as Baker (2018) notes, is that collocations are flexible in form and at least one component word retains its individual meaning. Idioms, by contrast, resist both formal variation and compositional interpretation. Fixed expressions fall between these poles, allowing little or no variation in form while their meanings remain relatively transparent.

Regarding metaphors, Leah (2014) observes that repeated use can diminish the conceptual force of a comparison until it passes below the threshold of conscious awareness. McCarthy and O'Dell (2008) argue that a metaphor used so frequently that its original comparative force is lost may be considered an idiom, though Flavell and Flavell (2011) caution that the boundary between the two is often indeterminate.

The distinction between idioms and proverbs merits particular attention, as proverbs constitute a category that is closely related to idioms in form but distinct in function and structure. Proverbs are "special, fixed, unchanged phrases which have special, fixed, unchanged meanings" (Ghazala, 1995, as cited in Leah, 2014, p. 2). More specifically, proverbs encode shared cultural wisdom and convey traditional values or experiential knowledge; they are "cultural models, giving information about how to behave or which values are upheld in a given culture" (Dobrovol'skij & Piirainen, 2005, p. 52). Proverbs are typically clause-length or sentence-length structures, often with a narrative or propositional form, and they may be shortened in actual use as a form of allusion. Crucially, while proverbs may resist paraphrase and vary little in form, properties they share with prototypical idioms, their meanings are generally recoverable from their components and from cultural context. This partial transparency, combined with their distinct discursive function as carriers of cultural or moral judgment, is what distinguishes them from idioms proper. As will be argued in the framework section below, proverbs represent a boundary category that an operational definition of idioms must explicitly address in order to avoid over-inclusion.

The extent of disagreement in the field is illustrated by the observation that "there is much debate in the literature over precisely where on the continuum a collocation ends, and an

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

idiom begins, and what exactly constitutes an idiom" (Glucksberg et al., 2001, p.76). This observation applies equally to the boundary between idioms and proverbs, and between idioms and conventionalized metaphors.

2.4. Limitations of Existing Frameworks

Although numerous definitions and classifications of idioms have been proposed, little consensus exists regarding the criteria that distinguish idioms from other multi-word expressions. This is not merely a terminological problem; divergent definitions lead researchers to identify different sets of expressions as idioms, which impedes cross-study comparisons and affects corpus annotation, dictionary compilation, and the design of pedagogical materials. Grant and Bauer (2004) themselves acknowledge that overly broad definitions risk including expressions with insufficient idiomaticity, while highly restrictive definitions exclude many conventionalized expressions that speakers process and acquire as formulaic units.

Grant and Bauer's (2004) framework is particularly vulnerable on the criterion of figurativeness, since determining whether a given expression is figurative depends on analyst judgment and lacks an independent verification procedure. Furthermore, by focusing on vivid, fully non-compositional structures, the framework is poorly suited to the analysis of idioms in academic or specialized genres, where idiomaticity is often more subtle. The encoding/decoding distinction proposed by Makkai (1972), while useful as a heuristic, similarly lacks the operationalization needed to function as a consistent identification procedure.

Taken together, these limitations indicate that idiom identification remains one of the most contested problems in phraseological research. The present study responds to this situation by proposing a more operational framework designed to increase consistency and reproducibility in idiom recognition across different research and pedagogical contexts.

3. Methodology

This study adopts a conceptual and critical review approach to the problem of idiom definition. Rather than collecting new empirical data, the research synthesizes existing theoretical accounts and empirical findings from the literature on idiomaticity to construct a principled definitional framework. This approach is appropriate given that the central problem is

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

one of conceptual clarity, namely the absence of a consistent, operationalizable set of criteria for distinguishing idioms from adjacent categories, rather than a shortage of linguistic data.

The methodology is grounded in conceptual analysis as practiced in theoretical linguistics and applied language studies. The goal is not to generate new corpus evidence but to critically evaluate and synthesize competing theoretical positions in order to produce a framework that is both theoretically defensible and practically applicable. The research question, what constitutes the necessary and sufficient criteria for identifying an expression as idiomatic, is therefore best addressed through structured engagement with the existing scholarly literature rather than through primary data collection.

