CHANGES IN THE USAGE OF LEXICAL BUNDLES IN THE FIELD OF EDUCATION DURING THE LAST 20 YEARS

Radmila PALINKAŠEVIĆ, Ph.D.,
Preschool Teacher Training and Medical College, Vršac, Serbia,
palinkasevic@gmail.com

Jelena PRTLJAGA, Ph.D.,
Preschool Teacher Training and Medical College, Vršac, Serbia,
jpivan@sezampro.rs

Abstract: *Formulaic language contributes to demonstrating membership in a specific discourse community. However, the norms of a specific language community tend to change over time, which adds to the complexity of academic writing. Therefore, the aim of this paper is to explore the changes in the usage of lexical bundles in scientific papers in the field of education over the last 20 years. The focus is on lexical bundles, multi word sequences that recur frequently and are distributed widely across different texts. As a prominent feature of any text, exploring their usage and change over time brings insight into the changes which occur in the academic writing style. The results showed a decrease in the usage of lexical bundles over time. Statistically significant differences were found in the structural and functional usage of lexical bundle tokens for both groups and subgroups. This indicates that the concrete bundles used over time have not significantly changed but only the frequency of their usage.*

Keywords: *lexical bundles; corpus linguistics; Academic English.*

1. Introduction

With his work Swales (1990) rekindled the interest of the linguistic academic community in exploring the concept of genre which in turn made analysis of academic discourse a popular topic once again. As a result, structures and functions of academic discourse became a prominent subject of analysis in contrastive and corpus linguistics. Apart from word lists and collocations, lexical bundles became a significant element in the literature which deals with the analysis of academic discourse.

Grammar is based on the principle of open choice of lexical units, but there is another important principle for combining words - the idiom principle, which indicates that a large number of preconceived word

combinations exist within a certain register (Sinclair 1991 according to Nam, 2017). Collocations represent the exploration of lexical combinations which are idiomatic in nature while the study of lexical bundles enables us to explore lexical combinations which are idiomatic and which are not (Nam, 2017). Lexical bundles in different communication registers open up a possibility to gain further insight into how the language is formed in certain genres, disciplines and linguistic communities. Their function was further clarified by Biber and his colleagues in the following way: „In general, these lexical bundles serve as discourse framing devices: they provide a kind of frame expressing stance, discourse organization or referential status, associated with a slot for the expression of new information relative to that frame“ (Biber et al. 2004: 400).

Lexical bundles have been defined in different ways:

- “extended collocations: bundles of words that show a statistical tendency to co-occur” (Biber et al. 1999: 989);
- “recurrent lexical sequences (e.g., take a look at, know what I mean) identified through corpus analysis that includes specific frequency thresholds and dispersion requirements” Pan et al. (2016: 60);
- “extended collocations which appear more frequently than expected by chance, helping to shape meanings and contributing to our sense of coherence in a text” (Hyland 2008a:41).

The initial studies on lexical bundles analysed their presence and usage in different registers and genres. Biber at al. (1999) explored the usage of lexical bundles in conversation, fiction, press and academic discourse. They found that lexical bundles are present to varying degrees in each of these registers and that their structural and functional usage is also different. The second wave of studies focused on analyzing the usage of lexical bundles in different academic fields. Hyland (2008b) explored the usage of lexical bundles in academic texts (academic papers, Master’s theses and doctoral dissertations) in four different fields of study (electrical engineering, biology, business studies and applied linguistics). All of the fields showed different tendencies for using lexical bundles, which further highlighted the role that lexical bundles play in expressing belonging to a certain language community. After Hyland’s (2008b) pioneering work other researchers explored how lexical bundles are used in almost all other academic fields such as medicine (Jalali et al. 2014), chemistry (Kashiha & Heng 2014), psychology (Esfandiari & Barbary 2017), physics (Farvardin 2012), etc.

Once the characteristics of lexical bundle usage in various academic

fields had been determined, the next logical questions posed by researchers were how do varying levels of language proficiency influence the usage of lexical bundles and do native and non-native speakers differ in their usage. Numerous studies (e.g. Cortes 2004, Staples et al. 2013, Bychkovska & Lee 2017) looked at learner usage of lexical bundles in comparison to expert writers, but also compared the output of learners with varying levels of proficiency in regards to their lexical bundle usage. It has been established that learner proficiency influences the usage of lexical bundles. Higher proficiency learners tend to use fewer bundles and lower proficiency learners tend to gravitate towards using more verb-based bundles which are usually the least frequent structural category in the academic context (Staples et al. 2013). The studies which explored how native and non-native speakers use lexical bundles also started from exploring lexical bundle usage in language learners (Appel & Murray 2020, Cortes 2004, Bychkovska & Lee 2017) but later widened their field of research to include expert non-native speakers (Lazić 2017, Pan et al. 2016, Salazar 2014). Most studies of expert native vs. non-native speaker lexical bundle usage showed that non-native speakers use more lexical bundles and that the main differences in structural and functional usage tends to be in the prominence of various subcategories.

