Bulletin of the *Transilvania* University of Braşov
Series IV: Philology and Cultural Studies • Vol. 13(62) Special Issue
https://doi.org/10.31926/but.pcs.2020.62.13.3.8

Lexical bundles in PhD dissertations and master theses: A comparative inquiry

Omid REZAIE[1], Mehrdad Vasheghani FARAHANI[2], Millad MASOOMZADEH[3]

The objective of the current research was to analyze and compare the 4-gram lexical bundles from the three part-genre corpora. For the theoretical framework, Hyland's (2008) taxonomy of lexical bundles was exploited. The corpus of the study consisted of 100 Master theses and PhD dissertations. The study was focused on abstract, introduction, and conclusion only and did not enter into other parts of the Master theses and PhD dissertations. The abstract part-genre contained noticeably more bundles than the other two part-genres. In addition, introduction and abstract part-genres contained the highest amount of research-oriented bundles. As far as participant-oriented bundles are concerned, it is the conclusion part that includes noticeably more of these bundles than the other two part-genres. Overall, findings of this research showed that in three major part-genres of the academically key genre of PhD dissertations and M.A. theses, i.e. abstract, introduction, and conclusion, it is the abstract that enjoys a high amount of formulaic language in the form of lexical bundles.

Key-words*: lexical bundles, master theses, PhD dissertations, comparative study*

1. Introduction

The linguistic analysis of words and phrases within mostly written texts in various contexts dates back to register analysis. As stated by Widdowson (2007), register analysis exclusively focused on words and sentences in written genres and was

[1] Isfahan University; Iran, omidrezaei.rezaei99@gmail.com
[2] Leipzig University; Germany, mehrdadfarahani1365@gmail.com
[3] Independent Researcher, milad.masoomzadeh@gmail.com

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

later replaced by discourse analysis that put the limelight more on levels of analysis beyond the sentence level used in the real communication. It was within the upheaval of discourse analysis which genre-based studies gained ground, that mostly centered on the formal as well as functional analysis of communicative events whose participants share the same communicative purpose in mind (Swales 1990). This line of research has been quite widespread in the study of second language learning (SLA) during the last 40 years or so (Hyland 1999; Samraj 2005).

As mentioned above, most studies have particularly focused on grammatical and functional structures of multi-word expressions and formulaic language in major genres despite the fact that these genres have been made of some smaller-scale sections called part-genres (or sub-genres) in the literature (Samraj 2005). Different names have been suggested for these multi-word expressions, including lexical phrases, formulas, fixed expressions, prefabricated patterns, lexical bundles and academic clusters (Biber 2006; Hyland 2008). The variety of names shows that it seems that in each line of study different underlying approaches have been selected to define them in specific ways, making them reflect different perspectives on the use of these multi-word sequences. For example, some studies regard idiomatic phrases (e.g. to get to the heart of the matter) and some others have focused on non-idiomatic but salient sequences (e.g. as it went above).

Among all of these names and perspectives, one more recent line of research has centered on the study of lexical bundles which happen to occur in both written and oral discourse more frequently than by chance, and help to form the text meaning and lead to better understanding of particular genres (Biber et al., 1999; Biber and Reppen 2002; Hyland 2008). These particular strings of words have been the center of attention in discourse and genre studies during the past two decades (Chen and Baker 2010).

The literature in the study of lexical bundles confirms that these word strings differ in terms of frequency, functional types, and grammatical formation across different genres; academic disciplines and L1/L2 writing (see Biber et al. 2004, Cortes 2007, Chen 2008, Hyland 2008, and Ädel and Erman 2012, among others). However, it is still not clear if the use of these word combinations is sensitive to various sub-genres within particular genres as well or not.

As stated by Haswell (1991), the smooth use of such word strings directly leads to a native-like performance by non-native writers and shows to what degree they are familiar with particular genres. In the same vein, Adolphs and Durow (2004) mention that if students are more integrated into the L2 environment they

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

can acquire and use such word bundles more naturally. Therefore, a focus on the frequency, functions and grammatical types of bundles used in part-genres of main academic genres (e.g. theses and dissertations) might prove fruitful in showing how genres are made out of their part-genres. Accordingly, it is still not clear if the use of these word combinations is sensitive to various sub-genres within particular genres as well or not.

Conducting a rigorous analysis of lexical bundles in terms of their functional and structural characteristics could shed some light on how less experienced academic writers form and shape up the main sub-genre sections of dissertations so that future participants of applied linguistics' discourse community (i.e. university students) could be more aware of specific formulaic language.

1.1. Research questions

1. Is there any significant difference among three part-genres (i.e. abstract, introduction, and conclusion) of applied linguistics' dissertations and theses in terms of the frequency of lexical bundles?
2. Is there any significant difference among three part-genres (i.e. abstract, introduction, and conclusion) of applied linguistics' dissertations and theses in terms of the functional categorizations of lexical bundles?
3. Is there any variation among three part-genres (i.e. abstract, introduction, and conclusion) of applied linguistics' dissertations and theses in terms of the structural categorizations of lexical bundles?

1.2. Null-hypotheses

1. There is no significant difference among three part-genres of applied linguistics' dissertations and theses in terms of the functional categorizations of lexical bundles?
2. There is no significant difference among three part-genres of applied linguistics' dissertations and theses in terms of the frequency of lexical bundles?
3. There is no variation among three part-genres of applied linguistics' dissertations and theses in terms of the grammatical structure of lexical bundles?

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

2. Literature review

2.1. Lexical bundles

Biber et al. (1999, 990) define this special type of formulaic language as "sequences of word forms that commonly go together in natural discourse". They identified and extracted these bundles in their quantitative analysis of the 40 million word Longman Spoken and Written English Corpus. The findings of their study were very interesting and remarkable. About 28% of the words in conversation occurred in 3- and 4- word lexical bundles, and 20% of words in academic prose. As the key concept of the research is lexical bundles, I am going to have a look at the defining features of lexical bundles through the following sections. As defined by Hyland (2008) these are word strings that happen more frequently than expected by accident in the text and not only help to shape the meaning but also help the readers to distinguish across different genres.

