E·S·P·Today
JOURNAL OF ENGLISH FOR SPECIFIC PURPOSES AT TERTIARY LEVEL

Vanessa Leonardi*

Sapienza University of Rome, Italy
Faculty of Economics
vanessa.leonardi@uniroma1.it

IDIOMS IN ACADEMIC WRITING: A CORPUS-BASED STUDY OF NATIVE ENGLISH SPEAKERS ACROSS ACADEMIC LEVELS AND DISCIPLINES

Abstract

This study explores the use of idiomatic expressions in academic writing, comparing texts written at undergraduate and master's level in the Cambridge Academic English (CAE) Corpus with an *ad hoc* corpus of 30 PhD theses (COPLUS EN) written by native English speakers across three European Research Council (ERC) domains: Social Sciences and Humanities (SH), Physical Sciences and Engineering (PE) and Life Sciences (LS). While idioms are often studied in spoken language, this research focuses on their role in formal academic writing and seeks to address three primary objectives: (1) to determine the extent to which native English speakers use idiomatic expressions in academic writing, particularly in PhD theses compared to undergraduate and master's level texts, (2) to assess how proficiency level and linguistic awareness influence the frequency and functions of idiomatic expressions at different stages of academic study, and (3) to identify whether certain idioms are more prominent in specific academic disciplines. A mixed-methods approach combines quantitative analysis of idiom frequency with qualitative analysis of their functions within each ERC domain to examine whether idiomatic usage varies across academic levels and across disciplines. Pedagogical implications include recommendations for English for Academic Purposes (EAP) educators to integrate idiomatic expressions into teaching, focusing on their role in improving students' ability to construct persuasive arguments, establish epistemic positioning and communicate with greater precision in academic contexts.

Key words

idiomatic expressions, academic writing, corpus-based analysis, disciplinary variation, English for academic purposes (EAP).

* Corresponding address: Vanessa Leonardi, Via del Castro Laurenziano 9, 00161 Rome, Italy.

1. INTRODUCTION

Idioms are a pervasive feature of both spoken and written English, adding nuance, cultural context and rhetorical variety to communication. Despite their association with informal or colloquial registers, idiomatic expressions are also prevalent in academic contexts. Research conducted by Simpson and Mendis (2003) and Miller (2020) reveals that idioms are more frequently associated with spoken academic discourse, such as lectures and seminars, yet they also feature in written academic texts, though less frequently. This interesting intersection of idiomatic language and formal writing invites further exploration into how these expressions enhance the clarity and expressiveness of scholarly communication.

Scholars such as Fernando (1996), Wray (1999, 2000, 2002) and Schmitt (2000) have underscored the importance of idiomatic expressions in achieving advanced language proficiency. These studies highlight how idioms contribute to fluency and successful communication, particularly for native speakers, who often incorporate them seamlessly into their language use. Idioms, far from being merely decorative, serve key discourse functions, enhancing coherence, rhetorical impact and cultural resonance. For learners of English, developing both receptive and productive knowledge of idioms is vital for full participation in academic discourse. The ability to use idioms appropriately in academic writing, for instance, can signify linguistic and cultural competence, enabling students to engage more effectively with the conventions of their academic community. Moreover, understanding idioms in academic texts and lectures supports students' ability to interpret complex ideas, particularly where idioms are used to simplify abstract concepts.

However, idiomatic language can present significant challenges for students. Idioms are often highly context-dependent and culturally embedded, making them difficult to decode without prior exposure. Unlike more transparent academic vocabulary, idioms do not always lend themselves to straightforward inference, which can hinder comprehension for those unfamiliar with them. Miller (2020) notes that in spoken academic contexts, lecturers often signal idioms using phrases such as *so to speak* or *as it were*, guiding listeners toward nonliteral interpretations. In written academic texts, idioms may similarly be flagged with phrases such as *as they say* or through the use of quotation marks, indicating their figurative or idiomatic nature. Nevertheless, such signposting is inconsistent, creating further barriers to comprehension.

Despite their role in facilitating communication and adding rhetorical variety, idioms remain underrepresented in the teaching of English for Academic Purposes (EAP). This gap is particularly evident in the context of academic writing, where research on idioms is relatively sparse. Popular EAP textbooks, such as *Oxford EAP* by de Chazal and Moore (2013), dedicate minimal space to idioms, while O'Dell and McCarthy's *English idioms in use: Advanced* (2010) includes only a brief section on idiomatic expressions relevant to academic writing. This limited focus overlooks the potential of idioms to enrich academic texts by enhancing their readability,

coherence and engagement. A corpus-based study on Business English textbooks further confirms that idioms are often inconsistently represented across educational materials, leading to gaps in formal academic writing instruction (Parizoska & Rajh, 2017).

In academic writing, formal, precise language is often prioritised to maintain clarity and credibility. Idiomatic expressions, defined as fixed or semifixed phrases whose meanings cannot be directly inferred from their individual components (Moon, 1998), occupy an ambiguous space within this context. While traditionally regarded as informal, idioms may nevertheless play strategic roles in academic writing. The growing body of research on idiomatic language has largely concentrated on casual spoken communication, media texts and the challenges faced by nonnative speakers in acquiring idioms. By contrast, relatively little attention has been paid to the use of idioms by native English speakers in formal academic writing, particularly within structured, argument-driven genres such as PhD theses.

This study addresses this gap by focusing on idiomatic expressions in academic writing by native English speakers, comparing writing at undergraduate and master's level, and doctoral theses across three disciplinary domains, namely Social Sciences and Humanities (SH), Physical Sciences and Engineering (PE) and Life Sciences (LS). The decision to focus on native speakers is deliberate, as idioms form an integral part of their linguistic repertoire. Native speakers may employ idioms not only to add rhetorical flair to their writing but also strategically, enhancing the clarity and persuasiveness of academic arguments. Furthermore, this study examines the extent to which the use of idioms varies between academic disciplines, reflecting the different conventions and rhetorical expectations of each domain. It posits that idioms may serve distinct functions across disciplines and that their frequency and prominence will differ accordingly.

