



TWIST

Journal homepage: www.twistjournal.net

Lexico-Grammatical Features in MA Dissertation Introductions: A Corpus Analysis of Moroccan EFL Students' Writing

Abderrazzak Belbouah*

Language and Society, Faculty of Languages, Letters and Arts,
Ibn Tofail University, Kenitra, Morocco
[*Corresponding author]

Mounir Chibi

Language and Society, Faculty of Languages, Letters and Arts,
Ibn Tofail University, Kenitra, Morocco

Abstract

This study is an inductive, qualitative corpus analysis of MA dissertation introductions written by Moroccan students. Guided by the Systemic Functional Linguistics framework, it examines 105 texts from the academic years 2022–2023 and 2023–2024, focusing on three key lexico-grammatical components: lexical bundles, type-token ratio, and lexical density. The findings reveal low lexical density (41%–50%), limited lexical richness (≈ 0.170), and a dominance of the lexical bundle 'noun phrase + of,' indicating challenges in meeting academic writing conventions. The findings of this study provide pedagogical implications for EFL Moroccan instructors and learners to consider lexicogrammar as a key component in their instruction and learning processes.

Keywords

Lexicogrammar, EFL Writing, Corpus analysis, Applied linguistics, MA students

INTRODUCTION

Research on writing in English as a Foreign Language (EFL) contexts has long been a major focus of numerous studies. These studies have examined writing from various perspectives and methods, with the primary—if not sole—objective of improving the quality of EFL writing instruction and learning processes. Accordingly, studies have investigated writing from the scope of intercultural rhetoric (Connor, 1996; Yusal, 2008; Chibi, 2021), instructional approaches and pedagogy (Gabrielatos, 2002; Faraj, 2015; Jashari, 2019; Belbouah et al., 2023), and, last but not least, genre studies and rhetorical structures (Swales, 2004; Zheng, 2013; Belbouah & Bouguerba, 2025).

The persistent interest, or rather concern, of researchers in how EFL learners write highlights the challenges these learners face in producing written texts and identifies pertinent strategies to overcome them. However, the mediocrity of EFL writing is still widely documented in a plethora of research—across different levels, various fields, and diverse contexts. One such area is the writing of MA dissertations by applied linguistics students at Moroccan universities, which is the focus of this study.

This gap in the literature constitutes a key motive for conducting this study. Other motivations include the need to examine the lexico-grammatical patterns in students' academic writing, identify recurring difficulties in the use of academic language, and provide a database that future research can use to implement strategies for the improvement of students' writing. More specifically, by analyzing a corpus of students' dissertation introductions, this study aims to provide insights into their lexico-grammatical choices and offer recommendations for improving writing pedagogy in applied linguistics programs. Accordingly, the study investigates three lexico-grammatical components lexical bundles, Types-Token Ratio, and lexical density. The research question it attempts to answer is:

- How effectively do MA Moroccan applied linguistics students use lexicogrammar in their dissertation writing?
- By responding to this question, this study aims to highlight areas of lexicogrammar instruction and learning that Moroccan EFL instructors and learners should focus on for improvement.

LITERATURE REVIEW

The analysis of written texts, at least from within the perspective of Systemic Functional Linguistics, departs from and is guided by the view this perspective has about writing genres. Its focus is on the relationship language has with the context where it occurs (Royce, 2008), and on language as a means of communication in definite cultural situations (Hasan & Halliday, 1985). It is, Martin (1997) proposes, a social activity that is oriented towards the achievement of a goal.

Of equal importance to the work of SFL genre analysis are the strata of language. These are, according to Hjelmslev (1961), arranged under two components: expression and content. While the expression component includes “sounding (for phonology), wording (for lexicogrammar), and meaning (for discourse semantics), the content stratum is for “lexicogrammar (or grammar) and discourse semantics (or semantics).” The following figure represents these levels:

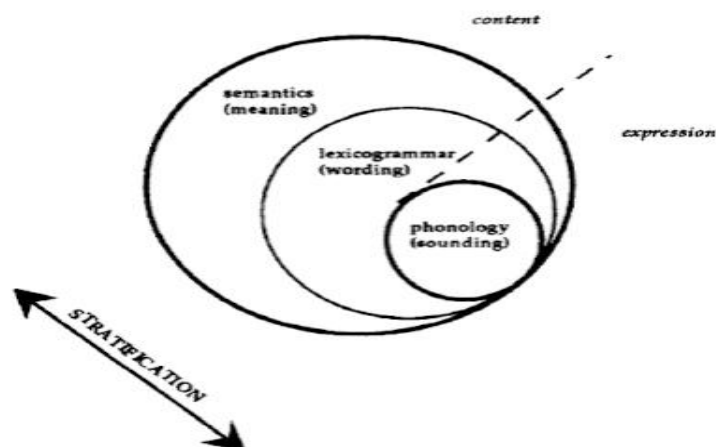


Fig. 1 Language as a tri-stratal system (Halliday & Matthiessen, 1999, p.5)

Halliday and Matthiessen (1999) note the representation of these levels in circles means they are interconnected and that they are all governed by the same organizational norms. In addition, this representation connotes lexicogrammar belongs to the sphere of semantics and simultaneously constitutes the sphere of phonology. In brief, their claim connotes phonology is a component of expression, and it is at this level that language takes its physical manifestation in the form of a phonic component. However, the focus here is on lexicogrammar and semantics because these constitute an important part of SFL genre analysis of texts written by both native and foreign learners of English. Also, the two strata constitute the core SFL analysis this paper conducts.