The study of lexical bundle usage over time is a branch of research which has still not been well explored (Hyland & Jiang 2018). This paper aims to enrich this direction of lexical bundle research by looking at the changes in lexical bundle usage in educational scientific articles over a period of 20 years, namely from 2001 to 2020. Apart from the fact that our paper explores the changes of lexical bundles over time in a context which has not been researched in this way before, we consider our paper to be significant because it takes into account whether articles were written by native or non-native speakers during the corpus formation phase, which has not been done before in studies which explore lexical bundle changes over time.

2. Previous research of lexical bundles in the field of education

The usage of lexical bundles in the field of education has been studied from a variety of different perspectives:

- the usage of lexical bundles in different educational registers such as classroom teaching, textbooks and academic prose (Biber et al. 2004),
- the usage of lexical bundles in learner writing (Chen & Baker 2010, Nam 2017, Shin 2019),
- the differences in the usage of lexical bundles is scientific articles in the area of education by native and non-native speakers (Güngör & Uysal 2016, Güngör 2016, Güngör 2019),

- paradigmatic influences on the usage of lexical bundles in scientific articles in the field of education (Candarli & Jones 2019),
- changes in the usage of lexical bundles over time in doctoral dissertations in the field of education (Cui & Kim 2021) etc.

As it has been previously shown, lexical bundles greatly vary when it comes to genre and educational field (Hyland 2008b). The study conducted by Biber and colleagues (2004) showed that differences are also prominent when we analyze different registers in the same field. In their study they compared lexical bundles used in classroom teaching and textbooks which they compared to bundles found in conversations and academic prose. The results showed that classroom teaching uses a larger number of different lexical bundles while conversation uses a smaller set of bundles but much more frequently. The bundles present in classroom teaching are associated with both spoken and written registers. A smaller number of lexical bundles is used in textbooks than in classroom teaching and this may be attributed to the mode of presentation i.e. oral or written. This research and section of research in general shows us the importance of analyzing each segment of academic prose for gathering a more comprehensive understanding of its functioning.

In his doctoral dissertation and two more studies Güngör (Güngör & Uysal 2016, Güngör 2016, Güngör 2019) explored the usage of lexical bundles in scientific articles in the field of education by native and non-native speakers. In all of the above-mentioned studies the L1 of the non-native speakers is Turkish. Here in more detail, we will describe the study Güngör & Uysal (2016) since it is the most similar to our study in terms of research design. Two corpora of about half a million words were formed for the needs of the research. One consisting of academic articles in the field of education written by native speakers of English and the other consisting of academic articles written by non-native speakers. In the native speaker corpus 32 four-word lexical bundles were identified while in the non-native corpus the number of identified lexical bundles was 98.

The results indicate that the most prominent structural category in the native speaker corpus was the prepositional phrase-based bundles (50%) followed by noun phrase-based bundles (40.63%), while verb phrase-based bundles and other structures were less frequent (6.25% and 3.12% respectively). In the non-native speaker corpus, the division of lexical bundles in structural categories looks quite different. The most prominent structural category was verb phrase-based bundles (33.67%) closely followed by noun phrase-based bundles (31.63%), while preposition-based phrases and other structures were less

frequently used (24.50% and 10.20% respectively). The functional analysis also showed differences in the usage of lexical bundles by the two groups of speakers. Namely, native speakers used research-oriented bundles (68.8%) most frequently, while non-native speakers used text-oriented bundles most frequently (66%). Both corpora share 13 lexical bundles. It may be concluded that although the non-native speakers used more bundles, they did not use the bundles present in the native speaker corpus frequently enough.

Candarli & Jones (2019) explored whether, in the field of education, there exist differences in the usage of lexical bundles in studies which implement the qualitative or quantitative approach i.e. explored the effect of paradigmatic influences. As Gray (2015:6 as cited in Candarli & Jones 2019) stated "little attention has been paid to the possibility that research articles themselves are not a monolithic concept" and the aforementioned study aimed to contribute to the correction of this oversight by revealing intra-disciplinary variations in the use of lexical bundles. The results of the research indicated that a larger number of lexical bundles were found in the corpus with quantitative research articles than in the corpus with qualitative research articles. The structural analysis showed differences in the two corpora. Namely, in the quantitative corpus the most prevalent bundles were PP-based followed by VP-bundles and in the third position NP-bundles, while in the qualitative corpus PP-bundles were followed by NP-bundles and VP-bundles in the third place. In regards to the functional distribution of lexical bundles it was found that in both corpora referential expressions were the most prevalent followed by stance expressions and discourse organizers. All three functional categories were significantly more frequent in the quantitative corpus. The findings of this study indicate that the paradigm variable is significant when the usage of lexical bundles in research articles is explored and that there might be a need to include it as a factor in further studies.

