Research Journal in Translation, Literature, Linguistics, and Education (RJTLLE)
https://scientific-board.com/index.php/rjtlle

https://doi.org/10.64120/1d0rsa51

RESEARCH ARTICLE **Section:** *Language and Linguistics*

Published by **Scientific-Board LLC** in the Research Journal in Translation, Literature, Linguistics, and Education (RJTLLE). **Volume 1, Issue 4, 2025**.

ISSN: *3067-6290 (Online)*
3068-8450 (Print)

Article Details

Received: 01 August 2025
Revised: 30 August 2025
Accepted: 10 September2025
Published: 29 October 2025

Conflict of Interest: The author/s declared no conflict of interest.

Formulaic language in high-stake research writing: Investigating the semantic implications of collocations and fixed expressions in postgraduate dissertation

Naeem Fatima[1]* & **Sohail Ahmad**[2]
[1]Air University, Islamabad, Pakistan
[2]SSE English School Education Department (SED), Govt. of Punjab, Pakistan
https://orcid.org/0000-0001-8710-3237
Corresponding author's email: *ahmad.sohail664@gmail.com*

Abstract
The use of formulaic sequences and consistent combinations of words proves valuable in enhancing the sophistication, organisation, and connection to a specific field of academic writing. The goal of this study was to look into the semantic value, distribution, and ways of understanding formulaic sequences in postgraduate dissertations in Linguistics and Education by blending document analysis with the results of a questionnaire survey. Our approach included studying 113 postgraduate dissertations to find, classify, and examine formulaic patterns in different dissertation sections, as well as surveying 113 post graduate (PD) students to know their level of awareness, frequency, perceived challenges, and interpretations, which were then analysed using varying statistical and testing approaches. There were many instances of formulaic sequences found in the literature review (30.5%) and discussion (19.3%) sections, as well as a lot of vague and improper use of these formulaic structures, mostly in parts of the text that are meant to be argumentative or evaluative. This result was supported in the survey, where participants reported having a moderate understanding (M = 3.08) and facing difficulties when choosing the right order of actions (71.7%) and making precise language choices (68.1%) in the use of formulaic sequence. Additional analysis showed that students in Master's programmes were more aware than those in Doctoral programmes, while there were also major variations in how the fixed expressions were used, depending on the discipline ($p < 0.05$). By examining these data sets together, it was clear that researchers' awareness of pragmatic uses of language lagged behind their actual practises, requiring attention in language classes to specific formulaic language uses. The study concluded by suggesting that targeted phraseological practise, scaffolded guidance, and genre-targeted academic writing support should be part of postgraduate education. Furthermore, the paper recommended more in-depth research to learn more about the processes involved in using formulaic language in academic writing.

Keywords: Formulaic sequences, words proves, sophistication, organisation, academic writing

How to Cite:
Fatima, N., & Ahmad, S. (2025). Formulaic language in high-stake research writing: Investigating the semantic implications of collocations and fixed expressions in postgraduate dissertation. *Research Journal in Translation, Literature, Linguistics, and Education, 1*(4), 36-47. https://doi.org/10.64120/1d0rsa51

1. Study Background and Context

Research writing is a fundamental high-stake component of the global higher education system, especially in post doctorate programs. Academic dissertations invariably require scholars to use formulaic language in a well-developed manner. Using standard language sequences, connecting phrases, and certain fixed phrases contributes to coherence, easy reading, and making the argumentation stronger (Wray, 2002). With these devices, researchers present ideas in a clear and organised manner, making them crucial tools for creating academic arguments (Simpson-Vlach & Ellis, 2010). Advanced studies that lacked these features or include proper usage of the tools can weaken and even invalidate an academic text (Biber, Conrad, & Cortes, 2004).