2.2. Lexical bundles' characteristics

Bundles are one type of multi-word formulaic expressions which were first identified in the work of Altenberg (1998) who learned that they enjoy three common features: semantic transparency, fragmental grammatical structures, and pragmatic specialization. He examined these word combinations in the London-Lund Corpus of Spoken English in order to determine whether there is a certain amount of formulaic language in the spoken corpus. The cut-off point frequency he set was 10 times per million and decided in his investigation to recurrent word combinations of three and more words. For example, I do not think, do you know, on the other hand, yes I did were identified as recurrent word clusters. A look at these chunks shows that most of them are transparent in meaning, unlike pure idioms; these bundles cannot be paraphrased with a single word as is the case for many idioms (i.e., look forward to = anticipate). It was also found that those combinations obviously do not have complete grammatical structures. Altenberg's results showed that only ten percent of recurrent word combinations took the form of full clauses, and the majority of word combinations have incomplete grammatical structures, with 76% being clause constituents and 14% incomplete phrases.

In fact, it was Altenberg's work that inspired Biber et al. (1999) who introduced for the first time the concept of lexical bundles. They asserted that, apart from the frequency characteristics of these bundles mentioned by Altenberg

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

(1998), there should be a distributing criterion to evade any individual idiosyncrasy in the identification of lexical bundles. It is clear that lexical bundles are a modified version of recurrent word combinations, and hence they inherit their three main features discussed above (e.g. semantic transparency, fragmental grammatical structures, and pragmatic specialization).

2.3. Lexical bundles' grammatical features

According to Biber et al. (1999), these bundles in various genres are likely to have particular structural types. They argued that their structural forms are sensitive to different genres or registers. About 44% of lexical bundles in conversations take the form of verbal and clausal units, as in I don't know why and I thought that was. However, lexical bundles in academic prose are more likely to be nominal usually in the form of noun phrases and prepositional phrases such as *the nature of the* and *the size of the* in academic writing. Despite variations in grammatical structures of lexical bundles, they share another feature: the incompleteness in their grammatical structures. As Biber et al. (1999, 991) noted, "Lexical bundles extend across structural units".

2.4. Functional load of lexical bundles

As for the functional categorization of lexical bundles, Biber et al. (2004) introduced the first taxonomy for four registers of classroom teaching, textbooks, conversation, and academic prose. They made use of two corpora, the TOEFL 2000 spoken and written academic corpus and three main subcorpora of Longman Spoken and Written English corpus. They only focused on functional roles of 4-word bundles with the frequency of higher than 20 times per million words, but present at least in 5 different texts. Finally, they ended up with 3 main categories in their overall taxonomy: (1) stance bundles, (2) discourse organizers, and (3) referential expressions. They defined stance expressions as being used to "express attitudes or assessments of certainty that frames some other proposition" (2004, 384). They divided these bundles into two major semantic categories: attitudinal/modality stance bundles and epistemic stance bundles. The former refers to "speaker attitude toward action or event described in the following proposition" (Biber et al. 2004b, 390).

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Discourse organizing bundles "reflect relationships between prior and coming discourse" (Biber et al., 2004), and are categorized as topic introduction/focus bundles which are used to introduce a new topic in the classroom (i.e., *want to talk about*), and elaboration/clarification bundles which speakers or writers use to further explain topics (i.e., *as well as the*, *and on the other hand*).

Referential bundles are employed to "make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important" (Biber et al., 2004a, 384). The pragmatic functions of these bundles are divided into five groups: (1) identification/focus referential bundles which are used to focus readers'/listeners' attention on the noun phrase after the bundle (i.e., for those of you who came late I have the quiz), (2) imprecision bundles which indicate vagueness of reference in classroom teaching and conversation such as and thing like that, (3) attribute specifying bundles that can specify the quantity of the following noun phrase (i.e., a lot of the, than or equal to) or frame the tangible/intangible properties of the noun phrases such as the nature of the, the size of the, (4) time/place/text referential bundles which are used to indicate specific places, times or parts of the text itself, and (5) multi-functional bundles which may have time/place/text referential bundles at the same time.

2.5. Lexical bundles and research findings

Levy (2003) set out to analyze the variations of lexical bundles between two groups of writings, essays written by professional writers and those written by students, hence analyzing the role of proficiency in and familiarity with particular genres reflected in the use of lexical bundles. Students' essays were categorized based on their writing levels. Her findings revealed that both professional and student writers used more academic bundles in their written samples than conversational bundles, and that both professional and proficient students, those scoring higher in the level-determination test, used more lexical bundles to structure discourse than non-proficient college students.

Therefore, it could be stated that the role of language proficiency and genre familiarity in the use of lexical bundles is prominent, hence revealing that bundles deserve to be considered as a valid indicator of students' proficiency in natural language use. The findings of her research also point to the fact that proficient

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

student writers tended to quote and paraphrase the source material more than the non-proficient ones.

Cortes (2004), first investigated 4-word bundles in the published journals of two academic disciplines (history and biology), and categorized them structurally and functionally, and then focused on students' use of these lexical bundles. The results showed that different lexical bundles have been used to perform the same function. In the case, referential bundles, all bundles belonging to time markers in the history corpus referred to years or time periods in which historical happenings took place, whereas in the biology corpus those bundles were used to refer to different periods of time, main stages in the evolutionary or developmental processes of different biological phenomena.

In the same vein, Biber et al. (2004) analyzed the frequency of lexical bundles in four different academic genres: classroom teaching, conversation, academic prose, and textbooks. The analysis of the frequency of lexical bundles across these for genres shows an outstanding difference among them, with classroom teaching using lexical bundles more than other genres and academic prose the least Figure 2.1.

Later they (Biber et al. 2004) presented a functional taxonomy for the extracted bundles, comprising 'stance expressions, discourse organizers, and referential expressions'. They found that the use of lexical bundles in classroom settings is quite different in terms of frequency and function. According to their study, classroom teaching uses more 'stance' and 'discourse' organizing bundles than conversation does, but at the same time, classroom teaching uses more referential bundles than academic prose.