A basic definition of lexicogrammar would describe it as encompassing both the lexical and grammatical aspects of a text. However, SFL provides a deeper understanding, asserting that lexicogrammar represents the standard and natural use of both lexis and grammar within specific communication contexts, bridging the gap between sound and meaning (Halliday & Matthiessen, 2014). Another perspective defines lexis as an organized system of signs structuring a language's vocabulary, while grammar governs how these signs are organized within a text (Gledhill, 2011). Moreover, Gledhill emphasizes the interdependent nature of these two components, which collectively form a single stratum in language. It's worth noting that lexicogrammar plays a crucial role in maintaining language's potential for constant creativity (Egins, 1994).

According to Gledhill, an SFL genre analysis focuses on investigating lexico-grammatical patterns, which encompass various features including:

- Predictability and productivity of how signs can be sequenced;
- Possibility to be made of lexes that vary in the degree of abstractness;
- Ability to restructure and to integrate more patterns into continuous discourse;
- Extensionality along the parts of the same text.

These patterns can be analyzed at two levels. The first level, paradigmatic, refers to the connections between words, where one word can be substituted for another. The second level, syntagmatic, concerns the relationships between items in terms of how they are typically used together (Halliday & Webster, 2008, p. 63). Building upon these two levels, Sardinha (2019) identifies four combinations: “syntagmatic + lexical, syntagmatic + grammatical, paradigmatic + lexical, and paradigmatic + grammatical.” Sardinha enumerates several concepts within SFL genre analysis, including “collocations, colligation, phraseology, lexical pattern, chunk, lexical bundle, formulaic language, and lexical frame, among others.”

Numerous studies have explored the impact of implementing SFL GW pedagogy on the awareness of language learners regarding the writing patterns within their genre communities. Overall, these studies consistently emphasize the effectiveness of SFL writing pedagogy in enhancing learners' awareness of the language features prevalent in their writing communities. This heightened awareness empowers learners, regardless of their disciplines and levels of education, to generate high-quality written texts.

Sachiko Yasuda (2011) concludes SFL genre writing pedagogy can help students improve their awareness of their genre community, and simultaneously develop their ability and perception of how writing within their community operates. Yasuda's study surveys how Japanese fresh-year FL students enhance their understanding of genre, knowledge

of language, and writing competencies in a genre writing program focusing on email-writing. Along a period of fifteen weeks, the researcher tried to implement a study wherein the students' objective was to learn the targeted writing components through a series of well-planned genre tasks. The aim of these tasks was mainly to enable the learners to be aware of and master the ways "in which different genres are shaped by different linguistic resources" (p. 111). The study proposes the implementation of GW pedagogy because, as the author provides, it is key for the creation of a fundamental knot connecting writing as a social product to the available writing choices, especially in terms of language appropriateness.

In another study, Yasuda (2015) used the SFL approach to explore the development of foreign language (FL) students' rhetorical and lexico-grammatical choices in writing summaries. Through qualitative analysis, the research tracked the progress of 30 undergraduates over one semester by comparing the summaries they produced at the beginning of the study with those they wrote at its conclusion. The findings revealed a significant improvement in the students' awareness of the rhetorical requirements of summary writing across interpersonal, ideational, and textual dimensions. Additionally, participants demonstrated more genre-appropriate lexico-grammatical patterns. The study also noted that the extent of students' awareness change depended on their proficiency level, with those having higher proficiency in English exhibiting better progress. Overall, the author concluded the role of SFL pedagogy is of paramount importance in enhancing the learning of foreign language writing.

Another notable study is "The SFL genre-based approach to writing in EFL Contexts" by Akiko Nagao (2019). In this study, the author examined the development of the cognitive attributes of Japanese EFL learners concerning the lexico-grammatical patterns and standard organization of discussion essays. Nagao implemented "text-based writing lessons" and observed the learners' progress over a three-month course. The lessons were structured into four stages: (1) exploration of model texts, (2) writing of targeted essays, (3) analysis of peers' texts, and (4) production of a written analytical reflection. The findings indicated that teaching EFL writing using SFL pedagogy can effectively assist learners in enhancing their writing skills by fostering awareness of how texts within their communities are organized through the exploration of target texts' writing styles.

In a distinct study, Zhang (2019) investigated how SFL writing curricula can influence the assumptions of EFL learners regarding writing. The author conducted an analysis of interviews, written texts, and feedback from professors and classmates involving four Chinese university students. A key finding of the study was the students' successful navigation of various challenges, both within and outside the classroom, leading to a significant awareness of writing as a sense-making process. This shows the participants' ability to cultivate a new conceptualization of writing, diverging from their previous beliefs about writing in their native language. Consequently, the study concluded the utilization of SFL pedagogy enhances learners' efficacy in academic text composition.

In 2019, Forey published an interesting study wherein she emphasized the importance of the explicit teaching of linguistic patterns. The study's aim was to respond to the question as to how teachers explicitly develop students' awareness of language for curriculum learning through the implementation of SFL pedagogy. As for the conduction of the study, the author collected data about how register, nominalization, and periodicity were integrated in learning through "classroom observations, teacher and student interviews, and students' texts" (p.1). The findings confirmed the effectiveness of the explicit SFL pedagogy in developing both the teachers and students' awareness of "disciplinary literacy" and the linguistic patterns of language curriculum. The author concludes the SFL approach she implemented in her case-study is highly applicable for all education levels in the UK and other countries. It is yet only successful if teachers employ SFL metalanguage explicitly, engage into constructive negotiation with learners about how language develops within various writing communities, and give due and relevant metalanguage criticism of their output. It is of special importance, the study recommends, to implement more SFL studies, especially with the goal to analyze how this theory could be more effective in all school levels.

Beyond the conclusions drawn from these studies, along with others sharing similar claims and findings, it becomes evident that the use of SFL Genre Writing (GW) pedagogy enhances students' writing skills and empowers them to produce texts that align with the lexico-grammatical patterns of their respective writing communities.

