



A Corpus-Based Investigation of Idioms in Applied Linguistics Research Articles

Elham Farahani^{1,*} and Hooshang Yazdani²

¹ *Corresponding Author: Assistant Professor of TEFL, Department of English Language and Literature, Faculty of Letters and Languages, Arak University, Arak, Iran; ORCID: 0000-0001-9068-755X
Email: e-farahani@araku.ac.ir*

² *Associate Professor of TEFL, Department of English Language and Literature, Faculty of Letters and Languages, Arak University, Arak, Iran
ORCID: 0000-0003-3628-1046
Email: h-yazdani@araku.ac.ir*

Abstract

Idioms are typically considered as elements of informal, vivid, and entertaining language, and are generally thought to be unsuitable for formal contexts. However, recent studies show that idioms are increasingly used in formal communication and academic writing. In this context, this corpus-based study aims to offer new insights into English idioms by examining them within a corpus consisting of 5,675,554 tokens from 640 randomly selected applied linguistics research articles (RAs). To achieve this, the researchers developed a working definition and framework for identifying academic idioms within written genres like RAs. Using this framework, they compiled a list of the most common academic idioms found in the genre. Additionally, the study explored whether the frequency of these idioms changed over the period from 1980 to 2020. The findings revealed that idioms are indeed present in applied linguistics RAs and merit appropriate attention. No significant correlation was found between time and idiom usage frequency, while the writing style of authors appeared to significantly influence how often idioms are used. By identifying and analyzing the most frequent academic idioms in applied linguistics RAs, this study's results may be valuable for EFL writing instructors, graduate and postgraduate students in applied linguistics, and developers of teaching materials.

Keywords: idioms, research articles, corpus, applied linguistics

ARTICLE INFO

Research Article

Received: Sunday, November 2, 2025

Accepted: Sunday, March 15, 2026

Published: Wednesday, April 1, 2026

Available Online: Sunday, March 15, 2026

DOI: <https://doi.org/10.22049/jalda.2026.31176.1916>

Online ISSN: 2821-0204; Print ISSN: 28208986



© The Author(s)

Introduction

RAs “the written form of the academy” (van Enk & Power, 2017, p.1), are the subject of numerous English for Research Publication Purposes (ERPP) and English for Academic Purposes (EAP) studies. In fact, RAs have become “the pre-eminent genre of the academy” (Hyland, 2009, p. 67). This is due to the significance of RAs in dissemination of knowledge in academic communities. In this regard, Hyland & Paltridge (2011) state: “publishing is the main means by which academics establish their claims for competence and climb the professional ladder” (p. 173). The saying: “publish or perish” accurately describes the vital role of RAs publication in the academic lives of scholars and graduate students.

English as a Foreign Language (EFL) graduate students and teachers are expected to publish RAs in English to be accepted in their academic communities. This requirement seems to be less challenging for applied linguistics graduate students since they have passed some English writing courses. However, these courses focus on general writing skills, not to RAs. Therefore, their academic writing skills usually lag behind their level of conversational skills. This is unsurprising since even native speakers of any language require formal training to write academic RAs.

As a result, EAP researchers aim to investigate the features of academic writing and the variables contributing to successful composition of RAs acceptable to members of their scientific communities. Using corpus-based methodology, this study investigates a neglected aspect: *academic idioms*.

The investigation of idioms in RAs may seem odd since they are often assumed to be more prevalent in informal communications, only oiling the wheels of friendly and conversational language (Miller, 2020; Simpson & Mendis, 2003). O’Keeffe et al. (2007, p. 90) state that an academic “context is not immediately associated with the occurrence of idiomatic expressions in most people’s minds”. They indicate that “idioms have functions which create and reinforce interpersonal relations, projecting informality, camaraderie and social bonding” (p. 88).

An examination of ten EAP books revealed that most of them warn learners about the use of idioms in their academic writing. For example, Sowton (2012) states “since one of the main goals of academic writing is to avoid ambiguity, idioms should be avoided.” (p. 89). Likewise, Bailey (2015) in his EAP book states that “there is no correct style of academic writing, but in general it should be accurate, impersonal and objective. For example, personal pronouns like I and idioms (i.e., informal language) are used less often than in other kinds of writing” (p. 138).

What is more, in coursebooks on idioms too, academic idioms are ignored implying learners should avoid using idioms in their academic discourse. As an example, in McCarthy and O’Dell’s (2010) self-study book on idioms, only one lesson is devoted to the formal idioms that can safely be used in writing.

However, it has been found that idioms which may strike an informal note in dry and academic contexts are frequent in spoken academic discourses (Miller,

2020; Simpson & Mendis, 2003). For instance, O’Keeffe et al. (2007) give examples of the use of idioms in academic situations in which people feel the necessity to make use of idioms to project a more friendly relationship to express meanings non-threateningly.

Related to this discussion is Biber et al.’s (1999) massive reference book which is a corpus-informed investigation on written and spoken English in four registers (conversation, fiction, news, and academic prose) from American and British English. Interestingly, one part of this influential book is devoted to idioms. In this part, they point out that a large number of idioms in English are verb-based (i.e., *bear in mind*, *take into account*, *stand a chance*), and that idioms are generally rare in conversation, and more common in fiction, a finding surprising those believing idioms belong to daily talk.

Biber et al. (1999) confirm that idioms are mostly colloquial. Crucially, they report that although idioms are generally colloquial and more frequent in other registers, “a few of these expressions are used ‘primarily’ in academic prose or other kinds of expository writing” (p.1026). They add: “although earlier studies have claimed that formulaic language is dominant in conversation, idiomatic phrases with the verbs *have*, *make*, and *take* are by far more common in the written registers. In fact, several of these phrases are notably common only in news reportage and/or academic prose.” (p. 1028). Idioms like *have an/no/the effect*, *make use of*, and *take the form of* are reported to be used only in academic discourse.

