STRUCTURAL ANALYSIS OF LEXICAL BUNDLES IN GRADUATE STUDENTS' RESEARCH PAPER ABSTRACTS

*Dragana **Gak**, Faculty of Technical Sciences, University of Novi Sad, Serbia, dgak@uns.ac.rs*

*Vesna **Bogdanović**, Faculty of Technical Sciences, University of Novi Sad, Serbia,*

*Vesna **Bulatović**, Faculty of Technical Sciences, University of Novi Sad, Serbia*

Abstract: Lexical bundles, as a recurrent sequence of words, are an important component of both spoken and written academic corpora, used both by professional academics and student writers in order to contribute to better structural organization, efficient communication, and well-established credibility. These expressions are formulaic and highly frequent, with rather clear discourse function and communicative purpose. The most frequently used structures of lexical bundles include NP-based, VP-based, and clause-based bundles. The aim of the paper is to explore the lexical bundles on a learner corpus of 448 abstracts written by non-native English speakers in English during their Masters studies. Structural analysis is performed in order to observe which structures students prefer and utilize while writing abstracts. The corpus was subdivided into two sub-corpora, depending on whether graduate students received any formal instruction on academic writing before publishing their abstracts, to explore whether instructions influence the use of lexical bundles with novice writers. The results offer the most common structures of lexical bundles that graduate students use, as well as the analysis of the types and diversity of lexical bundles used based on whether or not students had academic writing instructions.

Keywords: abstract, academic writing, lexical bundles, structures of lexical bundles

1. Introduction

Over the past decades, lexical bundles (a term first introduced by Biber & Conrad, 1999) or formulaic sequences (Wray, 2002) have been treated as an umbrella term referring to idioms, phrases, collocations, or multi-word units/expressions (Wang, 2018: 13; Buerki, 2016: 17; Durrant & Mathews-Aydinli, 2011: 59; Wray, 2002: 8; Pan, et al., 2016: 60; Liu, 2012: 25). Consequently, the interest into lexical bundles has grown into one of the rapidly growing areas in applied

linguistics and academic writing (Bestgen, 2017:.65, Wray, 2002; Hyland, 2008; Wang, 2018: 12). Researchers emphasize that formulaic sequences have 'an especially strong relationship with each other in creating their meaning' (Wang, 2018) and view lexical bundles (LBs) to be recurrent lexical sequences which are identified through corpus analysis that includes specific frequency thresholds and dispersion requirements (Hyland, 2008: 4; Pan, et al., 2016: 60). LBs in academic writing have been explored from different perspectives: their use in different genres (e.g. abstracts, introductions, articles, dissertations, textbooks), their frequencies and length, their structures and functions, position in sentences (Li, et al, 2020: 86).

In the academic context, research article abstracts have become an increasingly important genre in all knowledge fields (Jiang & Hyland, 2017: 1) because of their unique communicative purpose. They represent "mini-texts" and "screening devices", giving a brief summary of the whole article (Huckin, 2001: 93; Lorés, 2004: 281; Bhatia 1993: 82) without developing a detailed argument (Jiang & Hyland, 2017: 3). It is widely-accepted that abstracts frequently represent readers' first and only encounter with a research article (Jiang & Hyland, 2017: 3), after which the reader decides whether or not to take the time to go further into the paper itself (Jiang & Hyland, 2017: 1; Hyland, 2000; Lorés, 2004: 281). Therefore, acceptance or rejection of the paper may depend on how well the abstracts are written (Huckin 2001).

In general, L2 students may find academic writing quite challenging and even intimidating (Shin, 2020: 45). In addition to the disciplinary content, non-native-speaking students need to pay attention to structural organization and linguistic features, as well as to be able to choose appropriate lexical features (LBs included) to make their abstracts understandable to the target discourse community (Kanoksilapatham, 2013: 2).

The adequate use of LBs is one of the signals that students have mastered academic writing. According to Hyland (2008), LBs are an important component of fluent linguistic production (Hyland, 2008: 4) and one of the most noticeable features of academic written texts (Hyland, 2019: 383). LBs are familiar to text users and have customary pragmatic and discoursal functions (Hyland, 2019: 383). They are seen as a way of assisting communication to facilitate pragmatically efficient communication and a way to make the language more predictable to the reader (Wang, 2018: 357), signaling the appropriate use of a disciplinary code (Cortes, 2006; Li et al., 2020).

The majority of literature published on LBs in academic discourse focuses on longer pieces of writing, e.g., research articles (Hyland 2008), bachelor theses (Dontcheva-Navratilova, 2012), and Ph.D. dissertations (Yakut et al., 2021), or literature reviews (Wright, 2019). However, to the author' best knowledge, very few studies have been undertaken to investigate LBs in students' abstracts, especially in Serbia. The lack of academic writing courses available to Serbian

students creates confusion about academic writing conventions and feeling of insecurity (Blagojević, 2014: 77).