METHODOLOGY

This study employs a qualitative descriptive research design, which involves the "accurate description of something or someone... The something may be an event, phenomenon, or characteristic. The someone may be an individual, group, or community" (Dulock, 1993, p. 154). Although descriptive research design is commonly used in quantitative research to numerically describe data, qualitative researchers can also use it to describe, explain, and interpret their findings (Erickson, 1985; Sandelowski, 2000).

It is worth noting that qualitative analysis possesses several characteristics that align well with the scope and nature of the current study. One such characteristic is that it encompasses aspects of quantitative content analysis, as the reading of every text is inherently qualitative, even when the goal is to quantify certain patterns (Krippendorff, 2018).

Another characteristic is that this type of analysis can be either deductive or inductive (Azungah, 2018). When analyzing data inductively, researchers identify the analysis' thematic units along the research process; however, analyzing data deductively means analysts begin their analyses with clearly set and decided upon themes and patterns (Creswell, 2009). Inductive analyses are often used when researchers lack or have limited knowledge about the themes the text analysis could yield; on the other hand, a deductive analysis is used when theories and previous findings are

available before the conduction of research (Hsieh & Shannon, 2005). It is also possible for qualitative content analysts to employ both inductive and deductive techniques concomitantly (Krippendorff, 2018). It is within this framework that the current research has inductively and deductively approached the identification of the lexicogrammar patterns in the analyzed introductions.

Furthermore, qualitative content analysis proves valuable when the researcher's objective is to examine the features of texts in terms of the communicative purposes they serve (Lindkvist, 1981; Tesch, 1990). This aligns closely with the goal of this study as it aims to analyze the introductions of the participants by exploring their use of lexicogrammatical-patterns. Then, content analysis, in its generality, is guided by the goal of reaching a systematic description of the analyzed material (Krippendorff, 2004). Key to the identification of this goal is the analyst's need to specify from which angle he/she plans to study their data, and simultaneously be open to altering their goal along the analysis process (Krippendorff, 2018). Of equal importance is the possibility of qualitative content analysts to make conclusions that reach beyond the physical description of data (Krippendorff, 2004). They, in other words, can subjectively infer and/or evaluate meanings texts express either explicitly or implicitly (Hsieh & Shannon, 2005; Schreier, 2012).

Taking these characteristics in consideration, the current study has conducted an inductive content analysis to explore the **lexico-grammatical patterns common** in the students' introductions in abidance by Halliday's levels of the context of situations (1999). The analysis of lexicogrammar in the introductions of dissertations Moroccan applied linguistics students write should provide a rigorous understanding of the **generic identity**, ways a text is similar to others of its genre, of the writing of Moroccan EFL applied linguistics students (Imtihani, 2010).

Collecting data from the target Moroccan universities has proceeded quite smoothly, notwithstanding the few challenges the researcher had faced to access data sources. After obtaining official authorization to conduct the research both within and outside Ibn Tofail University, Kenitra, where this study was conducted, the researcher collected texts from the Departments of English Language and Literature at Ibn Tofail University, Kenitra; Mohamed V University, Rabat; University Hassan II, Casablanca; Chouaib Doukkali University, El Jadida; Moulay Ismail University, Meknes; Sidi Mohamed Ben Abdellah University, Fes; and Cadi Ayyad University, Marrakech.

To broaden the scope of the corpora significantly, the current study employed the snowball sampling technique. The researcher reached out to their professors, colleagues, classmates, and friends, requesting them to share the introductions of their dissertations and to extend the request to their own networks. Although this approach was time-consuming, it resulted in additional 40 introductions and extended the study's reach to include texts from universities that would have been difficult for the researcher to visit. It is worth noting that, despite criticism of snowball sampling for its perceived inability to produce generalizable results (Morgan, 2008), the technique garners support from qualitative experts who argue that the aim of qualitative research is not necessarily to generalize to broader populations, but rather to explore phenomena in depth (Creswell, 2002; Krippendorff, 2018). The table below provides details on the number of texts collected through in-person visits to university libraries and via snowball sampling.

Table 1 Body of texts collected from EFL departments of Moroccan universities

University	Number of texts
ibn tofail - kenitra	35
Mohamed V - Rabat	30
HASSAN II - Casablanca	30
CHOUAIB DOUKKALI - Eljadida	25
MLY ISMAINL – Meknes	30
SIDI MED BEN ABDELLAH - Fes	20
ABDELMALEK ESSAADI - Tetouan	20
Total:	210

The selection of texts (introductions) for analysis in this study was guided by several key criteria. Firstly, only dissertations authored by Moroccan MA applied linguistics students were considered eligible for analysis. Secondly, dissertations written within the academic years 2021 and 2022 were included. Thirdly, no restriction was on the topic of the dissertations as long as they related to applied linguistics. Lastly, the attainment of data saturation determined the number of texts utilized in the study. Overall, the analysis of the texts within the framework of these criteria was guided by the question of whether “the data provide useful evidence for testing hypotheses or answering research questions” (White & Marsh, 2006, p.27).

The collection of the body of texts encountered several challenges. Firstly, most Moroccan universities lack a centralized database for storing previous MA dissertations in an organized manner. Instead, dissertations from various disciplines are often stored together in mixed boxes, making it difficult for researchers to locate specific EFL applied linguistics dissertations. Secondly, librarians are often unable to grant access to students' previously written works, particularly when researchers request to photograph specific pages needed for their study. Consequently, researchers are left with no choice but to explain the purpose of their research in detail, seek permission, and assure confidentiality. Additionally, some universities do not maintain records of their students' previous MA dissertations, necessitating direct

contact with professors to inquire if they possess any copies. Fortunately, this approach proved successful in gathering significant amount of data for the current paper. Last, of the 210 sampled introductions, only 105 were considered for analysis as a result of data saturation. In fact, the whole data collection endeavor was conducted towards the objectives of the study, as Patton (2002) recommends, and guided by the structure of qualitative content analysis as outlined by Krippendorff (2018).