Researchers (McCarthy, 1998; Moon, 1998; Simpson & Mendis, 2003) conclude that the use of idioms are not merely alternatives to their more formal and literal counterparts. In fact, idioms are found to perform discourse roles and important pragmatic functions in academic contexts.

Therefore, it stands as a reason to conclude that, although idioms in academic English lack the frequency of informal English, they have important functions in this genre. As Miller (2020) states, low frequency does not equal low importance since students may miss or misunderstand text due to unfamiliar idioms. For example, ignorance of the idiomatic meanings of *that said*, *much as*, or *nothing if not* as academic idioms can lead to total misunderstanding of the sentences. Another example is *in view of something* that literally has nothing to do with the meaning of the whole: *because of something or as a result of something*.

In this study, idioms are not restricted to the vivid and colorful examples. Many phrases qualify as idioms because their form is fixed, and/or they show distinctive grammar. For example, *for example* is regarded as an idiom in this study. Though semantically transparent, 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 would say ‘*as an example*’, which is quite straightforward both semantically and grammatically, and therefore it is treated not as an idiom.

In sum, it is safe to say that idiomaticity is not a binary quality – it exists on a cline. Some expressions are more idiomatic than others and deciding whether a

particular expression is or is not an idiom may depend on where exactly you draw the line; this in turn may depend on what a researcher's purpose is.

Taken all these points into account, one of the primary purposes this study aims to accomplish is to challenge one of the long-held assumptions and stereotypes attested to idioms that they are not qualified enough to be used in academic writing. Moreover, attempts are made to present an operational definition and a framework for the detection of academic idioms in this corpus-informed study to develop a list of genre and discipline-specific academic idioms which are compiled from applied linguistics RAs published in eight professional journals in the field from 1980 to 2020. This list is the first attempt in the literature and is hoped to be insightful for EAP and EFL writing courses. Finally, any possible changes in idiom frequency used in applied linguistics RAs over the time span of 1980-2020 are investigated. This study aims to answer the following research questions:

1. How can academic idioms be defined to distinguish them from other multiword expressions?
2. What is the frequency of idioms in applied linguistics RAs?
3. Are there differences in idiom frequency in applied linguistics RAs from 1980 to 2020?

In order to address the third research question, the following null hypothesis is constructed to be tested out:

H01: There are no differences in idiom use in applied linguistics RAs from 1980 to 2020.

Literature Review

Definitions of Idioms

The term "idiom" derives from Greek, meaning "own, private, and peculiar" (Simpson & Weiner, 1989, p. 624), referring to phrases with unique meanings or structures (Panou, 2014). Idiomaticity describes expressions that are grammatical and native-like or non-compositional, such as *a dog's breakfast* (meaning a mess) (Richards & Schmitt, 2010). Sinclair's (1991) idiom principle posits that language users often select pre-fabricated phrases for efficiency.

Many researchers have long tried to provide a good definition for idioms. At one extreme, Hockett (1958) includes single morphemes to entire texts, while Grant and Bauer (2004) limit idioms to 140 "pure" cases. Weinreich (1969) and Fraser (1970) emphasize non-compositional meanings, and Makkai (1972) distinguishes transparent (encoding) and opaque (decoding) idioms. Nunberg (1978) categorizes idioms as decomposable (e.g., *pop the question*) or non-decomposable (e.g., *by and large*). Grant and Bauer (2004) classify idioms into core (non-compositional, non-figurative), figurative, and ONCE idioms. Gramley and Patzold (2004) highlight semantic contrast, dual meanings, and syntactic irregularities as defining traits.

Significance of Idioms

Idioms are integral to natural language, adding color and variety (Liontas, 2017; Cooper, 1998). With over 10,000 idioms in American English (Brenner, 2011), they enhance communication efficiency by replacing lengthy explanations (e.g., *hot potato*) (Khonbi & Sadeghi, 2017). Idioms contribute to fluency by reducing cognitive load through fixed phrases (Pawley & Syder, 1983, as cited in Hinkel, 2017). Idiomatic competence, developed through exposure, marks fluency and cultural integration (Liontas, 2017). However, misinterpreting idioms can lead to comprehension issues (Boers et al., 2007).

Properties of Idioms

Idioms are characterized by multi-wordiness, non-compositionality, fixedness, and institutionalization. They typically involve multiple words (Dobrovolskij & Piirainen, 2005). Non-compositionality, where meaning isn't derived from individual words, is debated, as many idioms have metaphorical or historical origins (Boers et al., 2007). Fixedness is also contested, with corpus studies showing syntactic flexibility (Moon, 1998; Glucksberg et al., 2001). Institutionalization reflects idioms' acceptance as conventional expressions (Fernando, 1996).

Idioms vs. Other Multi-Word Expressions

Idioms differ from other multi-word expressions in their fixed, non-compositional nature, which makes their meanings less predictable. For instance, collocations, such as *strong tea*, are transparent and flexible in meaning, allowing for easier interpretation compared to idioms like *call into question*, which rely on fixed phrasing (Baker, 2018; Gramley & Patzold, 2004). Metaphors, like *guiding light*, implicitly compare distinct concepts, while similes use explicit markers such as "like" or "as" for comparison, yet both contrast with idioms' non-literal rigidity (Leah, 2014; McCarthy & O'Dell, 2008). Proverbs, such as *the nail that sticks up*, convey cultural wisdom with relatively transparent meanings, unlike the non-narrative nature of idioms (Dobrovolskij & Piirainen, 2005). Clichés, like *have a good day*, are fixed yet transparent, setting them apart from the opacity of idioms (Gramley & Patzold, 2004).

