https://doi.org/10.18485/lsp.2025.6.ch24

Mirna Varga*
Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences in Osijek
Croatia
https://orcid.org/0000-0003-2318-4490

Višnja Pavičić Takač
Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences in Osijek
Croatia
https://orcid.org/0000-0001-8318-0235

THE PHRASEOLOGY OF RHETORICAL MOVES IN L1 AND L2 ENGLISH MASTER'S THESIS WRITING

Abstract: Academic texts are commonly associated with the structure of rhetorical moves that can be signalled by the use of formulaic multi-word sequences. Hence, rhetorical knowledge and the use of formulaic academic language may indicate writers' discourse competence, which is particularly important for L2 English novice writers. While the structure and discourse functions of multi-word sequences have been rather extensively examined in research paper writing in English, research on university student writing has been limited.

This study aims to identify four-to-six-word lexical bundles in the sub-corpora of the Discussion and Conclusion sections of masters' theses (n=40) written by Croatian and Serbian graduates and their UK- and US-based peers majoring in applied linguistics. The bundles are further analyzed based on the distinct rhetorical purposes of the moves in which they occur. The combined analysis of the move-bundle relationship may inform academic writing instruction, particularly in non-Anglophone settings.

Keywords: L1 English, L2 English, master's thesis, lexical bundles, rhetorical moves

* mvarga@ffos.hr

1. INTRODUCTION

One of the fundamental linguistic requirements for university students is the acquisition of disciplinary literacy (Hyland 2017, 20-21), which broadly includes mastering the disciplinary discourse. For master's thesis writers, this entails the development of proficiency in the rhetorical organization of the content and the fluent use of the disciplinary language, which includes not only the subject-specific vocabulary but also the genre-based formulaic language (Cortes 2004, 398; Hyland 2008a, 5; Lu and Deng 2019, 21). In thesis writing, the latter is commonly associated with the use of multi-word sequences, often termed *lexical bundles* (Biber et al. 1999, 989) or academic clusters (Hyland 2008b, 41), such as *on the other hand*, *previous studies have shown,* etc. However, for most university students, the use of formulaic language in academic writing may be challenging, particularly for L2 English students writing a master's thesis in English. An additional constraint often encountered in L2 English academic contexts, particularly at the (under)graduate level, is insufficient genre-based academic writing instruction that can allow students to acquire adequate rhetorical knowledge.

It is, therefore, justified to explore how L2 English majors use formulaic multi-word sequences across the thesis structure and to compare it with its congruent use by their L1 peers, which may broadly be considered the expected native-like language use. Examining and comparing L1 with L2 writing practices may reveal important insights into the distinct use of formulaic patterns across two language groups, which may be particularly beneficial for tailoring target academic writing instruction, most notably in English L2 academic settings.

2. GENERAL FEATURES OF LEXICAL BUNDLES

One of the essential features of formulaic multi-word bundles or clusters is that they are stored and retrieved from memory as wholes, which allows for easier and more efficient communication (Wray and Perkins 2000, 1; Hyland, 2018, 2). According to Hyland (2008b, 5-8), these strings of words co-occur more frequently than not, perform a range of discourse functions, and shape the distinctive characteristics of a particular genre. Biber et al. (1999, 989) define lexical bundles as recurrent word sequences of three or more words with a statistical tendency of co-occurrence. Most bundles are structurally incomplete and tend to vary across registers, meaning that different bundles are typically used in spoken (e.g. *I don't know why*) and written discourse (e.g. *on the other hand*) (Biber et al., 1999, 989-991). Bundles also show structural variations, in that those used in academic discourse mainly consist of noun phrases, prepositional phrases or anticipatory-*it* phrases, while bundles typically encountered in spoken discourse are predominantly verbal or clausal phrases (Biber et al. 1999, 991; Hyland 2018, 8).

To ensure that a word string is a lexical bundle rather than a product of an idiosyncratic use, researchers establish threshold criteria, defining a minimum number of occurrences and a range of texts in which the bundle needs to occur to be categorized as such (Cortes 2023, 220-221). For instance, Biber et al. (1999, 992) assigned the status of a bundle if a word sequence occurred at least ten times per million words in at least five different texts comprising a corpus of a particular register type. Although cut-off points may vary depending on the corpus size and the length of the lexical bundles (Biber et al. 1999, 992), frequency and range remain their defining features (Cortes 2004, 400; Cortes 2023, 221).

Lexical bundles are often identified and analyzed based on the number of their constituents, with three-word bundles being the shortest and most frequent sequences. They are usually integrated in four-, five- and six-word bundles, as in *the findings suggest (that)* (Biber et al. 1999, 990). In addition to the structural features, lexical bundles may vary depending on the functions they perform in the text. Thus, Cortes (2004, 401) identified several functions of bundles, including referential, referring to time and place (e.g. *at the same time*); text organizers that express meanings such as contrast, inference, etc. (e.g. *on the other hand*) and interpersonal functions which are concerned with expressing stance (e.g. *are more likely to*) or various other interactional meanings (e.g. *thank you very much*). One of the most cited taxonomy of bundle functions, relevant to the purposes of the present study, was proposed by Hyland (2008b, 49) who distinguished between research-oriented bundles, signalling different research activities (e.g. *a wide range of, at the same time*), text-oriented bundles, dealing with the textual organization of the content (e.g. *in the case of, in addition to the*) and participant-oriented bundles that are concerned with writer's attitudes and readers' engagement (e.g. *as can be seen, it is possible that*). As the present study deals with the use of lexical bundles in academic writing, the section that follows outlines previous findings relevant to its scope.