2.1.Previous research of lexical bundles over time

Hyland & Jiang (2018) were one of the first researchers to explore how lexical bundles change over time. They explored the changes in the usage of lexical bundles over a span of 50 years by selecting three specific years in which the usage of bundles was analyzed: 1965, 1985 and 2015. The usage of lexical bundles was investigated in scientific articles in four different fields: applied linguistics, sociology, electrical engineering and biology. The corpora for the study were built by selecting 30 articles from each field for each of the three time periods from five prominent journals in the specific field.

The results of the study showed a decrease in the usage of lexical bundles in the in the fields of sociology and biology, while an increase

was noted in the fields of applied linguistics and engineering. The structural analysis of the data showed that even though noun/preposition-related bundles accounted for the overwhelming proportion of bundles, they were also the category which had shown the largest overall decline, while the number of bundles containing verb phrases had risen. The most prominent change in the functional usage of lexical bundles in this study was the decrease of research-oriented bundles, increase of participant-oriented bundles in the hard sciences and a decrease of the usage of participant-oriented bundles in the soft sciences.

Another article which studied the changes in the usage of lexical bundles over time is Cui & Kim (2021). This study is closely related to our research for two reasons: 1. because it explored the usage of lexical bundles over the same two time periods that will be explored in this study, namely, 2001-2010 and 2011-2020; 2. because it investigated the field of education. However, their research corpora were comprised of doctoral dissertations in the field of English language teaching while the corpora in this study consist of scientific articles in the field of education. The corpora for each time period are comprised of 30 dissertations in the field of English language teaching. The 2001-2010 corpus consists of 1,200,689 words, while the 2011-2020 corpus consists of 1,189,404 words. A peculiar feature of this research is the length of lexical bundles explored. The researchers opted to analyse 3-, 4- and 5-word lexical bundles, even though the analysis of 4-word lexical bundles is the most common in this field (Chen & Baker 2010, Pan et al. 2016, Bychkovska & Lee 2017). For the identification of lexical bundles, the minimal frequency of 20 occurrences, dispersion in at least 5 texts and an MI score no lower than three were the selection criteria. The implementation of the MI score for the identification of lexical bundles is also not frequently implemented in lexical bundle research.

Lexical bundles in the two time periods were compared in reference to the bundle length, structural and functional categorizations. A statistically significant difference was found for all three lexical bundle lengths. Three- and five-word lexical bundles became more frequently used over time while the usage of four-word lexical bundles decreased.

As in the previous studies which analysed the structure of lexical bundles in educational texts the results of this study showed that noun and preposition-based bundles were more commonly used than verb-based bundles. In the 2001-2010 Corpus noun-based bundles made up 53% and preposition-based bundles made up 35% of all bundles, while in the 2011-2020 Corpus noun-based bundles made up 57% and preposition-based bundles 29% of bundles. The study has shown that the number of nouns based and verb-based bundles has increased in the

last twenty years while the usage of preposition-based bundles has decreased.

When it comes to the functional analysis of the lexical bundles the results showed that in both corpora research-oriented bundles were the most frequent (2001-2010 Corpus 68.3%, 2011-2020 Corpus 72.7%) followed by text-oriented bundles (2001-2010 Corpus 30.4%, 2011-2020 Corpus 26.6%). Participant-oriented bundles were the least frequent in both corpora (2001-2010 Corpus 1.3%, 2011-2020 Corpus 0.7%). As may be seen from the data above, over the last two decades the usage of research-oriented bundles has increased while the usage of text and participant-oriented bundles has decreased. Cui & Kim (2021) analysed the 100 most frequent lexical bundles in more detail and found that only 59 were found in both corpora. Such a finding indicates that the bundles in the area of education are changing quickly in response to new conditions and contexts.

3. Corpus and Methodology

The aim of the study is to explore the change in the usage of lexical bundles over time in a specific scientific register and field. The selected register is scientific papers and the field is education. The explored time period is the past two decades from 2001 to 2020.

Two corpora were created for the purposes of the study. Each of the corpora cover a period of 10 years, the first corpus covers the time period from 2001 to 2010 while the second corpus covers the time period from 2011 to 2020. The corpora are comprised of scientific articles in the field of education. Care was taken that both corpora be of the approximately same size (half a million words) and consist of a similar text number so that the results gathered from each corpus could be easily comparable. Please find detailed information regarding corpora size in Table 1 below.

Table 1. *Corpus information*

	2001-2010 Corpus	2011-2020 Corpus
Word tokens	500,142	500,985
Number of articles	91	90
Average article length	5497	5567
Standard deviation of article length	2227	1799

Specific for the corpora in this study is the fact that the factor of native vs. non-native speaker was taken into account in the corpus formation process. 23% of each corpus consists of scientific articles written by non-native speakers and 77% of articles written by native speakers. The identification of the authors' mother tongue was done using the

working definition proposed by Wood (2001) which indicates that the author must have a typical name for the target language and be associated with an institution in the country where the target language is the official language. For articles with more than one author the name and institution of the first author was checked and if the first author met the abovementioned criteria the paper was included in the corpus.