4. Results and Discussion

Based on the review of literature and considering the shortcomings of the definitions and frameworks provided, the following framework for idiom recognition was developed. This framework includes two major stages as follows:

4.1. The First Stage of the Framework of Idiom Recognition

In this framework, initially, an expression must have three properties to be a candidate for further analyses. If an expression does not show all these three features together, it should be excluded from further analyses. The three features are: *multi-wordiness*, *institutionalization*, and *restrictive flexibility*.

4.1.1. Multi-wordiness

By being multi-worded, it is meant that an expression should be at least two words and less than a sentence in length. It should be mentioned that, in line with previous studies (e.g., Dobrovol'skij & Piirainen, 2005; Liontas, 2017; Miller, 2011; Miller, 2020), the size of idioms was set to be at least two words and less than a complete sentence, and ignoring the idiomatic compound words were ignored, which are included in some studies on idioms (e.g. Katz & Postal, 1964; Makkai, 1972). Moreover, idiomatic phrasal verbs are also excluded since, as Hinkel (2017) argues, "their numbers are so large that their exact or even proximate counts are unknown" (p. 49), in contrast to Liu (2003), who considers phrasal verbs in his study of idioms.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

4.1.2. Institutionalization

The second feature in this stage of the framework, institutionalization, refers to "the conventionalization of what was initially an ad hoc, novel expression" (Fernando, 1996). In other words, institutionalization means that users of language accept an expression as part of their language (Miller, 2011; Moon, 1998). Whether an expression is institutionalized or not can be partially assessed by checking dictionaries or checking the results of corpus-based studies in which the occurrences of words and expressions are dealt with (Moon, 2015).

In this framework, if an expression is included at least in one of the dictionaries in English, it is assumed to be institutionalized.

4.1.3. Restrictive Flexibility

As for restrictive flexibility, the third feature in this framework, a point should also be mentioned here. The point is that idioms should not be considered as totally fixed or frozen structures, as it has been one of the stereotypes attached to them for a long time before corpus-informed studies could well shed light on the variations they can undergo in the real use of English (Glucksberg et al., 2001; Langlotz, 2006; Lynn, 2016; Kyriacou et al., 2020; Moon, 1998). Accordingly, in this study, the use of the word fixedness is avoided since idioms are found to be used creatively by language users. A language user can say: *Who let the cat out of the bag?* and receives an answer like this one: *"It was let out by Old George, of course!"* (Glucksberg et al. 2001, p.69). Idioms are also different in the degree of variation they can tolerate. Idioms with identical forms and structures can be different in terms of modification. The idioms *spill the beans* and *kick the bucket* have the same structure, but the first can be passivized whereas the second cannot (Kyriacou et al., 2020), without any clear logical or structural reason.

That said, idioms have a very important and distinguishing feature in terms of inflexibility in the words which constitute them; that is to say, the words in an idiom cannot be replaced with other synonyms and similar words and still retain their idiomatic meanings.

In fact, by taking this feature into account, one can easily differentiate idiomatic collocations from idioms. That is to say if the words in an expression, although idiomatic at first sight, can be replaced with other synonyms and still retain their meanings, it is not an idiom, but a collocation. Consider the example to break free from something. Here this expression, although having the first two features, is not an idiom because break can be replaced with walk, run, roam

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

and other synonyms. Free has also its literal meaning since one of the meanings of this word is not being controlled.

The expression *to keep the sight of something* is also a collocational pattern because the word keep can be replaced with have, get, catch, etc. However, lose the sight of something is an idiom, not a collocation since lose cannot be replaced with other synonyms and still keep the idiomatic meaning of *to forget.*

Similarly, the expression *to throw into sharp relief* is not an idiom since the word *throw* can be replaced with many other words such as *bring*, *put*, and *stand out*. The word *relief* also has its literal meaning.