Following her previous investigations, Cortes (2007) set out to investigate the role of explicit teaching of lexical bundles in native university students' writing performances. To compare the effectiveness of instruction on the use of such bundles, pre, and post-instruction analyses were carried out on students' class assignments. In addition, counterpart linguistic expressions functionally similar to lexical bundles were analyzed in students' final written production for the course. Based on the results of her study, she argues that students' use of target bundles (those selected for the study) in their writing for the history class both before and after instruction was extremely rare, hence explicit instruction did not make any significant difference in students' writing performances. Although, through emails and interviews, students mentioned that these instructions had increased their awareness towards the function, and the role of these bundles in their successful writing performance.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Chen (2008), in another study, investigated any significant difference between Electrical Engineering introductory textbooks and counterpart ESP textbooks in terms of the lexical bundles and their pragmatic functions. In order to conduct the comparison, he used Electrical engineering introductory textbook corpus (EEITC) and the English for Specific Purposes textbook corpus (ESPTC). The functional taxonomy of bundles is that of Biber et al. (2004). Her research results suggest that lexical bundles, functioning as discourse builders, serve to construct the body of knowledge and establish the reader/author relationship in Electrical Engineering introductory books. The results also demonstrate that there is a significant variation between the entry-level discipline and ESP textbooks.

As far as grammatical taxonomy is concerned, EEITC tends to use more phrasal lexical bundles than clausal bundles that help to construct explicit and exact meaning in introductory textbooks. And functionally speaking, EEITC made use of 'referential bundles' more frequently than 'stance bundles' and 'discourse organizers'. Probing the difference between two textbook genres, there is not any significant difference between ESP textbooks and Electrical Engineering textbooks in terms of the functional types used, though ESPTC has a much narrower scope of functional subcategories than EEITC; ESPTC covering 8 subcategories and EEITC 13. However, considering the frequency of forms, ESP textbooks use much fewer bundles than Electrical Engineering textbooks.

3. Methodology

3.1. The corpus of the study

The overall corpus of the study is comprised of 100 PhD dissertations and Masters theses in the field of applied linguistics published freely—either in part or whole—in the freely accessible internet data sources or received from some colleagues in person. Since the focus of this study is comparing part-genre sections on the one hand, and because, as mentioned by Cortes (2004), it should be number of words which are identical not the number of texts, the researcher ended up with 80 abstract sections (252,750 words), 150 introduction sections (252,219 words) and 200 conclusion sections (252,084 words). It should be mentioned that in order not to interfere with the natural presentation of language in these sections, all the comprising part-genre sections are in the original complete state with no part or sentence deleted out of the original texts.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

3.2. Instrumentation

In order to extract 4-gram lexical bundles (in line with the literature focusing mostly on 4-word bundles) from the three part-genre corpora of the study, the KfNgram software (version 2002-2007, by W. H. Fletcher) was employed. This is really sophisticated but free software which singles out bundles which run only on text-formatted files. This software not only presents the list of extracted bundles but also gives their frequency (i.e. both types and tokens).

One of the important criteria in the relevant literature is the frequency cut-off point of extracted bundles, meaning that bundles with what frequency count should be qualified as lexical bundles. A range of 20 to 40 per million words has been used in the literature with 40 per million being quite a conservative strict approach (see Hyland 2008). In the same vein, the frequency cut-off point of 40 per million was selected for the purpose of this study because in this way only highly frequent bundles are extracted which make the data more valid in nature. Therefore, as the number of words in each part-genre section of this study is about 250,000 words, the ratio is equal to 10, that is only bundles occurring at least 10 times in each part-genre corpora were considered for further analysis.

3.3. The functional framework

As stated above, the functional framework proposed by Hyland (2008) is preferred to that of Biber et al. (1999), for, as put forth by Hyland, Biber's framework is more useful for huge spoken and written corpus rather than smaller genre-based analyses of written texts. Hence, in this study, Hyland's functional framework has been adopted, categorizing bundles as follows:

Research oriented bundles: helping writers to structure their activities and experiences of the real world.
(a) Location: indicating time/place (at the beginning of, at the same time)
(b) Procedure (the use of the role of the purpose of the)
(c) Quantification (the magnitude of the, a wide range of)
(d) Description (the structure of, the size of the)
(e) Topic (related to the field of research)

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Text-oriented bundles: dealing with the organization of the text and its meaning as a message or argument.

(a) Transition signals: establishing additive or contrastive links between elements (in addition to the, on the other hand)
(b) Resultative signals: marking inferential or causative relations between elements (as a result of, it was found that)
(c) Structuring signals: text-reflexive markers which organize stretches of discourse or direct reader elsewhere in the text (in the present study, in the next section)
(d) Framing signals: situating arguments by specifying limiting conditions (in the case of, with respect to the, on the basis of).

Participant oriented bundles: focusing on the writers or reader of the text.

(a) Stance features: conveying the writer's attitudes and evaluations (are likely to be, may be due to, it is possible that)
(b) Engagement features: addressing readers directly (it should be noted that, as can be seen)

4. Data analysis

After the coding of the data (i.e. categorization of extracted bundles according to the frequency, functions and grammatical structures in the three part-genre corpora), the data were analyzed in two interpretation fashions. As for the frequency and functional comparisons of the three part-genres of abstracts, introductions, and conclusions, the statistical test of Chi-square was employed to probe any significant differences across the three part-genres. However, these three part-genres were compared descriptively as far as grammatical structures of bundles were concerned because there is no rigorous grammatical framework based on which we could build a statistical analysis.

4.1. Procedure

After each part-genre section of the study was selected out of all dissertations and theses, they were combined together to form a unified text consisting of, as stated

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

above, 100 abstract, introduction and conclusion sections. As a matter of fact, these three unified files comprised the studies three part-genre corpora. Later these files were turned into text files so that the employed software could accept it as an input.

Defining the required above-mentioned criteria (i.e. the frequency cut-off point and the number of word strings) in the software, the 4-word bundles occurring at least 10 times in each part-genre corpus were extracted and identified. Next, these extracted bundles were coded and categorized for functional and grammatical comparisons. Finally, to see if the use of lexical bundles is sensitive to different part-genres in terms of the frequency and functional types, the statistical test of Chi-square was employed in SPSS software.

5. Results

5.1. The Frequency comparisons across part-genres

As mentioned above in the methodology section, first, the three part-genres of abstract, introduction, and conclusion were compared in terms of the frequency of both types and tokens of 4-word lexical bundles occurring at least 10 times in each part-genre corpus (see Appendix 1 for the list of extracted bundles in all three part-genres).

Table 1 shows the frequency of extracted 4-word bundles across the three part-genres of applied linguistics' dissertations and theses. As it is depicted in the table, abstract enjoys the highest amount of bundles with regards to both the types of bundles and the tokens of bundles. This part-genre was identified in 186 bundle types (i.e. different bundles) and 3475 tokens in total (i.e. the total number of types with their frequency).