The corpus analysis program used in this study is AntConc (Anthony, 2013). For the identification of four-word lexical bundles the minimal frequency of 20 occurrences in a million words is set together with a minimum dispersion in 10% of texts (Hyland, 2008). Once the lists of bundles for each corpus are finalized common lexical bundles are identified. Statistical analysis is implemented to check whether possible differences in the number of lexical bundles used in the two time periods are significant.

The next step in the research is the division of the identified lexical bundles into structural and functional categories and the performance of statistical tests to see whether the differences identified are significant. For the structural classification of bundles this study implements the taxonomy proposed by Biber et al. (1999) while the taxonomy proposed by Hyland (2008a) is used for the functional classification. There is often a functional overlap between lexical bundle categories (Biber et al 2004, Chen & Baker 2010) and the same lexical bundle may be categorized in more than one functional category. Because of this, the bundles in our study are categorized in accordance with the most prominently used function of the particular bundle in our corpora. This was accomplished through a detailed analysis of the concordance lines of the identified lexical bundles.

In order to examine if there are any differences, i.e., association, between different corpora in structural and functional distribution of the bundles, a chi-square test of independence is used. For all significant tests, adjusted standardized residuals are calculated in order to determine which cells have the highest deviations from expected frequencies and an absolute value of 1.96 is used as an indicator of statistically significant residuals. In total, 8 chi-square tests were run, for structural groups and corpora, for structural subgroups and corpora, for functional groups and corpora and for functional subgroups and corpora. Tests were run for tokens (total frequencies of bundles) as well as for types (number of specific bundles) appearing in different corpora.

4. Results

Once the identification criteria had been implemented 96 lexical bundle types and 1911 tokens were identified in the 2001-2010 corpus,

while 89 lexical bundle types and 1730 tokens were identified in the 2011-2020 corpus. The lists were then refined by removing content specific bundles and merging four-word bundles which form one five-word bundle. In the 2001-2010 corpus 5 lexical bundles were removed because they were content specific (*the ministry of education, in the united states, education in New Zealand, in the New Zealand, department of education and),* 2 bundles were removed because they were incorrectly identified due to the fact that the program AntConc does not take into consideration numbers (*in the early s, in the late s*), 13 four-word-bundles were merged into 5 five-word-bundles and 1 six-word-bundle (*at the end of (the), (of) the quality of the, (the quality) of teaching and learning, in the context of (the), to the development of (the), (to) meet the needs of*) and 2 bundles were merged because they were the same except a different article was used (*in the/a number of).* In the 2011-2020 corpus 4 lexical bundles were removed because they were content specific (*the ministry of education, in the united states, in the New Zealand, ministry of education and),* 1 bundle was removed because it was incorrectly identified (*et al found that*), 14 four-word-bundles were merged into 7 five-word-bundles (*at the end of (the), as a result of (the), it is important to (note), (for) the purpose of this, at the beginning of (the), are more likely to (be), is one of the (most)*), 4 bundles were merged because they were the same except a different article was used (*as well as the/a, the development of the/a)* and 2 bundles were merged because they were the same except for the singular or plural version of the auxiliary verb (*have/has the potential to).* The final number of lexical bundles in the two corpora can be seen in Table 2 below. From these results we can see that the usage of lexical bundles has decreased over time in scientific articles in the area of education, which is in line with previous research which showed that the usage of 4-word-bundles decreased over time in doctoral dissertations in the field of English language teaching.

Table 2. *Number of lexical bundle types and tokens in the two corpora*

	2001-2010 Corpus	2011-2020 Corpus
Lexical bundle types	83	73
Lexical bundle tokens	1621	1427

Corpus specific bundles, i.e., bundles only present in one corpus, and bundles which were found in both corpora are presented in Appendix 1. There were 31 identical shared bundles, 52 bundles specific to the earlier corpus and 42 bundles specific to the later corpus. However, to the shared bundles we would also add the ones which differ due to the merging of bundles during the refinement process. For example, the bundle *as a result of* in the 2001-2010 Corpus and *as a result of (the)*

in the 2011-2020 Corpus would be added to the list of shared bundles. In this way 14 lexical bundles were added to the shared bundles list (they are put in italics in the appendix 1.). Therefore, the final list consists of 45 shared lexical bundles, 38 bundles specific to the 2001-2010 Corpus and 28 bundles specific to the 2011-2020 Corpus.

During the comparison of the two lists, it was interesting to find two pairs of bundles with the same meaning but which differ stylistically. In the earlier corpus the lexical bundles *a wide range of* and *the degree to which* were identified, while in the later corpus *a broad range of a*nd *the extent to which* were identified. Such details might give us insight into how the popularity of certain expressions is changing.

Differences between corpora based on structural groups – types

There were no significant differences between corpora on structural groups based on bundle types (Table 3), χ^2 (3, N =156) = 1.95, p = .583.