In an attempt to fill this gap, we compared lexical bundles used in two groups of abstracts produced by graduate students. One sub-corpus consists of research paper abstracts written by graduate students who had academic writing (AW) instructions and the other sub-corpus of abstracts by graduate students who did not have any formal instructions in abstract writing or academic writing in general. These abstracts were written by non-native L2 English students at the end of their Master's studies. Given the high-stake nature of abstracts, this study aims at identifying and investigating the lexical bundles students most frequently use, focusing on their structural diversities. More specifically, the study poses the following research questions:

- What are the most frequent structures of LBs used by Master students in the discourse organization of their abstracts in English?
- What differences, if any, are there in the structure of the lexical bundle used in Master students' abstracts between the two analyzed groups – students who had AW instruction and students who had no AW instruction?

2. Literature review – LBs in academic writing

Corpus studies conducted have demonstrated that LBs are pervasive in natural language use (Wang, 2018: 5; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman &Warren, 2000; Buerki, 2016: 15). They are seen as a marker of proficiency and an important component of fluent linguistic production. As a result, they are critical to the success of communicative acts and key to rapid language processing (Buerki, 2016: 15; Hyland, 2008: 4; Bestgen, 2017: 65; Laufer &Waldman, 2011: 648).

A significant body of research shows that LBs are widely used in academic writing and are known to be an important aspect of EAP writing development (Staples, et al., 2013: 214; Biber, 2006; Biber, 2009; Biber & Barbieri, 2007; Biber, Conrad & Cortes, 2004). Furthermore, research demonstrates that bundles are central to the creation of academic discourse since they provide the basic building blocks for coherent discourse (Hyland, 2008: 4), and represent a large stock of prefabricated phrases (Hyland, 2019: 354). Consequently, professional academics and student writers alike draw on formulaic resources to 'develop their argument, establish their credibility and persuade their readers' (Hyland, 2008: 59; Wang, 2018: 12).

The predominant trend in research studies is to take a frequency-based approach, relying on the computer to identify frequent recurrent uninterrupted linguistic forms in a given corpus (Wang, 2018: 12-13; Bestgen, 2017: 66). Some researchers claim that frequency is "the ultimate characteristic that defines lexical bundles" (Cortes, 2013: 34). From a methodological point of view, the

frequency-based approach has the advantage of being straightforward and consistent and can be scaled up to extensive datasets (Wang, 2018: 13).

However, frequency is not by itself a sufficient guide to how strongly a word is associated with its context, and it is also not a satisfactory guide to how well a phrase is associated with its communicative function (Durrant & Mathews-Aydinli, 2011). It is important to note that most of the lexical bundles identified by frequency alone are structurally/semantically incomplete (Carter & McCarthy, 2006: 828), and therefore, to determine relevant and pedagogically compelling lists of LBs additional analyses are required. To address this issue, Simpson-Vlach and Ellis (2010) developed an instrument for identifying useful LBs. Their target corpora consisted of an academic speech corpus (1.7 million words) plus BNC files of academic speech (431,000 words) (British National Corpus 2006), and the academic writing corpus consisting of Hyland's (2004) research article corpus (1.2 million words), plus selected BNC files (931,000 words) sampled across academic disciplines using Lee's (2001) genre categories for the BNC. Firstly, they looked for the most frequent LBs and determined a Mutual Information (MI) score. By doing this, they came up with two sets of data, one with the most frequent LBs but less functional LBs, and the other one with more functional and structurally complete LBs. This was followed by further analysis by experienced ESL instructors to get additional insight into whether a LB is a fixed phrase or expression and whether it is worth teaching. After applying these analyses, the authors were able to compile a list of the top 200 LBs in spoken academic English and a list of the top 200 LBs in written academic English.

One of the major characteristics of LBs is the fact that they do not represent a complete structural unit (Biber and Barbieri, 2007: 270). There are LBs that are complete phrases or clauses (e.g., *I want to know, as well as*); however, most LBs bridge two structural units (e.g., *the fact that the, the aim of the*). Structure-wise LBs in academic writing most frequently are prepositional phrases with -of fragments (*as a result of*), noun phrase + *of* fragments (*the nature of the*) or anticipatory *it* fragments (*it is argued that*) (Pan et al., 2016: 61). Together, these three forms comprise over 70% of 4-word patterns in research journals spaced at three periods over the past 50 years: 1965, 1985 and 2015 in four disciplines: applied linguistics, sociology, electrical engineering and biology (Hyland and Jiang, 2018: 391).

Chen and Baker (2010: 35) propose three broad structural categories: NP-based bundles, containing a noun phrase with a post-modifier fragment (*the role of the)*, PP-based bundles, which begin with a preposition followed by a noun phrase fragment (*at the end of*), and VP-based bundles as word combinations with a verb component (*was one of the*). They also included a number of subcategories in their taxonomy. The same taxonomy was used in the research

by Pan et al. (2016), as well as Hyland and Juang (2018), though the latter had more subcategories to make a clearer division.