Table 1. The frequency of 4-world lexical bundles across three part-genres

part-genre	abstract		introduction		conclusion	
Type/Token	Type	Token	Type	Token	Type	Token
Frequency	186	3475	101	1861	133	2089

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

The other parts, however, are closer and much less than the abstract part in terms of the frequency of extracted bundles, with the introduction sections having 101 bundles in types (1861 in tokens), and the conclusion section having 133 bundles in types (2089 in tokens). These results show that unlike the abstract part-genre, the other two major part-genres of dissertations and theses, i.e. introduction and conclusion, use more or less the same amount of formulaic language. Figure 1 shows also the bar graph presenting a better picture of the results.

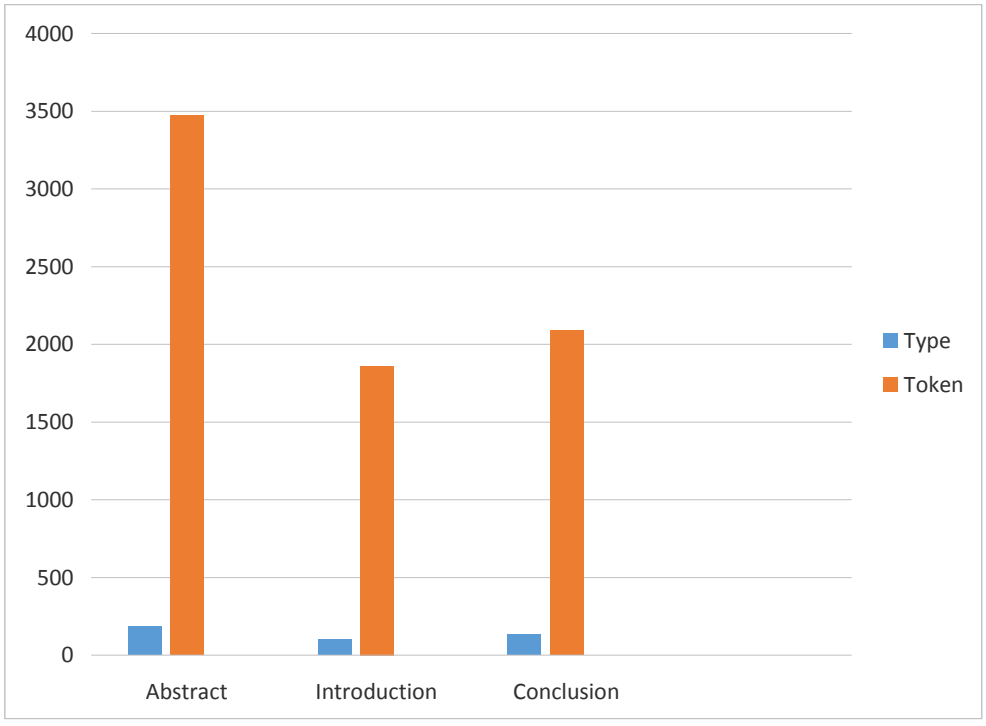


Figure 1. The frequency of bundles across three part-genres.

Nevertheless, it should be noted that the results of Chi-square statistical test reveal that there is no statistically significant difference among these three part-genres in terms of the frequency of 4-word bundles (see Table 2 for the results of the Chi-square test).

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Table 2. The results of Chi-square test for significant frequency variations

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.000[a]	10	.285
Likelihood Ratio	13.183	10	.214
Linear-by-Linear Association	1.367	1	.242
N of Valid Cases	6		

According to the statistical findings in Table 2, the first null-hypothesis is not rejected (p-value >0.05), meaning that, unlike the noteworthy difference between the abstract and the other two part-genres, there is not any statistically significant difference between the three relevant part-genres in terms of the frequency of 4-word lexical bundles. However, the lack of statistical difference does not mean that the observed differences could be easily ignored (as will be discussed in section 6) because the number of groups with their observed frequency was very low. Therefore, apart from the statistical interpretation of the quantitative data, a descriptive interpretation of the findings would be more promising.

5.2. The frequency comparisons across part-genres

The extracted bundles were functionally categorized based on Hyland's (2008) functional taxonomy. As with the frequency variations, there are also interesting functional variations across the three part-genres of applied linguistics' dissertations and theses (see Table 3 below). As it is depicted in the table, because the number of extracted bundles was different in each part-genre, the categorized bundles were then turned into percentages (as an example, 102 types of lexical bundles in the abstract part-genre corpus belong to research-oriented bundles out of 186 bundles in total, accounting for 55%).

Table 4. The functional distribution of bundles across three part-genres

Part-genre	Research Oriented	Text Oriented	Participant Oriented
Abstract	102 (55%)	77 (41%)	7 (4%)
Introduction	65 (65%)	27 (26%)	9 (9%)
Conclusion	59 (43%)	52 (40%)	22 (17%)

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

As the distribution of lexical bundles' functions in Table 4 shows, in general, the abstract part-genre enjoys a balanced presentation of different functions, while the introduction part-genre includes the highest amount of research-oriented bundles (65%), and the conclusion part-genre includes the highest amount of participant-oriented bundles (17%). As for the text-oriented bundles, the abstract and conclusion part comprise more or less the same amount of text-oriented bundles (41% and 40%, respectively). Figure 2 also shows the bar graph representation of the functional distributions of bundles.