Table 3. Crosstabs between corpora and structural groups (types)

	2001-2010 (%)	2011-2020 (%)
Noun phrase based	30.1	24.7
Prepositional phrase based	49.4	46.6
Verb phrase based	18.1	23.3
Other structures	2.4	5.5
Total	100.0	100.0

Differences between corpora based on structural groups – tokens

There were statistically significant differences in total bundle frequency (tokens) between corpora on structural groups (Table 4.), χ^2 (3, N =3048) = 34.54, p < .001.

Table 4. *Crosstabs between corpora and structural groups (tokens)*

	2001-2010 (%)	2011-2020 (%)
Noun phrase based	31.1	22.9
Prepositional phrase based	49.2	50.9
Verb phrase based	16.2	20.5
Other structures	3.6	5.7
Total	100.0	100.0

Interpretation of adjusted standardized residuals (Table 5.) indicates that there are differences on all structural groups except prepositional phrase-based bundles which are represented in both corpora in very similar proportions. The 2001-2010 Corpus had more noun phrase-based bundles, while later corpus had more verb phrase based and other structure bundles.

Table 5. *Adjusted standardized residuals between corpora and structural groups - tokens*

	2001-2010	2011-2020
Noun phrase based	5.1	-5.1
Prepositional phrase based	-0.9	0.9
Verb phrase based	-3.1	3.1
Other structures	-2.9	2.9

Differences between corpora based on structural subgroups – types

There was no significant association between corpora and structural subgroups based on bundle types (Table 6.), χ^2 (11, N =156) = 7.61, p = .748.

Table 6. *Crosstabs between corpora and structural subgroups (types)*

	2001-2010 (%)	2011-2020 (%)
Noun phrase with of-phrase fragment	24.1	17.8
Other noun phrases	6.0	6.8
Prepositional phrase with embedded of-phrase fragment	42.2	35.6
Other prepositional phrase fragments	7.2	11.0
Anticipatory it + verb phrase/adjective phrase	6.0	5.5
Passive verb + prepositional phrase fragment	0.0	2.7
Copula be + noun phrase/adjective phrase	4.8	4.1
(Verb phrase) + that clause fragment	2.4	2.7
(Verb phrase) + to clause fragment	2.4	4.1
Pronoun/noun phrase + be	1.2	0.0
Other verb fragments	1.2	4.1
Other structures	2.4	5.5
Total	100.0	100.0

Differences between corpora based on structural subgroups – tokens

Chi-square test, χ^2 (11, N =3048) = 102.33, $p < .001$, indicated that there was a statistically significant association between corpora and structural subgroups (Table 7.).

Table 7. *Crosstabs between corpora and structural subgroups (tokens)*

	2001-2010 (%)	2011-2020 (%)

Noun phrase with of-phrase fragment	23.4	16.5
Other noun phrases	7.6	6.4
Prepositional phrase with embedded of-phrase fragment	39.2	37.8
Other prepositional phrase fragments	9.9	13.1
Anticipatory it + verb phrase/adjective phrase	5.6	6.4
Passive verb + prepositional phrase fragment	0.0	1.7
Copula be + noun phrase/adjective phrase	4.4	3.3
(Verb phrase) + that clause fragment	2.7	2.9
(Verb phrase) + to clause fragment	2.1	3.5
Pronoun/noun phrase + be	0.8	0.0
Other verb fragments	0.6	2.7
Other structures	3.6	5.7
Total	100.0	100.0

Adjusted standardized residuals (Table 8.) indicate that noun phrase with of-phrase fragment and pronoun/noun phrase + be bundles were represented more in the earlier corpus, while other prepositional phrase fragments, passive verb + prepositional phrase fragment, (verb phrase) + to clause fragment, other verb fragments and other structures were more frequent in the later corpus. There were no significant residuals for other subgroups.

Table 8. *Adjusted standardized residuals between corpora and structural subgroups*

	2001-2010	2011-2020
Noun phrase with of-phrase fragment	4.8	-4.8
Other noun phrases	1.3	-1.3
Prepositional phrase with embedded of-phrase fragment	0.8	-0.8
Other prepositional phrase fragments	-2.7	2.7
Anticipatory it + verb phrase/adjective phrase	-0.9	0.9
Passive verb + prepositional phrase fragment	-5.2	5.2
Copula be + noun phrase/adjective phrase	1.6	-1.6
(Verb phrase) + that clause fragment	-0.5	0.5
(Verb phrase) + to clause fragment	-2.4	2.4
Pronoun/noun phrase + be	3.4	-3.4
Other verb fragments	-4.5	4.5
Other structures	-2.9	2.9

Differences between corpora based on functional groups – types

Even though there were more research-oriented bundle types in earlier corpus and more text-oriented bundles in later corpus (Table 9.), the chi-square test did not reach statistical significance, χ^2 (2, N =156) = 3.06, p = .216.

Table 9. *Crosstabs between corpora and functional groups (types)*

	2001-2010 (%)	2011-2020 (%)
Research oriented	49.4	37.0
Text oriented	42.2	56.2
Stance oriented	8.4	6.8
Total	100.0	100.0

Differences between corpora based on functional groups – tokens

Crosstabulation between corpora and functional groups is presented in Table 10. Chi-square test was significant, χ^2 (2, N =3048) = 44.35, p < .001, indicating that there was a statistically significant association between corpora and different functional groups.