2.1. Research on lexical bundles in academic writing

Empirical studies on the use of lexical bundles have been mainly targeted at the identification of their structural and functional features, with the ultimate aim of illuminating the specific aspects of formulaic language use in particular genres, disciplines, or writer groups. Cortes (2004, 397-421) compared the use of lexical bundles in research articles and student papers in biology and history. The study revealed considerable variations in the use of bundles by writers in the two disciplines. Specifically, student writers used very few bundles identified in professional writing, which underscores the importance of instruction on the use and functions of target disciplinary phraseology. Hyland (2008b, 41-60) compared the use of 4-word academic clusters in professional and student writing, specifically in research articles, PhD dissertations, and master's theses

from different disciplines. The results showed that professional and student writing are characterized by the use of different types and structures of clusters that also perform distinctive discourse functions across different genres. For instance, students used a number of clusters that were not used by professionals or were used with lesser frequency. Unlike other writer groups, master's students used more clusters, which also showed greater variations. This may reflect student writers' (over)reliance on the conventional language use than is the case with professional writers who do not seem to strictly follow the expected norms in structuring arguments. The genres under study also differed in terms of the prevalent functions of bundles. Thus, research-oriented bundles were most frequently used in master's theses, suggesting student writers' focus on showing competence and control over different aspects of their research. By contrast, PhD dissertations and research articles showed similarity in terms of greater frequency of text-oriented bundles, which may indicate writers' greater concern with the organization of the text, aligned with the disciplinary conventions.

Apart from genre variations, previous research on the use of bundles has shown distinctive patterns in writings by English L1 and L2 student writers. Comparing the use of bundles between native professional writers and L1 English and L1 Chinese student writers writing in English, Chen and Baker (2010, 43) found that L2 student writers used fewer stance bundles than native writers, both students and experts. This suggests their lesser concern with expressing caution in constructing argumentation, which has been identified as a general tendency of L2 English writing (Chen and Baker 2010, 43). Lu and Deng (2019, 21-36) explored the use of lexical bundles in English-medium PhD abstracts written by Chinese and L1 English students. The findings showed that Chinese students used more bundles than their L1 peers in writing. However, in terms of the structure, most bundles were clausal rather than phrasal, suggesting L2 students' insufficient familiarity with academic writing conventions in English. Concerning the use of stance bundles, the analysis of the lexical choices revealed Chinese students' preference toward a more assertive tone as compared to L1 students who adopted a more tentative stance in their writing.

2.2. Lexical bundles as lexical signals of rhetorical moves

A more recent line of lexical bundle research has been oriented towards the relationship between the use of lexical bundles and the rhetorical structure of academic texts, in particular, the role of bundles in signalling rhetorical moves. Introduced in Swales' (1990, 141) seminal work on the rhetorical move structure of the Introduction section of the research article, rhetorical moves denote semantic and functional units of the text that perform distinct communicative purposes (Kanoksilapatham 2007a, 23). A move may consist of sub-units, termed steps that realize more specific functions of its overarching rhetorical purposes (Cortes 2013, 35). Corpus-based analysis on the rhetorical move structure,

primarily in research articles, has often resulted in move-step frameworks that may facilitate understanding of the typical content organization in disciplinary writing. Relevant to the present study is Yang and Allison's (2003, 376-379) move-step model of research articles in linguistics which was used in the previous analysis of the rhetorical structure of the Discussion and Conclusion sections in master's theses (Varga and Pavičić Takač 2024, 555-572). The move-step framework of the given sections is presented in Figure 1.

Moves and steps in Discussion
Move 1—Providing background information
Move 2—Reporting results
Move 3—Summarizing results
Move 4—Commenting on results
- S4.1 Interpreting results
- S4.2 Comparing results with the literature
- S4.3 Accounting for results
- S4.4 Evaluating results

Move 5—Summarizing the study
Move 6—Evaluating the study
- S6.1 Indicating limitations
- S6.2 Indicating significance/advantage
- S6.3 Evaluating methodology

Move 7—Deductions from the research
- S7.1 Making suggestions
- S7.2 Recommending further research
- S7.3 Drawing pedagogical implications

Moves and steps in Conclusion
Move 1— Summarizing the study
Move 2— Evaluating the Study
- S2.1 Indicating significance/advantage
- S2.2 Indicating limitations
- S2.3 Evaluating methodology

Move 3— Deductions from the research
- S3.1 Recommending further research
- S3.2 Drawing pedagogical implications

Figure 1. Move-step framework in the master's thesis in Applied Linguistics (Varga and Pavičić Takač 2024, 560, adapted from Yang and Allison, 2003. 376-379)

Even early descriptions of the rhetorical move structure of research articles (Swales 1990, 144) recognized that rhetorical moves can be signalled by the recurring linguistic patterns that may re-occur. For instance, Swales (1990, 155) noted that in indicating a gap in previous research in the Introduction section

(Move 2_Establishing a niche), writers typically use lexical negation by means of verbs such as *fail, lack,* or negative quantifiers, such as *no, little*, etc. Subsequent studies on the lexico-grammatical features of moves have provided more detailed linguistic accounts of the rhetorical move structure predominantly in research articles (e.g. Kanoksilapatham 2007b, 73-121).