Corpus-Based Studies on Idioms

Corpus studies illuminate idiom usage. Moon (1998) found 40% of idioms in an 18-million-word corpus showed variation, though her corpus was journalism-heavy. Simpson and Mendis (2003) identified idioms in academic speech, noting functions like evaluation and emphasis. Liu (2003) reported rare pure idioms in spoken English. Grant and Nation (2006) listed 100 core idioms, emphasizing their scarcity. Miller (2020) confirmed idioms' presence in academic writing, identifying 43 frequent idioms in the Oxford Corpus of Academic English, suggesting their relevance for RAs. Alizadeh et al. (2024) compared the distinctive use of lexical bundles in soft and hard science publications. However, no study specifically addresses idioms in applied linguistics RAs, a gap this research aims to fill.

The literature reveals diverse idiom definitions, their significance in fluency and cultural integration, and their debated properties. Corpus studies highlight idioms' variability and presence in academic discourse, but their role in applied linguistics RAs remains underexplored. This study seeks to address this gap, providing insights into idioms' academic utility.

Method

Corpus Design

To investigate idiomatic expressions in RAs within applied linguistics, a specialized corpus was developed due to the absence of an existing online corpus tailored for this purpose. The corpus comprised 640 RAs randomly selected from eight high-ranking applied linguistics journals, identified using Scimagojr.com based on impact factor. Journals were included if published from 1980 to 2020, ensuring a 40-year span. The selected journals were *Applied Linguistics*, *ELT Journal*, *Journal of Communication*, *Journal of Research Literacy*, *Language Learning*, *TESOL Quarterly*, *The Modern Language Journal*, and *The Journal of Child Language*. For each journal, two RAs per year were randomly downloaded, yielding 80 articles per journal and a total of 5,675,554 tokens.

Data Analysis Tool

The corpus was analyzed using #LancsBox (Brezina et al., 2018, 2020), a sophisticated corpus analysis tool developed by Lancaster University. #LancsBox was chosen for its advanced filtering capabilities, ability to process various file formats (e.g., .txt, .pdf, .docx), and efficient statistical analysis of large datasets. The KWIC and Whelk tools provided concordance lines, frequency, relative frequency, and distribution data across the corpus, facilitating idiom identification.

Data Analysis Procedure

Given the complexity of idioms, automatic corpus searches were impractical. Instead, a two-stage approach was adopted, inspired by Miller (2020) and Simpson and Mendis (2003).

Stage One: Manual Analysis

In the first stage, 240 RAs (30 per journal) were manually analyzed to develop an operational definition and framework for identifying academic idioms. These articles, covering 1980–2020, were read to extract multi-word expressions, which were then evaluated against idiom dictionaries and criteria from the literature. Three phraseology experts provided feedback to refine the framework. The extracted idioms (6,834) were categorized by decade (1980–1989, 1990–1999, 2000–2009, 2010–2020) and sorted by frequency using a coding system, resulting in a list of 567 idioms.

Stage Two: Corpus-Wide Search

In the second stage, the 567 idioms were searched across the full 640-article corpus using #LancsBox. Advanced filtering accounted for idiomatic variations (e.g., *along/on (the) lines*), ensuring all forms were captured. For instance, filtering for *to a/an extent* reduced 1,049 concordance lines to 101. Idioms with dual meanings (e.g., *on the ground*) were manually checked to distinguish idiomatic from literal uses. Frequency, relative frequency, and range data were recorded for each idiom.

Results

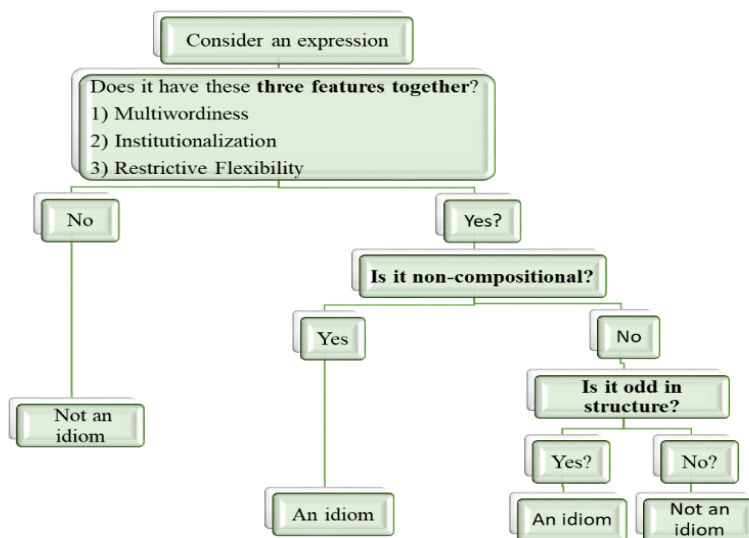
This study investigated idiomatic expressions in applied linguistics RAs using a corpus-based approach, focusing on their formality and usage patterns. A corpus of 640 RAs (5,675,554 tokens) was analyzed to address three research questions: defining academic idioms, determining their frequency, and examining diachronic trends from 1980 to 2020.

Research Question 1: Defining Academic Idioms

To differentiate academic idioms from other multi-word expressions, 240 RAs were manually analyzed, and over 700 expressions were evaluated using dictionaries and expert input. A framework for idiom identification was developed, as shown in Figure 1.