As another example, the expression *to serve the interests of somebody*, although labelled as an idiom in Merriam Webster (n.d.), is not an idiom because it lacks restrictive flexibility: *serve* can be replaced with *prompt*, *defend*, *advance*, or *protect*.

4.2. The Second Stage of the Framework of Idiom Recognition

If an expression has the three characteristics together mentioned above, it can go to the next stage to be checked for further analysis. At this stage, the expression should be either non-compositional or odd in structure to be labeled as an idiom. Some idioms can also be both non-compositional and odd in structure.

4.2.1. Non-compositionality

Non-compositionality refers to the fact that the individual words in an idiom do not function independently to produce the whole meaning of the expression. In fact, idioms have long held the position of being totally non-compositional. However, it seems, now, very simplistic and naive to define idioms based on this long-held assumption since it has been found that only a small number of them can be called totally non-compositional, or as Grant (2003) puts it, '*pure idioms'*.

The notion of non-compositionality has been challenged for the following reasons. First, it has been found that not all idioms are non-compositional given that most of them have a story or an origin (Boers et al., 2007). Flavell and Flavell's (2011) comprehensive dictionary of the origins of idioms can throw into relief the history and story behind most of the idioms in English.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

Second, the fact that many native speakers can decompose and metaphorically analyses the individual components of an idiom has been ignored (Kyriacou et al., 2020; Moon, 1998; Nunberg et al., 1994).

In the same vein, Kyriacou et al. (2020) state that idioms are similar to and behave like compositional expressions in many ways. Idioms like compositional expressions follow grammatical rules such as subject and tense agreement (He kicked the bucket) or undergo syntactic transformations such as passivation (The beans were spilt).

Considering all these points, it seems that non-compositionality in idioms is a matter of degree and should be considered as a continuum (Geeraert, 2016; Gluckberg et al., 2001; Langlotz, 2006; Moon, 1998; Taylor, 2012; Wulff, 2008; Zyzik, 2011). The researchers now agree that the distinction between non-compositional and compositional expressions is not clear-cut, and that it is not possible to draw a dividing line between them. As Wulff (2008) illustrates, the three expressions take the plunge, see a point, and write a letter have the positions of full non-compositionality, partial non-compositionality, and compositionality on the continuum of non-compositionality.

In all, only a small number of idioms are pure in the sense that they are totally non-compositional. In most of them, a connection between the literal meaning of words and the idiomatic meaning as a whole can be established.

Regarding the non-compositionality feature in this framework, another point of particular note is that in checking the non-compositionality of expressions, one must pay close attention to the different senses a word can have in English. Some expressions might seem to be idioms at first sight, but when different senses of each word are checked in the dictionaries, one might find out that the supposed idiomatic meaning is actually one of the different senses of the word.

For example, *to bring something into the open* is not an idiom, since one sense of the noun *open* is a public or unconcealed condition; therefore, the expression is compositional.

Along the same lines, the expression *to give a voice to somebody* is not an idiom, considering that one of the meanings of *voice* is the right to express one's opinions. Therefore,

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

this expression is not non-compositional at all. However, the expression *to give voice to somebody* is an idiom, although the difference might escape notice.

Similarly, *to get a grip on something* may appear to be an idiom; however, this is not the case because none of its components is used idiomatically, since one sense of the word *grip* is a mental grasp. However, the following expressions included in Table 1 have the three initial features in the first stage of this framework and are totally non-compositional idioms:

Table 1

Some Examples of Non-compositional Idioms

Expressions	Multi-wordiness	Institutionalization	Restricted flexibility	Non-compositionality	Structural oddity	Non-compositionality and structural oddity
To run very deep	✓	✓	✓	✓	–	–
Have a bee in one's bonnet	✓	✓	✓	✓	–	–
To dance to one's tune	✓	✓	✓	✓	–	–
To nail one's colors to the mast	✓	✓	✓	✓	–	–
Of course	✓	✓	✓	✓	–	–
To pick something off the shelf	✓	✓	✓	✓	–	–
Bring the house down	✓	✓	✓	✓	–	–
all the same	✓	✓	✓	✓	–	–
To cut no ice	✓	✓	✓	✓	–	–
in the scheme of things	✓	✓	✓	✓	–	–
As it were	✓	✓	✓	✓	–	–