Drawing on previous research, Hyland (2018: 391) proposes the classification of 4-word LBs structures, also focusing on three main categories: verb phrase-related bundles (passives, copular be, and imperatives), clause-related bundles (revolving around: anticipatory it, abstract subject, human subject, as-fragments, if-fragments, wh-fragments, that-fragments), and noun/preposition-related bundles (combinations of a noun phrase with of-phrase fragments, noun phrase with other post-modifier fragments, prepositional phrase expressions, and comparative expressions).

In one of the most recent studies, Yakut et al. (2021) decided to incorporate all separate bundle groups into their research and the three-category criticism by Lu and Deng (2019), who argued that PP and NP bundles should be distinct categories in order to observe all distinctive patterns. They suggest four major categories, namely noun-phrase bundles, prepositional-phrase bundles, verb-phrase bundles, and clause-related bundles, each of them classifying the most frequent sub-structures. The same taxonomy was adopted in this research.

An important contribution to the analysis of LBs used in the academic writing of Serbian researchers is a Ph.D. thesis by Lazić (2017) compiling and examining three corpora of research papers in the field of biotechnology: the first consisting of papers written by native speakers of English, the second written in English by speakers of the Serbian language and the third written in the Serbian language. Apart from specifying the most frequent LBs and analyzing their functions, structure, and frequency, this thesis examines the influence of the Serbian language on the use of LBs when writing in English, providing insights into the use of English LBs used by native Serbian speakers. The findings show that these LBs mainly correspond to the LBs the native Serbian speakers use when writing in the Serbian language. Furthermore, Lazić (2017) compiled a list of biotechnical LBs and exercises to be used in writing courses.

3. Methodology
3.1. Data collection
The research was conducted on a learner corpus (Table 1) made of 448 abstracts (total of 29,558 words) written by Master students at the Faculty of Technical Sciences, University of Novi Sad, Serbia, who were required to write a research paper before defending their Master theses with an abstract in both L1 Serbian and L2 English, and after a review process, it was published in the open-access Proceedings of the Faculty of Technical Sciences. The abstracts written in English and published in the Proceedings between 2016 and 2021 (four issues per year) were compiled to form a research corpus. These student-produced

abstracts were not written for the purpose of this or any other similar research and therefore represent an authentic use of the language. It is also important to notice that not all abstracts followed the word limitation, and some students wrote shorter abstracts than requested. The authors were aware of this, but regardless of that, included these abstracts since this analysis targeted the use of lexical bundles, not the structure of the entire abstracts.

Table 1 – Corpora word count

Students' abstracts	No. of abstracts	Word count	Average word no. per abstract
Sub-corpus 1 With AW instructions	219	16,841	76.89
Sub-corpus 2 Without AW instructions	229	12,717	55.53
Total	448	29,558	65.98

Students who wrote the abstracts analyzed in this paper took between 2 and 4 obligatory English language courses (each with four teaching hours per week) during undergraduate studies. For students having only two language courses, both courses are in general English language, the first being at the B1 level and the second one at the B2 level. Apart from two general English courses, students with 3 or 4 semesters of English have 1 or 2 courses focusing primarily on profession-related topics and vocabulary. None of these courses include explicit instructions in academic writing. After finishing undergraduate studies, it is expected that the students can understand the main ideas of complex texts, communicate fluently and clearly and explain a viewpoint on a topic (CERF, 2020: 48), and the students who had profession-related language course(s) are knowledgeable of the variety of topics relevant for the field the major in.

Since all students were majoring in engineering or related disciplines, the authors divided them into two groups based on whether or not they had any course in academic writing during their master's studies. The first sub-corpus of 219 abstracts was produced by students who had instruction in academic writing. These graduate students enrolled in the fields of architecture, graphic engineering and design, and electrical and computer engineering. For students who had a course in academic English, this was a one-semester course in the Master studies, two-hour classes per week, covering topics such as abstract writing, research paper writing, IMRAD structure, CARS model, transitions, metadiscourse markers, formal and academic vocabulary. During the course,

students needed to submit two practical tasks: an abstract and an introduction to a research paper. However, the course description does not include information on instructions on lexical bundles. We can only assume that they were mentioned while dealing with transitions, metadiscourse, and structures of abstracts and research introductions.

The second sub-corpus included 229 abstracts written by students without academic writing instruction while studying mechanical engineering, civil engineering, surveying, industrial engineering and management, safety at work engineering, traffic engineering, and software engineering. Since all abstracts belong to engineering sciences, and students have a number of interdisciplinary courses, the assumption is that all students have similar general knowledge, and hence abstracts cannot be divided according to disciplines.

As can be observed, the abstracts were written by novice L2 English writers. Hence, the authors did not focus on grammatical forms and errors but strictly on general patterns and structures of lexical bundles found in the corpus. As for the expertise lexical bundles, these students, during their third and fourth year of undergraduate studies, as well as their master studies, are obliged to read in both Serbian and English; hence, their professional knowledge is bilingual in the typical topics of their field of study.