The lexical bundle approach, however, is more specific in its focus, as it aims to illuminate the formulaic patterning characterizing the underlying rhetorical purposes of the moves. According to Cortes (2013, 35-36), the relevance of connecting studies of linguistic features and the text structure of academic genres is reflected in the fact that both serve as building blocks in the construction of discourse, which justifies the need to examine the relationships between moves and bundles in which they occur. In other words, lexical bundles serve as the characteristic lexico-grammatical sequences, whereas rhetorical moves serve as the structural building blocks used to organize the schematic structure of a text (Cortes 2013, 35-36). Cortes (2013, 36) examined the use of lexical bundles of various lengths across the rhetorical structure of Introduction sections in research articles related to a range of disciplines. The findings showed that certain bundles can be used as triggers introducing a rhetorical move or step, such as *little is known about* lexical bundle to announce the *Indicating a gap* move. By contrast, other bundles function as comments that complement other expressions that signal the moves (e.g. *in the sense of*). Yang (2022, 147) focused on the 4-word bundles across the IMRaD structure of research articles in two branches of psychology. Based on Hyland's (2008b, 49) classification of lexical bundle functions, the results showed functional variabilities of bundles in different sections, reflecting the distinctive communicative purposes of the respective moves (Yang 2022, 152). Thus, text- and participant-oriented bundles were most salient in Introductions in Discussions, while research-oriented bundles prevailed in the two middle sections. Additionally, the findings illuminated high-frequent bundles signalling distinctive moves across the main article sections (Yang 2022, 156-158). To illustrate, Move 1 (Introduction) *Establishing a territory* showed the highest frequencies of bundles, such as *studies have shown that*, *there is evidence that*, etc., which are typically used to point to previous research findings.

Overall, in combining corpus- and genre-based analysis, the bundle-move research has been recognized as particularly valuable for academic writing research and teaching purposes (Cortes 2013, 42; Lu, Yoon and Kisselev 2021, 64-65). This is primarily reflected in its potential to illuminate the characteristic phraseology in the given genre, which in turn may provide authentic language data to inform writing instruction relevant to students' disciplinary needs. However, despite its pedagogical significance, the move-bundle research has been relatively scarce (Yang 2022, 145) and mainly focused on research articles (Lu, Yoon and Kisselev 2021, 64). By contrast, student genres have been largely understudied, particularly from the L1 English vs. L2 English writing perspective.

2.3. Present study

To fill the above-mentioned research gap, the present small-scale study aimed to explore the use of lexical bundles across the move structure in the Discussion/Conclusion sections in the L1 English and L2 English master's thesis in applied linguistics. Based on our previous research (Varga and Pavičić Takač 2024, 555-572), these sections were selected for the analysis due to their argumentative and interpretative rhetorical purposes, which is reflected in high density of rhetorical moves.

Due to the small corpus size (N=40 D/C sections), the present study is of a preliminary nature. It represents the first step of what attempts to be an extensive examination of the move-bundle relationship across the concluding sections of master's theses written by L1 and Croatian L2 graduates writing in English. The main aim of the present research is to initiate the identification of the patterns of similarities and differences in L1 English and L2 English student writers' use of academic formulaic expressions to signal the rhetorical moves in the given thesis sections. Against this background, the study aims to address the following research questions:

1. Which lexical bundles are used across the move structure in the Discussion/ Conclusion sections in the L1 vs. L2 master's theses?

2. Which discourse functions do lexical bundles serve and what is their distribution across the two sub-corpora?

3. METHODOLOGY

3.1. Corpus sample

The present study is based on the corpus previously compiled for the analysis of the rhetorical structure of the Discussion (D) and Conclusion (C) sections in L1 and L2 master's theses (Varga and Pavičić Takač 2024, 558-559). The study was a part of a more extensive research project on the use of metadiscourse in L1 English vs. L2 English academic writing (Varga, Kovačević and Molnar 2020, 145-154). The present corpus consists of two sub-corpora, each consisting of 20 D/C, totalling nearly 100,000 words (see Table 1).

Table 1. Word count in L2 sub-corpus and L1 sub-corpus

	L2 sub-corpus		**L1 sub-corpus**	
	number of words	average per section	number of words	average per section
Discussion	19309	965.45	56740	2837
Conclusion	9166	458.3	13182	659.1
Total	**28475**	**1423.75**	**69922**	**3496**

The original L2 corpus was collected from the personal folders of L2 students' supervisors, comprising unrevised thesis drafts. This methodological decision allowed us to explore L2 students' linguistic and rhetorical choices before supervisors' interventions. L1 corpus consists of the samples of defended MA theses in applied linguistics that were extracted from open-access institutional repositories across British and American universities. The collected theses were based on the empirical studies, and all followed the conventional IMRaD structure with the section headings Discussion and Conclusion. As for the terminological qualification of L1 English and L2 English student, we followed in footsteps of our previous study (Varga, Kovačević and Molnar 2020, 151-152; Varga and Pavičić Takač, 2024, 559). The term "L1 student/writer (English)" is used here to denote writers whose L1 background is unknown to us. Yet, their language proficiency is officially acknowledged based on their thesis defence in an Anglophone university setting. Conversely, "L2 student/writer (English)" refers to EFL graduates who submitted their theses at a university in Croatia and Serbia.

3.2. Procedure

Based on the previously designed corpus, we first compiled the textual files of all identified moves and steps in D/C sections and uploaded the files in the Sketch Engine. To extract the lexical bundles in each rhetorical move, we used the N-gram tool. In line with prior research (Cortes 2013, 37), the present study focused on the use of four-, five- and six-word bundles, while shorter bundles were excluded from the analysis given their embediment in the structure of longer ones (Biber et al., 1999: 995).

