

Vilnius University
Faculty of Philology
Department of English Philology

Gintarė Strunkytė and Eglė Jurkūnaitė

**Written Academic Discourse: Lexical Bundles in Humanities and
Natural Sciences**

Thesis submitted in partial fulfilment of requirements for the degree of BA in
English Philology

Supervisor: Dr Augustina Stungienė

2008

Acknowledgements

We are particularly grateful for the advice and encouragement of our supervisor Dr Augustina Stungienė. We have greatly profited from her material provided to us, which has contributed enormously to the comprehensiveness of this research. We also like to thank Dr Inesa Šeskauskienė, whose course on Academic Discourse was very helpful in arranging our research paper. Our thanks also go to our colleagues for useful comments on earlier drafts.

We would like to thank Dr Mike Scott from the University of Liverpool, who kindly provided us with the license for his WordSmith Tools computer programme. We also wish to thank Dovilė Savickaitė for contacting with Viviana Cortes, who allowed to use her research articles. All the errors that remain, however, are our own.

Contents

1. Introduction.....	5
2. Data and Methods	9
3. Results and Discussion	10
3.1 Distribution of lexical bundles.....	10
3.2 Structural analysis.....	12
3.3. Functional analysis	19
3.3.1. Stance bundles	20
3.3.2. Text organizing bundles.....	21
3.3.3. Referential bundles	21
3.4 The relationship between structural and functional categories.....	26
4. Conclusions.....	28
5. Summary in Lithuanian	30
6. References.....	31

Abstract

This paper investigates the use of frequent word combinations referred to as ‘lexical bundles’ in research articles in two broad disciplinary domains: humanities, represented by research articles in linguistics and education, and natural sciences, represented by research articles in physics and astronomy. The aim of this study is to compare the use of lexical bundles in the two domains in terms of the frequency of occurrence and distribution of different structural and functional types across the subject areas. Lexical bundles in forty research articles (each domain represented by 20 articles) were identified using WordSmith Tools computer programme. The structural and functional types of lexical bundles found in the articles under analysis were described and compared. It was found that research articles in humanities contain a larger stock of lexical bundles than those in natural sciences. There are three main structural groups of lexical bundles that dominate in two domains: *noun phrase with of-phrase fragment*, *noun phrase with other post-modifier fragment* and *prepositional phrase expression*. As regards functional types, stance and referential bundles occurred more frequently in the articles in humanities, while discourse organizing and referential bundles prevailed in natural sciences articles. The study of the lexical bundles used in the research articles in humanities and natural sciences provided additional evidence for their importance as building blocks of discourse associated with basic communicative functions.

1. Introduction

Linguists have been interested in sequences of words that tend to co-occur for quite a long time. Linguists have used different terms to refer to recurrent word combinations. One of the first to draw attention to recurrent word combinations was Firth (1957). He called them collocations but did not provide an explicit definition. Halliday et.al. (1964:33) gave a more detailed definition of a collocation which emphasized ‘the tendency of a lexical item to co-occur with one or more other words’. Other terms suggested by linguists were *conventionalized language forms* (Yorio 1979); *speech formulas* (Pawley 1985), *ready-made expressions* and *multi-word units* (Cowie 1988), and *fixed expressions* (Moon 1992) (all quoted in Cortes 2004). Although word sequences have been studied under many rubrics, linguists do not agree on the defining characteristics distinguishing one type of word combinations from another.

Nattinger and DeCarrico (1992) also focused on prefabricated language which they found helpful because it can be easily memorized and retrieved when the situation calls for it. According to Yorio (1979), this ability helps to speak with greater fluency and facilitates interaction. It permits both speakers and listeners to direct their attention to the language structure of the discourse rather than to keep focused on individual words. Such forms make communication more orderly: prefabricated language organizes reactions, facilitates choices and reduces the complexity of communicative exchanges. According to Drazdauskienė (1981), lexical bundles perform the phatic function, which means that these bundles help to establish contacts and to get involved in communication. Thus, recurrent word combinations in the phatic function are the result of the demands of the recurrent context of situation realized as a habitual need to express the meaning of sharing and involvement known as a standard of civil behaviour.

A significant step in the investigation of frequent word sequences was made by Biber et al. (1999). They applied new computer programmes to identify recurrent word combinations in texts. By means of these programmes it became possible to identify and count word units of varying length. Biber et al. (1999:990) proposed to call these newly found units *lexical bundles* and defined them as recurrent expressions that usually co-occur in natural language use, regardless of their idiomaticity and their lexical status.

Cortes (2002) and Biber (1999) agree that the main difference between lexical bundles and other recurrent word sequences is the way in which lexical bundles are identified. This identification needs no previous knowledge or intuition of which word units can occur frequently. Lexical bundles are restricted to those word combinations that occur over 10 times per one million words, and that repeat in 5 or more texts. Thus, word sequences which meet this frequency criterion are considered to be lexical bundles (Biber et al. 1999).

These prefabricated sequences differ among themselves in form. According to Nattinger and DeCarrico (1992), 4 structural criteria characterize prefabricated sequences. The first has to do with their length and grammatical level, thus lexical bundles can be of word level (e.g. *all in all*) or sentence level (e.g. *Can I help you?*). The second signifies whether the phrase has canonical (e.g. *on the other hand; a year ago*) or non-canonical (e.g. *want not; by and large*) shape. The third distinguishes whether the phrase is variable (e.g. *this is a X*) or fixed (e.g. *what is this?*). And the fourth shows whether the unit is continuous or discontinuous, that is, whether it consists of an unbroken sequence of words or is interrupted by variable lexical fillers. In applying these criteria, it is necessary to think in terms of a continuum, since it is sometimes difficult to draw a boundary between the categories. Syder and Pawley (1983:38) warn about this difficulty: ‘Again we would assert that this feature of graduation is a fact of language, and in speaking discrete classes we are in danger of misrepresenting the nature of the native speaker’s knowledge’. However, the taxonomy provided by Nattinger and DeCarrico contains as much confusion in the attempt to classify prefabricated sequences structurally as is in the terminology used to refer to these combinations.

Nattinger and DeCarrico (1992) further classify lexical bundles into polywords, institutionalized expressions, phrasal constraints and sentence builders. They understand polywords (e.g. *by the way*) as short phrases that function very much like individual lexical items. They can be both canonical and non-canonical; they allow no variability and are continuous. Institutionalized expressions (e.g. *How are you?*) are combinations of sentence length that are canonical and invariable. Phrasal constraints are constructions of short or medium length, which can be canonical or non-canonical, allow variation and are mostly continuous (e.g. *on the first/second point; as far as I know/can tell*). Sentence builders, according to Nattinger and DeCarrico (1992), are lexical phrases that provide the framework for the whole sentence. They allow considerable variation of phrase and clause elements, and can be canonical and non-canonical, continuous and

discontinuous (e.g. *the fact that; it points the importance of*). As can be seen, this classification is based on a variety of insufficiently defined principles and is rather confusing. Sometimes it is not quite clear to which class a prefabricated sequence should be ascribed, thus, for example, it is not clear whether phrases like *what on earth?; what, me worry?* should be assigned to the category of polywords or institutionalized expressions.

Moreover, prefabricated sequences differ among themselves in functional characteristics. Nattinger and DeCarrico (1992) single out social interactions, necessary topics and discourse devices. Social interactions are understood as phrases that are markers describing social relations, which show how conversations begin, continue and end (e.g. *have you heard about?; thanks very much*). According to Nattinger and DeCarrico (1992), necessary topics are those phrases which mark topics about which learners will be asked, or those that are necessary in daily conversation (e.g. time, weather). Discourse devices stand for phrases that connect the meaning and structure of the discourse. As can be seen, the classification of the functions performed by prefabricated sequences is not based on a single principle and is rather complicated. Some prefabricated sequences are classified according to the functions they play in the logical structure of the text, i.e. in creating its textuality, for example spatial connectors, temporal devices; other lexical units are classified according to the pragmatic functions they perform, for example, fluency devices.

Type of discourse also accounts for differences in the use of lexical bundles. According to Hill (1985, quoted in Bhatia 1993), type of discourse is responsible for the linguistic variety and determines differences in the use of linguistic items in terms of their structure and function. Many linguists concentrated on exploring lexical bundles in scientific discourse. This type of discourse contains a lot of declaratives, highly routinised and formulaic descriptions of procedures, results and discussions (Hyland 1998). Halliday (1988) also points out two most prominent features of scientific discourse, namely the occurrence of nominalizations and the prevalence of relational intensive processes, realized by verbal elements which relate nominalized processes externally (e.g., happening A *causes* happening B) or internally (e.g., happening A *causes me to think* happening B). According to Cortes (2005), 60% of all lexical bundles identified in academic prose, are parts of noun phrases (e.g. *as a result of, the nature of the*).