Table 10. *Crosstabs between corpora and functional groups (tokens)*

	2001-2010 (%)	2011-2020 (%)
Research oriented	45.5	34.3
Text oriented	46.1	57.7
Stance oriented	8.4	8.2
Total	100.0	100.0

Adjusted standardized residuals are presented in Table 11., and results indicate that there were more research-oriented bundles in the earlier corpus, while there were more text-oriented bundles in the later corpus (compared to the expected frequencies).

Table 11. *Adjusted standardized residuals between corpora and functional groups*

	2001-2010	2011-2020
Research oriented	6.3	-6.3
Text oriented	-6.4	6.4
Stance oriented	0.5	-0.5

Differences between corpora based on functional subgroups – types

Chi-square test was not significant, χ^2 (8, N =156) = 5.52, p = .700,

indicating that there was not a significant association between corpora and functional subgroups on bundle types (Table 12.).

Table 12. *Crosstabs between corpora and functional subgroups (types)*

	2001-2010 (%)	2011-2020 (%)
Location	6.0	6.8
Procedure	8.4	6.8
Quantification	15.7	8.2
Description	19.3	15.1
Transition sig.	7.2	12.3
Resultative sig.	7.2	11.0
Structuring sig.	1.2	0.0
Framing sig.	26.5	32.9
Stance features	8.4	6.8
Total	100.0	100.0

Differences between corpora based on functional subgroups – tokens

Crosstabulation between corpora and functional subgroups is presented in Table 13. Chi-square test was significant, $\chi 2$ (8, N =3048) = 112.04, $p < .001$, indicating that there was a statistically significant association between corpora and different functional subgroups.

Table 13. *Crosstabs between corpora and functional subgroups (tokens)*

	2001-2010 (%)	2011-2020 (%)
Location	6.5	9.1
Procedure	6.4	6.5
Quantification	16.2	6.9
Description	16.3	11.8
Transition sig.	10.9	15.0
Resultative sig.	7.6	11.5
Structuring sig.	0.8	0.0
Framing sig.	26.8	31.3
Stance features	8.4	7.9
Total	100.0	100.0

Adjusted standardized residuals indicate that there were no significant differences on two subgroups, procedure and stance features, while there were significant differences on all other subgroups (Table 14). Quantification, description and structuring sig. bundles were more frequent in earlier corpus, while location, transition sig., resultative

sig., and framing sig. were more frequent in the later corpus.

Table 14. *Adjusted standardized residuals between corpora and functional subgroups*

	2001-2010	2011-2020
Location	-2.6	2.6
Procedure	-0.1	0.1
Quantification	8.0	-8.0
Description	3.5	-3.5
Transition sig.	-3.4	3.4
Resultative sig.	-3.7	3.7
Structuring sig.	3.4	-3.4
Framing sig.	-2.7	2.7
Stance features	0.5	-0.5

5. Discussion

The statistical analysis of this study showed that no statistically significant differences in the type structural and functional categories and subcategories occurred, but the differences were significant for tokens. Such results indicate that over time the concrete lexical bundles used have not changed in a significant way but that the frequency of their usage has changed. The fact that the number of lexical bundle types and tokens has decreased over the years may also indicate that academic prose in the field of education is becoming less rigid in structure and that other less formulaic structures are becoming more prominent for expressing the same meanings.

The structural categorization of lexical bundles into categories showed that noun phrase and prepositional phrase bundles make up the majority of the corpora. This is a typical characteristic of academic writing which uses these two groups of structures to focus on the discourse itself rather than to referential content or participants (Hyland & Jiang, 2018). In our study the most prominent structural category is noun phrase bundles followed by prepositional phrase bundles and verb phrase bundles. The same structural category distribution is present in the corpus of educational research articles written by native speakers in the Güngör & Uysal (2016) study. In the Cui & Kim (2021) study the most prominent structural category is the noun phrase bundles followed by prepositional phrase bundles. Such a discrepancy in the results can be explained by the fact that the Cui & Kim (2021) study explored lexical bundles in doctoral dissertations and not research articles and by the fact that her corpus covered a narrower field of education, namely only English language teaching.

The results of the token structural category analysis indicated that usage of noun phrase structures significantly decreased over time while

the usage of verb phrase structures and other structures increased. Verb phrase bundles were shown to increase over time in both Hyland & Jiang (2018) and Cui & Kim (2021). Such a shift might indicate a movement towards a more direct writing style in the scientific papers in the field of education.

In Hyland & Jiang (2018) a preference was shown for the subcategories of *passive verb + prepositional phrase fragment* and *copula be + noun phrase/adjective phrase* within the verb phrase-based bundle category. Our study also shows a significant increase in the usage of the *passive verb + prepositional phrase fragment* over time. Bundles in this subcategory typically mark locative or logical relations (Hyland, 2008b). With their usage writers aspire to guide readers through the text:

(1) The results of these analyses *are shown in Table* 2. (2011-2020 Corpus).