Post to the identification of the units of analysis, elaborated upon in details in the Data Analysis section below, the study commenced with data collection. Initially, the researcher visited the library of Mohamed V University in Rabat, accompanied by two other PhD students, to select and Camscan introductions from MA applied linguistics dissertations authored by Moroccan students in the academic years 2021-2022 and 2022-2023. These scanned texts were organized into a single folder, labeled with the university name, student gender, and submission year. Subsequently, the entire body of texts was converted into an editable Word document and preserved for analysis. Similar procedures were replicated to gather data from other universities; however, this time, the author of the current study conducted the collection of data solely.

Secondly, the researcher reached out to various university professors, colleagues, friends, and members of Facebook PhD groups, requesting access to MA applied linguistics dissertations/introductions they possessed. This supplementary procedure enriched the dataset with a substantial number of texts that would otherwise have been inaccessible. Additionally, the internet was utilized to search for additional introductions, resulting in the inclusion of a few more texts in the corpus. As with the Camscanned texts, each soft and hard copy received was converted into an editable Word document, annotated with the university name, student gender, and submission year. It is noteworthy that each introduction was assigned a representative number to facilitate differentiation from others in the body of texts.

Systemic Functional Linguistics (SFL) lexico-grammatical content analyses of written texts can be approached in various ways and for different purposes. Central to these is the SFL assertion that lexis and grammar constitute a unified linguistic entity that must be analyzed within its contextual use (Hasan, 1987; Halliday & Webster, 2008). Reinforcing this perspective is the SFL argument that the text, whether spoken or written, serves as the fundamental unit of analysis, necessitating a functional examination of the lexicogrammar that backs it, and that while describing a text without considering its textual features is unproductive, it becomes meaningless to analyze it without contextualizing it within its language load (Halliday, 1985).

Thus, to examine the lexicogrammar patterns within the students' introductions, this study employed an inductive content analysis approach. This method does not predefine the patterns to be considered beforehand (Baayen, 2008). However, the researcher conducted the analysis with reference to lexicogrammar patterns explored in previous Systemic Functional Linguistics (SFL) genre writing research, which could hold significance for the quality of the population's genre writing. This analytical process was guided by Krippendorff's model of qualitative content analysis:

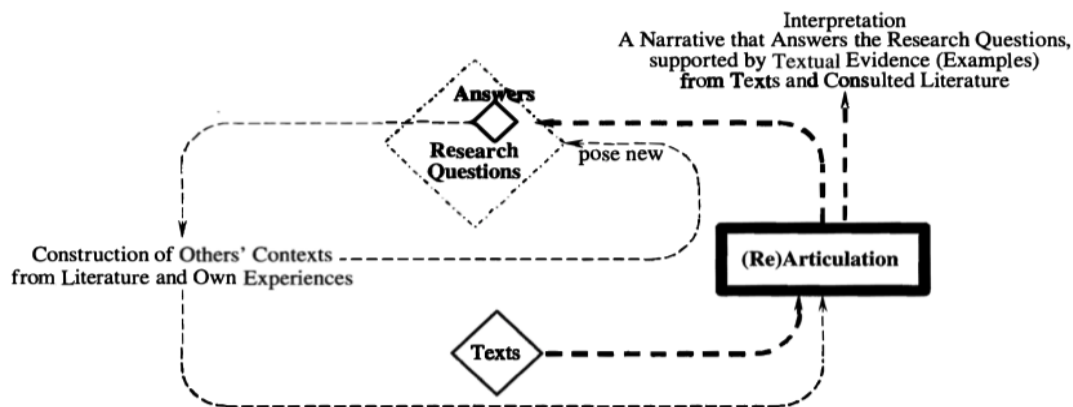


Fig. 2 Qualitative Content Analysis Model

As a preliminary step, the researcher, along with two research assistants— a PhD holder and a PhD candidate—conducted three weekly meetings. The purpose was twofold: to ensure their familiarity with Systemic Functional Linguistics (SFL) lexicogrammar theory and practice, and to compile a list of potential lexicogrammar patterns for exploration in the study. Twenty texts were randomly chosen to kickstart the analysis at this stage.

Subsequently, each rater had three weeks to independently analyze the selected texts. Following this period, the three analysts met to discuss and agree upon the lexicogrammar patterns that entailed further examination. As a result of the discussions between the researcher and both assistants, three key patterns emerged as focal points for the subsequent analysis:

1. Lexical Bundles
2. Types-Token Ratio
3. Lexical Density

For the analysis of lexical bundles, the study utilized AntConc 3.5.9 for Windows and 4.1.4 for Mac. Available for free, this software provides usages, including “a concordance module, a file viewer, and even a cluster tool, and will allow users to search documents for common words or phrases” (Bytesin, n.d.). Lexical density was analyzed using Ure’s framework (1971), which employs the formula below:

$$\text{Lexical Density} = \left(\frac{\text{Number of Content Words}}{\text{Total Number of Words}} \right) \times 100 \quad (1)$$

The exploration of the types (the total number of unique words in a text) and tokens (the total number of words in a text, including repetitions) employed a formula that has been used in previous research, such as that of van Gijesel et al. (2005) and Tarone and Swierzbins (2009):

$$\text{Lexical Richness} = \frac{\text{Number of Tokens}}{\text{Number of Types}} \quad (2)$$

Key to this analysis is that the more types a piece of writing has, the higher its quality is (Ji-Young, 2014).