4.2.2. Structural Oddity

The second possible option for an expression to be labeled as an idiom is to be odd in structure. Being odd in structure means that an expression deviates from the conventional grammatical rules of English. In other words, if the expression is not non-compositional but idiosyncratic in terms of its structure, it is considered an idiom. For example, *for example* is regarded as an idiom in English. Although the meaning in for example is pretty transparent, so a separate definition may not in fact be required, but the grammar is idiomatic because it treats 'example' as if it were an uncountable noun, which in general it is not. The totally unidiomatic expression would be 'for an example', but it is not used by English speakers. English speakers

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

would say '*as an example*', which is quite straightforward both semantically and grammatically, and therefore it is treated not as an idiom.

The following expressions are idioms, since first they have the three characteristics in the first stage of this framework and are odd in structure although the meaning is clear:

Table 2

Some Examples of Idioms with Structural Oddity

Expression	Multi-wordiness	Institutionalization	Restricted flexibility	Non-compositionality	Structural Oddity
To give someone to understand	✓	✓	✓	–	✓
In turn	✓	✓	✓	–	✓
In truth	✓	✓	✓	–	✓
Word for word	✓	✓	✓	–	✓
Do someone proud	✓	✓	✓	–	✓
In detail	✓	✓	✓	–	✓
In fact	✓	✓	✓	–	✓
Do the dirty on someone	✓	✓	✓	–	✓
Last but not least	✓	✓	✓	–	✓
In sum	✓	✓	✓	–	✓
At best	✓	✓	✓	–	✓
In essence	✓	✓	✓	–	✓
To date	✓	✓	✓	–	✓

4.2.3. Non-compositionality and Structural Oddity

Some expressions can also enjoy both the feature of *non-compositionality* and *structural oddity*. Some examples of the idioms belonging to this group are: *by and large*, *on end*, and *at large*.

Table 3

Idiom Identification Test

Expression	Test
To place the blame on somebody	**1.** Is this expression at least two words and less than a sentence? **YES** **2.** Is this expression institutionalized? **YES** **3.** Does it have restrictive flexibility? **NO**, the word *place* can be replaced with other synonyms such as *put, lay, pin, and attach* Therefore, **NOT** an idiom. It is a **collocation.**
To put another little brick in the wall	**1**. Is this expression at least two words and less than a sentence? **YES** **2**. Is this expression institutionalized? **No** Therefore, **NOT** an idiom. It is a **metaphor.**
To work their way up to the academic ladder	**1.** Is this expression at least two words and less than a sentence? **YES** **2.** Is this expression institutionalized? **YES** **3.** Does it have restrictive flexibility? **YES** **4.** Is it non-compositional? **NO**

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

	One of the meanings of ladder in English is: "A series of levels which someone moves up and down within an organization" **5.** Is it odd in structure? **No.** Therefore, **NOT** an idiom.
To hold water	**1**. Is this expression at least two words and less than a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3.** Does it have restrictive flexibility? **YES** **4**. Is it non-compositional? **YES** Therefore, **an idiom**.
Then again	**1.** Is this expression at least two words and less than a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4.** Is it non-compositional? **YES** Therefore, an **idiom**.
To give weight to something	1. Is this expression at least two words and less than a sentence? **YES** 2. Is this expression institutionalized? **YES** 3. Does it have restrictive flexibility? **NO**, the word *give* can be replaced with other synonyms such as *put, lend, carry, etc.* Therefore, **NOT an idiom**.
To open a window for something	**1**. Is this expression at least two words and less than a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4**. Is it non-compositional? **NO** *One of the meanings of window in English is "a short period of time that is available for particular activity" and one of the meanings of open "to have a chance to do something"* Therefore, **NOT an idiom**.
To open the door to something	**1**. Is this expression at least two words and less that a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4**. Is it non-compositional? **YES** Therefore, **an idiom**.
At any rate	**1**. Is this expression at least two words and less than a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4**. Is it non-compositional? **YES** Therefore, **an idiom**.
For instance	1. Is this expression at least two words and less than a sentence? **YES** 2. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4.** Is it non-compositional? **No** **5**. Is it odd in structure? **YES** Therefore, **an idiom**.
In the final analysis	**1**. Is this expression at least two words or more but less than a sentence? **YES** **2**. Is this expression institutionalized? **YES** **3**. Does it have restrictive flexibility? **YES** **4**. Is it non-compositional? **YES** Therefore, **an idiom**.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