Figure 1

Framework for Idiom Definition and Recognition in Written Academic Corpora



Based on this framework, an expression must meet three initial criteria to be considered an idiom: multi-wordiness (two or more words, less than a sentence), institutionalization (listed in at least one English dictionary), and restrictive flexibility (words cannot be replaced with synonyms without altering meaning). Expressions passing these criteria are further evaluated for non-compositionality (meaning not deducible from components) or odd structure (deviation from grammatical norms). Examples include:

- **Non-compositional idioms:** *hold water, then again, open the door to* (e.g., “The tendency of linguistics applied will be to dance attendance to whatever tune is currently in theoretical fashion,” Edge, 1989, p. 48).

- **Odd in structure idioms:** *for example, in turn, by and large* (e.g., “By and large, although with some differences, they accepted our interpretation,” Alderson & Clapham, 1992, p. 154).
- **Non-idiomatic examples:** *place the blame* (replaceable words), *build on sand* (not institutionalized).

Considering the different nature of idioms in written academic genres such as RAs, in contrast to other more informal genres, this framework can be used for idioms and other similar multiword expressions to be differentiated. Table 1 presents the summary of all the points referred to above in practice by way of examples from the corpus of RAs used in this study.

Table 1

Idiom Identification Test in Academic Genres

Expression	Multi-word	Institutionalized	Restrictive Flexibility	Non-compositional	Odd Structure	Idiom
Place the blame on sb	Yes	Yes	No (replaceable words)	-		No
Put another brick in the wall	Yes	No	-	-		No
Work way up the academic ladder	Yes	Yes	Yes	No		No
Hold water	Yes	Yes	Yes	Yes		Yes
Then again	Yes	Yes	Yes	Yes		Yes
Give weight to sth	Yes	Yes	No (replaceable words)	-		No
Open a window for sth	Yes	Yes	Yes	No		No
Open the door to sth	Yes	Yes	Yes	Yes		Yes
At any rate	Yes	Yes	Yes	Yes		Yes
For instance	Yes	Yes	Yes	No	Yes	Yes
In the final analysis	Yes	Yes	Yes	Yes		Yes

This framework distinguishes idioms from collocations (e.g., *break free*), metaphors (e.g., *a recipe for failure*), and other expressions by emphasizing their unique properties in academic contexts.

Research Question 2: Frequency of Idioms

Using the developed framework, 567 idioms were identified in 240 RAs and searched across the full corpus using #LancsBox. A total of 42,499 idiom occurrences (564 unique idioms) were found. Table 2 lists idioms with a relative

frequency (RF) ≥ 0.10 per 10,000 words, including frequency (f) and range (r, number of texts).

Table 2

Most Frequent Academic Idioms in Applied Linguistics RAs

Idiom	f	rf	r
For example	3565	6.28	407
As well as	2019	3.56	376
In addition	1560	2.75	327
In terms of	1475	2.60	331
In order to	1469	2.59	348
At least	1247	2.20	332
In fact	969	1.71	308
On the other hand	661	1.16	258
In particular	585	1.03	243
For instance	568	1.00	172

Notably, *for example* is idiomatic due to its grammatical deviation (lacking an article before “example”). Non-compositional idioms, such as *shed/shine/throw/cast light on* (f=120, rf=0.21, r=87), were less frequent but significant (Table 3).

Table 3

The Most Frequent Non-Compositional Idioms Used in RAs

Idioms	f	rf	r
Shed/shine/ throw/ cast light on	120	0.21	87
In the light of	60	0.11	45
By/in virtue of	57	0.10	36
In keeping with	41	0.07	33
In the hands of /In sb’s hands	36	0.06	32
Come/ bring into play	28	0.05	26
By and large	26	0.05	17
Let alone	24	0.04	20
Come to terms with	20	0.04	11
Bring/come to light	14	0.02	12
Bring sth home	10	0.02	9
Rule(s) of thumb	9	0.02	6
Part and parcel of	8	0.01	7
Be in tune with	8	0.01	8

It stands to reason	8	0.01	8
Come/get to grips with	8	0.01	6
Take sth on board	6	0.01	6
Sweep sth under the carpet	6	0.01	3
Take stock of sth	6	0.01	4
Sit comfortably/uncomfortably/well/with	6	0.01	6
By heart	5	0.009	3
In a nutshell	5	0.009	4
pay lip service to sth	4	0.007	4
Scrape the barrel	4	0.007	1
Shake one's head at sth	3	0.005	3
Stand a chance	3	0.005	3
Leave sb's mark on sth/sb	3	0.005	3
The other way round	3	0.005	3
Fall into the trap of	3	0.005	3
Order of the day	3	0.005	3
Find your voice	3	0.005	3
As much	3	0.005	3
On the right track	3	0.005	3
On a par with	3	0.005	3
Come of age	3	0.005	3

Table 4 lists pedagogically relevant idioms for EAP writing, such as *in terms of* and *in light of*, which are formal and suitable for academic contexts.