This research is guided by three principal questions:

1. To what extent are idiomatic expressions employed in academic writing, particularly in PhD theses compared to texts produced at the undergraduate and master's levels authored by native English speakers?
2. In what ways do proficiency levels and linguistic awareness influence the frequency and functions of idiomatic expressions within academic writing at the undergraduate and master's levels versus doctoral writing?
3. Are certain idioms more prevalent in specific academic domains within doctoral theses, and if so, what are the underlying reasons for this phenomenon?

To address these questions, this study employs a mixed-methods approach, combining quantitative analysis to measure the frequency of idioms in academic texts with qualitative analysis to examine their rhetorical and functional roles. The findings are expected to provide new insights into the use of idiomatic language in

academic writing, offering practical recommendations for EAP educators and highlighting the potential of idioms to enhance scholarly communication.

The subsequent sections of this paper are organised as follows: the literature review explores previous research on idiomatic expressions in both academic and general discourse. The methodology section outlines the corpus and analytical framework employed in this study. The results and discussion section presents and interprets the findings in relation to the research questions. This is followed by an exploration of the pedagogical implications of the outcomes. Finally, the concluding section summarises the key findings, addresses the limitations of the study and suggests avenues for future research on idioms in academic writing.

2. LITERATURE REVIEW

Early research on idiomatic expressions predominantly focused on their semantic and syntactic properties, as seen in foundational works by Weinreich (1969) and Makkai (1972). These studies laid the foundation for structural analyses of idioms, but more recent scholarship has expanded to explore the broader communicative and pragmatic functions of idiomatic language. A significant shift towards understanding idioms as essential tools for communication is evident in the work of Strässler (1982), Fernando (1996), McCarthy (1998), and Moon (1998), who argue that idioms play a vital role in communication beyond their linguistic structure.

McCarthy (1998) contends that idioms should not be regarded merely as linguistic anomalies but as functional elements of language that facilitate interaction. This perspective resonates with a broader research trend on formulaic language, which highlights fixed expressions as integral to achieving native-like fluency. In this regard, contributions from Nattinger and DeCarrico (1992), Wray (1999, 2000, 2002), and Moon (1998) collectively emphasise that formulaic expressions, including idioms, are crucial for developing communicative competence. Wray (1999, 2000, 2002) further argues that idioms aid both comprehension and production, serving as holistic units that are retrieved from memory, rather than constructed analytically. This shift in perspective underscores the importance of mastering idioms for effective communication in diverse social and academic contexts.

Fernando (1996) and Wray (1999) echo this view in the context of language learning and translation, stressing that idioms must be prioritised in pedagogy to ensure natural language use. They align with Pawley and Syder's (1983) observation that learners struggle to distinguish between idiomatic expressions and grammatically correct but nonidiomatic phrases. Nattinger and DeCarrico (1992) contribute to this discussion by classifying these expressions as lexical phrases and proposing pedagogical strategies to integrate them into second language instruction, thereby enhancing learners' ability to communicate effectively.

In more recent studies, idioms have garnered increasing attention within the field of EAP due to the challenges they pose for nonnative speakers. Idioms in academic discourse often convey complex meanings that cannot be understood by analysing individual components, which makes them particularly challenging for learners seeking fluency and comprehension in academic contexts. Data-driven learning (DDL) methods, particularly corpus-based instruction, have been suggested as effective tools for helping learners acquire multiword expressions (MWEs) and idioms (Marín-Pérez & Aguado Jiménez, 2024).

Several studies have explored the use of idioms in academic discourse. Simpson and Mendis (2003) conducted an early investigation using the Michigan Corpus of Academic Spoken English (MICASE), identifying 238 distinct idioms, though only 32 were found to appear recurrently across contexts. While their work highlighted the pragmatic importance of idioms in academic discourse, it did not provide an in-depth analysis of their distribution across different academic settings. Liu (2003) expanded on this by analysing MWEs across three American English corpora, identifying 302 highly frequent expressions, thus offering a more frequency-based perspective on the use of idioms in academic contexts. In contrast, Grant and Bauer (2004) focused on core idioms, those which are noncompositional and opaque, identifying a set of expressions with high instructional value. Grant (2007) further refined this by examining idiom frequency in British and American corpora, noting the need to exclude vague expressions and phrasal verbs to identify idioms of higher instructional relevance.

Simpson-Vlach and Ellis (2010) introduced a phraseological approach, compiling an Academic Formulas List to identify key idiomatic expressions for academic contexts based on frequency and mutual information measures. Liu (2012) contributed to this work by examining MWEs in academic writing across the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), highlighting the cross-disciplinary relevance of idioms for EAP instruction. Martinez and Schmitt (2012) further developed these ideas by distinguishing fixed from semifixed expressions, offering insight into the distinctions between idioms and collocations, which is valuable for learners. Martinez (2013) extended this work, proposing that the instructional value of idioms should be assessed not only on their frequency but also on their transparency, arguing that both factors are important for effective teaching.

The ongoing debate between discipline-specific and general academic idiom lists has also attracted attention. While Hyland and Tse (2007) argue for the need for discipline-specific vocabulary, Ackermann and Chen (2013) support the use of cross-disciplinary lists. Liu (2012) argues for a general academic vocabulary that can be adapted to specific fields, reflecting the need for a balance between general and specialised language resources in EAP instruction.

More recently, Miller (2020) challenged the notion that idioms are too informal for academic contexts by identifying 545 frequently used idioms in the BASE and OCAE corpora. His study found a higher prevalence of idioms in the Social

Sciences, which contrasts with Simpson and Mendis's (2003) claim that idioms are not dominant in any academic discipline.

While previous studies have explored idiomatic expressions in diverse academic contexts, this study fills a key gap by focusing specifically on the use of idioms by native English speakers. Unlike prior research, it provides a detailed examination of idiomatic variation across disciplines, analysing both the frequency and functions of idioms within the specialised context of doctoral theses. Furthermore, it compares idiomatic usage across proficiency levels, investigating differences between undergraduate, master's, and doctoral writing. This allows an assessment of how idiomatic expressions evolve with academic and linguistic advancement among native English speakers.