3.2 Identification of lexical bundles

Having created the corpus, the abstracts were analyzed for lexical bundles used. Since it could not be readily determined whether students would use longer or shorter lexical bundles, all 3-, 4- and 5-word lexical bundles were collected using the corpus analysis software AntConc version 3.5.8 (Anthony 2019). It automatically retrieved multi-word units with frequency and distribution criteria specified, after which the bundle-driven approach (Li et al., 2017) could be used for the structural analysis.

Two-word bundles were not included in the research due to the fact that they are very common, and analyzing them would not be very beneficial (Kopaczyk, 2015). On the other hand, very long lexical bundles (more than 5 words) are rather rare, and they often contain shorter ones; thus, they were not included in the analysis. According to previous research (Hyland, 2019; Cortes, 2004; Li, 2020), the most frequently studied lexical bundles are 4-word bundles, and they are 10 times more frequent than 5-word sequences, offering more structures to analyze. Furthermore, there is a number of 3-word bundles (e.g. *on the other*, *it can be*), which frequently expand into the 5-word bundles (*on the other hand the*, *it can be seen that*), supporting Cortes (2004) observation that many four and five word strings 'hold 3-word bundles in their structure'. This research explored the word sequences as the frequent combinations used together and includes 3-word lexical bundles when they make a structural and

functional unit (e.g., *the paper concludes*), but counts them as 4-word or 5-word bundles if that is the whole structure used as such (e.g., *if and only if, it can be concluded that*).

Since the corpus is rather small (Table 1), the frequency per million words could not be applied. Thus, raw frequency count (Chen and Baker 2010) was used to identify the type (T), i.e., the number of structures utilized at least once, and tokens (N), i.e., the number of all occurrences of lexical bundles used in the corpus. Then, the authors decided to set the range threshold at three texts, and thus examples occurring once or twice are not listed in the paper. Range threshold is set to make sure that the variety of lexical bundles is used in a variety of texts rather than by individual authors in a text. Most research studies adopt the Hyland's threshold of at least 10% of texts; however, this cannot be applied to small-scale research. The overlapping lexical bundles constituting the complete overlap (Chen and Baker 2010: 33), meaning that 3-word and 4-word bundles are a part of the 5-word bundles identified, were manually looked for in the AntConc final bundle lists and removed in order not to influence the quantitative results (following the solution by Chen and Baker, 2010).

An extensive body of research on lexical bundles deals with the structures of lexical bundles (Biber et al., 2004; Hyland, 2008, 2019; Chen & Baker, 2010, Li et al., 2020). Although they investigated different corpora, they mainly concluded that the most frequently used structures are: NP-based, PP-based, VP-based, and clause-based bundles. That taxonomy was used in this research.

4. Findings and discussion

Following research by Chan and Baker (2010), Hyland and Jiang (2018), and Yakut et al. (2021), the results are presented in Table 2. Lexical bundle structures and sub-structures were listed, with the types (T) that have been found, as well as their row frequency (N) and individual frequency (No). Only bundles occurring three or more than three times in the corpus were listed. Structures occurring only once or twice are not listed in Table 2.

Table 2. Structural distribution of LB in two sub-corpora and LBs used

Major structures	Sub-structures	Sub-corpus 1				Sub-corpus 2			
		T	N	Bundles	No.	T	N	Bundles	No
Noun-phrase	with of-phrase fragment	44	91	*the development of the*	15	31	74	*the part of the*	7
				the research of the	13				

				the analysis of the	11			the development of the	6
				the use of the	9				
				the part of the	8			the use of the	5
				the design of the	7				
				the description of the	6			the result of the	3
				the result of the	4				
				the design of the	4			the search of the	3
				the basis of the	3				
Prepositional-phrase	with embedded of-phrase	7	18	as a result of		4	11	as a result of	
				for the purpose of				for the purpose of	
				with the aim of				with the aim of	
				in terms of the				at/in the end of	
				at the end of					
				on the basis of					
				in the context of					
	other prepositional phrases	4	7	on the other hand		2	6	at the same time	
				at the same time					
				in the present study				in the present study	
				with the aim of					

	comparative expressions	1	62	*as well as*	62	1	56	*as well as*	56
Verb-phrase	with passive verb forms		363					298	
	be+ Noun/Adjective phrase		29	*is the same as*				18	*is the same as*	
				is a matter of					*is a matter of*	
				is due to the					*is due to the*	
				is/was the result of					*is/was the result of*	
	modal verb forms		24	*could be initiated*				16	*can be planned*	
				can be considered					*can be discussed*	
				can be analyzed					*can be analyzed*	
				should be equipped					*should be located*	
				can be achieved						
				should be demolished						
Clause-related	Abstract subject	24	96	*this paper presents the*	37	22	64	*this paper presents the*	19
				this paper describes the	17			*this paper describes the*	11
				this paper deals with	14			*this paper analyzes the*	6
				this paper analyzes the	11			*this paper deals with*	3
				this paper examines	7				
				this paper contains the	4				
	human subject	3	20	we	15	3	13	*we*	11