Cortes in her study *Lexical Bundles in Freshman Composition* (2002) compares the use of lexical bundles in the corpus of freshman writing, academic prose and conversation. She infers that the

main function of lexical bundles consists in helping students to speak and write with greater fluency and accelerating language acquisition process. Biber (1999) in his study compared lexical bundles in conversation and academic prose, while Cortes, Biber and Conrad (2005) investigated lexical bundles in student disciplinary writing (history and biology). Biber and his colleagues concluded that lexical bundles function as basic building blocks of the discourse.

Although considerable research has been carried out into lexical bundles in various discourses, rather little attention has been paid to analyzing lexical bundles in research articles in different disciplines. This paper is meant to fill in this gap. It examines lexical bundles used in humanities (education and linguistics) and natural sciences (physics and astronomy), namely, in research articles. Thus, the purpose of this paper is to conduct a cross-disciplinary analysis of lexical bundles. The following research questions helped to focus the investigation: 1. What are the most frequent lexical bundles in humanities and natural sciences? 2. Are there any structural and functional differences between lexical bundles in the two domains?

The remainder of this paper is divided into three sections. Section 2 describes the method and data used in the research. Section 3 presents a list of grouped and categorized lexical bundles, as well as a comparison between lexical bundles used in humanities and natural science articles. Besides, this section includes a discussion of process analysis, indicates the limitations of the research and gives recommendations for further research. Finally, section 4 provides conclusions drawn about the differences between lexical bundles used in the research articles in humanities and natural sciences.

2. Data and Methods

In order to investigate lexical bundles in the two broad disciplinary domains, humanities and natural sciences, 40 research articles have been randomly selected from such well known journals as *English for Specific Purposes*, *PMC Physics A* and some others. In this research articles in humanities are represented by linguistics and educology and articles in natural sciences are represented by physics and astronomy. Each domain is represented by 20 articles, published in the period between 1998-2008. The articles ranged in length from 2.069 to 22.053 words, with a mean of 8181 words. The total number of words in the articles in each domain accounted for approximately 130.000 words.

Lexical bundles were identified using WordSmith Tools computer programme. This determined the quantitative mode of our research. The computer programme identified every 3, 4 and 5-word sequences in the research articles. Then every article storing recurrent sequences of words has been read by this programme. Each time a sequence was identified, it was automatically checked against previously identified sequences, and a running frequency count showed how often each sequence was repeated. The identification of lexical bundles was based on orthographic word units that repeated in five or more texts. Lexical bundles that contained a turn boundary or a punctuation mark were excluded. It should also be said that the computer programme not being created for the identification of lexical bundles, had to be specially adjusted to our needs.

To limit the scope of the investigation, only recurrent sequences of 3, 4 and 5-word bundles were analyzed in detail. They have been classified according to their structure and function. In the structural and functional classifications of lexical bundles we followed Biber, Conrad and Cortes' taxonomies (2005). Then a frequency driven approach was applied for the comparison of lexical bundles in the two disciplinary domains: structural and functional types prevailing in each domain were singled out, described and compared.

The research method used was primarily inductive. We grouped together bundles that served similar functions based on their typical meaning and use. We applied concordance listings for analyzing the use of each bundle in its discourse context.

3. Results and Discussion

3.1 Distribution of lexical bundles

Figure 1 gives the overall distribution of lexical bundles in the two domains. The total number of all lexical bundles was 2179. The research articles in humanities account for 1251 bundles, while the research articles in natural sciences contain 928 bundles. The research articles in humanities contain a larger stock of lexical bundles than the research articles in natural sciences. This suggests that the language of humanities is more varied.

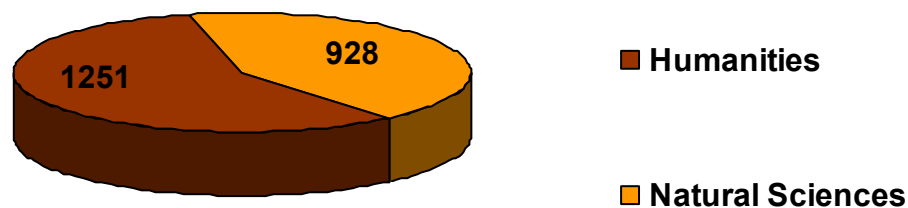


Figure 1. Overall distribution of lexical bundles in the two domains

Figure 2 below shows the distribution of 3, 4 and 5-word bundles in the two domains. 3-word bundles turned out to be the most common in both domains as they account for 82% of the lexical bundles in the research articles in humanities and for 76% of the lexical bundles in the articles in natural sciences. As regards 4-word bundles, they are less numerous than 3-word bundles. They are more common in natural sciences. However, the difference between their percentages is very slight - it amounts to 1%. 5-word bundles are the least numerous ones as they stand for only 2% of the lexical bundles in humanities and 7% in natural sciences. Their low frequency of occurrence could be explained by the complexity of their production. It takes the writer more effort and time to produce a 5-word bundle than 3 or 4 – word bundles. Moreover, the writer who cares for his readership and wants the reader to grasp the meaning of his text easily, tries to use less complex expressions.

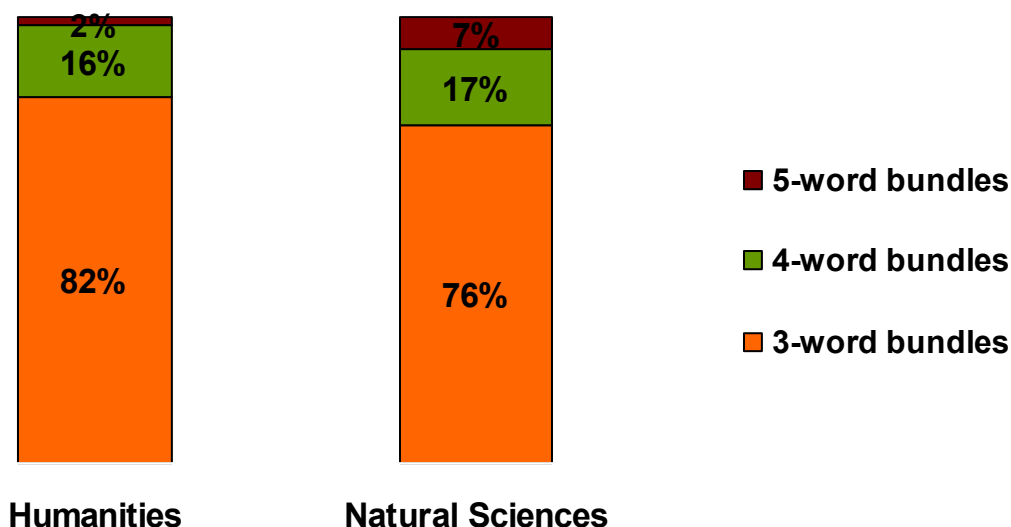


Figure 2. Distribution of 3, 4 and 5-word bundles in the two domains

Table 1 below presents the most frequent lexical bundles in humanities and natural sciences. Their frequencies of occurrence are indicated in brackets. Overall, according to the frequency of occurrence, 3-word bundles in the two domains take the first place. *The use of*, which occurs 142 times, is the most frequent lexical bundle within the group of 3-word bundles in humanities, while *the number of*, which occurs 83 times, is the most frequent in the group of 3-word bundles in natural sciences. 4-word bundles take the second place in both domains. The most frequent 4-word bundle in humanities is *on the other hand* (44) and *in the case of* (49) in natural sciences. The last place is taken by 5-word bundles. The highest frequency of occurrence is displayed by *the difference in the use* (10) in humanities and *the maximum extent of the* (13) in natural sciences. However, only a few lexical bundles occur with very high frequencies. An interesting phenomenon is that the most frequent 3, 4 and 5-word lexical bundles in natural sciences all belong to the structural category called *noun phrase with of-phrase fragment*.

Table 1. The most frequent lexical bundles in humanities and natural sciences

Humanities	Natural sciences
3-word bundles:	3-word bundles:
<i>the use of</i> (142)	<i>the number of</i> (83)
4-word bundles:	4-word bundles:
<i>on the other hand</i> (44)	<i>in the case of</i> (49)
5-word bundles:	5-word bundles:
<i>the differences in the use</i> (10)	<i>the maximum extent of the</i> (13)

3.2 Structural analysis

All lexical bundles have been classified according to the structural taxonomy suggested by Biber, Conrad, and Cortes' (2005). This taxonomy falls into 4 structural categories, each with its sub-categories. They are presented in Tables 1, 2, 3, 4 below and illustrated by the examples from the examined research articles in humanities and natural sciences.