Since dissertations have greater demands in both knowledge and communication, formulaic sequences become more important in postgraduate research writing. Such documents are a good example of writing that demands researchers to follow research rules and write in a strong academic way (Hyland, 2008). According to some studies, the majority of non-native English postgraduate writers find it difficult to handle these types of linguistic resources, which results in writing materials that are not rich in vocabulary or accurate in content and context (Chen & Baker, 2010; Pérez-Llantada, 2014; Al-Awawdeh et al., 2023). A critical search of the literature unveiled that there is limited research on how the semantic qualities of formulaic influence and guide how scholars respond appropriately to various situations during the documentation of their dissertations (Jiang & Nekrasova, 2007). Most studies in the field are centred on the number of words and word patterns (Cortes, 2004). Nesselhauf, 2005), often neglecting the nuanced semantic appropriations and pragmatic intentions that inform learners' usage.
As these gaps exist, the present study applies a triangulated mixed-method approach, analysing documents from postgraduate dissertations and giving postgraduate students in Linguistics and Education a structured survey questionnaire. This way, it helps to find commonly used expressions in authentic academic writings and critique researchers' opinions and attitude of using them (Hyland & Tse, 2007; Durrant & Schmitt, 2009; Al-Awawdeh et al., 2023). Analysing different aspects of writing and learners' views is essential to find out how their perceptions influence the use of formulaic sequences in high-stakes research writing.

Thus, the goal of this study is to collect and study how formulaic sequences, collocations, and fixed expressions are used in postgraduate dissertations. The intent of the study is also to look at how postgraduate writers use and understand specific linguistic constructions. The paper went on to outline certain factors that affect whether researchers rely on formulaic language or misuse it. The study provided suggestions for using these resources to enhance students' strategic skills during their studies.
The following research questions are posed to guide this paper:

a. What types and patterns of formulaic sequences, collocations, and fixed expressions are predominantly employed in postgraduate dissertations in Linguistics and Education?
b. How do postgraduate students in Linguistics and Education interpret the semantic appropriateness and contextual nuances of the formulaic sequences and collocations they use in their dissertations?
c. What are the perceived challenges and influencing factors that mediate the use, misuse, or underuse of formulaic sequences and fixed expressions by postgraduate students in their academic writing?

2. Review of Related Studies

2.1. Conceptualizing Formulaic Language, Collocations, and Fixed Expressions in Academic Writing

Formulaic language, made up of collections of words that are often used together, is necessary for effective writing in academics (Wray, 2002). Formulaic language is based on ready-made phrases that are stored in the mind and quickly used for speaking efficiently and with less effort (Schmitt & Carter, 2004). In this approach, a collocation occurs when certain words are very likely to appear together, whereas a fixed expression is an idiomatic or semi-fixed form whose meaning is sometimes easy to understand and

sometimes unclear (Nesselhauf, 2005; Martinez, 2013). At the same time, lexical bundles are common sequences of two or more words that regularly show up in a certain area of study or subject (Biber, Conrad, & Cortes, 2004).

In essence, such formulations play a greater role than only improving the vocabulary of texts. Instead, they are used to help clarify the writer's meaning, connexions between ideas, and their principles (Cortes, 2004; Paquot, 2010). According to Biber et al. (2004), these sentences demonstrate positioning in arguments and communicate points within a discipline. Hyland (2008) reiterates the importance of this language in helping researchers from diverse fields follow the research community's norms. Formulaic expressions make it easier for authors to keep their text coherent and organised, allowing them to create strong arguments (Altenberg & Granger, 2001).

Even so, the obscure semantics of typically used formulaic sequences may lead to misunderstandings, particularly for people writing in a non-native language (Wray, 2008). According to Chen and Baker (2010), several learners tend to overgeneralize, match words incorrectly, or understand sentences in a direct rather than accurate way, leading to poor understanding. Jiang and Nekrasova (2007) also acknowledged that second language (L2) writers tend to avoid formulaic language sequences, making it harder for them to write naturally and with ease. Pérez-Llantada (2014) also suggests that many students with English as an L2 language struggle to enhance the persuasive elements in their writing because they lack an adequate number of formulaic phrases needed for such effects (Al-Awawdeh et al., 2023).

In addition, since the same formulaic expressions commonly appear in research writing, the meanings assigned might depend on the genre, requirements of the discipline, and unique circumstances in the text (Simpson-Vlach & Ellis, 2010). Oftentimes, an inaccurate grasp of these small but powerful tricks by students leads them to use words or phrases that do not really match the point (Shin & Nation, 2008). Such discrepancies underscore the necessity for pedagogical interventions that not only emphasize the formal properties of these sequences but also cultivate learners' metapragmatic awareness of their contextual and semantic affordances (Ahn, 2012; Paquot, 2010).