3. METHODOLOGY

This study adopts a corpus-based, mixed-methods approach to explore the use of idiomatic expressions in academic writing across different disciplines and academic levels. The list of idiomatic expressions analysed in this study is based on Miller's (2020) influential research and, more precisely, on those idioms which were found to be more frequently used in writing than in spoken academic language. In this respect, Miller identified idioms that appeared with a frequency of at least 2.00 occurrences per million words (pmw), prioritising expressions that are particularly salient in written academic contexts. This frequency-based methodology provided a solid framework to isolate idiomatic expressions that play a significant role in academic writing. From Miller's analysis, a list of 43 idioms was adopted and forms the basis for the present study (see Table 1).

To explore the use of these idiomatic expressions in academic writing, two specific corpora were used: the Cambridge Academic English (CAE) Corpus and the COPLUS EN corpus. CAE was filtered to include only writings by native English speakers at the undergraduate and master's levels. This corpus serves as a baseline to examine idiomatic usage at lower academic levels. COPLUS EN is an *ad hoc* corpus constructed for this study, which is made up of 30 PhD theses written by native English speakers. These theses were sourced from the EThOS database of the British Library, ensuring a high standard of academic quality and representativeness. The doctoral theses in the COPLUS EN corpus were drawn from three academic domains as defined by the European Research Council (ERC), namely SH, LS and PE. Each domain contributed ten theses to ensure balanced representation across disciplines, enabling meaningful comparisons of idiomatic usage. The COPLUS EN corpus was constructed with a focus on authenticity and contemporary relevance. Only theses published between 2019 and 2024 were included to ensure alignment with current trends in academic writing. Standardisation was a critical aspect of corpus design, particularly in terms of length and format, to minimise biases that might arise from discrepancies in token counts. As a result, the final corpus comprises approximately

3 million tokens, which may be deemed sufficient to capture the complexity and variability of idiomatic usage while maintaining comparability across domains. The texts were preprocessed and uploaded into Sketch Engine, which was selected for its advanced capabilities, including lemmatization and part-of-speech tagging among others.

The analysis was conducted in two distinct stages. The first stage involved a comparison of idiomatic expressions in the CAE and COPLUS EN corpora. The aim was to assess whether idioms are used more frequently or differently at the undergraduate/master's and doctoral levels, or if their usage remains consistent across these stages of academic progression. Several factors were considered, including relative frequency, which was calculated through normalisation per million tokens (pmt), diversity and contextual application to determine how idiomatic usage might evolve alongside increasing linguistic proficiency and academic expertise. In the second stage, the focus shifted exclusively to the COPLUS EN corpus, which was subdivided into three academic domains corresponding to the ERC categories. The idioms were analysed for their relative frequency and distribution within each domain using corpus analysis tools to compute normalised frequencies (pmt) and manual coding to identify contextual and rhetorical patterns. This part of the analysis aimed to ascertain whether certain idiomatic expressions are more prevalent in specific disciplines, reflecting variations in rhetorical conventions and linguistic preferences. Both stages of analysis employed a mixed-methods approach. Quantitative techniques included relative frequency analysis, which involved normalising occurrences per million tokens (pmt). This method was specifically chosen to ensure comparability between corpora and subsets of varying sizes, thus enabling reliable cross-corpus and cross-domain analysis. Qualitative analyses complemented this by examining the rhetorical and contextual functions of idiomatic expressions within texts. This combination provided a comprehensive understanding of idiomatic usage, capturing both numerical trends and their practical roles in academic writing.

4. RESULTS AND DISCUSSION

The comparative analysis of idiomatic expressions between CAE and COPLUS EN highlights significant differences in their usage by undergraduate/master's students versus PhD candidates. These differences reflect not only the varying levels of linguistic proficiency but also the distinct rhetorical and stylistic demands at different academic stages.

Table 1 shows the results of the first analysis comparing idioms in CAE and COPLUS EN corpora:

Idiom	CAE (pmt)	COPLUS EN (pmt)
on the other hand	45.21	62.22
in the light of	17.92	15.38
on the one hand	9.36	15.02
(take) a step back/further	5.89	8.39
driving force	4.82	3.14
on the other [hand]	4.01	6.7
trial and error	4.01	0
golden age	3.48	7.69
in its own right	3.21	2.45
in the hands of	2.94	5.6
bear in mind	2.94	4.54
come into play	2.94	0.35
beg the question	2.68	4.19
come to light	2.68	1.75
in the early days	2.68	0.35
on one hand	1.87	4.54
along the lines of	1.61	4.9
on the face of it	1.61	2.8
across the board	1.61	0
the big(ger) picture	1.34	3.14
from scratch	1.07	2.45
get to grips with	1.07	2.1
in the long run	1.07	0.7
pros and cons	1.07	0
bridge the gap	0.8	3.84
balance of power	0.8	3.5
raison d'être	0.8	1.4
behind the scenes	0.8	0.7
the high point	0.54	0.7

the whole story	0.54	0.35
last resort	0.27	6.30
at the end of the day	0.27	1.4
rule of thumb	0.27	0.7
state of the art	0.27	0.7
bad news	0.27	0
the bottom line	0.27	0
gold standard	0	5.24
track record	0	4.19
go hand in hand with	0	2.1
one's fair share	0	0.7
in the short run	0	0.7
the good life	0	0
win win	0	0

Table 1. Idiomatic expressions in academic writing: CAE vs. COPLUS EN

The idiom *on the other hand*, for instance, has a frequency of 45.21 pmt in CAE and 62.22 pmt in COPLUS EN. This phrase, which contrasts two opposing ideas or arguments, is a cornerstone of structured academic reasoning. Its increased frequency in COPLUS EN reflects the heightened need for refined argumentation in doctoral-level writing, where presenting alternative perspectives is critical to constructing complex and balanced arguments. Similarly, *on the one hand* and *on the other [hand]*, used together to introduce contrasting points, also appear more frequently in COPLUS EN. Their respective frequencies of 9.36 pmt versus 15.02 pmt and 4.01pmt versus 6.7 pmt suggest that PhD students engage more systematically with multifaceted arguments, a hallmark of advanced scholarly discourse.