			I	4				
anticipatory *it*	7	11	*it is possible to*	4	4	7	*it is necessary to*	4
that + fragments		124			59			
there + fragments		19				10		
wh- + fragments		104				62		
in order to + fragment		18	*in order to determine*	4		14	*in order to improve*	3

As can be observed from Table 2, students' use of LBs is rather limited and primarily includes *that + fragments* and *wh- + fragments*, together with *VP with passive verb form, NP with of-phrase fragment,* and *abstract subjects* are lexical bundles being used in abstracts. It is also clear that students in the group without AW instructions (sub-corpus 2) wrote shorter abstracts and used even fewer LBs. This was to be expected since the absence of formulaic sequences in language production signals the "lack of mastery of a novice writer in a specific disciplinary community" (Pérez-Llantada, 2014: 85).

The results also concur with the research arguing that L2 learners have a more limited repertoire of lexical bundles that they tend to repeat often (Adel & Erman, 2012; Chen & Baker, 2010). This can, for example, be observed in the use of only one comparative expression found in the corpus, *as well as.* Learners, such as graduate students, tend to have a small inventory of formulaic sequences that they overuse (Wray, 2012: 235). The reason why there are more different LB types in the sub-corpus 1 is related to the fact that only very advanced learners (or in this case, learners with the appropriate course and learning material) can be expected to use bundles similar to that of native speakers (Boers & Lindstromberg, 2012; Bestgen, 2017: 66). More detailed analysis of the LBs used by the students is provided in the following sections.

4.1 Noun phrases with of-phrase bundles

In academic writing, noun phrases with of-phrase fragments are frequently used to specify the attributes of what is being discussed, identify quantity, place, and size; mark existence, or highlight qualities (Hyland and Jiang 2018). Hyland (2008) and Yakut et al. (2021) found NPs with of-phrase to be the most dominant LBs.

In this study, both groups of graduate students used NPs, especially ones *with the of-phrase fragment*. Since research paper abstracts include introducing background or a problem, presenting current research, describing methodology, and reporting results, it was no surprise that the most frequent NPs with of-phrase fragments used by the students were *the development of the*, *the research of the, the analysis of the, the design of the, the description of the, the results of the*, etc.

The use of these NPs can be seen in both corpora:

(1) After a short introduction, the steps of *the development of the* applications are described, as well as the end results. (aw24)

(2) The research is based on sample scanning, and *the analysis of the* obtained results shows the resistance of the print to rubbing, depending on the substrate. (aw188)

(3) The task of this research is to analyze *the development of the* postal sector. (no aw216)

(4) This paper presents *the part of the* building in Zrenjanin which was assessed and reconstructed. (no aw73)

The results show that in sub-corpus 1, students with academic writing courses used significantly more of these bundles in comparison to students without these lectures. Bearing in mind that these courses focused on the abstract structure, the function of key abstract elements (moves), and on the essential vocabulary utilized for introducing the research problem, describing methodology and results, it was expected to find a larger number of NPs and more diverse NPs in abstracts written by students who had instructions in academic writing.

In sub-corpus 2, there are only several LBs used three or more times, with the most common being *the part of the* and *the development of the* (3, 4), used to identify quantities and processes. The lack of LBs in sub-corpus 2 clearly underlines the fact that academic writing courses are essential, even for short pieces of writing such as abstracts.

4.2 Prepositional phrase bundles

After noun phrases with of-phrase fragments, prepositional phrase expressions present the largest share of bundles in academic writing. Bundles containing PPs can make up to 40% of all bundles (Hyland and Jiang 2018:393). These bundles are especially used to facilitate the exploration of possibilities and identify logical relationships in an argument. Gil and Caro (2019: 79) state that the most common structures in the graduate theses' corpus include *of this paper, according to, in this study* and *of the most*.

However, our research presented completely different results. As it seems students did not prefer using bundles with prepositions. There were several examples of bundles with embedded of-phrase related to the purpose of their writing (e.g., *as a result of, for the purpose of, with the aim of*) (5, 6), and only a few with other prepositional phrases. In sub-corpus 2, the number of diverse types is really small, almost to be neglected.

(5) This paper presents the application of reinforcement learning *for the purpose of* training an agent capable of playing the video game Road Fighter. (aw134)

(6) *For the purpose of* this application the results of measurements obtained by the Festo Air Box portable laboratory are collected. (no aw113)

The authors of this research believe that students avoiding using PPs can be contributed to several factors. PPs, such as *for the purpose of*, *in the context of*, *in terms of the*, etc., are used to emphasize a logical relation or give further details about a problem, study, or results. Students in sub-corpora 1, although they had academic writing instructions, focused more on the essential moves in abstracts and not on providing further details in their abstracts. Furthermore, students in both sub-corpora tend to write short abstracts and did not include explanations of logical relations and additional details. Further reasons for the lack of PP bundles may lie in the fact that students tend to translate directly from Serbian. In those cases, NP bundles or PP bundles do not occur in the Serbian text.