2.2. A Review of Empirical Studies

In the last decades, there has been a growing interest in the use of formulaic sequences in important writing assignments (Hyland & Tse, 2007). In studying dissertations, Hyland (2008) points out that many students frequently use the same words and patterns to organise their thoughts, make their arguments clear, map out their stances, and connect with others' ideas. By studying postgraduate writers' work, he observed clear patterns in how they use certain expressions to build their stand and authority in the specific field. Furthermore, Pérez-Llantada (2014) looked into the way expert and novice writers use routine phrases in academia, finding that L1 expert writers are more skilled at using them and can control their use strategically.

According to Biber et al. (2004), non-native academic writers use lexical bundles differently and far less often compared to those who are native. It was established that L1 writers use more formulaic expressions and use them in a way that fits the context well. Also, in constructing the Academic Formulaic sequences List (AFL), Simpson-Vlach and Ellis (2010) mentioned how essential these sequences are in academic writing and how they are challenging for L2 learners to master. In their view, education programmes should be designed to improve this language gap.

Researchers working in education have found that explicit teaching of set phrases in EAP can help students overcome the same basic challenges. Ahn (2012) demonstrated that systematic instruction of collocations and formulaic sequences significantly improved learners' written fluency and lexical sophistication, albeit with varying degrees of retention and transferability. Arifin et al. (2025) further extended this inquiry by exploring the potential of AI-assisted models in personalizing academic writing instruction, positing that such technologies could facilitate individualized feedback on formulaic

language usage and enhance learners' metalinguistic awareness. Nonetheless, these interventions have largely centred on form-focused instruction, with limited emphasis on fostering learners' interpretive and contextualized use of these sequences within authentic academic texts.

Different studies have dealt with formulaic speech using both longitudinal and corpus-based methods. Li and Schmitt (2009) did a study over time that showed learning phrases in academic writing takes a prolonged and can be irregular. Laufer and Waldman (2011) showed, by analysing learner writing, that learners routinely have issues using verb-noun collocations in a useful and appropriate style. In short, these findings point out the various issues that come with using formulaic sequences in important academic work.

Despite these scholarly contributions, it is pertinent to note that much of the extant literature has adopted predominantly quantitative, corpus-driven methodologies, often eschewing qualitative insights into learners' subjective experiences, interpretative strategies, and perceptions regarding their use of formulaic sequences (Martinez, 2013). This methodological orientation has inadvertently marginalized the learner's voice, thus rendering the cognitive and metapragmatic dimensions of formulaic language use in academic writing relatively underexplored. In view of this, using a mixed-method approach that connects analysis of documents with learners' ideas about their skills can better describe this phenomenon (Revier, 2014).

2.3. Gaps in the Literature

A careful examination of the current literature reveals that not many studies have looked into the meanings and contexts of formulaic language in postgraduate dissertations. While corpus-based research describes how words often combine in general, its analysis at this level fails to reveal much about the contextual and pragmatic aspects (Durrant & Schmitt, 2009; Pérez-Llantada, 2014). The lack of information on how writers at this level perceive and use these phrases with respect to their discipline is what prompted this research (Jiang & Nekrasova, 2007; Simpson-Vlach & Ellis, 2010).

Essentially, research on formulaic sequences mostly overlooks learners' personal perceptions and usage patterns, especially in analysing written dissertations where the importance of language is much greater (Li & Schmitt, 2009; Hyland, 2008). Studies in this area rarely look at what affects learners' decisions and strategies, focusing mainly on analysing texts (Kalsoom et al., 2025). For this reason, the different ways learners apply formulaic sequences in their writing still need further study.

Additionally, there are few research studies that include document analysis and student opinions when examining academic writing (Laufer & Waldman, 2011; Revier, 2014). Existing corpus studies, while illuminating in their quantitative rigor, often fail to capture the interpretative processes and contextual reasoning that underpin learners' usage patterns, thereby presenting a partial and, at times, reductionist account of formulaic language use.