It is important to note that the present study is descriptive and preliminary in that it aims at testing feasibility and refining the methodology of data analysis. Its results would subsequently be applied in the study involving a larger corpus, allowing for a more rigorous examination of the move-bundle relationship and leading to generalizable conclusions. Additionally, given the small corpus, we could not follow some conventional cut-off points for assigning the status of a lexical bundle (e.g. 20 tokens per mil words) and those for a range (e.g. 5 texts). Instead, the cut-off point established for extracting the bundles was min. 3 tokens per 2 texts. By the same token, conducting inferential statistical analysis was not feasible. We calculated the raw frequencies of the lexical bundles meeting the set criteria in each rhetorical move across the two sub-corpora. As the frequencies of bundles were very low in individual rhetorical moves in Conclusion sections, we calculated their frequencies at the level of the Conclusion section as a whole. Following Hyland's (2008b, 49) taxonomy, we also assigned the discourse functions of bundles and calculated their distribution (%) in both sub-corpora.

4. RESULTS AND DISCUSSION

Raw frequencies of lexical bundles in the Discussion/Conclusion sections in both sub-corpora are shown in Table 2. The results indicate that L1 students used more bundles when writing the given sections than their L2 peers. However, when it comes to the centrality of bundle types, the results showed similar tendencies in that four-word bundles were predominant in D/C sections in both sub-corpora, while the frequencies of five- and six-word bundles were considerably lower. At the most general level of analysis, this distribution is expected as it reflects the general features of bundles whose frequencies decrease with their length (Cortes 2013, 37), resulting in longer bundles being generally less common in language use (Biber et al. 1999, 992).

Table 2. Raw frequencies of lexical bundles (tokens) in L1 and L2 sub-corpus

	L1 sub-corpus			**L2 sub-corpus**		
	4-word bundles	5-word bundles	6-word bundles	4-word bundles	5-word bundles	6-word bundles
Discussion	541	115	115	241	69	15
Conclusion	42	4	4	32	32	11
Total	**583**	**119**	**119**	**273**	**101**	**26**

The distribution of the lexical bundles across the rhetorical moves in the Discussion and Conclusion sections in L1 and L2 sub-corpora is presented in Figure 1 and Figure 2, respectively, while the raw frequencies of bundles per moves are displayed in Appendix 1. As can be seen, the breakdown of the results shows the uneven distributions of bundles across the rhetorical move structure in the Discussions in both writer groups. In other words, the use of bundles was generally associated with specific rhetorical moves, while in other moves, they were rarely or not used at all. Despite this shared tendency in both sub-corpora, bundles were more dispersed across the rhetorical moves in L1 Discussions. More specifically, four-word bundles were absent in 2 moves (M 6.1. and M7.1), and longer bundles in 7 moves/steps. By contrast, in L2 sub-corpus, four-word bundles were absent in 3 moves (M4, M 6.1. and M7.1). There were no exponents of five-word bundles in 9 moves/steps and six-word bundles were absent in 10 moves/steps. Though L2 writers used slightly more bundles to signal moves in the Conclusion, their distribution is similar in both sub-corpora.

As for the features of the use of bundles in L1 sub-corpus, the results showed the highest frequencies of bundles in *M2 Reporting results*, followed by *M1 Providing background information* (Figure 2). Additionally, a substantial use of bundles was reported in two steps realizing *M4 Commenting on results*, i.e. *M4.2 Comparing results with literature* and *M4.3 Accounting for results* as well as

in *M7.2 Recommending further research*. These results broadly correspond to the findings of our previous study on the rhetorical move structure in D/C sections that showed the centrality of identical moves (Varga and Pavičić Takač 2024, 563). High density of bundles in the most salient D/C moves provides further support to the claim that in discussing their results, master's students were mostly focused on reporting and commenting on their results, while evaluating the study and drawing deductions were less prominent purposes. The latter is reflected in low frequencies (< 20) of four-word bundles in most remaining steps/moves and almost no occurrences of longer bundles.

When it comes to the use of bundles across the move structure in L2 D/C sections (Figure 3), the results showed that writers used most bundles in *M2 Reporting the results*, which matches the L1 findings. However, unlike a relatively similar distribution of bundles across a few additional moves in L1 sub-corpus, in L2 sub-corpus bundles were predominantly clustered in a single move. Nevertheless, despite the low frequency of occurrences, the results showed that L2 writers generally followed similar patterns, with the most prominent use of bundles associated with moves similar to those observed in L1 writing.