Table 2 below presents the first structural category, which includes lexical bundles incorporating verb phrase (VP). It is divided into the following sub-categories: *passive verb + prepositional phrase fragment*; *copula be + noun phrase/adjective phrase*; *anticipatory it + verb phrase/adjective phrase*.

Table 2. Lexical bundles incorporating verb phrase (VP)

<i>passive verb + prepositional phrase fragment</i>	
Humanities:	Natural sciences:
1. are used to	1. is dominated by
2. is based on	2. is divided into
3. are shown in	3. is related to the
4. are expected to	4. are characterized by the void
5. are associated with	5. be discussed in
6. be related to	6. is defined as the
7. be seen as	7. is determined by the
8. be found in	8. be observed in
<i>copula be + noun phrase/adjective phrase</i>	
Humanities:	Natural sciences:
1. is an example	1. is the probability
2. are a number of	2. is the average
3. is consistent with the	3. is due to
4. are able to	4. is important to
5. is the fact	5. is related to
6. is the case	6. be able to
7. are the same	7. is a demonstration
8. be relevant to	8. is far from
<i>anticipatory it + verb phrase/adjective phrase</i>	
Humanities:	Natural sciences:
1. it is worth	1. it is assumed
2. it is possible	2. it is found
3. it may be that	3. it was determined
4. it appeared that	4. it is easy to
5. it seems that	5. it is practical
6. it comes to the	6. it would be

7. <i>it is important to</i>	7. <i>it should be noted</i>
8. <i>it is clear</i>	8. <i>it was interesting</i>

Table 3 below presents the second structural category. It includes lexical bundles incorporating noun phrase (NP). Its sub-categories are: *noun phrase with of-phrase fragment*; *noun phrase with other post-modifier fragment* and *noun/pronoun phrase + be (+...)*.

Table 3. Lexical bundles incorporating noun phrase (NP)

<i>noun phrase with of-phrase fragment</i>	
Humanities:	Natural sciences:
1. <i>the use of</i> 2. <i>the result of</i> 3. <i>the purpose of</i> 4. <i>context of the</i> 5. <i>role of the</i> 6. <i>a combination of</i> 7. <i>the use of a wide</i> 8. <i>the mean number of clauses</i>	1. <i>the number of</i> 2. <i>wide range of</i> 3. <i>view of the</i> 4. <i>the basis of the</i> 5. <i>the total number of</i> 6. <i>a comparison of the</i> 7. <i>percent of the total</i> 8. <i>the end of the</i>
<i>noun phrase with other post-modifier fragment</i>	
Humanities:	Natural sciences:
1. <i>context in which</i> 2. <i>degree to which</i> 3. <i>the belief that</i> 4. <i>the need to</i> 5. <i>the relationship between the</i> 6. <i>the lesson in</i> 7. <i>significant differences between the</i> 8. <i>the fact that the</i>	1. <i>the results for</i> 2. <i>the ratio between</i> 3. <i>signal in the</i> 4. <i>information about the</i> 5. <i>agreement with the</i> 6. <i>the ways in which</i> 7. <i>the contrast between</i> 8. <i>addition to the</i>
<i>noun/pronoun phrase + be (+...)</i>	
Humanities:	Natural sciences:
1. <i>quotation is a</i> 2. <i>syntactic structures are</i> 3. <i>they are not</i> 4. <i>it is also</i> 5. <i>this article is</i> 6. <i>there were no</i> 7. <i>this view is</i> 8. <i>there is some</i>	1. <i>there is a</i> 2. <i>the system is</i> 3. <i>this paper is</i> 4. <i>they will be</i> 5. <i>this is not</i> 6. <i>there is also a</i> 7. <i>there are a number</i> 8. <i>there are many</i>

Table 4 below presents the third structural category – lexical bundles incorporating prepositional phrase (PP). It falls into the following sub-categories: *prepositional phrase with embedded of-phrase fragment*; *prepositional phrase expressions*.

Table 4. Lexical bundles incorporating prepositional phrase (PP)

<i>prepositional phrase with embedded of-phrase fragment</i>	
Humanities:	Natural sciences:
<ol style="list-style-type: none"> 1. <i>by means of</i> 2. <i>on the part of</i> 3. <i>in the speech of</i> 4. <i>from the point of</i> 5. <i>on the basis of the</i> 6. <i>in the frequency of</i> 7. <i>to the use of</i> 8. <i>for the analysis of</i> 	<ol style="list-style-type: none"> 1. <i>in the form of</i> 2. <i>for the case of</i> 3. <i>by the presence of</i> 4. <i>at the level of</i> 5. <i>on the structure of</i> 6. <i>in the construction of</i> 7. <i>from the perspective of the</i> 8. <i>at the end of the</i>
<i>prepositional phrase expressions</i>	
Humanities:	Natural sciences:
<ol style="list-style-type: none"> 1. <i>in this article</i> 2. <i>for example in</i> 3. <i>from one language</i> 4. <i>on the grounds that</i> 5. <i>on the part</i> 6. <i>in the educational process</i> 7. <i>of the social embedding</i> 8. <i>with initial stress ending in</i> 	<ol style="list-style-type: none"> 1. <i>to the next</i> 2. <i>of the system</i> 3. <i>in addition to the</i> 4. <i>with the case of</i> 5. <i>about the characteristics</i> 6. <i>at the same time</i> 7. <i>in sections 2 and 3</i> 8. <i>of changes in the</i>

Table 5 below presents the fourth structural category – lexical bundles incorporating clause fragment (CF). It has 3 sub-categories: *(verb phrase+) that-clause fragment*; *(verb/adjective+) to-clause fragment* and *adverbial clause fragment*.

Table 5. Lexical bundles incorporating clause fragment (CF)

<i>(verb phrase+) that-clause fragment</i>	
Humanities:	Natural sciences:
<ol style="list-style-type: none"> 1. <i>to suggest that</i> 2. <i>show that the</i> 3. <i>to ensure that</i> 4. <i>may be that</i> 5. <i>indicate that the</i> 6. <i>was determined that</i> 7. <i>will have that</i> 8. <i>means that the</i> 	<ol style="list-style-type: none"> 1. <i>was found that</i> 2. <i>note that the</i> 3. <i>assume that the</i> 4. <i>will have that</i> 5. <i>found that the</i> 6. <i>have shown that</i> 7. <i>determined that the</i> 8. <i>reveal that the</i>
<i>(verb/adjective+) to-clause fragment</i>	
Humanities:	Natural sciences:
<ol style="list-style-type: none"> 1. <i>is likely to</i> 2. <i>be used to</i> 	<ol style="list-style-type: none"> 1. <i>is critical to</i> 2. <i>are difficult to</i>

3. <i>is possible to</i> 4. <i>be more likely to</i> 5. <i>is necessary to</i> 6. <i>proved to be</i> 7. <i>is due to</i> 8. <i>be useful to</i>	3. <i>is proportional to the</i> 4. <i>considered to be</i> 5. <i>is easy to</i> 6. <i>will be to</i> 7. <i>were able to</i> 8. <i>is proportional to the</i>
<i>adverbial clause fragment</i>	
Humanities:	Natural sciences:
1. <i>as mentioned earlier</i> 2. <i>as measured by</i> 3. <i>as regards the</i> 4. <i>as illustrated in</i>	1. <i>as indicated in</i> 2. <i>as seen in</i> 3. <i>as opposed to the</i> 4. <i>as discussed in</i>

Figures 3, 4, 5 and 6 below show the distribution of lexical bundles across the sub-categories of each structural category in each domain.

Figure 3 below shows the distribution of bundles across the sub-categories of lexical bundles incorporating verb phrase (VP) in the two domains. The number on the top of each column shows how many bundles belong to a certain sub-category. It turned out that the distributions of lexical bundles across the sub-categories in humanities and natural sciences do not differ dramatically. The first place in humanities is taken by *anticipatory it + verb phrase/adjective phrase*, while in natural sciences it is *passive verb + prepositional phrase fragment*. The second most dominating sub-category in humanities is presented by *passive verb + prepositional phrase fragment* and by *copula be + noun phrase/adjective phrase* in natural sciences. *Copula be + noun phrase/adjective phrase* takes the last place in humanities, while *anticipatory it + verb phrase/adjective phrase* appeared to be the least numerous in the language of the research articles in natural sciences.