3. Methodology

3.1. Research Design

This study involved mixture of data from document analysis and survey of PG students from the selected universities. Through this method, the study explored how frequently formulaic structures occur in post graduate dissertations. The use of two methods assisted in the gathering of overwhelming evidence of the students' writing capacity and creative ingenuity in the use of formulaic structures in dissertations. For the survey, the researchers gained insights into the participants' understanding and perception of formulaic sequences and considered the practical usage of these expressions in their dissertations (Jiang & Nekrasova, 2007). The study triangulated the results from the both methods which further validated the credibility of the findings in this paper.

3.2. Participants

For the survey, a total of 113 PG students in Linguistics and Education departments at various universities to participate in the questionnaire. All the PG students that were engaged in the study had the necessary research knowledge about the use of formulaic sequence in academic writing, although their language use is constrained as second language users. Only the students who are currently working on their dissertations participated in the survey.

3.3. Data Collection Methods

3.3.1. Document Analysis

In the study, the researchers analysed the selected dissertations of the participants to collect the case information. All the information was closely analysed to spot any aptly used formulaic sequences, widely used words together, or common phrases in the students' writing. Afterward, the analysis examined the suitability of each sequence based on what it meant and how it was used in the dissertations.

3.3.2. Survey

In addition to studying the documents, participants were surveyed on how comfortable they felt with using formulaic vocabulary and collocations in their academic essays. Survey participants responded using multiple-choice questions for the required information and were able to provide more details using some of the additional questions (Jiang & Nekrasova, 2007). The tool was designed following the models in previous studies on phraseological awareness (Martinez et al., 2013), so it would be effective and accurate for the project. The researchers investigated the amount and frequency of these expressions in dissertations and how well writers can use them for various purposes.

3.4. Data Analysis

The information discovered in the document analysis was sorted based on different taxonomies, categories, and whether it fit the current surrounding. Data was analysed with use of frequencies, percentages, and mean scores to find out how the participants use formulaic sequences, their general understanding, and their difficulties. Besides using descriptive statistics, chi-square tests were used to identify any links between the study participants' educational features and their behaviours and attitude.

4. Results and Discussions

4.1. Results

This section shows how data from two approaches, documents and surveys, was analysed and interpreted using mixed methods and triangulation. It aimed to recognise the common sequences, how they are used, what their meanings are, and when they are rightly used in postgraduate dissertations, together with their authors' experiences and views. In the early stages, the demographics of the survey participants are studied since this helps give a context to the findings that follow.

Table 1: Demographic Characteristics of Participants (N=113)

Demographic Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	48	42.5%
	Female	65	57.5%
Level of Study	Master's	71	62.8%
	Doctoral	42	37.2%
Field of Study	Linguistics	64	56.6%
	Education	49	43.4%

The majority of participants in this study were female (57.5%), indicating some gender imbalance when

it comes to the selection of faculties. Most respondents (62.8%) were doing their Master's degrees, indicating the early-career focus of advanced studies and the nature of their high-stakes academic writing. The results suggest that Linguistics takes up the greater portion (56.6%) in this study, with Education slightly trailing at 43.4%. This context therefore offers important insights into how linguistic competence, rules of each discipline, and formulaic language are used in postgraduate texts.

4.1.1. Analysis of Formulaic Sequences in Postgraduate Dissertations

The following section shows the organised analysis of formulaic sequences, collocations, and fixed expressions found in the studied postgraduate dissertations (N=113). The data were coded qualitatively, the different expressions were counted, and statistics were used to show how these sequences are used, where, and for what purpose in various subjects and types of writing. The findings in academic English dissertations are divided by categories that match the major topics found in dissertations, which helps to reveal how this type of writing uses certain language.