Table 4

Pedagogical Academic Idioms

In terms of	In line with	In view of
In light of	Shed/shine/ throw/ cast light on	At best
As it were	Bring/come to light	So to speak
With respect to	In a way (meaning to some extent)	In this regard
In part	So as to	Keep/bear sth in mind
With/in regard to	With reference to	Thus far
As such	In conjunction with	In favour of
As with sb/sth	In this respect	In sb/sth's own right
As yet	In effect	Above all
Bring/come to the fore	In and of itself	For one thing
In a positive/negative light	Sit comfortably/uncomfortably /with	All the same
At any rate	Be in tune with sth	It stands to reason
By way of (example, etc.)	Do justice to sth	Goes without saying...
By virtue of	In reference to	As yet

Give rise to	By and large	When it comes to
That said	By reference to	Then again
By definition	Anything but	For the sake of argument
In keeping with	It is safe to say	Not least
That/this is to say	Stand the test of time	In tandem
Less than	In the final analysis	Not to mention
Bring to bear on/upon	By the same token	What is more
Take issue with sth	Bring sth into focus	

Research Question 3: Diachronic Trends

To examine differences in idiom use from 1980 to 2020, 240 RAs (1,850,559 tokens) were analyzed manually, divided into four decades. Frequencies were normalized to 300,000 words (Table 5).

Table 5

Normalized Frequencies of Idioms per Decade (Per 300,000 Words)

Decade	Articles	Tokens	Idioms	Normed Tokens	Normed Idioms
1980–1989	60	385,194	1259	300,000	980
1990–1999	60	498,665	1882	300,000	1132
2000–2009	60	437,240	1711	300,000	1173
2010–2020	60	531,204	1982	300,000	1119

A Chi-Square test (Table 6) showed no significant difference in idiom use across decades ($\chi^2 (9) = 12.000$, $p = .213$), confirming the null hypothesis.

Table 6

Chi-Square Tests for Idiom Use per Decade

Test	Value	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.000	.213
Likelihood Ratio	11.090	.270
Linear-by-Linear Association	1.491	.222

The developed framework effectively identifies academic idioms, revealing their frequent use (42,499 occurrences) in applied linguistics RAs. Idioms like *for example* and *in terms of* are prevalent, while non-compositional idioms like *shed light on* are less common but pedagogically valuable. No significant diachronic changes were observed, suggesting consistent idiom use over time.

Discussion

Academic Idioms Redefined and Recognized

The main purpose of this study was to shed new light on idioms in RAs and develop an operational definition for academic idioms in RAs that have been ignored thus far in the literature. A framework for idiom identification and an operational definition was developed:

Idioms are dynamically institutionalized expressions of two words or more, but less than a sentence, whose individual words cannot be replaced with other similar words, and are either non-compositional or odd in structure or both.

The criteria which were focused on in the framework were: *multi-wordiness* (two words or more but less than a complete sentence), *institutionalization* (i.e., being included in at least one of the dictionaries available in English), *restrictive flexibility* (i.e. The words in the expression cannot be replaced with other synonyms and keep the idiomatic meaning), *non-compositionality* (i.e. the sense of the words in the expression is not included in any of the dictionaries), and/or *oddness in structure* (i.e. The expression deviates from syntactical rules).

It should be borne in mind that an expression must have the first three features together along with either of the last two features, or both of them to be labeled as an idiom.

This definition of academic idioms was attempted to consider the shortcomings of the previous definitions proposed in the literature (e.g., Fernando, 1996; Gramley & Patzold, 2004; Grant & Bauer, 2004). In other words, this definition takes into account a neglected aspect of this linguistic phenomenon which is related to the formal aspect of them. Most of the definitions proposed are related to idioms in informal genres, ignoring that idioms are also used in academic genres (Miller, 2020; Simpson & Mendis, 2003).

In line with the previous studies (e.g., Dobrovol'skij & Piirainen, 2005; Liontas, 2017; Miller, 2011; Miller, 2020), the size of idioms was opted for to be of at least two words and less than a complete sentence, and ignoring the idiomatic compound words which were included in some studies on idioms (e.g. Katz & Postal, 1964; Makkai, 1972). Moreover, idiomatic phrasal verbs were 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 considered phrasal verbs in his study of idioms.

Moreover, *meaning* has been the major criterion on which idioms have been so far defined (e.g., Fernando, 1996; Fraser, 1970; Gramly & Patzold, 2004; Grant & Bauer, 2004; Weinreich, 1969), ignoring the idiosyncrasy of idioms in terms of their syntactic structures. However, the emphasis on this aspect of idioms is in line with Flavel and Flavel (2011) and Gramly and Patzold's (2004) conceptions of idioms.

More importantly, the definition provided in this study corroborate the findings of the previous corpus-based studies which manifested that non-compositionality as a major feature of idioms should be reconsidered, that non-compositionality is a matter of degree, and some idioms are more non-compositional than others (Geeraert, 2016; Gluckberg et al., 2001; Langlotz, 2006; Moon, 1998; Taylor, 2012; Wulff, 2008; Zyzik, 2011). More to the point, in this definition, non-compositionality is redefined so that it is operational, without involving the subjectivity, as was the case with Grant and Bauer's (2004) framework of idioms. In this framework, the differences between figurative expressions and non-compositional idioms were based on the intuition of a language user.

This study also confirmed that idioms should not be considered totally fixed and frozen structures (Glucksberg et al., 2001; Kyriacou et al., 2020; Langlotz, 2006; Lynn, 2016; Moon, 1998). For example, one of the idioms found in the corpus: *to shed light on something*, was used with many variations by different writers in RAs:

*"Qualitative findings **shed further light on** the quantitative results"* (Tedick & Young, 2016, p.20).

*"That is, corpus-based research **sheds new light on** some of our most basic assumptions about English"* (Biber et al., 1994, p. 174).

*"There was a general agreement that research was necessary into a variety of areas in order **to throw more light on** the nature of language proficiency"* (Alderson & Clapham, 1992, p. 162).

*"The diaries **threw valuable light on** aspects of vocabulary teaching which had been covered on the TEFL course"* (Richards, 1992, p.148).