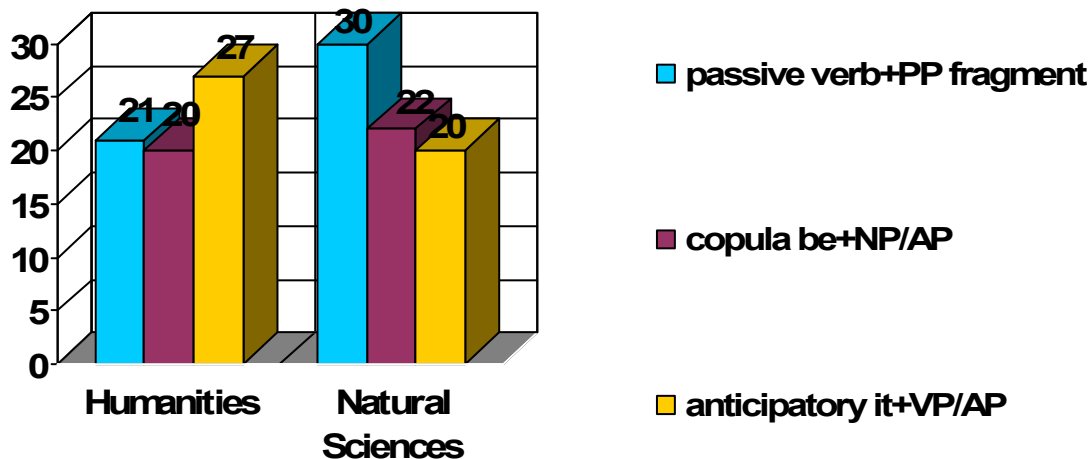


Figure 3. Distribution of lexical bundles across the sub-categories of lexical bundles incorporating verb phrase (VP) in the two domains

Figure 4 below shows the distribution of bundles across the sub-categories of lexical bundles incorporating noun phrase (NP) in the two domains. The ways this kind of lexical bundles are distributed across the domains are fairly similar. *Noun phrase with of-phrase fragment* is most dominating sub-category in both domains. It contains 357 lexical bundles in humanities and 445 bundles in natural sciences. The second position is occupied by *noun phrases with post-modifier fragments* and the third by *noun/pronoun phrases + be (+...)*.

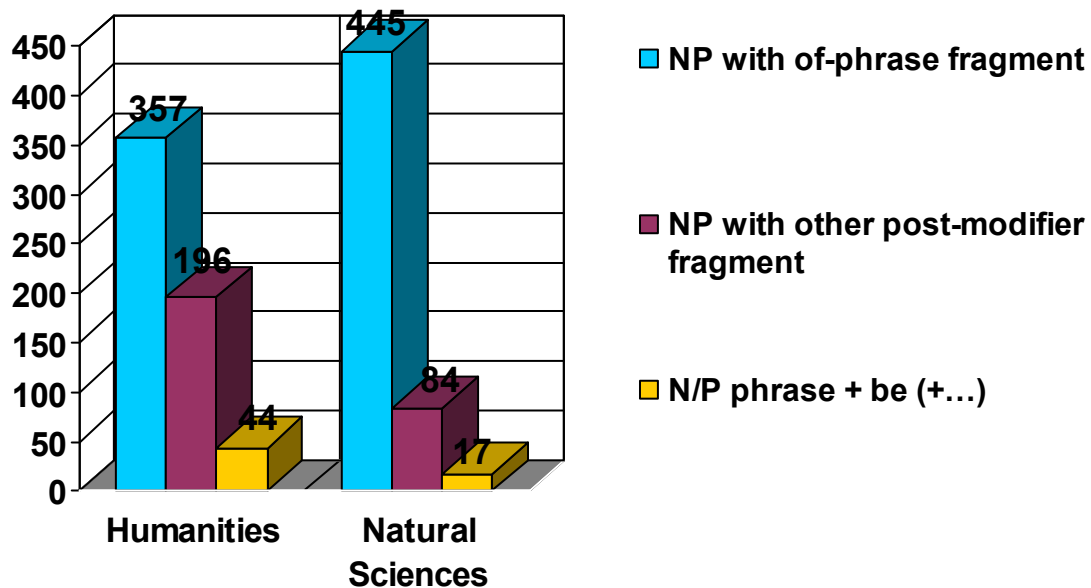


Figure 4. Distribution of lexical bundles across the sub-categories of lexical bundles incorporating noun phrase (NP) in the two domains

Figure 5 below shows the distribution of bundles across the sub-categories of lexical bundles incorporating prepositional phrase (PP) in the two domains. The numbers at the top of the columns suggest that *expressions with a prepositional phrase* take the leading position both in the research articles in humanities (418 bundles) and in the research articles in natural sciences (180 bundles). *Prepositional phrases with of-phrase fragment* are very much lagging behind as they amount to 64 bundles in humanities and 49 in natural sciences.

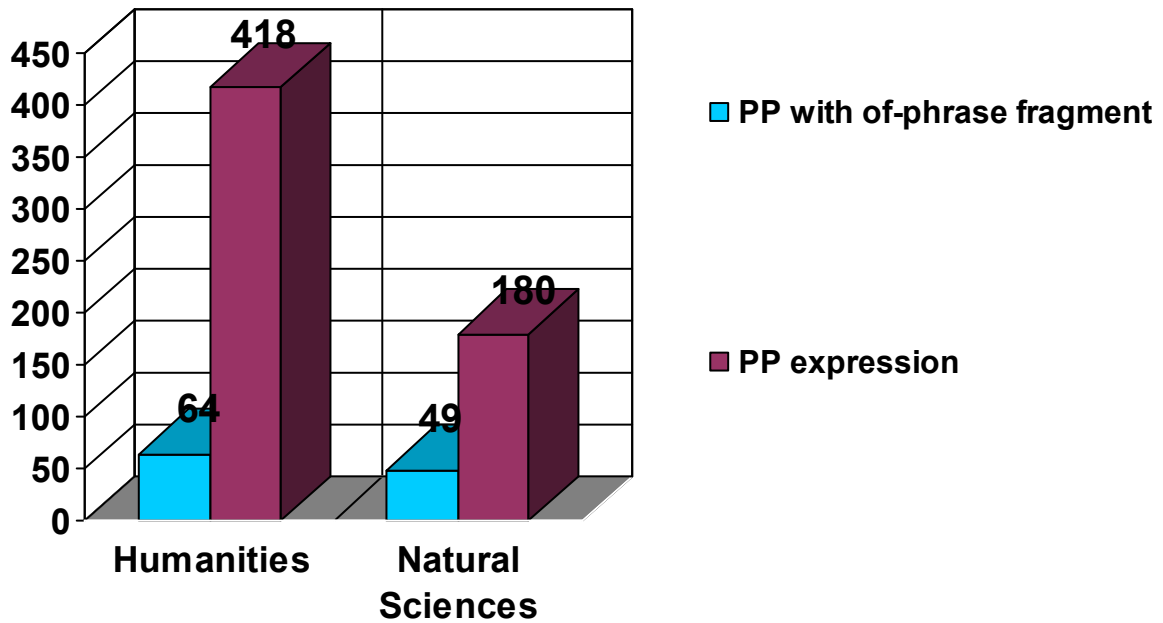


Figure 5. Distribution of lexical bundles across the sub-categories of lexical bundles incorporating prepositional phrase (PP) in the two domains

Figure 6 below shows the distribution of bundles across the sub-categories of lexical bundles incorporating clause fragment (CF) in the two domains. Their distribution is not as even as with lexical bundles incorporating noun phrase (Figures 4) and lexical bundles incorporating prepositional phrase (Figure 5). The first place is taken by *(verb/adjective +) to-clause fragment* in both domains. However, the next two places show a greater variability in the distribution. The second place is taken by *(verb phrase +) that-clause fragment* in humanities and by the adverbial clause fragment in natural sciences. The third place is taken by *adverbial clause fragment* in humanities, while in natural sciences it is taken by *(verb phrase +) that-clause fragment*.