Interestingly, the only bundle beginning with a preposition that students used often was a comparative expression *as well as* which was used more than 50 times in both sub-corpora. This may indicate that the students used it as a phrase they previously acquired in undergraduate English language courses, not necessarily as a PP lexical bundle.

4.3 Verb phrase-related bundles

Although present in academic writing, VP bundles are not the most frequently used type of bundles (e.g. in the research by Yakut et al. (2021: 485) with the exception of passives helping the writer to facilitate the presentation of propositional content, distance the author, and still make the basis of claims clear for readers (Hyland 2008).

However, this research demonstrates that Serbian students follow the traditional aspects of academic writing and overuse passive voice. Most abstracts contain at least one bundle with passive verb forms.

(7) The research *was based* on the measurement of the spectro-photometric values of prints for single classic and CFL bulb. (aw136)

(8) A proposal *was made* for measures to improve traffic conditions by variants, the results obtained were evaluated, and the best solution was proposed. (no aw228)

This shows that students in both sub-corpora continue to believe that by shifting the focus from the researcher to the object of the action, they are utilizing a more formal and more appropriate formulation. This result is surprising, especially for sub-corpus 1, since they had clear instructions emphasizing an apparent decline in the use of passive structures due to the increase in the use of self-mentions.

Verb-based LBs with modal verbs are also very rare, with only three modals: can, could, and should. Interestingly all modal verb forms are passives, such as: *could be considered*, *can be analyzed*, *should be located*, *can be achieved*, etc. All the modal verb examples help students express the ability or necessity of an action. Since modal verbs are taught in all general English courses, this leads to the conclusion that these were, again, pre-learned structures.

4.4 Clause-related bundles

The clause-related bundles are frequently used category of bundles in the corpus, including *that-* and *wh-* fragments, abstract subjects, anticipatory it, and *in order to-* fragment.

The most dominant subcategories are bundles with *that-fragments* and *wh-fragments*. Students' use of these bundles implies that they tend to use longer sentences with subordinate clauses to explain the research segments in more detail or provide clarification (9-12).

(9) A prediction technique *that produces sufficiently accurate predictive models* was selected. (aw48)

(10) There are a number of preventive measures *that can prevent illness and injury* while working on a computer. (no aw123)

(11) The output of a dynamic system is frequently estimated using learning-based algorithms *which are trained on some previous data*. (aw38)

(12) The first part consists of the standard introductory part, *which describes the problems of the subject of the research*, the goals of the work, and the methods used in the development of the work. (no aw117).

That + fragment bundle is the only type of bundle where students with academic writing instructions actually use a category twice more than those without the instructions. The frequency counts reveal that there are 124 *that-clauses* in the sub-corpus 1 and 59 bundles in the sub-corpus 2. Students having instructions utilized it more since they were instructed to elaborate their explanations and write longer sentences.

In addition to *that-* and *wh-* fragments, students also use abstract subjects. This sub-group of clause-related bundles is mainly found in the initial positions in sentences as a part of the purpose bundles. Although students with academic writing instructions utilized abstract subjects more, Table 2 demonstrates that similar types were used by both groups. Dominant abstract subject bundles in both sub-corpora are *this paper presents the* (13), and *this paper describes the* (14). This suggests that both groups of students are familiar with these particular structures and that they associate them with research paper abstract writing.

(13) *This paper presents the* role of the SCADA system in the Smart Grid. (aw160)

(14) *This paper describes the* production and assembly supporting structure of the bridge crane laboratory model, located in the Laboratory for mechanical structures. (no aw206)

Anticipatory *it* is a common clause-related bundle subcategory used for the purpose of disguising authorial interpretation (Hyland, 2008: 11) and focusing on the evaluation of the propositional content (Yakut et al., 2021: 486). As for two sub-corpora, it is evident that students do not write using anticipatory *it* bundles. There are only a few instances in both sub-corpora, and only one is used by more than four graduate students (15, 16).

(15) *It is possible to* select component of some type, and handle its data. (aw33)

(16) In order to achieve optimal organization and uniformity of workload of delivery areas, *it is necessary to* respect the general criteria and data available in application solutions and fieldwork. (no aw227)

Although anticipatory *it* is frequently seen in longer academic texts, such as research papers and theses, this limited usage of anticipatory *it* in abstracts suggests that it may not be typical for short and structurally pre-determined texts to include disguised authorial interpretations.

Finally, students in both sub-corpora rarely used one more clause-related bundle - *in order to* + fragment.

(17) Some of them are applied to data set *in order to determine* soil moisture behavioural regimes and patterns. (aw49).

(18) The problems encountered by the company and potential solutions were identified *in order to improve* the business and the transport itself. (no aw187).

In this study, students with academic writing instruction preferred *in order to determine* and *in order to create*, while the second group used *in order to improve* and *in order to increase*. In all examples, students used it to explain the selection of research methods and to present results. This shows that this clause-related bundle is not mentioned in the academic writing course.