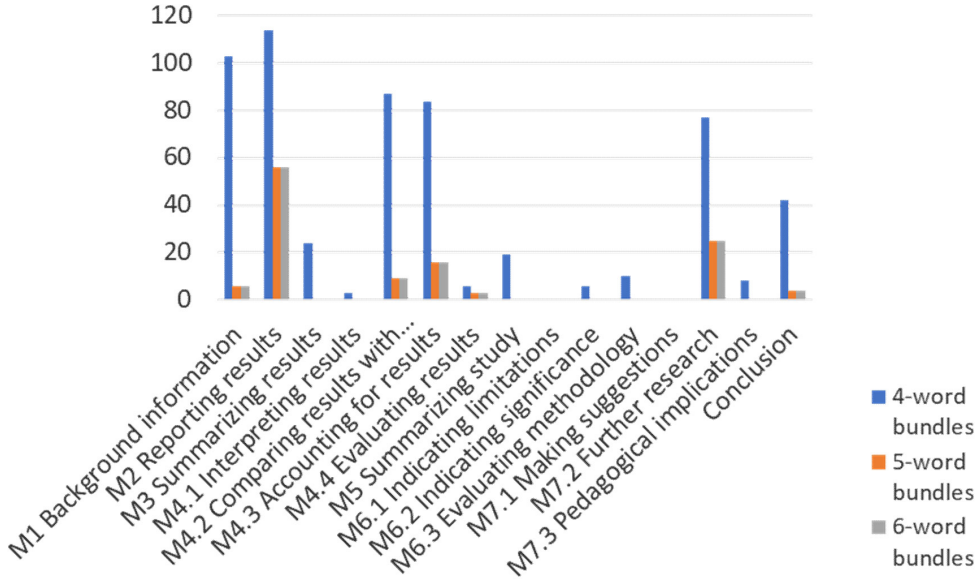


Figure 2. Lexical bundles across moves in L1 sub-corpus

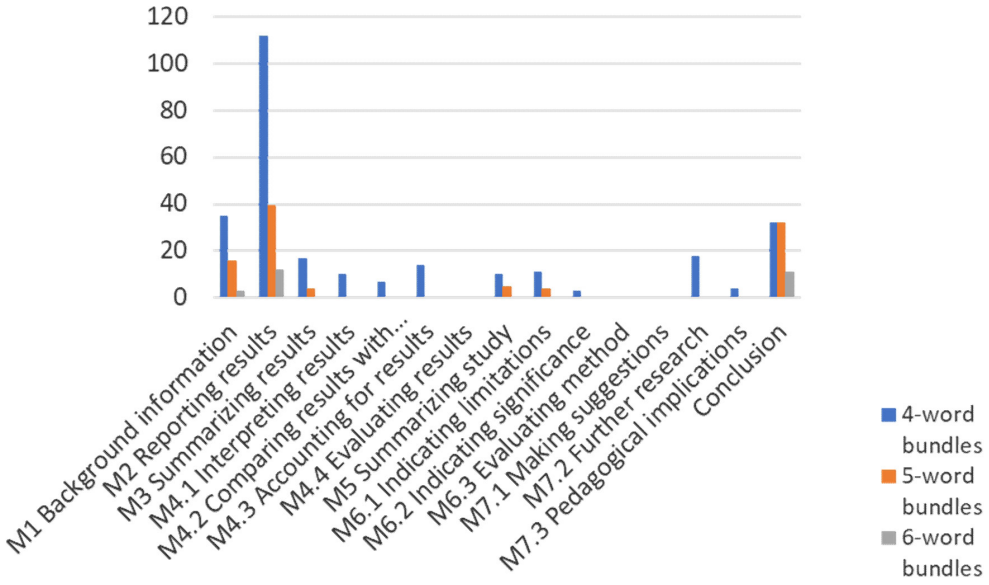


Figure 3. Lexical bundles across moves in L2 sub-corpus

Some of the exponents of lexical bundles used to signal the most prominent moves across both sub-corpora are provided below. We selected illustrative bundles that may be considered typical formulaic examples of the moves in which they were identified, and we excluded those that are specific to the topics under study (e.g. *English and Edinburgh English, induce faster translation latencies, the second language group,* etc.).

M1_Providing background information
L1: *in the present study, of the target word, this study was to, x is said to be, study was designed to, more exposure to the, used in this study, etc.*
L2: *aim of this research was to, x of this study was to, research question aimed at, main aim of this*
M2_Reporting results
L1: *on the other hand, can be seen in, it is important to, found to be statistically, the effect size between, percent of the respondents, etc.*
L2: *on the other hand, the results have shown that, results show that there is, the results show that, when it comes to, statistically significant correlation between, etc.*
M 4.2_Comparing results with the literature
L1: *the results of the, not support the prediction, is in accordance with, did not support the, this finding is in, is in line with, etc.*

L2: *is in line with, the results of the, are in accordance with, are similar to the, etc.*
M 4.3_Accounting for results
L1: *may have contributed to the, factor that may have contributed, it is possible that, to the fact that, the participants in the, in the present study, to focus on the, have been affected by, the fact that the, it is likely that, etc.*
L2: *it is logical that, their perception of their, in the fact that, etc.*
M7.2_Recommending further research
L1: *it would also be prudent to, it would be interesting to, it would also be, the use of the, research is needed to, needed to determine if, in a future experiment, be interesting to see, etc.*
L2: *it would also be, it would be useful, their willingness to communicate, etc.*

Conclusion
L1: *of this study was to, as a result of, in the case of, the early stages of, the effect of the, etc.*
L2: *it can be concluded that, the results of the research, when it comes to, a positive effect on, etc.*

Figure 4. Examples of lexical bundles across L1 and L2 sub-corpora

Due to space limitations, a complete account of the lexical bundles identified in each sub-corpus cannot be provided. However, even from the limited list provided in Figure 4, it is evident that student writers are broadly familiar with the phraseology commonly associated with the rhetorical functions of moves that are typical of the Discussion section. A case in point is the bundle *is in line with*, which is typically used to comment on results, particularly when establishing the relation between the obtained and previous findings, as reflected in M4.2_Comparing results with literature:

This finding is in line with La Heij et al. (1996) who found that the magnitude of the semantic effect in backward translation is greater than in forward translation. (ENG6_M4.2)

Another example is related to the use of bundle *it would be useful*, which is a common multi-word sequence when providing *Recommendations for further research* (M7.2).