Table 2: Distribution of Formulaic Sequences Across Dissertation Sections

Section	Total Sequences Identified	Percentage (%)	Mean per Dissertation	Standard Deviation
Introduction	812	18.4%	7.19	1.86
Literature Review	1,345	30.5%	11.9	2.43
Methodology	621	14.1%	5.5	1.73
Results	417	9.5%	3.7	1.12
Discussion	851	19.3%	7.53	1.95
Conclusion	439	9.9%	3.88	1.27
Total	4,485	100%	-	-

From the information in Table 2, it can be seen that most of the formulaic sequences in this paper occur in the Literature Review section (30.5%), with the Discussion and Introduction sections having the next highest percentages (19.3% and 18.4% respectively). This suggests that using formulaic sequences is essential for supporting the process of bringing together information, arguing an idea, and contextualising a study—the main types of rhetoric in these works (Hyland, 2008). Cortes, 2004). The reason for this could be that Results and Conclusion sections are mainly pushed to share facts and summaries in scientific researches, which requires less use of formulaic sequences compared to Introduction and literature reviews (Pérez-Llantada, 2014).

Table 3: Categorization of Formulaic Sequences by Function and Semantic Focus

Functional Category	Frequency (n)	Percentage (%)
Introducing Research (e.g., "The aim of this study is")	402	9.0%
Reviewing Literature (e.g., "Previous studies have shown")	1,131	25.2%
Methodological Procedures (e.g., "The data were analyzed using")	597	13.3%
Presenting Results (e.g., "The findings indicate that")	415	9.3%
Argumentation & Discussion (e.g., "It can be argued that")	985	22.0%
Stating Implications (e.g., "The study highlights the need for")	370	8.2%
Conclusion & Summary (e.g., "In conclusion")	278	6.2%
Total	4,178	100%

The numbers shown in Table 3 prove that formulaic sequences are essential in handling literatures reviews, arguments, and discussion (Biber et al., 2004; Hyland, 2008). Fewer formulaic expressions in the Conclusion and Summary category (6.2%) could mean that postgraduate writers do not use them as widely when they synthesise their ideas and describe their findings. This agrees with what earlier studies

say about novice academic writers lacking this type of phraseology in those sections (Nesselhauf, 2005).

Table 4: Top 15 Most Frequently Used Formulaic Sequences and Their Misuse Contexts

Rank	Formulaic Sequence	Frequency (n)	Correct Usage (%)	Misuse Context Example
1	*According to the literature*	298	89.3%	Misused without citations
2	*The aim of this study is to*	274	93.1%	Vague or circular aim statements
3	*Previous studies have shown*	263	86.7%	Lack of direct evidence
4	*The findings indicate that*	251	84.5%	Overgeneralization
5	*Based on the data collected*	233	88.4%	Inferences not supported by data
6	*It can be argued that*	221	81.2%	Unsupported claims
7	*In line with previous research*	213	90.7%	Misrepresentation of studies
8	*The results suggest that*	202	85.1%	Overstated implications
9	*The data were analyzed using*	196	92.0%	Lack of method clarity
10	*It is important to note that*	188	80.3%	Redundant or filler
11	*The participants were selected*	173	94.2%	Incomplete sampling details
12	*As a result of*	168	86.9%	Misleading causality
13	*The study contributes to*	157	88.5%	Unclear contribution
14	*The study highlights the need for*	149	84.2%	Vague recommendations
15	*It is widely accepted that*	146	78.6%	Presenting debatable claims as facts

Table 4 highlights that "According to the literature" was used the most (n=298) out of all the common formulaic patterns found in the corpus. Yet, some mistakes were found, mainly exhibited in using "It can be argued that," "The aim of this study is to," and "The findings indicate that," which reflected the habit of writers to repeat these moves without thinking about the evidence or context (Chen & Baker, 2010). Jiang & Nekrasova, 2007). This supports the previous idea that L2 writers often stick to using the same language patterns, even when the task requires them to think at a deeper level (Li and Schmitt, 2009).