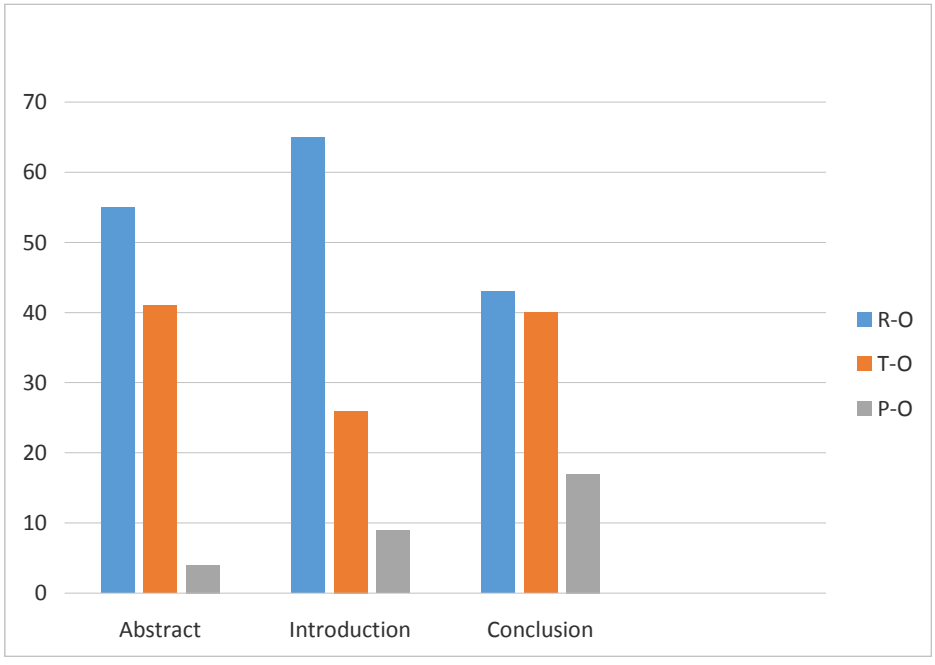


Figure 2. The functional distribution of lexical bundles across three part-genres.

As the overall functional distributions of bundles in Table 4 and Figure 2 show, in general, these three part-genres do not represent a high degree of functional variations. As a result, retaining the null-hypothesis, with the only explanation that conclusion parts enjoy a noteworthy difference in terms of the participant-oriented bundles, and introduction to lesser degrees the highest number of research-oriented bundles, leaving the abstract with a more balanced inclusion of all three functions.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

5.3. The structural comparisons across part-genres

Finally, three part-genres of the study were compared in terms of the structural patterns of the extracted 4-word lexical bundles. As stated before, the structural taxonomy proposed by Biber et al (1999) has been employed in this study as it is the only authentic available structural categorization in the research of lexical bundles.

Table 5 shows the structural distributions of bundles among the relevant part-genres of the study. As for the functional variations, since the number of extracted bundles occurring at least 4 times in each part-genre is different, the frequency of occurrence of structural types is then turned into percentages.

Table 5. The structural distribution of lexical bundles across part-genres

part-genres Structures	Abstract	Introduction	Conclusion
noun phrase + of	16 (10%)	2 (2%)	7 (6%)
other noun phrases	63 (34%)	41 (41%)	37 (28%)
prepositional phrases + of	30 (17%)	11 (11%)	17 (13%)
other prepositional phrases	19 (10%)	31 (30%)	33 (25%)
passive + prepositional phrase fragment	3 (1%)	0 (0%)	4 (3%)
anticipatory it + adj/verb	5 (3%)	1 (1%)	8 (8%)
be + noun/adjectival phrase	8 (5%)	3 (3%)	4 (3%)
others	37 (20%)	12 (12%)	18 (14%)

As it is depicted in Table 5, noun phrase structures (i.e. both noun phrase + of any other noun phrases) comprise the highest number of bundles in each part-genre (44% for abstract, 43% for introduction, and 36% for conclusion). However, as for both types of prepositional phrases, the abstract includes the lowest amount (27%), whereas introduction and conclusion part-genres have noticeably more prepositional phrases (41% and 38% respectively).

As for passive, anticipatory it, and be + noun/adjunctive construction, there is no noteworthy variations across the three-part genres. However, other types of constructions not included in the category seem to belong more to abstract part-genres rather than the other two (20% for abstract, but 12% and 14% for introduction and conclusion). Figure 3 could present a better picture for the structural distribution of lexical bundles.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

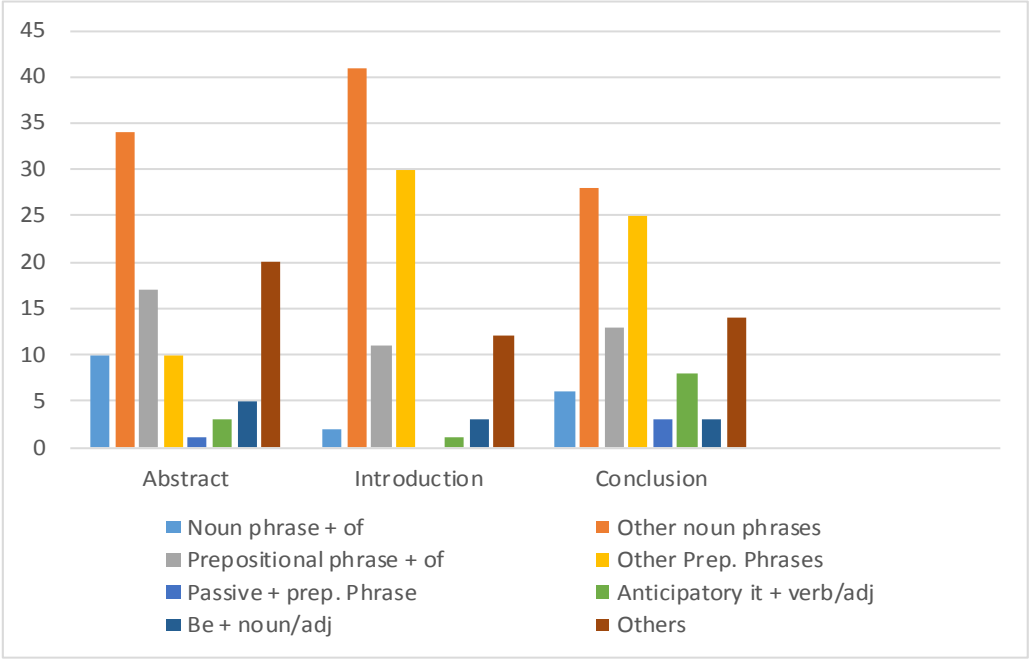


Figure 3. The structural distribution of lexical bundles across three part-genres

In general, the overall distribution of structural patterns of bundles across part-genres both in terms of the frequency of occurrence and percentages does now show significant variations, meaning that all patterns are more or less present in all part-genres. This point is also proved by the results of the statistical test of Kruskal-Wallis shown in Table 6 below (P-value of 0.784 > 0.5, hence retaining the null-hypothesis). It should be noted that although the types of bundles and their structural compositions used in each part-genre did not yield any significant statistical variations, it could be claimed that each part-genre actually enjoys specific lexical bundles' patterns, especially in the most frequent bundles. To clear the picture, Table 6 represents a sample of the most frequent 4-word lexical bundles extracted in this study.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Table 6. The most frequent lexical bundles of three part-genres of the study