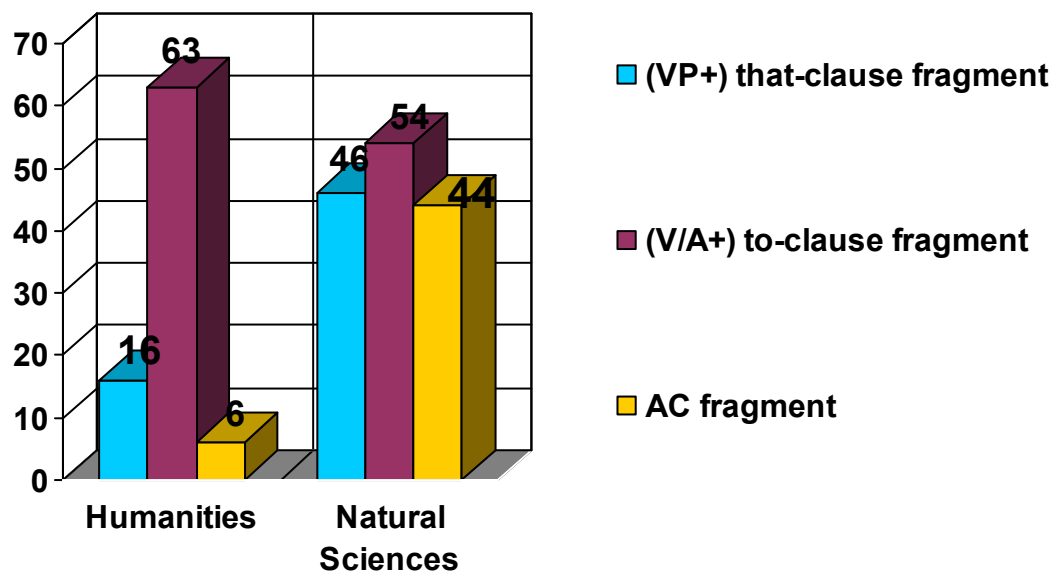


Figure 6. Distribution of lexical bundles across the sub-categories of lexical bundles incorporating clause fragment (CF) in the two domains

To limit the scope of the graphical presentation and to give the overall view of the distribution of lexical bundles, four major structural categories without the sub-category division are presented in Figure 7 below. The figures given here denote percentages. It turned out that lexical bundles incorporating noun phrase present the dominating structural category across the two domains. The incorporated noun phrase usually stands for some scientific term. This fact supports Halliday's (1988) claim that one of the most prominent features of scientific discourse is the occurrence of nominalizations. As can be seen from Figure 7 below that the percentages of lexical bundles incorporating noun phrase in the two domains differ - they account for 44 % in humanities and for 55 % in natural sciences. These figures suggest that the language of the research articles in natural sciences contains more nominalizations than the language of humanities. Lexical bundles incorporating prepositional phrase take the second place according to their frequency. The language of humanities tends to use more bundles of this type (39 %) than the language of natural sciences (26 %). The difference in the use shows that the language of humanities is more varied as it tends to include more expressions with prepositions. The third place is taken by lexical bundles incorporating verb phrase. Their percentages in the two domains differ slightly, i.e. 9 % in humanities and 10 % in natural sciences. The difference is not great and

shows that bundles play quite a significant role in both domains. According to Halliday (1988), verbal elements relate nominalized processes externally (e.g. happening A causes happening B) or internally (e.g. happening A causes me to think happening B) and are a significant element of scientific discourse. Lexical bundles incorporating clause fragment take the last place and do not differ much in their distribution across the two domains. They amount to 8 % in humanities and to 9 % in natural sciences.

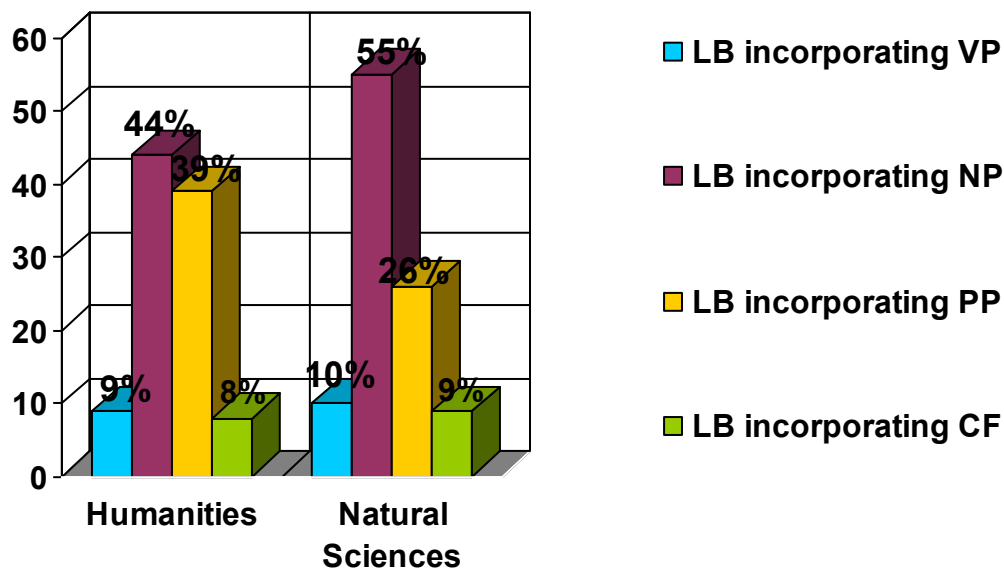


Figure 7. The number of lexical bundles across the structural categories in each domain

There is a greater variety of structural types of lexical bundles in humanities. This fact allows us to assume that the language of the research articles in humanities is more varied than the language of the research articles in natural sciences.

3.3. Functional analysis

As it was mentioned above, in the functional classification of lexical bundles we followed the taxonomy suggested by Biber, Conrad and Cortes (2005). This classification was chosen because its application is quite broad – it can be used to analyse functions realized in any discourse. All lexical bundles fall into three main functional categories: stance bundles, text organizing bundles

and referential expressions; each of the main categories falls into several sub-categories. In this study we discuss only those sub-categories which were found in the corpora.

3.3.1. Stance bundles

Stance bundles express attitudes or assessments of certainty that frame some other proposition. There are two groups of stance bundles: epistemic stance bundles and attitudinal/modality stance bundles. Epistemic stance bundles express some degree of certainty:

1. *These judgments were inevitably subjective, and it may be that in some cases an original, and simpler, version could be considered stylistically superior or more authentic in terms of voice.* (Humanities)

2. *It is possible that those grammatical structures which are most significant for students' writing development at any given stage can only be explored productively at a level below explicit awareness.* (Humanities)

Attitudinal/modality stance bundles express the speaker's/writer's attitudes towards the actions or events described in the following proposition. Two major categories of attitudinal/modality stance bundles are distinguished: obligation/directive and ability.

Obligations or directives are understood as personal expressions of stance, directing the listener/reader to carry out actions that the speaker/writer wants to have completed, or expressing predictions of future events that do not involve any participation of the speaker:

3. *Therefore, it is necessary to develop a very low noise read-out electronics characterised by high sensitivity in order to exploit the noise performances of the detector.* (Natural sciences)

4. *Although, it is worth noticing that the isotropically emitted Auger electron is not modulated by the X-ray polarization and therefore represents a disturbance, especially at low energy.* (Natural Sciences)

5. *In addition to the need for greater understanding of the use of languages in classroom activities, it would be a good idea to explore the demands and affordances of language learning in different curriculum areas.* (Humanities)

Ability bundles express ability and they are usually impersonal:

6. *Measuring only one or two channels, it is possible to measure response times of the order of 500 ms that are still limited by the read-out process.* (Natural sciences)

7. *If teacher-learners are not to be merely passive empty vessels into which knowledge is poured, then they need to be able to shape the course of the talk.* (Humanities)

3.3.2. Text organizing bundles

There are two major sub-categories included under text organizing bundles: topic introduction/focus and topic elaboration/clarification. Topic introduction bundles provide obvious signals that a new topic is being introduced. Many of these are expressions of intention or desire, but quite a number of them have a more specialized function, that of announcing the intention to begin a new topic. Consider the following examples:

8. *There is a variety of approaches to solve partial differential equations (PDEs) both analytically and numerically.* (Natural sciences)
9. *This is the region where most movements are background, and the peak with the higher intensity recorded whenever the <...>.* (Natural sciences)
10. *This is a growing concern in training where massive access to internet information, which is all too often of seriously inferior engineering quality [66], leads to the 'Internet Engineer' who in fact is deceived by Virtual Unreality [67].* (Natural sciences)

The second major sub-category of text organizing bundles is topic elaboration or clarification bundles. These bundles are used when additional explanation or clarification is required:

11. *The waveform recorder is a major challenge since it is required to operate at a rate of to record pulses with a tight power constraint.* (Natural sciences)
12. *The ANITA (Antarctic Impulsive Transient Antenna) experiment is designed to detect UHE neutrinos from GZK interactions of UHECRs with CMB photons, which requires an enormous target volume.* (Natural Sciences)
13. *This means that, for example, for the 80% of the day devoted to Spanish in kindergarten, the teachers give directions, speak, read books, and sing songs in Spanish only.* (Humanities)
14. *This suggests that the structural constraints from both languages are not only accessible but in simultaneous operation when code switching occurs.* (Humanities)

3.3.3. Referential bundles

The third functional category is referential bundles that generally identify an entity or single out some particular attribute of an entity as especially important. There are three major sub-categories included under referential bundles performing three major functions: identification/focus, specification of attributes and time/place/ text reference.