Table 5: Cross-tabulation of Formulaic Sequences Usage by Discipline

Formulaic Category	Linguistics (n=64)	Education (n=49)	χ^2 (Chi-square)	p-value
Literature Reviewing	684 (60.5%)	447 (39.5%)	5.43	0.019*
Argumentation & Discussion	558 (56.6%)	427 (43.4%)	3.92	0.048*
Methodological Procedures	330 (55.3%)	267 (44.7%)	1.83	0.176
Conclusion & Summary	153 (55.0%)	125 (45.0%)	1.72	0.190

There is a significant difference ($p < 0.05$) in using formulaic sequences for literature reviewing and argumentation between Linguistics and Education dissertations, suggesting that each discipline has its own pattern in using specific phrases. It is likely that these differences are due to the varying epistemological approaches and genres in the two fields. Differently from what is typically found in the humanities, Linguistics dissertations tend to structure literature and build arguments using much more formulaic language (e.g., Hyland, 2008). Pérez-Llantada, 2014). There were no statistically noticeable differences in how methodological or concluding formulaic sequences were used, which points to a convergence in both fields.

4.1.2. Survey Data Analysis

This part of the study looks at the survey results obtained from the 113 postgraduate students. Based on the survey, the study identified participants' level of awareness, their frequency of using these phrases, any challenges experienced, and what the phrases actually represent, within their school writing practises.

In analysing, descriptive statistics (frequencies, percentages, means, standard deviations) and inferential statistics (Chi-square tests, Independent Samples t-test, ANOVA) were used in suitable cases to study noticeable relationships and differences among demographic factors.

Table 6: Participants' Self-Reported Awareness of Formulaic Sequences

Awareness Level	Frequency (n)	Percentage (%)	Mean (Likert 1-5)	SD
Very High	15	13.3%		
High	33	29.2%		
Moderate	40	35.4%		
Low	19	16.8%		
Very Low	6	5.3%		
Overall Mean Awareness Score	-	-	3.08	0.91

Based on Table 6, it is clear that respondents, on the whole, report moderate knowledge of formulaic sequences (M = 3.08, SD = 0.91), while only 13.3% claim very high awareness and the rest report a moderate level. From this, it can be concluded that postgraduate students are less aware of metalinguistic aspects, which is very important since such sequences play a central role in important academic writing tasks (Simpson-Vlach & Ellis, 2010). This means some students are more aware than others, making it appropriate for teachers to use various educational strategies.

Table 7: Self-Reported Frequency of Use of Formulaic Sequences in Dissertation Sections

Dissertation Section	Mean Frequency (1-5 Likert)	SD
Introduction	4.1	0.83
Literature Review	4.5	0.72
Methodology	3.8	0.89
Results	3.2	1.02
Discussion	4.3	0.78
Conclusion	3.5	0.97

From Table 7, it is clear that the highest percentage of formulaic sequence use was seen in both the Literature Review (M = 4.5, SD = 0.72) and the Discussion sections (M = 4.3, SD = 0.78). The usage of formulaic sequences here is in line with what scientists show in a number of studies and confirms the strong preference for them in writing that requires putting together ideas and analysing them (Hyland, 2008). Alternatively, the Results section (M = 3.2, SD = 1.02) was used the least frequently, which may be related to the fact that many reports in this section rely on detailed description, where formulaic density often decreases (Pérez-Llantada, 2014).

Table 8: Participants' Perceived Difficulties in Using Formulaic Sequences (Multiple Responses Allowed)

Perceived Difficulty	Frequency (n)	Percentage (%)
Difficulty identifying contextually appropriate sequences	81	71.7%
Confusion over semantic precision	77	68.1%
Overgeneralization of sequences	64	56.6%
Uncertainty over disciplinary conventions	59	52.2%
Overreliance on basic sequences	55	48.7%

What the data in Table 8 show is that participants identified contextual suitability (71.7%) and the right use of the words (68.1%) as their biggest difficulties, in line with earlier studies that found L2

postgraduates often struggle with the precise forms of formulaic sequences in their writing (Chen & Baker, 2010; Jiang & Nekrasova, 2007). Such findings point out the importance of providing ways to guide learners in recognising the unique usage and subtle meanings of particular genres.