The idiom *in the light of*, which has a frequency of 17.92 pmt in CAE and 15.38 pmt in COPLUS EN, is employed to introduce justifications or reconsiderations based on new evidence or perspectives. Its slight decline in COPLUS EN may indicate a shift among PhD candidates towards more specialised or discipline-specific expressions, thus reflecting their ability to articulate complex ideas with greater precision. Conversely, the idiom *bear in mind*, with a frequency of 2.94 pmt in CAE and 4.54 pmt in COPLUS EN, retains its presence in both corpora. This phrase serves as a reminder to consider specific factors and its conversational undertones may make it slightly less common in academic writing; however, its presence in COPLUS EN suggests that it is still valued for its rhetorical function of guiding the reader's attention in dense, complex arguments.

Some idioms reveal stark contrasts in usage between the two groups. The phrase *gold standard*, for example, is absent in CAE but has a frequency of 5.24 pmt in COPLUS EN. As a formal idiom denoting the highest benchmark within a field, it aligns closely with the evaluative and methodological focus of doctoral research. PhD candidates often assess theories, methodologies or findings making such evaluative language essential. Similarly, *last resort* has a frequency of 0.27 pmt in CAE but increases sharply to 6.30 pmt in COPLUS EN. This idiom, which refers to an option pursued when all others have failed, is likely to reflect discussions of contingency strategies or problem-solving approaches. These elements are, undoubtedly, more common in the high-stakes and solution-oriented nature of PhD research.

The idiom *bridge the gap* also displays greater usage in COPLUS EN, rising from a frequency of 0.8 pmt in CAE to 3.84 pmt in COPLUS EN. This phrase, which describes efforts to reconcile differences or connect disparate ideas, reflects the integrative aims of PhD-level research, where candidates often work to address gaps in the literature or unify theories. In contrast, idioms like *trial and error* and *pros and cons* show a marked decline or complete absence in COPLUS EN. The former, which has a frequency of 4.01 pmt in CAE, refers to a method of solving problems through repeated attempts. Its absence in COPLUS EN suggests that PhD students, writing within a framework that values systematic and evidence-based approaches, avoid the informal and imprecise connotations of this phrase. Similarly, the idiom *pros and cons*, with a frequency of 1.07 pmt in CAE but not appearing in COPLUS EN, reflects its conversational nature and its unsuitability for the advanced and formal tone expected of doctoral theses.

The idiom *in its own right*, which highlights the intrinsic value of something, has a frequency of 3.21 pmt in CAE but only 2.45 pmt in COPLUS EN. While it is moderately formal, its decline in COPLUS EN may reflect the tendency of PhD candidates to use more precise or technical language. In contrast, *along the lines of*, which suggests approximation or comparison, increases from a frequency of 1.61 pmt in CAE to 4.9 pmt in COPLUS EN. This phrase may be particularly useful for PhD candidates to introduce analogies or illustrative examples, which are crucial in order to explain complex ideas. Meanwhile, *golden age*, which is generally used to denote a period of great achievement, appears more frequently in COPLUS EN, rising from a frequency of 3.48 pmt in CAE to 7.69 pmt. This may indicate a higher tendency among PhD students to engage with evaluative or historical discussions in their writing.

It is noteworthy to observe how certain idioms, particularly those with conversational or informal connotations, exhibit a marked decline in usage within the context of COPLUS EN compared to CAE. For instance, the idiom *come into play*, which has a frequency of 2.94 pmt in CAE, is notably lower at 0.35 pmt in COPLUS EN, thus indicating a shift away from informal expressions towards more formal language. This idiom typically refers to factors becoming relevant in a given context. Similarly, the phrase *in the early days* is present with a frequency of 2.68 pmt in CAE

but only 0.35 pmt in COPLUS EN. This disparity seems to suggest that PhD students tend to prefer more formal and precise expressions when discussing historical or developmental contexts, thereby minimising the use of generalised idiomatic language. This preference aligns with the academic rigour expected at the doctoral level, where clarity and specificity are paramount.

Conversely, the idiom *track record*, which is entirely absent in CAE but has a frequency of 4.19 pmt in COPLUS EN, underscores an emphasis on prior achievements and credibility, which are key themes that resonate with the evaluative and evidence-based nature of doctoral research. The presence of this idiom reflects a tendency among doctoral students to articulate their qualifications and past successes in a manner that is both formal and discipline-specific. In addition, the phrase *go hand in hand with*, which does not appear in CAE but has a frequency of 2.1 pmt in COPLUS EN, highlights advanced discussions of interdependence and causality characteristic of PhD theses. This idiomatic expression seems to suggest a clear understanding of relationships between concepts, which is a critical aspect of doctoral-level discourse. Interestingly, the idiom *at the end of the day*, often regarded as too informal for academic writing, has a frequency of 0.27 pmt in CAE but 1.4 pmt in COPLUS EN. Its presence in doctoral writing may indicate a deliberate, albeit infrequent, use for rhetorical emphasis, suggesting that even at advanced levels of study, some informal expressions may be employed strategically to enhance argumentation. These shifts in idiomatic usage between undergraduate and master's levels compared to PhD levels seem to suggest a broader trend towards minimising informal language and embracing more formal, discipline-specific terminology. This evolution reflects the increasing expectations for precision and professionalism as students progress through their academic careers.

The data presented in this study appear to indicate that idiomatic expressions fulfil varying functions depending on the academic level of the students involved. It could be suggested that undergraduate and master's students predominantly use idioms as accessible tools to structure arguments or introduce ideas. This tendency may reflect their developing linguistic proficiency, which is often characterised by a reliance on familiar expressions that serve to facilitate communication within their academic contexts. Such usage might be interpreted as an attempt to navigate the complexities of academic discourse while still grappling with the nuances of more sophisticated language.