Identification/focus bundles focus on the noun phrase following the bundle as especially important. In many cases, identification/focus bundles also have a discourse organizing function.

These bundles are often used after a lengthy explanation to emphasize or summarize the main point:

15. *Photons are emitted from a user-defined light source at Zs and their flux and time distributions are recorded and averaged in all spatial cells the photons traverse, one of which is shown as a shaded volume.* (Natural sciences)

16. *In the present design the detector has 12 vertical strings, each of which has an instrumented height of about 350 m and consists of 25 storeys with three optical sensors each.* (Natural sciences)

17. *“What is success” – that’s one of the questions I asked people in the first linguistic interviews I put together.* (Humanities)

In other cases, identification/focus bundles can be used to introduce a discussion by stating the main point first, and then giving the details:

18. *Two of those devices are part of the alpha particle X-ray spectrometers (APXS) on the Mars Rovers [4] analyzing the chemical composition of the Mars soil by measuring the alpha induced X-ray spectrum (Fig. 3).* (Natural sciences)

19. *Each of the three Wolter I mirror telescopes consist of 57 nested parabolic–hyperbolic mirror shells with 7.5 m focal length.* (Natural sciences)

20. *Those of the groups that is evidentials include stance expressions which show certainty or doubt or which evaluate the reliability of information.* (Humanities)

21. *One of the seven linguists was a goldsmith by trade, and did much of the work of things like plating royal artefacts.* (Natural Sciences)

The third sub-category of referential bundles identifies specific attributes of the following head noun. Some of these bundles specify quantities or amounts:

22. *The new generation large telescopes like IceCube and KM3NeT will be able to detect cosmic neutrinos of all flavors, over a wider range of energies and with an improved angular resolution.* (Natural sciences)

23. *At energies above, the degradation of the effective area is smaller and about one half of the peak at the zenith, which means that IceCube can observe a large part of the Galaxy, including the Galactic centre.* (Natural Sciences)

24. *There are a number of pairs where the noun ends in a voiceless and the verb in a voiced fricative.* (Humanities)

25. *Utterance-final position accounted the biggest part of all the cases provided below.* (Humanities)

Other bundles in this category describe the size and form of the following head noun:

26. *The pixel size of the array should be adequate to the required spectral resolution.* (Natural sciences)

27. *Due to the short absorption length of the vacuum UV (VUV) radiation in diamond (e.g. at 160 nm the absorption length is about 20 nm), the coplanar configuration of the electrodes forming each pixel is to be preferred because it maximizes the collection efficiency.* (Natural Sciences)

28. *These frames focus on the structure surrounding the stance expressions, such as preceding a verb, as well as the lexical context (great in Great Lakes is not stanced).* (Humanities)

Some specifying bundles can identify abstract characteristics:

29. *In all cases, the repeated nature of the PD projects and the knowledgeable people involved in eliciting the information flow dependencies reduce the risk of error in the construction of the PD networks.* (Natural sciences)

30. *In closing, I would like to emphasize the importance of the findings in this study: in terms of raw frequencies, men and women show little difference in affect, evidentiality and quantification, except in the lone subcategory of expletives.* (Humanities)

31. *Prior research indicates that speaker's sex affects the use of many lexico-grammatical features, including a few features identified here as varying across speaker's age.* (Humanities)

Finally, several referential bundles refer to particular places or locations in the text itself.

32. *Compared to XMM a very strong improvement of all key performance parameters has been reached as is demonstrated in Fig. 7 which compares the Carbon Ka spectrum with that obtained at XMM.* (Natural sciences)

33. *The Large Area Telescope (LAT) instrument of the Gamma-ray Large-Area Space Telescope (GLAST) mission described below can provide conclusive evidence on this matter since the spectral predictions by hadronic and leptonic models diverge below 100 GeV as presented in Fig. 18.* (Natural Sciences)

34. *This corpus was compiled in 1996 and 1997 and the section of the corpus used in this study was collected at 16 sites in the United States, and has speakers from 30 states.* (Humanities)

35. *We had intended to undertake an empirical study, but in the event this was not possible and in its place we held an invited seminar in Hong Kong, described in Davies et al. (2003), with representatives from Singapore, China, India, and Malaysia.* (Humanities)

Many of these bundles are multifunctional, that is, they are used for multiple purposes, and may refer to a place, time, and/or text deixis, depending on the particular context. For instance, the bundle *at the end of* (example 38) is used for time reference and reference to physical world.

36. *Nevertheless, the standard mode of operation with the acquisition of a set of pedestal values for all the 105 k channels at the beginning or at the end of a data taking run is still possible.* (Natural sciences)

37. *The detectors at the bottom of a tray combine with those on the top of the tray below to form a 90 stereo x, y pair with a 2 mm gap between them, and with the tungsten converter foils located just above.* (Natural Sciences)

38. *Fig. 7 shows the GEM signals and their pulse height distribution at the end of the 5 week observation period as recorded directly on the digital oscilloscope.* (Natural Sciences)

39. *Extract 1 (at the beginning of this section) is an example of such similarities in all of the stance categories. (Humanities)*

40. *Let us now look at some examples of small talk from the corpus (transcription conventions are given at the end of the article). (Humanities)*

Table 6. Distribution of stance bundles in each domain

1. Stance bundles:	No. of lexical bundles in Humanities	No. of lexical bundles in Natural sciences
Epistemic stance bundles	24	none
Attitudinal/modality stance bundles:		
a) obligation/directive bundles	7	13
b) ability bundles	1	6

Table 6 shows the distribution of the two singled out kinds of stance bundles. There are 24 epistemic stance bundles found in humanities, and there were no such stance bundles found in natural sciences. Attitudinal/modality stance bundles are present in both domains. Obligation/directive bundles amount to 7 bundles in humanities and 13 bundles in natural sciences, while the sub-category of ability bundles is represented by 1 bundle in humanities and by 6 bundles in natural sciences. This might imply that the need to establish a contact, to share and involve the reader in communication is slightly stronger in the research articles in natural sciences than in humanities.

Table 7. Distribution of text organizing bundles in each domain

2. Text organizing bundles	Number of lexical bundles in humanities	Number of lexical bundles in natural sciences
a) topic introduction/focus bundles	none	9
b) topic elaboration/clarification bundles	20	94

Table 7 illustrates the distribution of lexical bundles within the groups of text organizing bundles. Lexical bundles in the first sub-category (topic introduction/focus) are present only in natural sciences, which amounts to 9 bundles. Topic elaboration/clarification bundles are far more frequent in natural sciences (94) than in humanities (20). Text organizing bundles in natural

sciences are more numerous, which might suggest that the way in which lexical bundles connect the structural and rhetorical parts of discourse is more explicit in this domain.

Table 8. Distribution of referential bundles in each domain

3. Referential bundles	Number of lexical bundles in Humanities	Number of lexical bundles in Natural sciences
a) identification/focus bundles	905	24
b) bundles specifying attributes	267	739
c) time/place/text reference bundles	26	29

Table 8 presents the distribution of lexical bundles across the singled out sub-categories of referential bundles. Referential bundles of all sub-categories were found in both domains. As can be seen from this table, identification bundles in humanities account for a comparatively great number, that is, 905 bundles and only 24 bundles were found in natural sciences. The quantities of bundles specifying attributes also differ dramatically: 267 bundles in humanities and 739 bundles in natural sciences. There are no significant differences between the two domains when it comes to the use of time/place/text reference bundles: 26 were found in humanities and 29 in natural sciences. These findings might suggest that the research articles in both humanities and natural sciences contain high-level transactional information, such as exemplification, relationship between topics, evaluations, qualifications and asides.

Figure 8 below sums up the results of the functional categorization of lexical bundles in the two domains. As can be seen, stance bundles account for a comparatively small number: only 3% in humanities and 2% in natural sciences. Text organizers in humanities amount to 2% and natural sciences contain a greater number of text organizing bundles, they make up 10% of the bundles, whereas referential bundles account for the greater part of lexical bundles in both domains: 95% in humanities and 88% in natural sciences.