It would be useful to explore in future research whether there are differences in the extent to which these factors affect vocabulary learning through different social media. (ENG3_M7)

The comparison of the findings between two writer groups indicates that L1 and L2 student writers generally make use of a different set of bundles, which

may partly be attributed to different topics and scopes of their theses. This is reflected in very few overlapping instances of the bundles used in the same L1 and L2 moves, such as: *on the other hand* in *Reporting results* (M4.1), *it would also be* in *Recommending further research* (M7.2), and *is in line with in* in *Comparing results with the literature* (M4.2).

Apart from the use of substantially more bundles, L1 student writers also used more heterogeneous bundles to signal specific moves. This was particularly prominent in the use of a variety of bundles when *Accounting for results* (M 4.3), as shown in the following examples:

It is possible that these particular students still felt very dependent on outside sources when it comes to language learning. (ENG29_M4Acc)
Another potential factor that may have contributed to the noticeable improvement among participants from the experimental group is knowledge of L2 vocabulary. (ENG21_M4Acc)

A range of bundles marking stance in this move indicates that, apart from reporting results, L1 students were concerned with interpreting and evaluating their findings from different angles. This may suggest their higher awareness of the central rhetorical purposes of the given thesis section as compared to L2 peers who seem to place more importance on reporting rather than discussing results, as in:

The results show that English majors' perceived oral proficiency levels were significantly higher than the perceived oral proficiency levels of the other two subgroups of respondents, with an average perceived oral proficiency score of 4.57. (HR19_M2)

L1 students' higher proficiency in the use of the target bundles may be further corroborated by a greater frequency of overlaps with bundles reported in expert academic writing, such as research articles and doctoral dissertations (Hyland 2008b, 51). Unlike L2 writing, L1 writing under study, thus, seems to show more resemblance to professional, higher-ranked academic genres. Some overlapping instances of bundles used by student writers in the present study and those previously identified in research on research articles, PhD, and master's theses (Hyland, 2008b, 51) were underlined in Figure 4. To sum up, the findings suggest that compared to L2 writers, L1 writers seem to comply more with the rhetorical conventions of the target genre, showing more competence in using the formulaic phraseological patterns to signal distinctive rhetorical purposes of the moves specific to the given thesis sections.

Finally, following Hyland's (2008b, 49) taxonomy, we also identified the discourse functions of the bundles across the two sub-corpora.

Table 3. Discourse functions of the lexical bundles in L1 and L2 sub-corpora

	Research-oriented (%)	**Text-oriented (%)**	**Participant-oriented (%)**
L1 Discussion	40	21	39
L2 Discussion	42	38	20
L1 Conclusion	50	50	/
L2 Conclusion	/	50	50

As can be seen in Table 3, the functional categorization of lexical bundles showed both similarities and differences in the use of bundles between the two writer groups. While L1 writers used research-oriented and participant-oriented bundles in Discussions almost equally in terms of frequency rates, L2 writers showed a preference for research-oriented and text-oriented bundles. The centrality of research-oriented bundles in both sub-corpora shows that students' main focus was placed on various aspects of their study, such as research procedure, as in:

The main aim of this research was to explore correlation between WTC, the use of CSs, and the level of self-perceived speaking competence. (HR14_M1)

The finding is in line with previous research, which points to the dominance of research-oriented function of bundles particularly in master's theses (Hyland 2008, 54). As noted by Hyland (2008, 47), this is expected given the nature of the graded assignment in question, in which graduates need to show their knowledge of the subject matter and all relevant aspects of the research procedure. However, the frequencies of the participant-oriented bundles indicate that L1 writers were almost equally concerned with expressing evaluations or conveying stance towards the subject matter. This was significantly less prominent in writing by L2 writers who seemed to be less concerned with evaluating their findings and more with the organization of the content, as shown in the example below. This finding corroborates previous results on the less frequent use of stance bundles in writing by L2 student writers (Chen and Baker 2010, 41-42).

When it comes to intralingual errors, a lot of learners ignored rule restrictions or were unaware of them. (HR8_M2)

Different patterns in the use of bundles were also evident in the Conclusion sections, where L1 writers' focus was on research and textual organization, while stance was obviously conveyed earlier in the Discussion. By contrast, aside from the structural organization of the Conclusion section, L2 writers were particularly concerned with expressing stance and engaging with readers.

As noted in our previous study on the rhetorical structure of the Discussion and Conclusion sections in L1 and L2 master's theses, the observed differences in the linguistic choices between two writer groups may reflect students' previous exposure to academic writing instruction and familiarity with the conventionalized language use in academic genres. Without reliable data on students' educational experiences, we cannot argue but only assume that L1 students, immersed in the Anglo-American university settings that traditionally place emphasis on the development of academic language skills, were better trained and prepared for the genre writing under study. By contrast, graduating from an English L2 university context where genre-based writing instruction in English has not been systematically implemented in the curricula, L2 majors may lack the rhetorical knowledge and familiarity with the target formulaic language necessary for more proficient thesis writing.

These insights, however, should be viewed in light of the preliminary nature of the current study and its methodological limitations, which primarily stem from the small corpus size and low cut-off points for assigning the lexical bundle status. These constraints consequently determined the nature of the quantitative representation of the data. As previously noted, a further study is bound to encompass a substantially larger corpus that will allow us to examine more closely the use of bundles in signalling rhetorical moves in the given thesis sections, followed by a more rigorous data analysis.