It is important to note that the analysis conducted in the current study operated on a text-by-text basis. In simpler terms, each introduction collected underwent separate analysis before the resulting numerical findings were aggregated to derive cumulative results.

Lexicogrammar is the second component the content analysis of the current study has targeted in the introductions written by MA EFL students in Moroccan universities. To broaden the scope of this analysis, the current study has examined three lexicogrammar features: lexical density, types-token ratio, and lexical bundles. The primary objective is to establish a robust reading and assessment of the participants' introductions while concurrently gaining insights into their strengths and weaknesses from the perspective of Systemic Functional Linguistics (SFL).

FINDINGS

Lexical density

Lexical density as a linguistic text analysis tool aims to identify the proportion of content words in a particular text compared to its whole number of words (Halliday, 1985). It is, in other words, the ratio of content words (nouns, verbs, adjectives, adverbs) vis à vis function words (articles, prepositions, conjunctions) (Ure, 1971). To analyze lexical density of the collected body of texts, the current study has used Ure (1971) (see formula 1 above).

According to this formula, high lexical density is attained when a text's density rate is higher than 40% (Halliday, 1985). However, the measure employed in the current study is based on the one proposed by Ure (1971), which has been replicated in other studies by Johansson (2008) and Syarif (2018). This provides:

Lexical density analysis this research has conducted using Seoscout yielded some interesting findings. First, the gap amongst the introductions in terms of their lexical density is highly noticeable as the following two examples illustrate:

41-50%: not dense
51-60% : less dense
61-70% : dense
>70% : very dense

Table 2 Sample lexical density analyses of student's introduction

Word Count	376	Word Count	1410
Character Length	2359	Character Length	8774
Letters	1938	Letters	7161
Sentences	32	Sentences	75
Syllables	644	Syllables	2436
Average Words/Sentence	11.8	Average Words/Sentence	18.8
Average Syllables/Word	1.7	Average Syllables/Word	1.8
Lexical Density	48%	Lexical Density	54%

(a) Intro-013 content analysis

(b) Intro-058 content analysis

The examples above compare a text with the highest lexical density rate (Intro-058), which stands at 54%, to a text with the lowest lexical density (Intro-013), at 48%. Lexical density of other texts falls between these two rates. However, as the figure below indicates, lexical density in the great majority of these texts is below 50%.

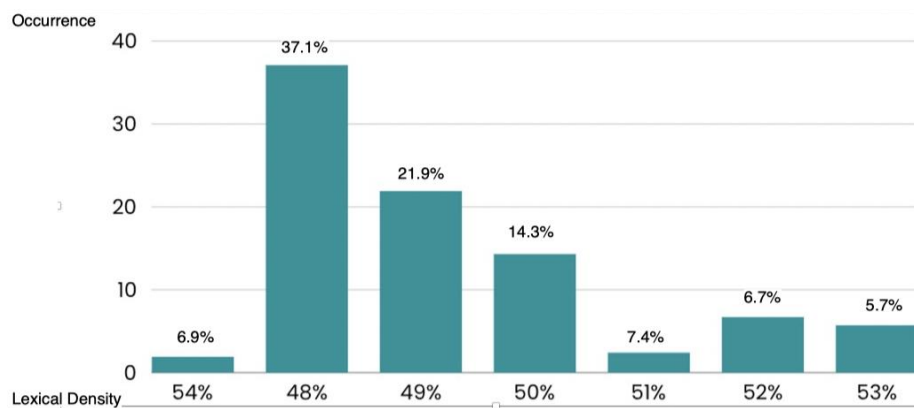


Fig. 3 Lexical density percentages in participants' introductions

The figure illustrates that most introductions written by the participants have a lexical density rate of 50% or less. Specifically, Participants write introductions with a lexical density rate of 48% in approximately 37% of cases, while they produce rates of 49% and 50% in about 22% and 14% of introductions, respectively. On the other hand, texts with lexical density rates of 51%, 52%, and 53% make up just over 14% of the entire body of texts. The remaining 6% corresponds to texts with a lexical density rate of 54%. When comparing these rates against the norm established by Ure (1971), as utilized by Johansson (2008) and Syarif (2018) in their analyses of students' writing lexical density, the current research concludes that the introductions written by Moroccan MA applied linguistics students in their MA theses are, on average, not dense, being within the range of 41% to 50%. This finding suggests that these texts lack a key characteristic of academic and advanced writing, as noted by Biber (2016). This is surprising, especially given that the participants are MA students on the cusp of becoming researchers, particularly if they plan to pursue PhD studies and have their research published. Nevertheless, this low lexical density rate aligns with several previous studies which have found EFL university students, in general, use a limited number of content words in their writing (Chaudron, 2003; Doughty & Long, 2005). Syarif and Putri (2018) have concluded similar findings as lexical density in the students' texts they analyzed were less dense with an average of about 50%. This, according to Syarif and Putri, could be the result of a lack of vocabulary (content words), using grammatically complex sentences, and/or writing in the spoken form.