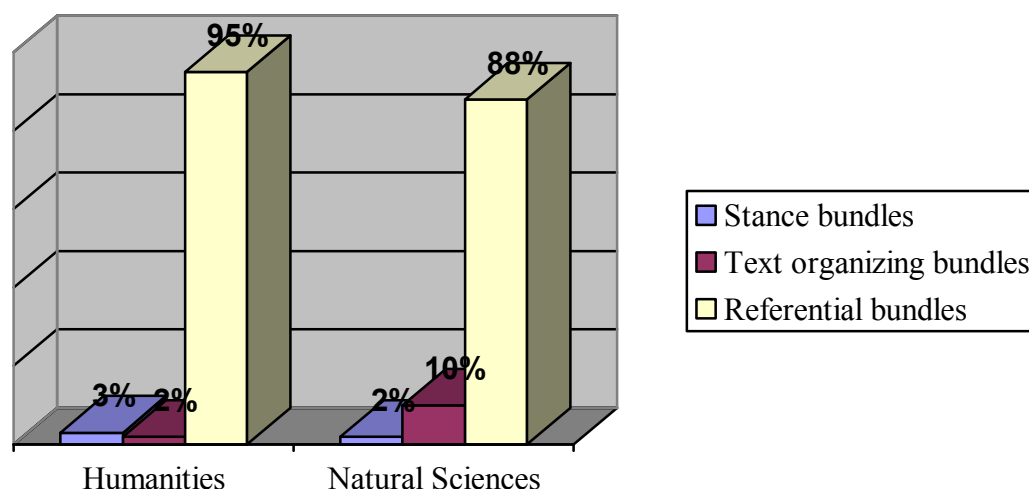


Figure 8. Distribution of lexical bundles across functional categories in each domain

According to Nattinger and DeCarrico (1992:75), to be aware of the text organizing bundles means to know organizational structure of the text and to comprehend it. The fact that text organizing bundles are more numerous in the articles in natural sciences, points to a greater need for the precision of text structuring in this domain. Also, referential bundles are more widely used in humanities than in natural sciences, thus research articles in humanities express a greater demand for the variety of relating facts and different ways for conveying all sorts of information. The findings of the study confirmed the previous findings of Biber, Conrad and Cortes (2005) which proved that recurrent word combinations play an important role in the construction of written academic discourse.

3.4 The relationship between structural and functional categories

Figure 8 below represents the distribution of structural types of lexical bundles across functional categories. As has already been mentioned, the overall number of lexical bundles in the two domains is 2179. This chart indicates that there is an overt relationship between structural and functional categories. Thus, stance bundles are composed entirely of verbal phrase fragments (57), while text organizing bundles are composed of noun phrase (118) or verb phrase (5) fragments. Referential bundles are the only functional category which is realized in all four structural types. The most common bundles are those incorporating noun phrase (1342); bundles

incorporating preposition phrase stand for 617 bundles. The least numerous structural types are bundles incorporating verb phrase (25) and bundles incorporating clause fragment (20).

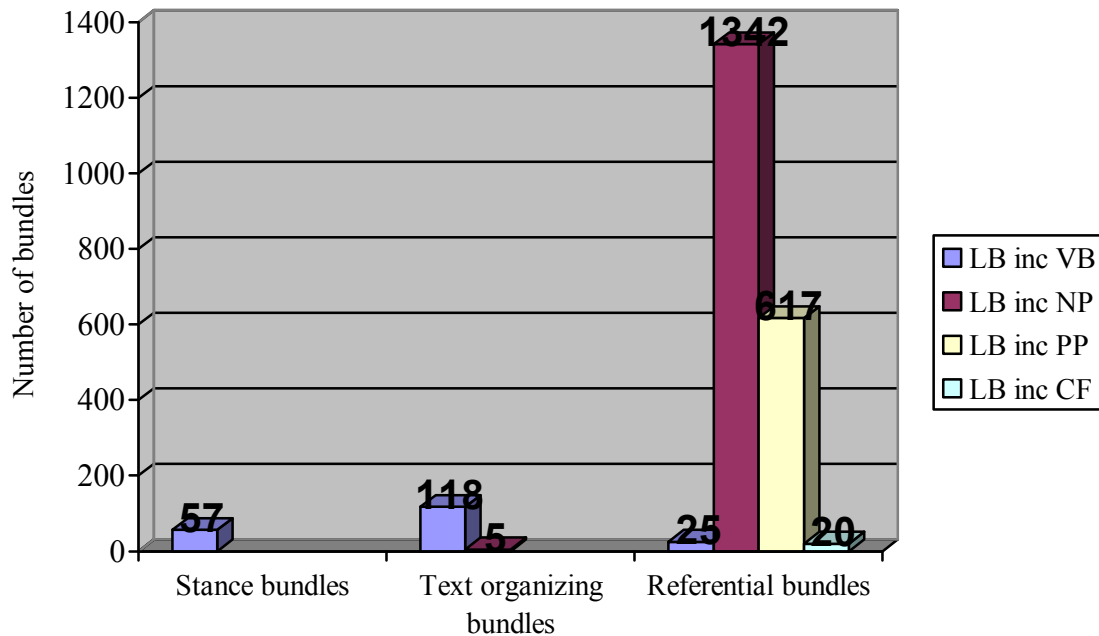


Figure 8. Interaction of structural and functional categories

The distribution suggests that there is a strong relationship between form and function of lexical bundles. For instance, combining nouns in noun phrases or prepositions with nouns in prepositional phrases were the most common devices employed in the construction of referential bundles, so it can be inferred that the most frequent recurrent word combinations from these structural categories would become fixed as referential lexical bundles. Moreover, bundles incorporating verb phrase are used for stance and text organizing functions, and can be interpreted as one of the most productive ways used to express stance and text organization in written academic discourse. Thus, it can be concluded that there is a close interrelation between structural categories and discourse functions.

4. Conclusions

The main objective of this study was to identify lexical bundles used in research articles in humanities, represented by linguistics and educology, and natural sciences, represented by physics and astronomy, and to compare the use of those bundles in terms of frequency of occurrence, distribution of different structural and functional types across the subject areas.

The analysis of the frequency of occurrence of lexical bundles indicates that lexical bundles in the research articles in humanities are used more frequently. The language of the research articles in humanities displays a greater variety of structural types of lexical bundles. This fact shows that the language of the research articles in humanities is more varied than the language of the research articles in natural sciences. As for the functional analysis, the findings revealed that the language of the research articles in natural sciences shows greater precision in text structuring as the articles in natural sciences contain a larger stock of text organizing bundles than the articles in humanities. Also, it shows that the need to establish a contact and to involve the reader in communication is slightly stronger in the research articles in natural sciences than in humanities. The analysis suggests that referential bundles in the research articles in both humanities and natural sciences contain high-level transactional information, such as exemplification, relationship between topics, evaluations, qualifications and asides. Since referential bundles are more frequent in humanities than in natural sciences, it could be said that the language of the research articles in humanities has a greater variety of ways to relate facts and to convey all kinds of information.

The relationship between structural and functional categories shows that the most common multi-word units within a structural category may become fixed and indicate a certain function. For instance, lexical bundles incorporating verb phrase usually perform stance or discourse organizing functions. Moreover, lexical bundles incorporating noun or prepositional phrases are strongly bound to the referential function. To know lexical bundles means to understand their communicative value in discourse and to improve the productive process of writing and the receptive process of reading.

The findings of this research contribute to the improvement of the ability to understand the language of written academic discourse. Lexical bundles serve as the basic building blocks of the

discourse and help to structure the lexicon in the text. What is more, the knowledge of the lexical items of the genre might help to understand and create text with greater ease.

However, this study has certain limitations. Disciplinary domains could have included a greater number of disciplines, thus making the results more trustworthy; the corpora could have been more extensive too.

Certain directions for further research can be suggested. Thus, it would be interesting to compare lexical bundles in research articles in humanities and natural sciences written by native and foreign speakers. Also, it would be useful to carry out a further research on lexical bundles in a variety of disciplines other than linguistics/educology and physics/astronomy in order to identify word combinations that may be discipline-bound.