Abstract		Introduction		Conclusion	
Bundles	F	Bundles	F	Bundles	F
as a foreign language	136	as a foreign language	52	on the other hand	60
English as a second	130	in the united states	49	the findings of this	51
English as a foreign	124	in the field of	45	findings of this study	47
as a second language	118	as a second language	44	in the united states	45
the purpose of this	90	English as a foreign	44	the results of the	43
a second language ESL	72	English as a second	43	in the present study	41
the results of the	72	the purpose of this	33	the results of this	38
a foreign language EFL	61	the significance of the study	31	in the present study	37
of English as a	55	at the same time	29	as well as the	34
of this study was	51	of the study the	28	results of this study	33
purpose of this study	50	as well as the	27	at the same time	30
as well as the	47	of this study is	27	in the current study	27
the extent to which	47	on the other hand	27	the findings of the	27
this study was to	46	one of the most	27	it is important to	26
in the context of	41	in the context of	26	the use of the	25
in the united states	41	for the purpose of	25	in terms of the	24
the findings of this	36	statement of the problem	25	with regard to the	24
of this study is	35	a second language ESL	24	in the process of	23
the findings of the	34	findings of this study	23	in the target language	23
in the field of	33	purpose of the study	23	in the use of	23
study was to investigate	33	purpose of this study	23	in the current study	23
findings of this study	32	to be able to	22	it was found that	22
the use of the	30	the findings of this	21	limitations of the study	22
the ways in which	28	the relationship between the	21	of this study was	22
results of the study	27	a large number of	20	the end of the	21
findings of the study	26	as a result of	20	in the case of	20
the relationship between the	25	is one of the	20	in the field of	20
this study investigated the	25	sense of efficacy in	20	on the part of	20
the results showed that	24	the extent to which	20	the fact that the	20
as a result of	23	on the basis of	19	at the beginning of	18

A careful look at this table reveals that less than 50% bundles in each part-genre are present among the 30 most frequent bundles of the other two part-genres. This

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

proves that part-genres are also more or less sensitive to the structural types of bundles they mostly employ.

6. Discussions

6.1. Discussions of the findings of frequency variations across part-genres

As the total frequency distribution of 4-word lexical bundles in all three part-genres of dissertations and theses showed, the abstract part-genre contained noticeably more bundles than the other two part-genres, in spite of the fact that this part-genre includes fewer words as compared with the others. The reason for this might be due to the fact that abstracts must follow a very constrained and controlled format of writing in terms of the number of words and the information they have to include in a very limited range of words (Hyland and Tse 2005), which is normally, as observed in the corpus of this study, 300 to 400 words in theses and 600 to 700 in dissertations.

According to Van Bonn and Swales (2007), abstract sections should present the overall type of information in a very limited condensed format which in turn makes them incline towards using fixed expressions with clear meanings. Therefore, it is possible to infer that it is due to this nature of abstract parts that the authors try to use as much formulaic language as possible so that they do not overstep the boundaries. However, the introduction and conclusion part-genres are not faced with such a limiting restriction as it is usually up to the writer to decide on the length of sentences they use to present their information. The findings of this study are also in line with Atai and Tabandeh's (2015) findings, stating that abstract part-genre enjoys the highest number of lexical bundles compared with other part-genres in applied linguistics internationally published articles. In conclusion, it is revealed that abstract part-genre in different academic genres, say articles, dissertations, and theses, uses the highest amount of formulaic language in the form of lexical bundles.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

6.2. Discussions of the findings of functional variations across Part-genres

The overall functional distribution of 4-word lexical bundles depicted in Table 4.3 shows that introduction and abstract part-genres contained the highest amount of research-oriented bundles (65% and 55%, respectively). Hyland (2008) defines this type of bundles as representing the content of the research and the structure of the research experience of the researchers. Hence, based on the nature of these bundles, it is not surprising that introduction part-genre in which the writer tries to establish a research basis for their research by referring to previously conducted related research includes this functional type of bundles more than the other part. As for the abstract part-genre, according to Hyland and Tse (2005), the introductory information is obligatory to set the scene for the reader; this makes abstract part-genre to also include a high number of research-oriented bundles.

Text-oriented bundles, on the other hand, are more prevalent in conclusion and abstract part-genres. These bundles, according to Hyland (2008), tend to reflect the discursive and evaluative patterns of argument and produce tolerance in readers through an ethical rather than cognitive progression. This definition is in line with the nature of the conclusion parts in which the writer tries to argue for their findings and keep the reader interested to follow their arguments through appropriately structured language.

Finally, as far as participant-oriented bundles are concerned, it is the conclusion part that includes noticeably more of these bundles than the other two part-genres (17% as compared with 4% in abstracts and 9% in introductions). Participant oriented bundles help the writer to express their epistemic and affective attitudes and judgments with regard to the conclusions they draw from the discourse (Hyland 2008). Therefore, it is not surprising that the conclusion part-genre in which the writer presents and summarizes the overall findings of their study enjoys the highest amount of participant-oriented bundles. It is through using these bundles that the writer could show how committed they are to what they say and claim by ‘stance bundles', and could address the readers to engage them with some special parts of the text by ‘engagement' bundles.

In conclusion, it is interesting to note that, with regard to the functional distribution of bundles, abstract part-genre are more or less in between in terms of three functional categories, meaning that they do not include any functions to the extreme degree. This could be attributed to the fact that this part-genre includes in nature some aspects of introduction and conclusion sections as well, or as stated

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

by Van Bonn and Swales (2007), it is an encapsulation of writers accompanying article or a type of advertisement inviting the readers to read the whole article.

6.3. Discussions of the findings of structural variations across part-genres

As it is shown in Table 4 there also exist slight, but noteworthy, structural variations across part-genres in terms of the grammatical patterns of extracted bundles. As for the noun phrase structures (both noun phrase + of structures and other noun phrases), all three part-genres used slightly the same amount of these structures (44% abstract, 43% introduction, and 34% conclusion). According to Hyland's (2008) functional taxonomy, one of the sub-categories of research-oriented bundles is *topic* which focuses on the field-related information in the research.
This type of bundles are not only in nature comprised of noun phrase structures but also based on the nature of the genre here (i.e. academic dissertations and theses) very prevalent in all parts of the genre especially in the abstract and introduction part-genre because, as stated above, the conclusion section usually tries to convey the overall findings of the study in the form of textual bundles for coercive arguments.