In contrast, at the PhD level, students seem to exhibit a greater degree of selectivity regarding their use of idiomatic expressions. This selectivity appears to favour idioms that align more closely with the precision, formality and evaluative tone expected in advanced academic writing. For instance, idiomatic phrases that might be considered too informal or overly general, such as *trial and error* or *pros and cons*, tend to be less frequently employed in this context. Instead, idioms that convey a sense of methodological or evaluative rigour, such as *gold standard* or *bridge the gap*, appear to gain prominence. This shift in preference could be

indicative of the evolving linguistic and rhetorical sophistication required as students progress through higher levels of academic discourse.

Furthermore, it may be worth noting that this evolution in idiomatic expression usage highlights not only a change in linguistic preferences but also a broader transformation in the expectations placed upon students as they advance academically. The increasing emphasis on precision and formal evaluative language at the doctoral level suggests a significant shift in how students are expected to engage with complex ideas and arguments. This could imply that as students progress through their studies, they are not only acquiring more advanced language skills but also adapting to the rigorous demands of scholarly communication.

To further investigate this phenomenon, the study moves into the second stage of analysis, aimed at exploring whether significant differences exist in the distribution and usage of idioms across various academic disciplines within the COPLUS EN corpus. The following table shows these patterns across three ERC domains, thereby providing deeper understanding of how idiomatic expressions function within distinct academic contexts. By examining these differences, it is possible to gain deeper insights into the role that idioms play in shaping academic discourse across different fields of study.

Idiom	**COPLUS EN** SH (pmt)	**COPLUS EN** LS (pmt)	**COPLUS EN** PE (pmt)
on the other hand	47.22	36.43	38.97
on the one hand	17.71	1.01	0
in the light of	14.33	10.12	0
golden age	9.27	0	0
(take) a step back/further	8.43	0	5.2
on the other [hand]	8.43	0	0
last resort	7.59	0	0
in the hands of	6.75	0	0
along the lines of	5.9	0	0
beg the question	5.06	0	0
track record	5.06	0	0
bridge the gap	4.22	1.01	0
on one hand	4.22	2.02	1.3
bear in mind	4.22	0	3.9
balance of power	4.22	0	0

on the face of it	3.37	0	0
the big(ger) picture	3.37	0	1.3
go hand in hand with	2.53	0	0
get to grips with	2.53	0	0
in its own right	1.69	2.02	1.3
from scratch	1.69	3.04	0
driving force	1.69	1.01	5.2
come to light	1.69	1.01	0
at the end of the day	1.69	0	0
raison d'être	1.69	0	0
in the long run	0.84	0	0
in the short run	0.84	0	0
rule of thumb	0.84	0	0
one's fair share	0.84	0	0
the high point	0.84	0	0
behind the scenes	0.84	0	0
state of the art	0	0	2.6
the whole story	0	0	1.3
in the early days	0	1.01	0
gold standard	0	9.09	8.1
come into play	0	1.01	0

Table 2. Distribution and frequency of idioms across disciplines in the COPLUS EN corpus

The analysis of idiomatic expressions in the COPLUS EN corpus across the three ERC domains shows significant variations in their distribution, thus suggesting that idiom usage reflects the epistemological priorities and stylistic conventions of these disciplines. This detailed discussion not only explores the general meanings and functions of the idioms but also incorporates examples from the corpus, providing a comprehensive examination of how these expressions are embedded in academic writing. The analysis considers the frequency of idioms in different disciplines, their rhetorical roles as well as their alignment with the conventions of academic discourse, thus offering insights into why certain idioms are more prominent in specific fields as compared to others.

The idiom *on the other hand* serves as a significant marker of contrast, frequently employed to introduce alternative perspectives or arguments. Its prevalence in SH is particularly noteworthy, as it has a frequency of 47.22 pmt. In comparison, its usage is considerably less frequent in the fields of LS and PE, where it has a frequency of 36.43 pmt and 38.97 pmt, respectively. This distribution could arguably reflect the argument-driven nature inherent in SH disciplines such as history, philosophy and linguistics, where the ability to balance multiple perspectives is often considered a core feature of scholarly writing. In contrast, the relatively lower frequency of *on the other hand* in LS and PE may be attributed to the linear, evidence-driven reasoning that these domains typically favour. Scholars in these areas often prioritise technical precision over explicit dialogic framing, which might explain why idiomatic expressions that signal contrast are less prevalent. While it is conceivable that scientists and engineers also engage in contrasting arguments, they seem to express this tendency through more formal or technical language, employing terms such as *however* or *alternatively*. This observation seems to suggest that the rhetorical strategies used in SH are distinctively different from those in LS and PE. In SH contexts, *on the other hand* frequently serves to introduce counterpoints or develop complex arguments, thus reflecting the emphasis placed on interpretative discourse within these academic fields, where exploring differing viewpoints is not only common but also encouraged. Moreover, the related idiom *on the one hand* is similarly concentrated within SH, where it has a frequency of 17.71 pmt, but only 1.01 pmt in LS and is absent altogether from PE. This imbalance may underscore a broader reliance within SH disciplines on explicitly structuring arguments using idiomatic pairings such as *on the one hand* and *on the other hand*. These expressions provide a rhetorical framework that facilitates the exploration of opposing viewpoints or comparisons between alternatives. Such a practice appears to be less common in LS and PE, where arguments tend to adhere more closely to a direct cause-and-effect logic. Likewise, the idiom *on the other [hand]* exhibits a comparable pattern; it has a frequency of 8.43 pmt in SH while being entirely absent from LS and PE. This complete absence in LS and PE could reflect a fundamental difference in how arguments are constructed across these disciplines. It implies that transitions between ideas are often made without the explicit use of idioms; instead, scholars in these fields may rely on data-driven or procedural framing to convey their points. While it may be tempting to generalise about the nature of argumentation across disciplines based solely on idiomatic usage, it is essential to recognise that these patterns are likely to arise from deeper epistemological differences. The distinct rhetorical strategies employed in SH versus LS and PE not only shape how arguments are framed but also influence how knowledge is constructed and communicated within these varied academic landscapes.