5. Conclusions and pedagogical implications

This research was designed to investigate the most frequently used LBs by Master students when writing research paper abstracts and to attempt to determine whether the number of LBs used, their structures, and diversity differ based on the formal academic writing instructions they had. Contrary to the authors' expectations, there were few significant differences in the use of LBs in both sub-corpora.

Abstracts written by students who had courses in academic writing demonstrated a more diverse use of noun phrases with of-phrase fragments. Students used them mainly to introduce the research background, explain the method, discuss results, or point out conclusions. This confirms that these LBs were discussed in detail while covering the abstract structure. Furthermore, this group of students used more prepositional phrase bundles and therefore were more able to provide explanations for logical relations. Also, more clause-related LBs with *that*-fragments were found in the abstracts by students in sub-corpus

1. The explanation for this may lie in their instructions to write sentences with specific explanations and clarifications.

Although the students who attended academic writing courses wrote better abstracts, this research also shows that these courses insufficiently focus on structures such as lexical bundles and consequently give students very little insight into the importance of using LBs properly.

The authors are also aware of the limitations of this research. Research paper abstracts written by students may not follow the requirements regarding word limitation, leading to writing shorter abstracts missing key moves. It is also difficult to prove whether or not they had expert help or used a translation application to write these abstracts. However, the authors believe that all the same, it is important to examine these abstracts in an attempt to be beneficial in identifying the most relevant and frequent LBs and in designing materials and activities to be covered in academic writing courses.

References:

Anthony, L. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software, 2019.

Appel, R., Murray, L. "L1 differences in L2 English academic writing: A lexical bundles analysis". *Journal of English for Academic Purposes*. 46, 2020. 1-15.

Bestgen, Y. "Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence." *System* 69, 2017. 65-78. http://dx.doi.org/10.1016/j.system.2017.08.004 .

Biber, D. *University language: a corpus-based study of spoken and written registers.* Amsterdam: John Benjamins, 2006.

Biber, D. "A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing." *International Journal of Corpus Linguistics*, 14.3, 2009. 275-311.

Biber, D., Barbieri, F. "Lexical bundles in university spoken and written registers." *English for Specific Purposes*, 26, 2007. 263-286. doi:10.1016/j.esp.2006.08.003.

Biber, D., Conrad, S., Cortes, V. "If you look at …: Lexical bundles in university teaching and textbooks." *Applied Linguistics*. 25.3, 2004. 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. *Longman Grammar of Spoken and Written English*. Harlow, Essex : Pearson Education Ltd., 1999.

Blagojević, S. "Conference abstracts written by Anglophone and Serbian authors: Contrastive analysis." *Proceedings dedicated to Draginja Pervaz: English language and Anglophone literature in theory and practice*. Faculty of Philosophy, Novi Sad. 2014. 65-80.

Buerki, A. "Formulaic sequence: a drop in the ocean of constructions or something more significant?" , *European Journal of English Studies* 20.1, 2016. 15-34, doi: 10.1080/13825577.2015.1136158.

Carter, R., McCarthy, M. *Cambridge grammar of English*. Cambridge, England: Cambridge University Press, 2006.

Common European Framework of Reference for Languages (CERF): Learning, Teaching, Assessment. Available at https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4 ,2020.

Chen, Y.H., Baker, P. "Lexical bundles in L1 and L2 academic writing." *Language Learning &Technology*. 14.2, 2010. 30-48.

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

Dontcheva-Navratilova, O. "Lexical bundles in academic texts by non-native speakers." *Brno Studies in English*, 38.2, 2012. 37-58. DOI: 10.5817/BSE2012-2-3.

Durrant, P. "Lexical bundles and disciplinary variation in university students' writing: Mapping the territories." *Applied Linguistics* 38.2, 2017. 165-193. doi:10.1093/applin/amv011 .

Durrant, P., Mathews-Aydinli, J. "A function-first approach to identifying formulaic language in academic writing." *English for Specific Purposes* 30, 2011. 58-72. doi:10.1016/j.esp.2010.05.002.

Gil, N. N., Caro, E;. M. "Lexical bundles in learner and expert academic writing." *Bellaterra Journal of Teaching & Learning Language & Literature* 12.1, 2019. 65-90.

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

Hyland, K., Jiang, F.K. "Academic lexical bundles: how are they changing?" *International Journal of Corpus Linguistics* 23.4, 2018. 383-407. DOI: 10.1075/ijcl.17080.hyl.

Jiang, F.K. & Hyland, K. "Metadiscursive nouns: Interaction and cohesion in abstract moves." *English for Specific Purposes* 46, 2017. 1-14. http://dx.doi.org/10.1016/j.esp.2016.11.001.

Kanoksilapatham, B. "Generic Characterization of Civil Engineering Research Article Abstracts." *3L: The Southeast Asian Journal of English Language Studies* 19.3, 2013. 1 – 10.

Kopaczyk, J. "Applications of the lexical bundles method in historical corpus research." In P. Pezik, *Corpus Data across Languages and Disciplines*, Berlin: Peter Lang. 2012. 83-95.