As an example, taking a look at Table 7 shows that the most frequent bundles in both abstract and introduction part-genres are mostly noun phrases referring to the topic (*as a foreign language, English as a second, English as a foreign, as a second language, a second language ESL* in abstract sections, and *as a foreign language, in the United States, in the field of, as a second language, English as a foreign* in the introduction sections). On the contrary, the most frequent lexical bundle in the conclusion section is a text-oriented prepositional phrase, *on the other hand*, accompanied by other types of research-oriented noun phrase bundles. This finding is also in line with Hyland's (2008) findings showing that noun phrase structures are the most prevalent lexical bundles' patterns in articles, dissertations, and theses of different academic disciplines.

Whereas noun phrase structures are more common in abstract and introduction sections, prepositional phrases (both prepositional + of and other prepositional phrases) are more common in conclusion and introduction part-genres, rather than in abstracts (41% in the introduction and 38% in conclusion, but only 27% in the abstract). This pattern belongs more to the text-oriented bundles in which textual transitions, structuring signals, and framing bundles are used to establish connections within the text, organize the presentation of the materials and limit the arguments respectively (Hyland 2008).

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

As a matter of fact, these functions are mostly performed by prepositional phrases, such as *on the other hand* in the present study, and *in the case of* in our bundle list. Therefore, more prevalent in the conclusion and introduction part-genres as abstract sections usually are so limited in the number of words that should directly refer to research findings by research-oriented bundles and noun phrase structures.

There are also some other types of structural patterns which, in line with previous studies (e.g. Hyland 2008), are comparatively less common in all three part-genres. For example, the passive construction and anticipatory it in abstract and introduction are rare (4% together in abstract and 1% in the introduction), but comprising 11% of the conclusion part-genre bundles. The use of the passive and anticipatory it construction more in conclusion might be due to the fact that these patterns are used to refer to findings of the study and the way the author presents their findings to the reader, as with Resultative text-oriented bundles which are more prevalent in conclusion part genres together with some participant-oriented bundles engaging the reader.

7. Summary and Conclusions

7.1. Frequency variations across part-genres

The first part of the study focused on the frequency of extracted 4-word lexical bundles occurring at least 10 times or higher across three part-genre corpora (i.e. abstract, introduction, and conclusion). The overall frequency distribution of bundles showed that the abstract part-genre outstandingly included more bundles than the other two part-genres, almost twice as many as the number of bundles in introduction and conclusion. As for the types of bundles, 186 4-word bundles with the frequency cut-off point of 10 were observed in abstracts, whereas the number for introduction was 101 and for conclusions was 133. This variation was also reflected in the number of tokens with abstract part-genre possessing 3475 bundles in total, while introduction and conclusion part-genres had 1861 and 2089 bundles in total, respectively.

Although surprisingly the statistical test of Chi-square did not yield any significant variations—which might be due to the very few dependent variable groups on the one hand and the unadjusted nature of statistical tests for corpora analysis (see Bestgen 2013, in this regard)—the prevalence of lexical bundles in abstract part-genre truly reflects the discursive nature of this main section of

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

academic writing (Van Bonn and Swales 2007). This part-genre of dissertations and theses consists of a very limited number of words, compared with other parts and, within such a limited linguistic context, the writer has to deliver the required fixed amount of information on the importance of the research, how it is conducted, and the major findings of the research. It seems that the best way to present the highest amount of information within a limited space is to rely on using formulaic language in the form of lexical bundles.

7.2. Functional variations across part-genres

The results of functional categorization of 4-word lexical bundles based on Hyland's taxonomy (2008) showed that research-oriented bundles are mostly prevalent in abstract and introduction part-genres. The nature of these bundles is in fact quite in line with this finding as these types of bundles focus on the content of the research and the structure of the area of investigation. Abstract sections should include information about the research itself and the way it is conducted. Introduction sections also try to provide a baseline for the research at hand by referring to the general research framework (Hyland and Tse 2005).

In contrast, conclusion part-genre comprised the highest number of text-oriented bundles. According to Hyland (2008), as already stated above, these lexical bundles represent the discursive and evaluative patterns of argument and produce tolerance in readers through an ethical rather than cognitive progression. Such characteristics is quite on a par with the nature of conclusion section in which the writer tries to wrap up the overall findings of their study and present the information in a way that the reader could easily follow the stream of information. This is the requirement that makes the writers try to establish a closer relationship with the readers in this part by using more participant-oriented bundles. That is why, according to the findings of this study, conclusion part-genre had the highest amount of participant-oriented lexical bundles.

In conclusion, in line with previous studies (e.g. Hyland 2008; Atai and Tabandeh 2015), as far as the overall functional distribution of lexical bundles shows, abstract part-genre employed all types of functions in a more balanced way than the other two parts. This might be due to the fact that this part-genre not only includes some aspects of introduction and conclusion but also it is representative of all parts of the research, hence including a balanced proportion of functions (Van Bonn and Swales 2007).

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

7.3. Structural variations across part-genres

The structural comparisons of lexical bundles in the three part-genres of applied linguistics dissertations and theses showed that abstract and introduction sections enjoyed the highest amount of noun phrase structures (44% and 43%, respectively) compared with the conclusion sections, having only 34% noun phrase bundles. This finding is because of the fact that these two part-genres also comprised research-oriented bundles which in nature belong more to noun phrase structures. As an example, the *topic* bundles—as one type of research-oriented bundles—is mostly represented through noun phrase structures (e.g. English as a second, English as a foreign). However, as for the lexical bundles made up of prepositional phrases, it is the conclusion part-genre which included the highest amount of these structural type of bundles because this sub-section of dissertations and theses consists of text-oriented bundles that are realized through prepositional phrase structures (e.g. *on the other hand*, *in the present study*, *in the case of*). In fact, these structures help the writer to present their findings in smooth fashion helping the reader to follow the flow of information (Hyland 2008).