Generally, the taxonomy of discourse functions applied (Hyland 2008, 49) proved useful in identifying the underlying functions of lexical bundles. However, there were a few exponents of bundles whose functions could not be assigned straightforwardly to the taxonomy. For example, similar to Hyland's (2008, 49) examples of bundles (*it was found that, these results suggest that, etc.), we assigned* the text-oriented function to the bundle *the result of the research.* Yet, as it refers to the aspects of the research procedures discussed, it could perform a research-oriented function. Attesting this finding calls for a study involving a larger corpus, which could also reveal other similar examples potentially requiring an intervention into the existing taxonomy of discourse functions.

However, even the insights gained from the present analysis may be used to address gaps in L2 student writers' use of phraseology in writing the given thesis sections. As Cortes (2013, 42) observed, the data obtained from corpus-based research may be valuable to instructors when creating teaching materials for genre-based writing courses. For instance, students could become more familiar with the concept of the rhetorical structure of the research article in their target discipline, as research articles are central to the subject-specific literature they must read and use when completing higher-level written assignments. Students may be exposed to different tasks aimed at identifying rhetorical moves across the IMRaD sections along with the target lexical bundles that mark their underlying communicative purposes. This may be followed by the tasks aimed at practising the use of a range of synonymous bundles to mark rhetorical moves, which

may assist students in broadening their knowledge of academic phraseology. Additionally, authentic samples of the students' work, such as the one building the present corpus material, may be adapted to teaching purposes. For example, students may be required to analyze the rhetorical and phraseological features of successful L1 student writing and to compare them with those of L2 writing. A tailor-made instruction of this kind may raise students' awareness of the structural, lexical and other discoursal components underlying successful target academic writing.

5. CONCLUSION

The present study represents the initial step of a further, more comprehensive study on the use of lexical bundles in L1 and L2 master's thesis writing. The analysis of the move-bundle distribution in the Discussion and Conclusion sections revealed that L1 student writers used more bundles than L2 student writers in structuring the content of the given sections. While both writer groups used most bundles in reporting results, higher frequencies of bundles dispersed over several moves in the L1 Discussion suggest that L1 student writers were more profoundly engaged in elaborating their research, which primarily concerned contextualizing and commenting on the results. In terms of the discourse functions of bundles, both writer groups showed preferences towards research-oriented bundles in both sections. However, L1 writers were more focused on the use of participant-oriented bundles, whereas text-oriented bundles prevailed in L2 Discussions. Reversed patterns of use were observed in Conclusions, where participant-oriented bundles were more frequent in L2 writing while text-oriented bundles were more common in L1 writing. A more advanced use of lexical bundles in signalling rhetorical moves suggests L1 student writers' higher levels of rhetorical and phraseological awareness as compared to L2 students. Besides overall language proficiency, this might be associated with students' different educational experiences and exposure to academic writing instruction. The present results may provide important insights for further explorations of university students' use of phraseology in writing academic genres. This could be especially important for promoting action aimed at enhancing genre-based writing instruction in L2 English higher education settings.

References

Biber, Douglas, Johanston, Stig, Leech, Geoffrey, Conrad, Susan, and Edward Finegan. 1999. *Longman grammar of spoken and written English.* Edinburg: Pearson Education.

Chen, Yu-Hua, and Paul Baker. 2010. "Lexical bundles in L1 and L2 academic writing."*Language Learning & Technology* 14(2): 30-49. http://llt.msu.edu/vol14num2/chenbaker.pdf

Cortes, Viviana. 2004. "Lexical bundles in published and student disciplinary writing: Examples from history and biology." *English for Specific Purposes* 23(4): 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, Viviana. 2013. "The purpose of this study is to: Connecting lexical bundles and moves in research article introductions." *Journal of English for Academic Purposes* 12(1): 33–43 https://doi.org/10.1016/j.jeap.2012.11.002

Cortes, Viviana. 2023. "Lexical bundles in EAP." In *The Routledge Handbook of Corpora and English Language Teaching and Learning*, edited by Reka R. Jablonkai and Eniko Csomay, 220-233. London and New York: Routledge Taylor & Francis Group.

Hyland, Ken. 2008a. "As can be seen: Lexical bundles and disciplinary variation." *English for Specific Purposes* 27(1): 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, Ken. 2008b. "Academic clusters: Text patterning in published and postgraduate writing." *International Journal of Applied Linguistics* 18(1): 41-62. DOI: 10.1111/j.1473-4192.2008.00178.x

Hyland, Ken. 2017. "English in the Disciplines: Arguments for Specificity." *ESP Today* 5(1): 5-23. DOI: 10.18485/esptoday.2017.5.1.1

Hyland, Ken, and Kevin Jiang. 2018. "Academic lexical bundles: How are they changing?" *International Journal of Corpus Linguistics* 23(4): 383-407. https://doi.org/10.1075/ijcl.17080.hyl

Kanoksilapatham, Budsaba. 2007a. "Introduction to move analysis." In *Discourse on the Move: Using corpus analysis to describe discourse structure,* edited by Douglas Biber, Ulla Connor, and Thomas A. Upton, 23- 41. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Kanoksilapatham, Budsaba. 2007b. "Rhetorical moves in biochemistry research articles." In *Discourse on the Move: Using corpus analysis to describe discourse structure,* edited by Douglas Biber, Ulla Connor, and Thomas A. Upton, 73- 121. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Lu, Xiaofei, and Jinlei Deng. 2019. "With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students." *Journal of English for Academic Purposes*, 39: 21-36. https://doi.org/10.1016/j.jeap.2019.03.008

Swales, M. John. 1990. *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Varga, Mirna, Kovačević, Predrag, and Draženka Molnar. 2020. "Establishing the Criteria for Building Comparable Corpora in Research on Academic Writing in L1 and L2 English. ", In *Jezici i kulture u vremenu i prostoru IX/1*, edited by Snežana Gudurić, Biljana Radić-Bojanić, and Predrag Mutavdžić, 145-154. Novi Sad: Filozofski fakultet Univerziteta u Novom Sadu.