In addition, the current study has conducted a parallel analysis to identify the type of content words that characterize the lexicon of the participants. The primary finding indicates a predominant use of nouns, followed by verbs, adjectives, and then adverbs. The following graph provides a detailed breakdown of these findings:

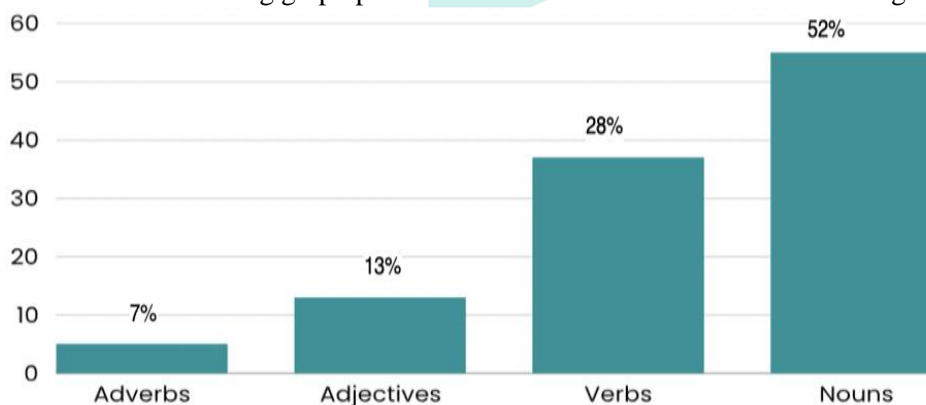


Fig. 4 Percentage of content words in students' introductions

The graph indicates a predominance of nouns in the participants' introductions, comprising nearly 52% of the content words. Following that are verbs at approximately 28%, and then adjectives and adverbs at 13% and 7%, respectively. Other studies have reached similar findings. One of these is a study Ustunalp and Esmer (2022) and which provides Turkish graduate students employ more nouns than verbs in their dissertations.

Types-token ratio

Another method employed in the current study to analyze the dissertations written by Moroccan MA applied linguistics students involves comparing the types and tokens they utilized. This analysis involves examining the number of types (the total number of unique words in a text) in relation to the number of tokens (the total number of words in a text, including repetitions). The aim is to determine the participants' ability to produce lexically rich writing. In this respect, the formula used is consistent with that adopted in many previous studies, including van Gijesel et al. (2005) and Tarone and Swierzbins (2009) (see formula 2 above). It is argued that a higher number of types indicates a richer lexicon and, consequently, greater writing quality among students' writing (Kim, 2014).

Employing Anthony's (2018) Antconc-4 software, each dissertation was analyzed independently before the percentages were aggregated to calculate the overall lexical richness rate. This approach was necessary because analyzing

the entire content of the text yielded a very low lexical density rate, which posed a significant issue against the validity of the findings. The analysis has revealed the participants' dissertations contain a total of 72,000 words, of which 16,000 are types and 56,000 are tokens.

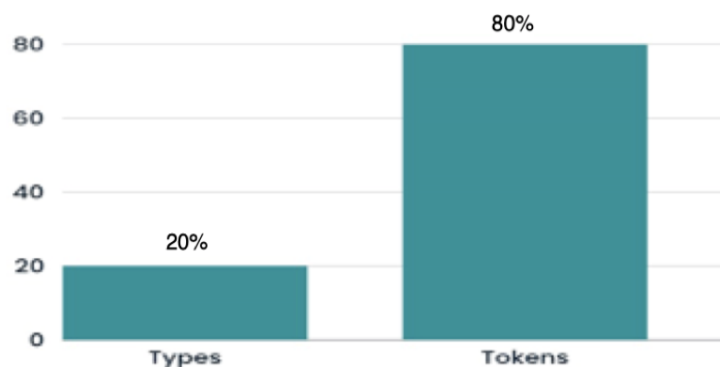


Fig. 5 Types and tokens in students' introductions

In terms of percentages, the bar graph above indicates 80% of the participants' words are tokens while the number of types drops down to about 20%. In other words, the number of types in the writing of the participants constitutes 25% of the number of tokens. From the perspective of the types-tokens measure, a value closer to 1 indicates that the text is more lexically rich. As for the lexical richness in the dissertations of the applied linguistics MA students the current study has investigated is 0.175. This finding is consistent with a study conducted by Djiwandono (2016), which provides the lexical richness of Indonesian EFL students was 0.170. Another study has identified that lexical richness in texts produced by Chinese EFL PhD candidates reaches about 0.258 (Lei & Yang, 2020). Then, A study conducted by Kathryn Marie Johnston at Portland State University (2017) has found types make about 0.230% of the writing of native English applied linguistics students.

The lexical richness in the texts produced by EFL MA applied linguistics students in Morocco is notably low at 0.175, which is most probably a consequence of their frequent repetition of words throughout their written texts. Henceforth, it is reasonable to conclude that the dissertations written by the participants in the current study may not be of a high writing quality because of their quite low lexical richness.

Lexical bundles

Studies have referred to recurring word sequences and / or patterns in written texts using a variety of labels; of these are "lexical phrases, formulas, fixed expressions, and pre-fabricated patterns" (Biber & Barbieri, 2007). However, the current study refers to these lexical recurrences as lexical bundles, which is the term adopted in previous research, including studies of Cortes (2002), Biber and Barbieri (2007), Hernandez (2013), and Mohamdar (2015). Researchers classify lexical bundles under various correlates, depending on their structures and functions (Biber et al., 1999; Cheng, 2006).

The current investigation has analyzed the participants' texts using Antconc 4, a free software developed by Anthony (2018). Additionally, it has employed Hyland's (2008) structural classification of Lexical Bundles (LBs), which categorizes them into noun-based bundles, preposition bundles, verb bundles, and bundles of other forms. In terms of the n-grams for the bundles, this study focused on the top 20 3-gram recurrences in the body of texts, as shown in the following two tables. The choice of 3-gram bundles this study has opted for is common and well- documented in the literature (Liu, 2012; Simpson-Valche & Ellis).