In the light of, another idiom with strong disciplinary associations, appears frequently in SH where it has a frequency of 14.33 pmt, but is considerably less common in LS with a frequency of 10.12 pmt and is entirely absent from PE. This idiom is typically employed to signal a shift in perspective or to reconsider a topic

in the context of new information or evidence. For instance, in SH contexts, it may introduce reevaluations of historical interpretations or philosophical arguments, thus aligning with the emphasis on the recontextualisation of ideas within these fields. Its limited use in LS may reflect the preference of these disciplines for more specific, empirical expressions of causality, such as *given that* or *because of*. The absence of this idiom in PE suggests that this domain favours technical or procedural alternatives that align more closely with the precise and mechanistic tone characteristic of engineering and physics writing. These patterns seem to indicate that the rhetorical choices made within each domain are influenced by their respective epistemological frameworks and communicative needs. In SH, the use of idioms like *in the light of* facilitates a more interpretative approach to knowledge construction, whereas LS and PE prioritise clarity and specificity in their discourse.

Bear in mind, a metadiscursive idiom used to caution readers or contextualise arguments, serves as another example of differential usage across disciplines. It has a frequency of 4.22 pmt in SH and 3.9 pmt in PE but it is totally absent in LS. This idiom encourages the audience to consider specific factors or limitations, often acting as a reminder of methodological constraints or broader implications. For instance, in a PE context, the idiom appears in the statement, "The trends measured may still be compared *bearing in mind* this additional uncertainty," where it signals an awareness of the uncertainty inherent in the data. Similarly, another PE example states, "We *bear in mind* that the modelled emissivity and reflected portion of DLR from the surface may also be contributing to model error," highlighting the need to consider potential sources of error when interpreting results. These instances reflect the methodological focus of these disciplines on precision and the importance of acknowledging potential limitations. The absence of *bear in mind* in LS might suggest that life scientists prefer more direct and quantified approaches to addressing uncertainty. Rather than highlighting such considerations rhetorically, they often integrate them into their methodological descriptions. Overall, these patterns indicate that the rhetorical choices made within each discipline are shaped by their respective epistemological frameworks and communicative needs, with SH employing idioms like *bear in mind* to engage readers more interpretatively, while LS and PE prioritise clarity and specificity in their discourse.

The idiomatic expression *come to light* exhibits a notable variation in usage across different academic disciplines. In SH, the idiom appears with the highest frequency at 1.69 pmt and this relatively high occurrence seems to suggest that the expression resonates well with the nature of research and discourse in social sciences and humanities. The metaphorical allusion of this idiom to the revelation of hidden truths aligns with the focus of the field on uncovering complex social phenomena, historical facts, and human behaviours. LS demonstrates a lower, yet still present, usage of the idiom at 1.01 pmt. This reduced frequency might reflect the more empirical nature of life sciences where discoveries are often described in more precise, technical terms. However, the presence of the idiom suggests that it still holds relevance in conveying the emergence of new scientific findings. An

example from the LS corpus provides valuable insight into the contextual use of this idiom: "The effect of both monovalent and divalent cationic salt concentrations could be explored at a later date, if there was sufficient scientific interest to do so, or if new evidence regarding the intracellular conditions *came to light*." In this instance, *came to light* is employed to express the potential emergence of new scientific evidence. The use of this idiom here serves multiple purposes: 1) it acknowledges the dynamic nature of scientific inquiry where new findings can emerge unexpectedly, 2) it emphasises the importance of evidence-based research in life sciences, and 3) it suggests a degree of uncertainty and openness to future discoveries, which is crucial in scientific discourse. Intriguingly, the idiom is entirely absent in PE. This absence could indicate a preference for more direct, unambiguous language in a field where precision and technical accuracy are paramount.

Conversely, *state of the art* is exclusive to PE with a frequency of 2.6 pmt. This idiom refers to the highest level of development achieved in a particular field, device, technique or scientific area at a specific time. It is often employed to describe technologies, methodologies or systems that incorporate the most advanced features and innovations available. For example, one might say, "The exhibition showcased the *state of the art* in robotics," indicating that the technologies displayed represent the pinnacle of current capabilities and standards. In the context of the PE domain, the use of this idiom reflects a pronounced focus on technological innovation and advancement. One instance drawn from the corpus illustrates this well: "A dedicated ZYNQ-7000 signal processing board acts as an intermediary between a *state of the art* camera and an Operating System (OS)." This usage not only highlights the integration of advanced technology within engineering applications but also suggests a broader narrative about the significance of contemporary technological frameworks in this discipline. Similarly, the other example from the corpus states: "We have explored a lot of the *state of the art* systems for adaptive control." Here, the idiom highlights the engagement with current technological benchmarks, thus reinforcing its relevance in discussions centred around innovation and efficacy. Nevertheless, it is important to consider that while this idiom does not appear in SH and LS, this absence may not necessarily imply a lack of relevance or engagement with technological concepts within those fields. Instead, it could indicate that these disciplines employ alternative expressions or frameworks that are more aligned with their specific emphases. For instance, SH might prioritise theoretical constructs and critical analyses over technological terminology, whereas LS may focus on biological or empirical discoveries rather than on cutting-edge tools. The limited occurrences of this idiom in PE suggest that while it highlights a certain aspect of technological discourse within that domain, these findings should be interpreted with caution. The small number of instances does not comprehensively capture the full spectrum of idiomatic usage within PE or serve as a definitive representation of disciplinary priorities. A broader analysis may reveal other idiomatic expressions that fulfil

similar roles in articulating advancements or methodologies across various academic contexts.