Varga, Mirna, and Višnja Pavičić Takač. 2024. "Rhetorical structure of the discussion and conclusion sections in master's theses written by English L1 and L2 students." *Journal of Teaching English for Specific and Academic Purposes* 12, 3: 555–572. https://doi.org/10.22190/JTESAP240530043V

Wray, Alison, and Michael R. Perkins. 2000. "The functions of formulaic language." *Language and Communication* 20: 1–28.

Yang, Ruiying, and Allison, Desmond. 2003. "Research articles in applied linguistics: Moving from results to conclusions. "*English for Specific Purposes*, 22: 365-385. https://doi.org/10.1016/S0889-4906(02)00026-1

Yang, Mei. 2022. "Connecting the Functions of Lexical Bundles and Moves in Published Research Articles: The Case of Developmental and Educational Psychology. "*Nordic Journal of English Studies* 21(1): 141–189. https://doi.org/10.35360/njes.743

DIE PHRASEOLOGIE DER RHETORISCHEN ZÜGE IN DIPLOMARBEITEN DER L1- UND L2-ENGLISCH SCHREIBENDEN

Zusammenfassung

Akademische Texte verbinden sich generell mit der Struktur rhetorischer Züge, die durch formelhafte Mehrwortsequenzen signalisiert werden können. Dementsprechend können rhetorisches Wissen und die Verwendung formelhafter akademischer Ausdrücke auf die Diskurskompetenz der Schreibenden hinweisen. Während die Struktur sowie die Diskursfunktionen von Mehrwortsequenzen in wissenschaftlichen Artikeln auf Englisch ziemlich ausführlich untersucht wurden, wurden studentische Texte nur begrenzt analysiert. Die vorliegende Studie beschäftigt sich mit vier- bis sechswortigen lexikalischen Bündeln in den Teilkorpora mit dem Diskussions- sowie Schlussteil von Diplomarbeiten (N = 40), die von kroatischen und serbischen Masterstudierenden sowie von ihren britischen und amerikanischen Kollegen verfasst wurden, die Angewandte Linguistik studieren. Die vorgenommene Analyse der Bündel basiert auf den unterschiedlichen rhetorischen Funktionen der Züge, in denen sie auftreten. Durch die kombinierte Analyse des Verhältnisses zwischen Zügen und Bündeln kann der Unterricht akademischen Schreibens wertvolle Impulse erhalten, besonders in nicht-anglophonen Kontexten. Die Ergebnisse suggerieren, dass studentische L1-Schreibende mehr lexikalische Bündel verwendet haben als L2-Schreibende, was auf eine bessere Fähigkeit zur Strukturierung von Argumenten hinweist. Die beiden Gruppen stützten sich bei der Darstellung der Ergebnisse auf lexikalische Bündel, aber L1-Schreibende verwendeten sie in größerem Umfang für verschiedene rhetorische Züge. Dies deutet auf eine tiefere Auseinandersetzung mit der Kontextualisierung und Kommentierung ihrer Ergebnisse hin und spiegelt eine tiefere Analyseebene wider.

Bezüglich der Diskursfunktion zeigten sowohl L1- als auch L2-Schreibende eine Präferenz für forschungsbezogene Bündel (engl. research-oriented) in beiden Textteilen. Doch verwendeten L1-Schreibende im Diskussionsteil häufiger teilnehmerbezogene Bündel (engl. participant-oriented), während L2-Schreibende verstärkt auf textbezogene Bündel (engl. text-oriented) zurückgriffen. Interessanterweise bevorzugten L2-Schreibende im Schlussteil eher teilnehmerbezogene Bündel, während L1-Schreibende eher textbezogene Bündel verwendeten.

Diese Ergebnisse legen nahe, dass Studierende, die in ihrer L1 schreiben, über ein tieferes Verständnis rhetorischer Strukturen und phraseologischer Konventionen verfügen. Dies scheint vor allem auf ihre höhere Sprachkompetenz und eine intensivere unterrichtliche Beschäftigung mit dem akademischen Schreiben zurückzuführen sein.

Schlusswörter: L1-Englisch, L2-Englisch, Diplomarbeit, leksikalische Bündel, rhetorische Züge

Appendix 1.
Raw frequencies of lexical bundles

	L1 sub-corpus			L2 sub-corpus		
	4-word bundles	**5- word bundles**	**6- word bundles**	**4- word bundles**	**5- word bundles**	**6- word bundles**
M1 Providing background information	103	6	6	35	16	3
M2 Reporting results	114	56	56	112	40	12
M3 Summarizing results	24	0	0	17	4	0
M4.1 Interpreting results	3	0	0	10	0	0
M4.2 Comparing results with the literature	87	9	9	7	0	0
M4.3 Accounting for results	84	16	16	14	0	0
M4.4 Evaluating results	6	3	3	0	0	0
M5 Summarizing the study	19	0	0	10	5	0
M6.1 Indicating limitations	0	0	0	11	4	0
M6.2 Indicating significance/ advantage	6	0	0	3	0	0
M6.3 Evaluating methodology	10	0	0	0	0	0
M7.1 Making suggestions	0	0	0	0	0	0
M7.2 Recommending further research	77	25	25	18	0	0
M7.3 Drawing pedagogical implications	8	0	0	4	0	0
Conclusion	42	4	4	32	32	11