Table 3 presents a test for the recognition of idioms. By applying this framework, one can systematically distinguish idioms from other multi-word expressions such as collocations, metaphors, and proverbs.

The proposed framework demonstrates that idiom identification cannot be reduced to the traditional criterion of non-compositionality alone. The analysis presented in this study indicates that institutionalization, restrictive flexibility, and structural oddity all play important roles in distinguishing idioms from other multi-word expressions. Consequently, expressions that have traditionally been excluded from idiom studies on account of their relatively transparent meanings may still qualify as idioms due to their structural peculiarities or restricted lexical variation.

The framework also offers a practical solution to the longstanding difficulty of differentiating idioms from collocations. While collocations may exhibit conventionalized word combinations, they generally allow greater lexical substitution than idioms. The criterion of restrictive flexibility therefore provides an additional diagnostic tool that complements semantic analysis.

From an applied linguistic perspective, the framework may facilitate more consistent identification of idioms in corpus studies, pedagogical materials, and lexicographic work. It may also assist language learners who often struggle to distinguish idioms from collocations, metaphors, and proverbs.

These findings also support broader calls within critical applied linguistics for reconsidering how linguistic categories are constructed and maintained. Rather than treating idioms as naturally bounded entities, the present framework suggests that idiomaticity is best viewed as a gradient phenomenon involving multiple interacting criteria. This approach encourages researchers and teachers to critically evaluate established classifications rather than accepting traditional definitions uncritically.

At the same time, it should be acknowledged that some cases remain borderline. Judgments regarding institutionalization and structural oddity may vary across dictionaries,

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

varieties of English, and speech communities. The framework should therefore be understood as a systematic heuristic rather than an absolute classification system.

5. Conclusion

This study developed a two-stage framework for the identification of idioms. In the first stage, a candidate expression must simultaneously exhibit three properties: *multi-wordiness* (comprising two or more words but falling short of a complete sentence), *institutionalization* (i.e., inclusion in at least one recognized English dictionary), and *restrictive flexibility* (i.e., the individual words of the expression cannot be substituted with synonyms while preserving the idiomatic meaning). Expressions that satisfy all three criteria proceed to the second stage, where they are evaluated for non-compositionality (i.e., the overall meaning is not derivable from the meanings of the component words) and/or structural oddity (i.e., the expression deviates from regular syntactic rules). If an expression exhibits one or both second-stage properties, it is classified as an idiom. Based on this framework, the following definition is proposed: *idioms are institutionalized expressions of two words or more, but fewer than a sentence, whose individual words cannot be replaced with other similar words, and are either non-compositional or odd in structure, or both.*

Beyond its theoretical contribution, the framework has practical implications for language education, lexicography, corpus annotation, and phraseological research. By offering explicit, multi-criteria guidance for identifying idioms, it promotes greater consistency in the treatment of multi-word expressions across domains of applied linguistics. In EFL and EAP contexts in particular, the framework equips teachers with a principled basis for helping learners distinguish idioms from collocations and proverbs, a distinction that has proven persistently difficult in classroom practice (Liontas, 2017; Miller, 2020; Rafatbakhsh & Ahmadi, 2019, 2020). Teachers are also encouraged to develop exercises drawing on the authentic examples analyzed in this study, as these illustrate the full range of multi-word expression types that learners encounter in academic English.