The other types of structural patterns were not that common in all three part-genres of the study (passive and anticipatory it constructions) with the only exception that conclusion used more of these bundles than the other two part-genres (11%). The use of passive and anticipatory *it* construction more in conclusion might be due to the fact that these patterns are used to refer to findings of the study and the way the author presents their findings to the reader, as with Resultative text-oriented bundles which are more prevalent in conclusion part genres together with some participant-oriented bundles engaging the reader. For example, bundles such as it is shown that; can be found in; it should be noted; it is possible that, it is important to are present in the corpus of the study.

8. Pedagogical implications

There are some pedagogical implications that could be derived from this research. First, the findings of this research showed that in three major part-genres of the academically key genre of PhD dissertations and M.A. theses, i.e. abstract, introduction, and conclusion, it is the abstract that enjoys a high amount of formulaic language in the form of lexical bundles. Using almost twice as many 4-word bundles as the introduction and conclusion part-genres in a very condensed

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

linguistic context, the abstract proves to present a summary of the research in as few words as possible. Therefore, providing novice M.A. and PhD students with the generic formulaic nature of the abstract part-genre though explicit presentation and instruction of lexical bundles could help them to become more familiar with how this important part-genre is developed.

Not only were there frequency variations, but also the three part-genres showed functional variations as well. Abstract and introduction part-genres included the highest amount of research-oriented bundles whereas the conclusion part possessed the highest amount of text-oriented and participant oriented bundles. If students are familiar with the functional characteristics of these three part-genres, they would probably be more successful in presenting their tentative readers with the information common and expected in each part-genre while reporting their research in every stage.

As an example, it is possible to explicitly disclose to students—in EAP writing courses—that the best place to engage directly with the readers is where they are presenting the overall main findings of their research—namely, in conclusion, part-genre—because it is in this section that, by using text-oriented bundles, writers put the flow of information in a logical framework, and by using participant-oriented bundles they engage with their readers to mitigate their findings. Finally, the structures of bundles found in each part-genre shed some light on the more formal aspects of formulaic language. Syllabus designers and university instructors in EAP writing courses could devote some time to focus on grammatical aspects of writing part-genres in detail by putting the limelight on the formal properties of lexical bundles. If students know that it is mostly a noun phrase structure that is used to realize research orientated bundles, they could practice writing and form these type of bundles in their free writing activities. In addition, knowing that text-oriented bundles are mostly comprised of prepositional phrase structures could help them to be aware that conclusion part-genre includes the highest amount of text-oriented bundles, therefore a comprehensive command of using prepositional phrases could lead them to be able to present their arguments better in the conclusion sections.

9. Suggestions for further research

In what follows, some possible tentative research avenues that could complete this research are introduced:

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

1. Some research has been conducted on the effects of L1 and L2 writing in different genres across different disciplines. However, a study focusing exclusively on the effects of L1 and L2 writing across part-genres could shed light on how native and non-native writers develop part-genres in isolation.
2. Investigating the role of explicit instruction in helping novice students and writers as future practitioners of the field to structure and develop main part-genres of different academic genres might also prove fruitful. This line of research is very important because it reveals whether all the findings in exploratory research of lexical bundles could actually be helpful in practical real-life situations.
3. Another interesting line of research—a more theoretical linguistic one though—is that lexical bundles across language could be compared in both spoken and written contexts across various genres and disciplines. Such a comparison shows how and to what degrees different languages make use of formulaic language in the form of lexical bundles to present information.

References

Ädel, Annelie and Britt Erman. 2012. Recurrent word combinations in academic writing by Native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes,* 31, 81-92.

Adolphs, Svenja and Valerie Durow. 2004. "Sociocultural integration and the development of formulaic sentences'". In *Formulaic Sequence,* ed. by Norbert Schmitt, 107-126. Amsterdam: John Benjamins.

Altenberg, Bengt. 1998. On the phraseology of spoken English: the evidence of recurrent word combinations. In *Phraseology: Theory Analysis and Applications*, ed. by Anthony Paul Cowie, 101-122. Oxford: Oxford University Press.

Atai, Mahmood Reza and Farhad Tabandeh. 2015. "Lexical Bundles in Applied Linguistics Articles: Exploring Writer, Sub-Discipline, and Sub-genre Variations." *Kharazmi University* 33-56. Tehran, Iran.

Bestgen, Yves. 2013. "The inadequacy of the chi-squared test to examine vocabulary differences between corpora." *Literary and Linguistic computing*, first published online April 15, 2013: doi:10.1093/llc/fqt020.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Grammar of Spoken and Written English*. Harlow: Longman.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)

Biber, Douglas and Randi Reppen. 2002. “What does frequency have to do with grammatical teaching?” *SSLA* 24: 199-208.

Biber, Douglas. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins.

Chen, Lin. 2008. *An investigation of lexical bundles in Electrical Engineering Introductory textbooks and ESP textbooks*. Unpublished Master’s Thesis. Canada Carlton University.

Chen, Yu-Hua and Paul Baker. 2010. “Lexical bundles in L1 and L2 academic writing.” *Language Learning and Technology* 14 (2): 30-49.

Cortes, Viviana. 2004. “Lexical bundles in published and student disciplinary writing: Examples from history and biology.” *English for Specific Purposes* 23: 397–423.

Cortes, Viviana. 2007. “Teaching lexical bundles in the discipline: An example from a writing intensive history class.” *Linguistics and Education* 17(4): 391-406.

Haswell, Richard. 1991. *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.

Hyland, Ken. 1999. “Talking to Students: Meta-discourse in introductory course books.” *Journal of English for Specific Purposes* 18 (1): 3-26.

Hyland, Ken. 2008. “As can be seen: Lexical bundles and disciplinary variation.” *Journal of English for Specific Purposes* 27 (2): 4–21.

Levy, Stacia A. 2003. *Lexical bundles in professional and student writing*. Unpublished Doctoral Dissertation, University of the Pacific Stockton, Canada.

Samraj, Betty. 2005. “An exploration of a genre set: Research article introductions in two disciplines.” *English for Specific Purposes* 24 (2): 141-156.

Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge University Press.

Van Bonn, Sarah and John Swales. 2007. “English and French journal abstracts in the language sciences: Three exploratory studies.” *Journal of English for Academic Purposes* 6 (2), 93–108.

Widdowson, Henry G. 2007. *Discourse analysis*. Oxford University Press.

BDD-A32206 © 2020 Transilvania University Press
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-15 20:40:41 UTC)