Table 9: Inferential Analysis—Differences in Awareness by Level of Study (Independent Samples t-test)

Group	N	Mean Awareness Score	SD	t	p-value
Master's	71	2.98	0.87	-2.67	0.009*
Doctoral	42	3.31	0.94		

It can be seen from Table 9 that Doctoral students had significantly more awareness of formulaic sequences than Master's students (t = -2.67, p = 0.009). This agrees with the idea that benefiting from academic lessons and practise at writing in school gradually makes learners more aware of language and its ways of use (Li & Schmitt, 2009). Hyland & Tse, 2007).

4.2. Discussion

The results of this study give useful knowledge about how phrases, collocations, and fixed expressions are used, understood, and interpreted in postgraduate dissertations in the field of Linguistics and Education. When looking at the combination of document and survey results, we discovered that participants handle and view these resources differently sometimes, but other times in similar ways. The paper centres on the study's research questions, with a main focus on how text resources and learners' perspectives are connected, resulting in a better overview of formulaic language in academic writing (Hyland & Tse, 2007; Simpson-Vlach & Ellis, 2010).

The analysis of the document revealed that one-third of the Literature Review and nearly one-fifth of the Discussion consists of formulaic sequences, underlining their importance in gathering other research, placing key ideas, and dealing with the main issues of students' disciplines (Hyland, 2008; Biber et al., 2004). Participants reported that they rely more on formulaic expressions in the literature review (4.5) and discussion (4.3) sections. It means these students understand the style they should use, since they have read many texts and take a positive approach towards them (Pérez-Llantada, 2014; Cortes, 2004). It can be seen that phrases such as "It can be argued that" and "The findings indicate that" appeared in documents without supporting evidence and thinking, pointing to an automatic and meaningless habit (Chen & Baker, 2010; Jiang & Nekrasova, 2007). Overall, 71.7% of people participating in the survey mentioned trouble recognising proper sentence patterns, and another 68.1% expressed difficulty with semantic accuracy. This disconnection between awareness and effective application underscores enduring challenges in learners' metapragmatic competence, as previously noted by Laufer and Waldman (2011) and Li and Schmitt (2009).

The first research question was addressed by the study, which found that phrases and formulaic sequences are distributed unevenly in certain topics and have specific ways of use, such as describing the literature, outlining the methods, and arguing the point. Even so, our qualitative analysis found that writers overly depended on a set of frequent patterns, using most of them in predictable ways without taking into account the rules of various types of writing (Simpson-Vlach & Ellis, 2010). There was a clear narrowing of phraseology in the Results and Conclusion sections, as both sections scored lower usage frequencies (M = 3.2 and M = 3.5 respectively), implying that writers do not use formulaic language much in parts that require the integration of results and highlighting their relevance.

Regarding the second research question, the findings revealed critical gaps in participants' semantic interpretations and contextual usages of formulaic sequences, especially among Master's students, who reported significantly lower awareness levels than their doctoral counterparts (t = -2.67, p = 0.009). Such findings indicate that there is a clear link between students' academic development and the way they learn and use phrases, which agrees with previous studies by Li and Schmitt (2009). The high

levels of standard deviation in awareness and frequency scores also point to the fact that students have differing skills, which requires special learning strategies for each one (Ahn, 2012; Kalsoom, Mujahid, & Khanam, 2025).

Participants in the study also reported, in triangulated analysis, that many of them find it difficult to know what kind of rules or phrases are common in different subjects. The challenge shown by students demonstrates the same outcome found in the document analysis, where they repeat old writing patterns into academic writing and use language in an unsuitable way. The results add support to Pérez-Llantada's (2014) statement that L2 academic writers have trouble grasping the particular phraseology used in their field, eventually affecting their writing by making it either word-light or lacking in certain social aspects.

Responding to the third question, the analysis found that participants said using language in the correct context (71.7%) and marking meaning clearly (68.1%) were their biggest challenges, just as Wray (2008) points out about how challenging it is for L2 writers to master the procedures involved in using language properly. It is noteworthy that a statistically significant difference in the usage of formulaic sequences could be found in literature reviewing and argumentation when comparing Linguistics and Education dissertations ($p < 0.05$). The result confirms that the field of study affects phraseological choices, similar to what Hyland observed in his work from 2008: academic disciplines shape their own way of communicating.