However, several limitations should be acknowledged. The framework has been developed conceptually and has not yet undergone large-scale empirical validation; in particular,

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

the claim that independent analysts will reach consistent classifications using the proposed criteria has not been tested through any inter-rater reliability procedure, and formal validation using a measure such as Cohen's κ\kappaκ remains a priority for future research. Judgments regarding institutionalization may also vary depending on the dictionaries or corpora consulted, since dictionaries differ considerably in their coverage of idiomatic expressions, meaning that an expression recorded in one source may be absent from another. Moreover, certain expressions may additionally occupy intermediate positions on the compositionality continuum, and boundary cases at the intersection of idioms and adjacent categories such as proverbs and clichés remain partially unresolved by the framework, despite the principled argumentation offered in the preceding sections. Finally, the framework was developed based on literature drawn predominantly from English, and its applicability to typologically different languages has not been examined.

Future research may test the framework through corpus-based analyses, inter-rater reliability studies, and cross-linguistic comparisons, which would clarify the extent to which different researchers and language users apply the proposed criteria consistently. Further work might also explore how the framework can be integrated into phraseological pedagogy and language assessment.

Declaration

We confirm that this work is original and is not under consideration by any other journal.

Transparency Statements

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgements

The authors would like to thank all those who contributed to the completion of this project.

Declaration of Interests

The authors declare that they have no conflict of interest.

Funding

No funding was received for this study.

Ethical Consideration

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

This manuscript adheres to the ethical guidelines provided by the Committee on Publication Ethics (COPE) for ensuring integrity and transparency in the research publication process.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

References

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Baker, M. (2018). *In other words: A coursebook on translation*. Routledge.

Bauer, L. (2018). Compounds and multi-word expressions in English. In B. Schlucker (Ed.) *Complex lexical units* (pp. 45-68). De Gruyter.

Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemoics? *Language Teaching Research*, *11*(1), 43-62. https://doi.org/10.1177%2F1362168806072460

Brenner, G. (2011). *Webster's new world American idioms handbook*. Mifflin Harcourt.

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, *102*(3), 280-298. https://doi.org/10.1016/j.jecp.2008.08.001

Dobrovol'skij, D. & Piirainen, E. (2005). *Figurative language: Cross-cultural and cross-linguistic perspectives*. Elsevier Ltd.

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

Flavell, L., & Flavell, R. (2011). *Dictionary of idioms and their origins*. Kyle Cathie Ltd.

Fotovatnia, Z., & Goudarzi, M. (2014). Idiom comprehension in English as a foreign language: Analysability in focus. *Procedia-Social and Behavioral Sciences*, *98*, 499-503.

Geeraert, K. D. (2016). *Climbing on the bandwagon of idiomatic variation: A multi-methodological approach*. [Doctoral dissertation, University of Alberta]. University of Alberta Libraries. https://doi.org/10.7939/R3222RB4F

Ghazala, H. (1995). *Translation as problems and solutions: A coursebook for university students and trainee translators*. Dar wa Maktabat Al-Hilal.

Glucksberg, S., McGlone, M. S., Grodzinsky, Y., & Amunts, K. (2001). *Understanding figurative language: From metaphor to idioms* (No. 36). Oxford University Press.

Gramley, S., & Pátzold, M. (2004). *A survey of modern English*. Routledge.

Grant, L (2003). *A corpus-based investigation of idiomatic multiword units*. [Doctoral dissertation, Victoria University of Wellington]. Victoria University of Wellington Research Archives. http://hdl.handle.net/10063/327

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, *25* (1), 38-61. https://doi.org/10.1093/applin/25.1.38

Hinkel, E. (2017). Teaching idiomatic expressions and phrases: Insights and techniques. *Iranian Journal of Language Teaching Research*, *5*(3), 45-59. http://dx.doi.org/10.30466/ijltr.2017.20304

Hockett. C. F. (1958). *A Course in modern linguistics*. Macmillan.

Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. M.I.T. Press.

Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, *63*(2), 404-435. https://doi.org/10.1177%2F0023830919847691

Langlotz, A. (2006). *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. John Benjamins Publishing Company.