Passive structures are recommended by the APA style guide for the description of the research/experimental set up of the study. The increase of passive bundles might indicate that researchers in the field of education are placing greater focus on the methodological section of their articles. Such a deduction is perhaps supported by the findings in Candarli & Jones (2019) which showed that studies which implemented the quantitative paradigm used more passive lexical bundles than studies with a qualitative paradigm. The increase of the passive lexical bundles in academic research articles in the field of education over time, therefore, might also indicate a paradigmatic shift in the articles but further research is needed for the exploration of that possibility.

In the verb-based bundle subcategories a significant increase was noted in the subcategory *(verb phrase) + to clause fragment* and a decrease in the *pronoun/noun phrase + be.* No change was shown in the usage of the subcategory *anticipatory it + verb phrase/adjective phrase* which indicates that writers continue to frontload statements and evaluate meanings with this structural subgroup of lexical bundles in the same way over time.

As in our study, noun phrase bundles showed a decrease in the usage over time in Hyland & Jiang (2018), but Cui & Kim (2021) noted an increase of usage in their study. Noun phrase lexical bundles are usually used to specify the characteristics of what is being discussed and their decreased use might indicate that authors are starting to employ different forms for conveying such meanings. When we look at the noun phrase token subcategories a significant decrease was determined in the *noun phrase with of-phrase fragment* category.

The functional analysis of lexical bundles in our study showed that in

both of the corpora text-oriented bundles were the most prevalent followed by research-oriented bundles and with participant-oriented bundles as the least frequent. Such results are not in line with previous studies of lexical bundles in the field of education since in both Güngör & Uysal (2016 in the native speaker corpus) and Cui & Kim (2021) the most prevalent functional group was research-oriented bundles followed by text-oriented bundles and with participant-oriented bundles as the least prevalent. Such discrepancies might be explained for Cui & Kim (2021) by the fact that the corpus consists of doctoral dissertations and not academic articles in the field of education and for Güngör & Uysal (2016) by the fact that the corpus of our study contains both native and non-native speaker texts. In the Güngör & Uysal (2016) study the non-native speaker corpus showed the same functional category distribution as in our study.

The statistical analysis of the functional distribution of lexical bundle types did not show any statistically significant results in the main categories or in the functional subcategories. However, statistically significant differences were found in the functional category and subcategory distribution of tokens. Such results indicate that the concrete lexical bundles used in the two time periods did not change but that the frequency of their usage did.

The usage of research oriented lexical bundle tokens showed a statistically significant decrease and it went from 45.5% in the 2001-2010 Corpus to 34.3% in the 2011-2020 Corpus. A decrease in the usage of research-oriented bundles over time was also found in biology and electrical engineering fields in the Hyland & Jiang (2018) study. Research oriented bundles convey real-world and empirically-focused meanings in a text. The research-oriented subcategories which significantly decreased were quantification and description while the usage of bundles in the subcategory of location actually increased. Some examples of lexical bundles which were identified in the 2001-2010 Corpus but which were no longer identified as bundles in the 2011-2020 Corpus are:

- in the quantification subcategory *a wide range of, the total number of, one of the most, the majority of the* etc.;
- in the description subcategory *of the importance of, were more likely to, the relationship between* etc.;

The usage of text-oriented bundle tokens significantly increased over time and went from 46.1% in the 2001-2010 Corpus to 57.7% in the 2011-2020 Corpus. The usage of text oriented lexical bundles helps the author navigate the reader through the text by providing warrants, making conclusions, specifying limitations and connecting ideas. Text oriented subcategories which showed a significant increase were

transition, resultative and framing signals while structuring signals significantly decreased. Some examples of lexical bundles present in the 2011-2020 Corpus but which were not identified in the 2001-2011 Corpus are:

- in the transition signal subcategory *when it comes to, as well as to* etc.
- in the resultative signal subcategory *research has shown that, are shown in table* etc.
- in the framing signal subcategory *to participate in the, in this case the* etc.

6. Conclusion

Like a live organism, language changes and evolves over time, lexical bundles as a part of this system are no exception. Our study contributed to the small but growing research tide which explores how lexical bundles are changing over time. In order to get the most concise information regarding the changes over time we sought to control many variables and focus on a narrow field. Namely the focus was placed on the usage of lexical bundles over time in research articles in the field of education. Our study also controlled for the native/non-native speaker factor which was show in previous studies to influence the number of identified lexical bundles and their usage. The results showed a decrease in the usage of lexical bundles and statistically significant changes in the structural and functional usage of lexical bundle tokens over time. Statistically significant differences were only present for the tokens which indicates that the usage of the specific lexical bundles has not significantly changed, only the frequency with which they are used. The usage of verb phrase bundles increased over time while the usage of noun phrase bundles decreased which might indicate a shift towards a less rigid writing style in educational research articles. The increase of text-oriented bundle and decrease of research-oriented bundles might indicate that writers are increasing their efforts to help readers navigate through the research which they are presenting.