Table 3 Twenty most common 3-gram bundles in the Moroccan MA applied linguistics dissertations

	Type	Rank	Freq	Range
1	in order to	1	49	1
2	a foreign language	2	46	1
3	as well as	3	45	1
4	as a foreign	4	43	1
5	one of the	5	41	1
6	in terms of	6	39	1
7	the language of	7	35	1
8	the present study	8	34	1
9	the field of	9	33	1
10	english as a	10	30	1
11	in the moroccan	10	30	1
12	the use of	10	30	1
13	of the study	13	28	1
14	the importance of	14	27	1
15	due to the	15	26	1
16	arabic as a	16	25	1
17	of this study	16	25	1
18	students and teachers	16	25	1
19	the role of	19	24	1
20	a number of	20	23	1

Two significant findings can be deduced from the table above. One is that the lexical bundle noun phrase + of reoccurs 172 times in the top 20 lexical bundles accounting for about 26% of the whole set of the 20 top frequent bundles. The frequent use of noun-phrase bundles can reflect the students' developed writing skills, especially that according to research, as students' "literacy grows their writing is to show ... increased noun phrase related structures" (Liu & Li, 2016, p.52). In addition, the most common noun phrase + 'of' bundle, which is used 35 times, is followed by the word 'instruction' 12 times, as the table below indicates.

Table 4 The language of lexical bundles in the Moroccan MA applied linguistics dissertations

	File	Left Context	Hit	Right Context
1	Antconc ...	with English was looked down onto as it was	the language of	a person who lacked the concept of
2	Antconc ...	led descriptions of the nature and features of	the language of	e-mail. Gains (1999) focuses on generic featur
3	Antconc ...	lism. However, the effectiveness of English as	the language of	economic power may have weakened Fus'ha,
4	Antconc ...	er examples of good practice by investigating	the language of	effective leaders who might potentially serve a
5	Antconc ...	ave I chosen the language of leadership? Why	the language of	female leadership? And finally, why the Arab
6	Antconc ...	ain. Secondly, the reason behind my focus on	the language of	female rather than male leaders or perhaps
7	Antconc ...	t Moroccans commonly believe that English is	the language of	future opportunities. Since Morocco is seeking
8	Antconc ...	economic power may have weakened Fus'ha,	the language of	government and public education. If the diagl
9	Antconc ...	e illusive language policy of higher education.	The language of	instruction remains one of the fuzziest areas
10	Antconc ...	owing hypothesis: 1. H0: There is no effect of	the language of	instruction on Moroccan EFL master students.
11	Antconc ...	e male speakers use more alveolar trill [r], yet	the language of	instruction being French for certain participan
12	Antconc ...	minology that – possibly due to French being	the language of	instruction in the field of science – is
13	Antconc ...	the students in the field of humanities due to	the language of	instruction being Arabic are more exposed to
14	Antconc ...	occan higher education maintained French as	the language of	instruction in the faculties of sciences, econo
15	Antconc ...	ducation. Arabic gradually replaced French as	the language of	instruction since 1965. By the year 1988, all th
16	Antconc ...	arabized and Modern Standard Arabic became	the language of	instruction and the Moroccan education syster
17	Antconc ...	and Scientific Research) to readopt French as	the language of	instruction for scientific subjects has revealed
18	Antconc ...	anguage in the country owing to its status as	the language of	instruction in the Moroccan schools. After inde
19	Antconc ...	ender and the field of study, the latter due to	the language of	instruction – that all have an impact on
20	Antconc ...	n different focus groups due to the impact of	the language of	instruction, as the students in the field

Other frequent noun phrase bundles include a foreign language (46 occurrences), the present study (34 occurrences), and English as (30 occurrences). They account for approximately 17% of the 20 top LBs.

In addition, prepositional phrases rank second to noun phrases among the top 20 most frequent lexical bundles in the body of texts this study has analyzed, with a count of 137. This represents about 21% of the 20 top LBs. The graph below summarizes the distribution of the most frequent lexical bundles in the participants body of texts.

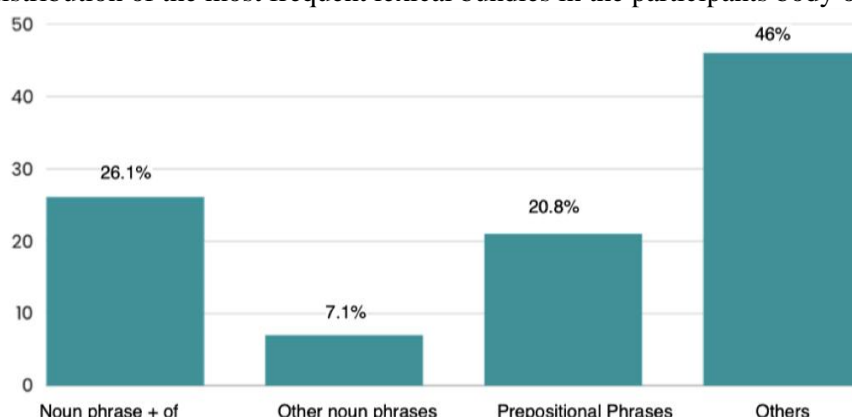


Fig. 6 Most common lexical bundles in the Moroccan MA applied linguistics dissertations

The breakdown of lexical bundles offers a comprehensive view of the intricate patterns within the analyzed dissertations. A notable 26.1% consists of noun phrase lexical bundles paired with the preposition of, reflecting a widespread syntactic structure often found in academic or formal discourse (Hyland & Jiang, 2018). Meanwhile, other noun phrase bundles constitute 7%, further highlighting the multi-layered nature of noun phrase LBs beyond the specific of pattern.

Additionally, prepositional phrase bundles account for 21%, emphasizing the significance of prepositions in establishing relationships and syntactic cohesion (Halliday and Matthiessen, 2014). Interestingly, the category labeled as 'Others' includes a substantial 46%, highlighting a vast array of diverse lexical combinations that contribute to the

richness and complexity of the language. These include LBs with anticipatory it (it is the (14 times), it is not (13 times)), passive verbs (is based on (13 times), is devoted to (13 times)), and that clause (that it is (13 times)). Collectively, this LBs distribution reflects the varied syntactic landscape inherent in the analyzed dissertations.