Leah, C. (2014). *Idioms-grammaticality and figurativeness*. http://www.theroundtable.ro/Current/Language/Claudia_Leah_Idioms_Grammaticality_and_Figurativeness.pdf

Lim, E. A. C., Ang, S. H., Lee, Y. H., & Leong, S. M. (2009). Processing idioms in advertising discourse: Effects of familiarity, literality, and compositionality on consumer ad response. *Journal of Pragmatics*, *41*(9), 1778-1793. https://doi.org/10.1016/j.pragma.2008.09.034

Liontas, J. I. (2015). Developing idiomatic competence in the ESOL classroom: A pragmatic account. *TESOL Journal*, *6*(4), 621-658. https://doi.org/10.1002/tesj.230

Liontas, J. I. (2017). Why teach idioms? A Challenge to the profession. *Iranian Journal of Language Teaching Research*, *5*(3), 5-25. https://dx.doi.org/10.30466/ijltr.2017.20302

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly, 38*(4), 671-700. https://doi.org/10.2307/3588217

Lynn, E. M. (2016). *Getting all the ducks in a row: Towards a method for the consolidation of English idioms.* [Master's thesis, Brigham Young University]. BYU Scholars Archive. https://scholarsarchive.byu.edu/etd/6014/

Maisa, S., & Karunakaran, T. (2013). Idioms and importance of teaching idioms to ESL students: A study on teacher beliefs. *Asian Journal of Humanities and Social Sciences* (AJHSS), *1*(1), 110-122.

Makkai, A. (1972). *Idiom structure in English*. De Gruyter Mouton.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

McCarthy, M. J. & O'Dell, F. (2008). *Academic vocabulary in use*. Cambridge University Press.

Merriam-Webster. (n.d.). *Merriam-Webster.com dictionary*. Retrieved June 16, 2025, from https://www.merriam-webster.com/

Miller, J. (2011). *Never judge a wolf by its cover*: *An investigation into the relevance of phrasemes included in advanced learners' dictionaries for learners of English as an additional language in Australia.* [Unpublished doctoral dissertation]. Flinders University

Miller, J. (2020). The bottom line: Are idioms used in English academic speech and writing? *Journal of English for Academic Purposes*, *43*, 1-14. https://doi.org/10.1016/j.jeap.2019.100810

Moon, R. (1998). *Fixed expressions and idioms in English.* Clarendon Press.

Moon, R. (2015). Multi-word items. In J.R. Taylor (Ed.). *The Oxford handbook of the word*. (pp.120-140). Oxford University Press.

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70* (3), 491-538. https://doi.org/10.1353/lan.1994.0007

Panou, D. (2014). *Idiom translation in the financial press: a corpus-based study*. Cambridge Scholars Publishing.

Rafatbakhsh, E., & Ahmadi, A. (2019). A thematic corpus-based study of idioms in the Corpus of Contemporary American English. *Asian-Pacific Journal of Second and Foreign Language Education, 4*(1), ٤-11. https://doi.org/10.1186/s40862-019-0076-4

Rafatbakhsh, E., & Ahmadi, A. (2020). The most frequent idioms used in contemporary American English: A corpus-based Study. *Applied Research on English Language*, *9*(2), 205-228. https://dx.doi.org/10.22108/are.2019.114449.1389

Simpson, A.J., & Weiner, E. S. C. (1989). *The Oxford English dictionary*. Oxford University Press.

Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford University Press.

Taylor, J. R. (2012). *Mental corpus: How language is represented in the mind*. Oxford University Press.

Teodorescu, A. (2015). Mobile learning and its impact on business English learning. *Procedia-Social and Behavioral Sciences*, *180*, 1535–1540.

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language and Communication, 20,* 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. A&C Black.

Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and Pedagogical sequencing. *Language Teaching Research*, *15*(4), 413-433. https://doi.org/10.1177%2F1362168811412025

[Downloaded from cals.gonbad.ac.ir on 2026-10-16]