Another idiom that is exclusive to PE is *the whole story*, which, although rare, carries significant implications in the contexts in which it is used. This idiom typically appears in discussions that acknowledge the limitations of a particular theory or explanation, thereby suggesting an awareness of the complexities inherent in scientific inquiry. The only instance found in the corpus reads as follows: "Though the band theory that is normally used to study condensed matter systems has existed since the very origins of quantum mechanics, it does not tell *the whole story*." This usage reflects a recognition within the PE domain of the necessity for a holistic understanding, even when operating within highly specialised theoretical frameworks. The phrase *the whole story* implies that while certain theories may provide valuable insights, they may also fall short of capturing the entirety of a phenomenon. In this context, it highlights the importance of considering multiple perspectives and additional factors that contribute to a comprehensive understanding of complex systems. Such an acknowledgment is particularly relevant in fields like engineering and physics, where multifaceted interactions often necessitate a broader analytical lens. However, the rarity of this idiom within the corpus suggests that expressions invoking narrative completeness may be less suited to the fragmented and modular nature characteristic of engineering and physics research. The emphasis in these disciplines often lies on precise measurements, specific methodologies as well as compartmentalised studies rather than on overarching narratives or holistic explanations. Consequently, while *the whole story* serves as a meaningful expression when discussing theoretical limitations, its infrequent occurrence may indicate a preference for more technical or specialised language that aligns with the analytical demands of PE. This observation raises important questions about how idiomatic expressions function within different academic contexts. The limited use of *the whole story* could suggest that researchers in PE might prioritise clarity and specificity over narrative cohesion when articulating their findings. As such, it may be worthwhile to explore whether other idiomatic expressions related to completeness or wholeness are similarly underrepresented in this domain.

The idiom *gold standard* also reveals disciplinary specificity with a frequency of 9.09 pmt in LS and 8.1 pmt in PE, but completely absent in SH. This idiom denotes a benchmark or ideal standard and is commonly applied in LS to refer to methods or practices against which others are measured. Its prominence in LS aligns with the reliance of this academic domain on validated, evidence-based methodologies, which are critical to ensure the credibility and reliability of research findings. For instance, references to the *gold standard* in contexts such as experimental design or diagnostic techniques underscore the rigorous evaluation processes that are inherent to the field. In LS, the term is often employed to signify methodologies that have been thoroughly tested and established as the most effective or reliable options available. This usage reflects a commitment to maintaining high standards of

scientific integrity and precision, which are paramount in life sciences research. In PE, the presence of *gold standard* similarly reflects a focus on achieving technical excellence, although its application appears to be less metaphorically flexible compared to its use in LS. While PE does use this term to denote high-quality engineering practices or technologies, it may not encompass the same breadth of metaphorical meaning that it holds within LS. In this context, *gold standard* may be more narrowly focused on specific technical benchmarks rather than encompassing a wider range of applications or methodologies. The absence of *gold standard* in SH seems to suggest a different orientation within that discipline, where the emphasis may be more on theoretical constructs or qualitative analyses rather than on establishing empirical benchmarks. This divergence could indicate that idiomatic expressions related to standards and benchmarks are less relevant or applicable in fields that prioritise narrative and critical discourse over empirical validation.

The idioms analysed in the COPLUS EN corpus reveal how disciplinary conventions shape idiomatic usage. The rhetorical, epistemological and stylistic priorities of SH, LS and PE determine the frequency and context of these expressions, illustrating broader differences in how knowledge is constructed and communicated across academic domains. This observation aligns with previous research on disciplinary writing, which has shown that argumentation structures and linguistic choices vary significantly depending on the field of study (Walková & Bradford, 2022).

The examples provided further contextualise these patterns, thus showing how idioms serve as both linguistic tools and reflections of disciplinary norms. This analysis underscores the importance of idiomatic choices in academic writing, highlighting their role in aligning language with the objectives and expectations of each discipline. By understanding how idiomatic expressions are employed differently across these scientific domains, scholars can gain deeper insights into the linguistic practices that underpin academic discourse. Such comprehension is essential to foster effective communication within and across disciplines, ultimately contributing to a more refined appreciation of how language shapes knowledge construction in diverse academic contexts.

5. PEDAGOGICAL IMPLICATIONS

The findings of this corpus-based investigation offer critical insights into the pedagogy of EAP, particularly in relation to the underexplored yet significant role of idiomatic language in academic discourse. The strategic use of idiomatic forms by native English-speaking academic writers, across disciplines and levels of study, suggests that such expressions are not peripheral but integral to effective academic communication. This challenges prevailing pedagogical assumptions and necessitates a reevaluation of how idiomatic competence is addressed within EAP instruction.

A primary implication is the need to reconceptualise vocabulary instruction in EAP curricula. Conventional methods often emphasise discipline-specific terminology, academic word lists, and formal registers, while relegating idioms to the periphery as informal or stylistically unsuitable. However, corpus evidence reveals that idiomatic language plays a significant role in academic writing. Consequently, EAP practitioners should broaden the scope of lexical teaching to include idioms, not as trivial embellishments, but as essential linguistic tools that enhance clarity, emphasis, and authorial stance.

In this context, the adoption of DDL methodologies offers considerable potential. By exposing students to concordance lines and authentic examples derived from academic corpora, educators can promote inductive learning of idioms within their natural contexts. Such empirical engagement allows learners to identify usage patterns and contextual norms, thereby enhancing both their comprehension and productive skills. However, employing DDL requires learners to have a certain level of independence and metalinguistic awareness, which may not always be present. This underscores the importance of providing guided instruction and designing scaffolded corpus-based activities.

Another important pedagogical aspect concerns the communicative functions that idiomatic forms fulfil. The findings demonstrate that idiomatic expressions often enhance argumentative flow, establish interpersonal relationships, and convey epistemic positioning. Yet these roles are seldom explicitly addressed in classroom teaching. EAP educators should therefore incorporate critical discourse analysis into their lessons, encouraging students to explore how idioms are used to construct persuasive arguments and assert credibility. This might involve activities such as annotating texts, conducting genre-based analyses, and comparing how idioms function across various academic genres and registers.

Moreover, the research reveals notable disciplinary variation in idiom usage, reinforcing the need for context-sensitive pedagogy. A one-size-fits-all approach to idiom instruction is insufficient; instead, teaching should be tailored to reflect the linguistic realities of specific academic domains. This can be achieved through the use of discipline-specific corpora and collaboration with subject specialists, enabling students to acquire idioms that are both frequent and functionally salient within their fields. Such alignment ensures that idiom instruction is relevant, purposeful and immediately applicable to students' academic contexts.