CONCLUSION

The lexicogrammar analysis of the introductions written by EFL MA applied linguistics students in Morocco has yielded several key findings. First, the texts exhibit a relatively low lexical density, with an average ranging between 41% and 50%. This suggests that the students rely on a significant proportion of function words, which may be indicative of less compact or less information-density in the writing of applied linguistics MA students. Second, the lexical richness of the introductions is also low, with an approximate value of 0.170. An implication of this is the students' limited vocabulary variation, which could affect the overall sophistication and expressiveness of their academic writing. Lastly, the analysis reveals that the most frequently used lexical bundle among the participants is the structure "lexical noun phrase + of." This recurring pattern suggests a preference for certain formulaic expressions, which may reflect common academic conventions or a lack of syntactic variety. These findings highlight areas for potential pedagogical intervention to enhance students' academic writing skills, particularly in terms of lexical diversity and density.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF CONFLICT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Anthony, L. (2018). *AntConc (Version 4.2.4) [Computer Software]*. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
2. Azungah, T. (2018). Qualitative research: Deductive and inductive approaches to data analysis. *Qualitative Research Journal*, 18(4), 383–400.
3. Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge University Press.
4. Belbouah, A., & Bouguerba, T. (2025). Examining the use of Swales' CARS model in MA dissertation introductions: A study of Moroccan applied linguistics students. *TWIST*, 20(1), 347–355.
5. Belbouah, A., Khotbi, T., & Bouguerba, T. (2023). Challenges of writing in Arabic as a foreign language context: Case of the Spanish school in Rabat. *European Journal of Contemporary Education and E-Learning*, 1(1), 3–19.
6. Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.
7. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
8. Chaudron, C. (2003). Data collection in SLA research. *The handbook of second language acquisition*, 762–828.
9. Cheng, A. (2006). Understanding learners and learning in ESP genre-based writing instruction. *English for Specific Purposes*, 25(1), 76–89.
10. Chibi, M. (2021). Collectivism and individualism as cultural aspects in Arabic/English argumentative writing by Moroccan students. *International Journal of English Literature and Social Sciences*, 6(4), 60–69.
11. Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge University Press.
12. Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (9th ed., pp. 131–145). John Benjamins.
13. Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative*. Prentice Hall Upper.
14. Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
15. Djiwandono, P. I. (2016). Character education in content courses: Self-scoring as a means for developing honesty in students. *TEFLIN Journal*, 27(2), 153.
16. Djiwandono, P. I. (2016). Lexical richness in academic papers: A comparison between students' and lecturers' essays. *Indonesian Journal of Applied Linguistics*, 5(2), 209–216.
17. Dulock, H. L. (1993). Research design: Descriptive research. *Journal of Pediatric Oncology Nursing*, 10(4), 154–157.
18. Eggins, S. (1994). *An introduction to systemic functional linguistics*. Printer Publishers.
19. Erickson, F. (1985). Qualitative methods in research on teaching (pp. 119–162). *Institute for Research on Teaching*.
20. Faraj, A. K. A. (2015). Scaffolding EFL students' writing through the writing process approach. *Journal of Education and Practice*, 6(13), 131–141.

21. Forey, G. (2020). A whole school approach to SFL metalanguage and the explicit teaching of language for curriculum learning. *Journal of English for Academic Purposes*, 44, 100822.
22. Gabrielatos, C. (2002). *EFL Writing: Product and Process*.
23. Gledhill, C. (2011). The lexicogrammar approach to analysing phraseology and collocation in ESP texts. *ASp. la revue du GERAS*, (59), 5-23.
24. Halliday, M. A. K. (1985). *An introduction to functional grammar*. Arnold.
25. Halliday, M. A. (1999). The notion of "context" in language education. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 1-24.
26. Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition*. Cassell.
27. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar*. Routledge.
28. Halliday, M. A. K., & Webster, J. J. (2008). *Complementarities in language*. The Commercial Press.
29. Hasan, R. (1987). The grammarian's dream: Lexis as most delicate grammar. *New Developments in Systemic Linguistics*, 1, 184-211.
30. Hasan, R., & Halliday, M.A.K. (1985). The texture of a text. *Language, context, and text: Aspects of language in a social-semiotic perspective*. Deakin University.
31. Hjelmslev, L. (1961). *Prolegomena to a Theory of Language*. The University of Wisconsin Press.
32. Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
33. Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407.
34. Imtihani, N. (2010). Genre analysis in the frame of systemic functional linguistics. *Jurnal Universitas Gadjah Mada*, 22(1), 86-93.
35. Jashari, L. T., & Fojkar, M. D. (2019). Teachers' perceptions of developing writing skills in the EFL classroom. *ELOPE: English Language Overseas Perspectives and Enquiries*, 16(2), 77-90.
36. Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433.
37. Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.
38. Lindkvist, K. (1981). Approaches to textual analysis. In K. E. Rosengren (Ed.), *Advances in content analysis*, 25(2), 23-41. Sage. <https://www.jstor.org/stable/4194399>.
39. 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.
40. Liu, L., & Li, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *Journal of Asia TEFL*, 13(1), 48.
41. Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage.
42. Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). University of Michigan Press.
43. Uysal, H. (2008). Tracing the culture behind writing: Rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context. *Journal of Second Language Writing*, 17(3), 183-207.
44. Zhang, X. (2019). EFL writers' reconstruction of writing beliefs in a functional linguistics-based curriculum: What does the trajectory look like? *SAGE Open*, 9(2), 2158244019853915.
45. Zheng, C. (2013). A structure analysis of English argumentative writings written by Chinese and Korean EFL learners. *English Language Teaching*, 6(9), 67-73.