As another example which supports the results of the previous studies in this regard, the idiom *make sense* was also found to be used with variations. For example,

*"Use of heavy to modify rain **makes perfect sense**"* (Liu, 2012, p.18).

*"Meanings of strong versus powerful tea **make very good sense**"* (Liu, 2010, p.18).

Similarly, the idiom: *with reference to* had many variations in its form:

*"**With specific reference to** L2 grammar teaching it does raise interesting questions about..."* (Borg, 1998, p.31)

*"I now turn to consider the description of collocations and idioms in specialized dictionaries, **with special reference to...**"* (Cowie, 1981, p. 225).

The researcher avoided to use the term *fixedness*; instead, she opted for the term *restrictive flexibility*. This term refers to the fact that words in an idiom cannot be replaced with other similar words and still keeps its idiomatic meaning. This feature is similar to what Gramly and Patzold (2004) refer to as *recurrent semantic contrast*.

As manifested by the examples taken from the corpus, idiomatic collocational patterns and idioms can be differentiated considering these features.

What makes this framework different from the previous frameworks available in the literature is that by using this framework, the neglected group of idioms, and the idioms with oddness in structure, such as *for example* can also be taken into account. This group of idioms has been neglected in previous studies (e.g., Miller, 2020; Simpson & Mendis, 2003).

A List of Academic Idioms

The second research question focused on the frequency of idioms based on the framework and definition developed. A list of the most frequent academic idioms was compiled.

The results indicated that applied linguistics writers of RAs make use of idioms in their RAs. Moreover, the results showed that proficient writers in the field have the confidence to make use of *informal idioms* as well, and this usage is well accepted. In contrast to what is usually stated in EAP and EFL coursebooks, avoiding learners from making use of idioms in written academic genres (Bailey, 2015; Sowton, 2012).

Comparing the list of idioms with the lists of idioms reviewed in the literature shows that there are major differences between them. From the most frequent idioms in academic spoken genres reported by Simpson and Mendis (2003), none of the idioms are in common with the most frequent idioms reported in this study. However, three idioms: *come into play*, *the bottom line*, and *draw a/the line* were common, although with a low frequency.

Regarding the most frequent idioms reported by Grant and Nation (2006), only two idioms: *as well* and *by and large* are in common with the idioms found in this study. Of Chen & Wang's (2016) idioms, only *Achilles' heel* was in common with the idioms in this study.

However, comparing the idioms in Rafatbaksh and Ahmadi (2019) and Rafatbaksh and Ahmadi (2020) with the idioms in this study showed some similarities. Many idioms such as: *in the long run*, *in the short run*, *in the light of*, *as it were*, *in the final analysis*, *by and large*, *pave the way*, and *in tandem* were in common. This finding can be justified by the point that in these two studies, in contrast to previous studies, a part of the corpus investigated is devoted to academic written genres. Therefore, it stands to reason that there are more similarities between their results and those of this study.

And finally, considering the most frequent idioms with a frequency over 4 pmw in the academic written corpus in Miller (2020) shows that there are many similarities between that list and the results of this study. This can also be explained by the point that one of the corpora used in this study is devoted to academic written genres, although it is not specific to RAs.

The results of the comparisons which were run between the different lists of idioms introduced in the literature and the list in this idioms can be justified by the point that the corpus used in this study is a specialized one, focusing only on one genre and one discipline, in contrast to previous studies which reported generic idiom lists by using different genres in different disciplines (Liu, 2003; Miller, 2020; Rafatbakhsh & Ahamdi, 2020; Simpson & Mendis, 2003).

Given these results, the findings of this study throw into relief the differences that exist between different registers, genres, and disciplines in terms of idiom usage (Alizadeh et al., 2024) and confirm Hyland and Tse (2007) and Durrant's (2009) cautiousness that generic idiom lists will not apply to all disciplines and lexical needs of students vary in different disciplines.

Moreover, the results confirm the results of previous corpus-informed studies making use of huge corpora that although idioms, in general, do not enjoy high frequency compared to the other genres such as fiction, there are idioms more frequent in written academic genres such RAs (Moon, 1998; Biber et al., 1999).

As Biber et al. (1999) state (p.1026), a portion of the idioms found in this study can be called academic idioms specific to academic genres. This finding manifests the importance of dividing idioms into different groups such as *informal idioms*, *formal idioms*, *frequent idioms in spoken academic English*, and *frequent idioms in written academic English*.

The results show that the most frequent idioms found in applied linguistics RAs are relational idioms, which are used as cohesive devices by the authors. It comes as no surprise considering the feature of this academic genre. Most of the idioms in this list might seem totally literal at first sight, but this is not the case with all of them. Some of them are not non-compositional, but they deviate from the grammatical rules of English.

Results show that idioms used in RAs can be divided into three groups: non-compositional idioms, idioms with oddness in structure, and full idioms. Non-compositional idioms are those whose meanings cannot be deduced from the words included. It should be kept in mind that to be included as a non-compositional idiom, an expression must have the three initial features introduced in the previous section.

There are also figurative idioms used by the authors, implying that informal idioms can also be used in this genre. Although these types of idioms have low frequency, as Miller (2020) argues, lack of high frequency is not equal to the lack of significance.

Moreover, comparing the idioms with the lists of idioms reviewed in existing literature, shows that although some idioms are in common in most of these lists, such as *shed light on something*, many of the idioms found in this study are not reported in previous findings. This can be justified by the point that the corpus used in this study is a specialized one, focusing only on one genre, in contrast to previous studies which reported generic idiom lists by using different genres in different disciplines (Liu, 2003; Miller, 2020; Rafatbakhsh & Ahamdi, 2020; Simpson & Mendis, 2003;).