Equally critical is the recognition of the challenges idiomatic language poses for nonnative English speakers. Idioms often defy literal interpretation, are culturally embedded, and can obscure meaning if unfamiliar. Therefore, explicit instruction in idiomatic meaning, form, and function is essential. Educators should provide glossaries, contextualised usage examples, and decoding strategies, while also fostering students' critical awareness of when idiom use is appropriate or potentially problematic. However, care must be taken not to encourage mechanical reproduction; instead, students should be empowered to use idioms judiciously, guided by considerations of audience, register, and communicative intent.

In conclusion, the findings of this study highlight the need for a revised approach to EAP instruction that acknowledges the complexity and contextual depth of idiomatic expressions. By integrating idiom teaching into corpus-informed, rhetorically aware, and discipline-specific frameworks, educators can more effectively prepare learners to meet the sophisticated demands of academic discourse.

6. CONCLUDING REMARKS

This study provided a detailed examination of idiomatic expression usage in academic writing, focusing on native English speakers across various academic levels and disciplines. By investigating the role of idioms in scholarly communication, this research sought to address a gap in the existing literature by offering insights into the linguistic strategies employed in formal academic texts. Furthermore, these findings may have practical implications for teaching academic writing styles and features, particularly to second language (L2) learners. Observing best practices reflected in authentic materials produced by native English speakers could assist L2 students in becoming familiar with idiomatic usage, thereby fostering deeper understanding of scholarly conventions.

Nonetheless, this study is not without its limitations. Its scope is restricted to native English speakers, which limits the generalisation of the findings to nonnative academic writers, who may exhibit distinctive patterns of idiomatic usage influenced by linguistic and cultural factors. Furthermore, the analysis is based on two specific corpora, namely CAE and COPLUS EN, which represent particular academic levels. While these provide valuable insights, their size and disciplinary focus may not fully encapsulate the wider spectrum of idiomatic variability across global academic contexts. Certain idiomatic expressions were rare or absent in the data, potentially reflecting corpus-specific biases rather than general trends.

Future research should seek to address these limitations and build upon the findings in several ways. Expanding the demographic scope to include nonnative English speakers and multilingual academic writers could uncover a broader range of idiomatic practices and reveal how idioms are adapted or avoided in multilingual academic contexts. Enlarging and diversifying the corpora to encompass additional academic levels, disciplines and cultural settings would provide a more comprehensive understanding of idiomatic variation. Furthermore, an in-depth examination of discipline-specific idiomatic usage, especially in underrepresented fields, could refine pedagogical recommendations and support more targeted instruction in EAP. Comparing the same idioms across spoken and written contexts, while accounting for proficiency level differences and disciplinary variations, could also prove valuable in uncovering how idiomatic usage shifts between modes of communication.

Finally, empirical studies exploring the impact of idiom-focused teaching interventions could yield practical strategies to enhance students' idiomatic competence, improving their ability to engage effectively with scholarly conventions and communicate their ideas with greater clarity and precision. By identifying effective instructional methods and materials, such research could provide valuable insights for educators aiming to support students in mastering the complexities of academic language. This, in turn, would facilitate more effective participation in academic discourse and contribute to students' overall success in their scholarly pursuits.

[Paper submitted 28 Nov 2024]
[Revised version received 25 Mar 2025]
[Revised version accepted for publication 11 Apr 2025]

References

Ackermann, K., & Chen, Y.-H. (2013). Developing the academic collocation list (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

de Chazal, E. M., & Moore, J. (2013). *Oxford EAP: A course in English for academic purposes*. Oxford University Press.

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

Grant, L. E. (2007). In a manner of speaking: Assessing frequent spoken figurative idioms to assist ESL/EFL teachers. *System*, *35*(2), 169–181. https://doi.org/10.1016/j.system.2006.05.004

Grant, L. E., & 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

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

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly*, *37*(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*(1), 25–35. https://doi.org/10.1016/j.esp.2011.07.002

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

Marín-Pérez, M. J., & Aguado Jiménez, P. (2024). DDL learning in ESP teaching: An exploratory study on multi-word unit usage in business letters. *ESP Today*, *12*(2), 245–268. https://doi.org/10.18485/esptoday.2024.12.2.3

Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, *67*(2), 184–198. https://doi.org/10.1093/elt/ccs100

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320. https://doi.org/10.1093/applin/ams010

McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge University Press.

Miller, J. (2020). The bottom line: Are idioms used in English academic speech and writing? *Journal of English for Academic Purposes*, *43*, Article 100810. https://doi.org/10.1016/j.jeap.2019.100810
Moon, R. (1998). *Fixed expressions and idioms in English*. Clarendon Press.
Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.
O'Dell, F., & McCarthy, M. (2010). *English idioms in use: Advanced*. Cambridge University Press.
Parizoska, J., & Rajh, I. (2017). Idiom variation in business English textbooks: A corpus-based study. *ESP Today*, *5*(1), 46–67. https://doi.org/10.18485/esptoday.2017.5.1.3
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (191–226). Longman.
Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
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
Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058
Strässler, J. (1982). *Idioms in English: A pragmatic analysis*. Gunter Narr.
Walková, M., & Bradford, J. (2022). Constructing an argument in academic writing across disciplines. *ESP Today*, *10*(1), 22–42. https://doi.org/10.18485/esptoday.2022.10.1.2
Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *The substance and structure of language* (23–81). University of California Press.
Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*(4), 213–231. https://doi.org/10.1017/S0261444800014154
Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489. http://dx.doi.org/10.1093/applin/21.4.463
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

VANESSA LEONARDI is Associate Professor of English Language and Translation at the Italian University of Sapienza in Rome, Italy. In 2004 she received her Ph.D. in Translation and Comparative Studies at the University of Leeds. She also holds a TEFL Diploma. Her research interests lie in the fields of Translation Studies, Gender Studies, Corpus Linguistics, ESP and EAP. Her recent publications include *Ideological Manipulation of Children's Literature Through Translation and Rewriting: Travelling Across Times and Places* (2020, Palgrave Macmillan) and *Legal Language Through the Ages: A Corpus-Based Investigation of Change and Continuity* (forthcoming). ORCID: 0000-0002-7857-7550