Our study contributes to the understanding of how the usage of lexical bundles changes over time in the field of education by exploring their change in academic articles. We believe that it is also significant because it takes into consideration the factor of native/non-native speaker during the exploration of the changes of lexical bundle usage over time, which to the authors' knowledge has not been done before.

When considering the limitations of the study it should be noted that the corpora used in the study are fairly small. It would be useful to conduct a similar study with larger corpora. Further research into how the usage of lexical bundle changes over time in the writing of native

vs. non-native writers has not been explored, but might provide useful insight into the topic.

References

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research*, *30*(2), 141-161.

Appel, R., & Murray, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes*, *46*, 100873.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, *30*, 38-52.

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

Biber, D. (2012). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), The Oxford handbook of linguistic analysis (pp. 159–191). Oxford: Oxford University Press

Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora*, *14*(2), 237-263.

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

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

Cui, X. J., & Kim, Y. J. (2021). A Comparative Study of Lexical Bundles in Academic Writings between 2001-2010 and 2011-2020. *Multimedia-Assisted Language Learning*, *24*(4), 10-29.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, *29*, 21-42.

Farvardin, M. T., Afghari, A., & Koosha, M. (2012). Analysis of Four-Word Lexical Bundles in Physics Research. *Advances in Digital Multimedia*, *1*(3), 2166-2916.

Güngör, F., & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, *9*(6), 176-188.

Güngör, F. (2016). Cross-linguistic analysis of lexical bundles in L1 English, L2 English, and L1 Turkish research]. *Unpublished doctoral thesis). Gazi University/Eğitim Bilimleri Enstitüsü,*

Ankara.

Güngör, F. (2019). It was found that…: Introductory it Patterns by Native and Non-Native Authors. *Eurasian Journal of Applied Linguistics*, *5*(3), 473-493.

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

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*(1), 4-21.

Hyland, K. & Jiang, K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics.* 23(4): 383-407.

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2014). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information Science and Management (IJISM)*, *13*(1).

Kashiha, H., & Heng, C. S. (2014). Structural analysis of lexical bundles in university lectures of politics and chemistry. *International Journal of Applied Linguistics and English Literature*, *3*(1), 224-230.

Lazić, K. O. (2017). *Učestali leksički spojevi u engleskom jeziku biotehničke struke: Korpusna analiza radova izvornih i neizvornih govornika* (Doctoral dissertation, Univerzitet u Beogradu-Filološki fakultet).

Nam, D. (2017). Functional Distribution of Lexical Bundle in Native and Non-Native Students' Argumentative Writing. *Journal of Asia TEFL*, *14*(4), 703.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60-71.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, *40*, 1-14.

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

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, *12*(3), 214-225.

Appendix 1. *Shared and corpora specific bundles*

Shared	2001-2010	2011-2020
and the development of	*(of) the quality of the*	(for) the purpose of this
and the use of	(the quality) of teaching and learning	a broad range of
as part of the	*(to) meet the needs of*	*are more likely to (be)*
as well as in	a large number of	are shown in table
at the end of (the)	a wide range of	*as a result of (the)*
at the same time	are less likely to	as part of their
at the time of	*are more likely to*	*as well as the/a*
for the development of	*as a result of*	as well as their
in a variety of	as one of the	as well as to
in addition to the	*as well as the*	*at the beginning of (the)*
in relation to the	*at the beginning of*	at the heart of
in terms of the	at the expense of	has been shown to
in the case of	at the level of	*have/has the potential to*
in the field of	at the university of	in a range of
in the form of	by the end of	in a way that
it is necessary to	for each of the	in the area of
on the basis of	for the purposes of	*in the context of*
on the development of	*has the potential to*	*in the number of*
on the one hand	improve the quality of	in this case the
on the other hand	in a number of	*is one of the (most)*
that there is a	in the absence of	*it is important to (note)*
the extent to which	*in the context of (the)*	it is likely that
the fact that the	in the development of	*meet the needs of*
the impact of the	in the light of	of this study was
the nature of the	in the process of	on the part of
the results of the	in the use of	over the course of
the results of this	*in the/a number of*	research has shown that
the role of the	in this paper we	research in this area
there has been a	is likely to be	*the development of the/a*
through the use of	it is difficult to	the findings of this
to be able to	*it is important to*	the first year of
	it is possible that	*the quality of the*
	it is possible to	the students in the
	of education and training	*the way(s) in which*
	of the importance of	there is a need
	of this paper is	this study was to
	one of the most	to focus on the
	the degree to which	to participate in the
	the development of a	*to the development of*
	the introduction of the	when it comes to

	the majority of the	with the exception of
	the needs of the	within the context of
	the number of students	
	the relationship between the	
	the teaching and learning	
	the total number of	
	the use of the	
	the ways in which	
	to ensure that the	
	to the development of (the)	
	were more likely to	
	with a number of	