The convergent findings across data sets collectively underscore a complex interplay between learners' phraseological awareness, textual practices, and disciplinary contexts, suggesting that formulaic sequences continue to be both a linguistic resource and a source of struggle for postgraduate writers in high-stakes academic writing settings. The results support the idea that using formulaic language helps in forming an academic voice, making writing coherent, and developing authority (Biber et al., 2004; Hyland & Tse, 2007), they also expose enduring challenges in learners' semantic appropriacy, rhetorical flexibility, and critical engagement with such sequences, necessitating more targeted and discipline-sensitive instructional frameworks (Ahn, 2012; Arifin et al., 2025).

5. Conclusion and Recommendations

This study carefully analysed the use of formulaic sequences, collocations, and fixed expressions in Linguistics and Education postgraduate dissertations using a mixed-methods approach that worked with document data and survey responses. The results confirmed that formulaic expressions are common in both the literature review and discussion sections, and that there is a large number of inaccurate statements and inappropriate wording, largely seen in argumentative and outcome-based sections. In addition, many students demonstrated awareness of certain sequences, but since there was lots of ambiguous understanding and general use of these, there were clear gaps in students' metapragmatic abilities, especially among those who had finished Master's degrees.

Taking these findings into account, one can see that while using formulaic phrases is important for achieving successful writing, postgraduate writers in a second language often find it very challenging. As a result, the study suggests providing proactive, subject-oriented phraseological classes to postgraduate students, along with meaningful practise and supervision, to improve their understanding of both the meaning and practical uses of formulaic expressions.

It is suggested that future research address the study's shortcomings by using long-term, controlled experiments to analyse the growth and use of formulaic language skills by postgraduate students in different disciplines. Accordingly, adopting high-end corpus tools, along with detailed interviews, can help to better understand interactions between learners' thinking and their writing habits, and so benefit the teaching of academic writing.

Top of Form

Bottom of Form

References

Ab Manan, N. A. (2016). Formula instruction in English language writing: Learnability, teachability and learner variables. *International Journal of Applied Linguistics & English Literature, 5*(3), 1–9.

Aghbar, A. A. (1990). Fixed expressions in written texts: Implications for assessing writing sophistication. *ERIC Document Reproduction Service No. ED321486*.

Ahn, H. (2012). Teaching formulaic sequences in the EFL classroom: Focusing on collocations. *English Teaching, 67*(1), 47–73.

Al-Awawdeh, N., Al-shaboul, I. A., & Saleem Khasawneh, M. A. (2023). ADVANCING FOREIGN LANGUAGE TEACHING WITH AI-ASSISTED MODELS; INSIGHTS FROM LECTURERS AND UNIVERSITY ADMINISTRATORS. *Journal of Namibian Studies*, *34*.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*(2), 173–195.

Arifin, R. S. F., Supriyadi, A. A., Harmaen, D., & Muslim, A. (2025). How lecturers in foreign languages can utilize the innovativeness of AI models to improve students' performance through personalized assessment and customized learning systems. *Research Journal in Translation, Literature, Linguistics, and Education (RJTLLE)*, *1*(1), 11-23.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30–49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157–169.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching, 47*(2), 157–177.

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41–62.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235–253.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91*(3), 433–445.

Kalsoom, T., Mujahid, A. H., & Khanam, A. (2025). The correlation between foreign language undergraduates' usage of social media short videos as digital assistant tutors and increase in their reading and writing proficiency. *Research Journal in Translation, Literature, Linguistics, and Education (RJTLLE)*, *1*(1), 1-10.

Kang, J. Y. (2013). The use of formulaic sequences in academic writing: A case study of Korean EFL learners. *English Teaching, 68*(4), 3–23.

Laufer, B., & Waldman, T. (2011). Verb–noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647–672.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85–102.

Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal, 67*(2), 184–198.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Continuum.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes, 14*, 84–94.

Revier, R. L. (2014). Testing knowledge of whole English collocations available for use in written production. *Unpublished Ph.D. dissertation*, Aarhus University.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. *The European Journal of English Studies, 8*(2), 135–144.

Shin, Y. K., & Nation, I. S. P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal, 62*(4), 339–348.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulaic sequences list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512.

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

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.