The results of this study throw into relief the differences that exist between different registers, genres, and disciplines in terms of idiom usage and confirm Hyland and Tse (2007) and Durrant's (2009) cautiousness that generic idiom lists will not apply to all disciplines, and that lexical needs of students vary in different disciplines. Moreover, the results confirm the results of previous corpus-informed studies making use of huge corpora that although idioms, in general, do not enjoy high frequency compared to the other genres such as fiction, there are idioms more frequent in written academic genres such RAs (Biber et al., 1999; Moon, 1998).

Any Diachronic Change in the Use of Idioms?

The results which were obtained in response to research question 3 indicate that the use of idioms seems to be a matter of style in writing rather than a matter of time since in each of the four decades, there were some RAs which were richer by far in the use of idioms. An article published in the *Journal of Applied Linguistics* in 2000 authored by Widdowson (1993) is the most idiomatic of all, including 121 idioms. The next richest RA in terms of idiom usage belongs to decade1, authored by Canale and Swain (1980).

Some writers have the confidence and creativity to make use of idioms in their writings, some do not. This finding might go back to writers' first language. As Baker (2018) argues, "a person's confidence in actively using the idioms and fixed expressions of a foreign language hardly ever matches that of a native speaker" (p. 70). More to the point, as Baker discusses, in some languages, there is a considerable difference between written and spoken registers. In languages such as Arabic and Chinese, language users tend to avoid using idioms in formal discourses since it is assumed that idioms belong to informal registers. Given this point, a writer's tendency to make use of idioms may be influenced by her/his mother tongue. Furthermore, disciplinary variations in considering pragmatic and stylistic aspects of academic writing cannot be neglected. For instance, Khaghaninejad et al. (2022) that objectivity in academic texts varies across academic disciplines.

Conclusion

This study reveals that idioms, often viewed as informal, play a significant role in applied linguistics RAs, a formal academic genre. Though less frequent than other lexical items, idioms are used purposefully and creatively by proficient writers, enhancing cohesion and expressiveness, thus challenging the notion that they are unsuitable for academic writing. A two-stage framework was developed to identify academic idioms (Figure1). Expressions must exhibit multi-wordiness (two or more words, less than a sentence), institutionalization (listed in at least one English dictionary), and restrictive flexibility (non-replaceable words). Qualifying expressions are then evaluated for non-compositionality (meaning not deducible from components) or odd structure (deviation from grammatical norms). This framework, exemplified in Table1, distinguishes idioms from collocations and metaphors.

The operational definition of idioms as dynamically institutionalized expressions of two or more words, non-replaceable, and either non-compositional or

odd in structure, provides clarity for corpus-based studies. Analysis of a 640-RA corpus (5,675,554 tokens) identified 42,499 idiom occurrences (564 unique idioms), with relational idioms like *for example* ($f=3565$, $rf=6.28$) and *in terms of* ($f=1475$, $rf=2.60$) serving as cohesive devices. Informal idioms (e.g., *Achilles heel*) were less frequent but accepted in prestigious journals, reflecting authorial style and idiomatic competence. No significant diachronic differences in idiom frequency were found across 1980–2020 ($\chi^2(9) = 12.000$, $p = .213$), suggesting usage depends on writers' stylistic preferences rather than temporal trends. The framework addresses a literature gap by distinguishing idioms from other multi-word expressions, enhancing understanding of their role in formal genres. For EFL/EAP contexts, it aids teachers in clarifying idioms versus collocations or proverbs, improving learners' idiomatic competence (Liontas, 2017). The idiom list, particularly formal ones like *in light of*, can be integrated into EAP writing courses and textbooks to enhance authenticity and reduce cognitive load (Adel & Erman, 2012). Authentic RA examples (e.g., Widdowson, 1993) model creative idiom use, and the list supports translation courses addressing idiom challenges. Limitations include the manual analysis of only 240 RAs due to time constraints and the lack of comprehensive inter-rater reliability due to idiom definition complexities. Future research could explore idioms in other disciplines, compare native and non-native writers, investigate idioms in other languages (e.g., Farsi), or examine novice versus proficient writers. Research on other academic genres (e.g., theses, editorials) and idiom functions (e.g., evaluation, emphasis) is also recommended.

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>
- Alizadeh, A., Kuhl, D., & Hadidi, Y. (2024). Structural and functional differences of lexical bundles between Hard Science and Soft Science researchers: A comparative corpus-driven study. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, 12(2), 87-114. <https://doi.org/10.22049/jalda.2024.29174.1630>
- Bailey, S. (2015). *Academic writing: A handbook for international students* (4th ed.). Routledge.
- Baker, M. (2018). *In other words: A coursebook on translation*. Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics?. *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.

- Brezina, V., Timperley, M., & McEnery, A. (2018). #LancsBox v. 4.x. [software package].
- Brezina, V., Timperley, M., & McEnery, A. (2020). #LancsBox v. 4.x. [software package]
- Chen, Y., & Wang, L. (2016). Opening Pandora's box: A corpus-based study of idioms in ELT materials. *The Journal of Asia TEFL*, 13(4), 381-39. <http://dx.doi.org/10.18823/asiatefl.2016.13.4.10.381>
- Cooper, T. C. (1998). Teaching idioms. *Foreign Language Annals*, 31(2), 255-266. <https://doi.org/10.1111/j.1944-9720.1998.tb00572.x>
- Dobrovol'skij, D & Piirainen, E. (2005). *Figurative language: Cross-cultural and cross-linguistic perspectives*. Elsevier Ltd.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157-169. <https://doi.org/10.1016/j.esp.2009.02.002>
- Fernando, C. (1996). *Idioms and idiomaticity*. Oxford University Press.
- Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of Language*, 6(1), 22-42.
- 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>
- 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., & 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>
- Grant, L., & Nation, I.S. P. (2006). How many idioms are there in English? *ITL-International Journal of Applied Linguistics*, 151, 1-14. <https://doi.org/10.2143/ITL.151.0.2015219>
- 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.
- Hyland, K. (2009). *Academic discourse: English in a global context*. A&C Black.
- Hyland, K. L., & Paltridge, B. (Eds.). (2011). *Continuum companion to discourse*. Continuum.

- Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, 41(2), 235-253. <https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>
- Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. M.I.T. Press.
- Khaghaninejad, M. S., Azarian, M., & Javanmardi, F. (2022). A corpus-based cross-disciplinary analysis of objectivity manifestations in academic texts. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, 10(1), 119-140. <https://dx.doi.org/10.22049/jalda.2022.27325.1333>
- Khonbi, Z., & Sadeghi, K. (2017). Improving English language learners' idiomatic competence: does mode of teaching play a role? *Iranian Journal of Language Teaching Research*, 5(3), 61-79. <https://dx.doi.org/10.30466/ijltr.2017.20305>
- 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
- 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>
- Liu, D. (2012). The most frequently-used multi word constructions in academic written English: A multi-corpus study. *English for Specific purposes*, 31, 25-35. <https://doi.org/10.1016/j.esp.2011.07.002>
- 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/>
- Makkai, A. (1972). *Idiom structure in English*. De Gruyter Mouton.
- McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge University Press.

- McCarthy, M. J. & O'Dell, F. (2010) *English idioms in use*. Cambridge University Press.
- McCarthy, M. J. & O'Dell, F. (2008). *Academic vocabulary in use*. Cambridge University Press.
- 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.
- Nunberg, G. D. (1978). *The pragmatics of reference*. Indiana University Linguistics Club.
- O'Dell, F., & McCarthy, M. (2017). *English collocations in use: Advanced, how words work together for fluent and natural English; self-study and classroom use*. Cambridge University Press.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- 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), 4-11. <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>
- Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics*. Longman.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37(3), 419-441. <https://doi.org/10.2307/3588398>
- Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford University Press.
- Simpson, A.J., & Weiner, E. S. C. (1989). *The Oxford English dictionary*. Oxford University Press.
- Sowton, C. (2012). *50 steps to improving your academic writing*. Garnet Education.

- Taylor, J. R. (2012). *Mental corpus: How language is represented in the mind*. Oxford University Press.
- Van Enk, A., & Power, K. (2017). What is a research article? Genre variability and data selection in genre research. *Journal of English for Academic Purposes*, 29, 1-11. <https://doi.org/10.1016/j.jeap.2017.07.002>
- Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *The substance and structure of language* (pp. 23-81). University of California Press.
- 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>

Corpus References

- Alderson, J. C., & Clapham, C. (1992). Applied linguistics and language testing: A case study of the ELTS test. *Applied Linguistics*, 13(2), 149-167. <https://doi.org/10.1093/applin/13.2.149>
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, 15(2), 169-189. <https://doi.org/10.1093/applin/15.2.169>
- Borg, S. (1998). Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly*, 32(1), 9-38. <https://doi.org/10.2307/3587900>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47. <https://doi.org/10.1093/applin/I.1.1>
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235. <https://doi.org/10.1093/applin/II.3.223>
- Edge, J. (1989). Ablocutionary value: On the application of language teaching to linguistics. *Applied Linguistics*, 10(4), 407-417. <https://doi.org/10.1093/applin/10.4.407>
- Liu, D. (2010). Going beyond patterns: Involving cognitive analysis in the learning of collocations. *TESOL Quarterly*, 44(1), 4-30. <https://doi.org/10.5054/tq.2010.214046>
- Richards, K. (1992). Pepys into a TEFL course. *ELT Journal*, 46(2), 144-152. <https://doi.org/10.1093/elt/46.2.144>

- Tedick, D. J., & Young, A. I. (2016). Fifth grade two-way immersion students' responses to form-focused instruction. *Applied Linguistics*, 37(6), 784-807. <https://doi.org/10.1093/applin/amu066>
- Widdowson, H. G. (1993). Proper words in proper places. *ELT Journal*, 47(4), 317-329. <https://doi.org/10.1093/elt/47.4.317>

Authors' Biography



Elham Farahani received her BA in English Language and Literature from the University of Qom, and her MA and PhD degrees in TEFL from Arak University. She has served as a faculty member from 2018 to 2022 in the Department of English Language at Mofid University. In 2023, she joined Arak University, where she is currently an Assistant Professor. Her areas of interest include Corpus Linguistics, English for Academic Purposes, English for Research Publication Purposes, and (Critical) Discourse Analysis.



Hooshang Yazdani, Associate Professor of TEFL at Arak University, studied TEFL as an undergraduate at Shiraz University and went on to do an MA in the same field. He continued the same field of study for PhD. He conducted a piece of research on a new way of evaluating EFL readers' reading comprehension abilities for his MA Degree and investigated the effects of different variables on reading strategy choice and comprehension for PhD Degree. He has taught different basic courses related to language, ESP and Applied Linguistics for about five years before starting his PhD abroad in 1993 and nineteen years since holding the degree in 2000. His research interests mainly lie in Discourse Analysis, Vocabulary Acquisition, Individual Differences & SLA.