Lazić, K. O. *Lexical Bundles in Biotechnical English: A Corpus Analysis of Articles Written by Native and Non-Native Speakers*. 2017. PhD Dissertation, University of Belgrade, Faculty of Philology, available at: https://nardus.mpn.gov.rs/bitstream/handle/123456789/8604/Disertacija.pdf?sequence=6&isAllowed=y

Li, L., Franken, M., Wu, S. "Bundle-driven move analysis: Sentence initial lexical bundles in PhD abstracts." *English for Specific Purposes* 60, 2020. 85-97.

Liu, D. "The most-frequently used multi-word constructions in academic written English: A multi-corpus study." *English for Specific Purposes* 31, 2012. 25-35.

Lu, X., & Deng, J. "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, 2019. 21-36. https://doi.org/10.1016/j.jeap.2019.03.008.

Pan, F., Reppen, R., Biber, D. "Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals." *Journal of*

    English for Academic Purposes 21, 2016. 60-71. http://dx.doi.org/10.1016/j.jeap.2015.11.003.

Pérez-Llantada, C. 'Formulaic language in L1 and L2 expert academic writing: Convergent and divergent use.' *Journal of English for Academic Purposes* 14, 2014. 84-94. http://dx.doi.org/10.1016/j.jeap.2014.01.002.

Shin, Yu Kyoung "Lexical Bundles in Argumentative Essays by Native and Non-native English-Speaking Novice Academic Writers." Dissertation, Georgia State University. https://scholarworks.gsu.edu/alesl_diss/47 2018.

Shin, Y.K. "Evaluate prosody and semantic preference: Extending the analysis of recurrent multi-word sequences." *English for Specific Purposes* 59, 2020. 45-58.

Simpson-Vlach, R., Ellis, N.C. "An Academic Formulas List: New Methods in Phraseology Research." *Applied Linguistics* 31.4, 2010. 487-512. doi:10.1093/applin/amp058 .

Staples, S., Egbert, J., Biber, D., McClair, A. "Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section." *Journal of English for Academic Purposes* 12, 2013, 214-225. http://dx.doi.org/10.1016/j.jeap.2013.05.002.

Wang, Y. "Formulaic sequences signaling discourse organization in ELF academic lectures: a disciplinary perspective. "*JELF* 7.2, 2018. 355-376.

Wray, A. *Formulaic Language and the Lexicon*. Cambridge University Press, 2002.

Wright, H. R. "Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions." *English for Specific Purposes* 54, 2019. 1-14. https://doi.org/10.1016/j.esp. 2018.09.001.

Yakut, I., Yuvayapan, F., Bada E. "Lexical bundles in L1 and L2 English doctoral dissertations." *The Journal of Teaching English for Specific and Academic Purposes* 9.3, 2021. 475-493. https://doi.org/10.22190/JTESAP2103475Y.

STRUKTURANALYSE LEXIKALISCHER BÜNDEL IN ABSTRACT VON DIPLOMSTUDENTEN DES MASTERSTUDIUMS

Lexikalische Bündel als wiederkehrende Wortfolgen sind ein wichtiger Bestandteil sowohl mündlicher als auch schriftlicher wissenschaftlicher Korpora, die sowohl von professionellen Wissenschaftlern als auch von studentischen Schriftstellern verwendet werden, um einer besseren strukturellen Organisation, einer effizienten Kommunikation und einer etablierten Glaubwürdigkeit beizutragen. Diese Ausdrücke sind formelhaft und sehr häufig, haben sowohl eine ziemlich klare Diskursfunktion als auch einen kommunikativen Zweck. Die am häufigsten verwendeten Strukturen lexikalischer Bündel umfassen NP-basierte, VP-basierte und Satz-basierte Bündel. Das Ziel der Arbeit ist es, die lexikalischen Bündel auf einem Lernerkorpus von 448 Abstracts zu untersuchen, die von englischen Nicht-Muttersprachlern während ihres Masterstudiums auf Englisch geschrieben wurden. Es wird eine Strukturanalyse durchgeführt, um zu sehen, welche Strukturen die Studierenden beim Verfassen von Abstracts bevorzugen und verwenden. Das Korpus wurde in zwei Teilen unterteilt, je nachdem, ob Diplomstudenten des Masterstudiums vor der Veröffentlichung ihrer Abstracts formale Anweisungen zum wissenschaftlichen Schreiben erhalten haben, dass man untersuchen könnte, ob Anweisungen die Verwendung von Lexik bei Anfängern beeinflussen. Die Ergebnisse

zeigen die häufigsten Strukturen von lexikalischen Bündeln, die Diplomstudenten des Masterstudiums verwenden, sowie abschließende Bemerkungen, dass Diplomstudenten unabhängig von den gegebenen Schreibanweisungen dieselben Strukturen verwenden und ähnliche Fehler machen.

Schlüsselwörter: Abstrakt, wissenschaftliches Schreiben, lexikalische Bündel, lexikalische Bündelstrukturen.