5. Summary in Lithuanian

Rašytinis akademinis diskursas: leksiniai junginiai humanitariniuose ir gamtos moksluose

Šis darbas – tai bandymas tirti leksinius junginius (angl. lexical bundles), kurie yra apibrėžiami kaip pasikartojantys žodžių junginiai, nepriklausomai nuo jų struktūros ir reikšmės, angliškuose moksliniuose straipsniuose. Tyrimui buvo pasirinkta 20 humanitarinių straipsnių iš lingvistikos ir edukologijos bei 20 gamtos mokslų straipsnių iš fizikos ir astronomijos. Buvo analizuojami 3, 4, 5 žodžių leksiniai junginiai, kurie iš straipsnių buvo išrinkti kompiuterinės programos „WordSmith Tools“ pagalba. Jie buvo suklasifikuoti, pasitelkus Biber, Conrad ir Cortes (2005) struktūros ir funkcijos taksonomiją rašytiniams tekstams.

Keliamas tikslas – palyginti ar skiriasi leksinių junginių vartojimas humanitarinių ir gamtos mokslų straipsnių kalboje pasikartojimo, struktūros ir funkcijos atžvilgiu. Ši hipotezė buvo įrodyta. Ištirta, kad humanitariniuose moksliniuose straipsniuose vartojama daugiau leksinių junginių nei gamtos mokslų straipsniuose. Rasti ir struktūros skirtumai, kurie rodo, kad humanitarinių straipsnių kalba yra įvairesnė, o gamtos mokslų straipsnių kalba yra griežtesnė ir tikslesnė. Funkcijos atžvilgiu, gamtos mokslų straipsniai taip pat pasižymi tikslesne ir griežtesne kalba. Ryšys tarp leksinių junginių struktūros ir funkcijos rodo, kad dažniausiai pasikartojantys žodžių junginiai struktūrinėje grupėje gali būti automatiškai susieti su tam tikra funkcija.

Leksiniai junginiai buvo gana mažai nagrinėti moksliniuose straipsniuose, parašytuose žmonių, kuriems teksto kalba yra gimtoji ir tiems, kuriems teksto kalba nėra gimtoji. Taip pat būtų įdomu palyginti leksinius junginius, vartojamus kituose nei lingvistika/edukologija ir fizika/astronomija straipsniuose.

6. References

Primary sources

- Allen, J.E. and A. Self. 2008. Analysis of the integration of knowledge and novelty in creative engineering design, *Journal of Aerospace Engineering*, 222/1:127-167.
- Amin, I.J. and A.J. Taylor. 2007. Driver tracking and posture detection using low-resolution infrared sensing, *Journal of Automobile Engineering*, 9:1079-1088.
- Barbieri, F. 2008. Patterns of age-based linguistic variation in American English. *Journal of Sociolinguistics*, 12/1:58-88.
- Bellazzini, R. 2007. A sealed Gas Pixel Detector for X-ray astronomy, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 579/2:853-858.
- Braha, D.D. and Y. Bar-Yam. 2007. The Statistical Mechanics of Complex Product Development: Empirical and Analytical Results, *Management Science*, 53:1127-1145.
- Chemela, J. M. 2004. The politics of language acquisition: Amazon. *Women and Language*, 27/10-26
- Cheshire, J. 2005. Syntactic variation and beyond: gender and social class variation in the use of discourse-new markers. *Journal of Sociolinguistics*, 9/4:479-508.
- Dixon, R. M. W. 2008. Deriving verbs in English. *Language Sciences*, 30:31-52.
- Francis, N. (2000). An examination of written expression in bilingual language: assessment of sense story structure and interlinguistic transfer. *International Journal of Applied Linguistics*, 10:187-220.
- Giovannini, M. 2007. Semi-analytical approach to magnetized temperature autocorrelations, *PMC Physics A*, 1/5:1-53.
- Gungor, R.M. et al. 2008. Analysis of electromigration and stress induced dynamical response of voids confined in metallic thin films, *Surface Science*, 602/6:1227-1242.
- Hamp-Lyons, L. and Davies, A. 2008. The Englishes of English tests: bias revisited. *World Englishes*, 27/1:26-39.
- Holmes, J. 2000. Talking English from 9 to 5: challenges for ESL learners at work. *International Journal of Applied Linguistics*, 10/1:125-140.

- Kapses, A. 2008. KM3NeT, a new generation neutrino telescope, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 588/2:86-91.
- Keen, J. 2004. Sentence-combining and redrafting processes in the writing of secondary school students in the UK. *Linguistics and Education*, 15:81-97.
- Leung, C. 2005. Language and content in bilingual education. *Linguistics and Education*, 16:238-252.
- Lindemann, S. 2005. Who speaks "broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics*, 15/2:187-212.
- Loebell, H. And Bock, K. 2003. Structural priming across languages. *Linguistics*, 41/5:791-824.
- Lunberg, J. et al. 2007. Light tracking through ice and water—Scattering and absorption in heterogeneous media with Photonics, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581/3:619-631.
- Lutz, G. et al. 2007. Spectroscopic silicon imaging detectors: Past achievements and new developments, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 580/2:960-967.
- Margiotta, A. 2008. Data from the ANTARES Neutrino Telescope, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 588/1:80-85.
- Menary, R. 2007. Writing as thinking. *Language Sciences*, 29:261-632.
- Mitchell, R.N et al. 2008. Oceanic anoxic cycles? Orbital prelude to the Bonarelli Level (OAE 2), *Earth and Planetary Science Letters*, 2:1-16.
- Oh, W. and S. Jeon. 2007. Membership Herding and Network Stability in the Open Source Community: The Ising Perspective, *Management Science*, 53:1086-1101.
- Precht, K. 2008. Sex similarities and differences in stance in informal American conversation. *Journal of Sociolinguistics*, 12/1:89-111.

- Randolph, S.J. et al. 2006. Focused, Nanoscale Electron-Beam-Induced Deposition and Etching, *Critical Reviews in Solid State and Materials Sciences*, 31/55:89.
- Rezakiany, G. 1998. English proficiency and academic achievement in relation to extraversion: a preliminary study. *International Journal of Applied Linguistics*, 8/1:113-130.
- Ruiying, Y. and D. Allison. 2003. Research articles in applied linguistics: moving from results to conclusions. *English for Specific purposes*, 22: 365-385.
- Siegel, J. 2006. Language ideologies and the education of speakers of marginalized language varieties: adopting a critical awareness approach. *Linguistics and Education*, 17:157-174.
- Singamneni, S.B. et al. 2007. Evaluation of the thermal behaviour of bevelled cutting inserts using a numerical approach, *Journal of Engineering Manufacture*, 9:1607-1616.
- Singh, G. And Richards, J. C. 2006. Teaching and learning in the language teacher education course room: a critical sociocultural perspective. *Regional Language Center Journal*, 37/2:149-175.
- Sio, D. et al. 2007. Towards UV imaging sensors based on single-crystal diamond chips for spectroscopic applications, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 583/1:125-130.
- Sonnabend, G. et al. 2008. Ultra high spectral resolution observations of planetary atmospheres using the Cologne tuneable heterodyne infrared spectrometer, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 109/6:1016-1029.
- Tagliamonte, S. And D'Arcy. He's like, she's like: the quotative system in Canadian youth. *Journal of Sociolinguistics*, 8/4:493-514.
- Tajima, H. 2007. Detectors for cosmic rays on ground and in Space, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581/1:12-24.
- de Vries, M. 2008. The representation of language within language: a syntactico-pragmatic typology of direct speech. *Studia Linguistica*, 62 (1):39-77.
- Vold, E. T. 2006. Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics*, 16/1:61- 87.

- Weber, T.P. 2007. Carl du Prel (1839–1899): explorer of dreams, the soul, and the cosmos, *Studies in History and Philosophy of Science Part A*, 38/3:593-604.
- Wolfgang, G. 2007. Ricci Flow Gravity, *PMC Physics A*, 1/3:1-13.
- Zongxian, Y. et al. 2008. Facilitated vacancy formation at Zr-doped ceria (1 1 1) surfaces, *Surface Science*, 602/6:1199-1206.

Secondary sources

- Biber, D. et. al. 1999. *Longman Grammar of Spoken and Written English*. Edinburgh: Longman.
- Biber, D., Conrad, S., and Cortes, V. 2005. If you look at...: lexical bundles in University Teaching and Textbooks. *Applied Linguistics* 25/3:371 - 405.
- Bhatia, V.K. 1993. *Analysing Genre: language use in professional settings*. London: Longman.
- Cortes, V. 2002. Lexical Bundles in Freshman Composition. . In R. Reppen, S. Fitzmaurice, and D. Biber (eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins.
- Drazdauskiene, M. L. 1981. On stereotypes in conversation, their meaning and significance. In: Coulmas, F. (ed.) *Conversational Routine*. The Hague-Paris: Mouton.
- Firth, J.R. 1957. *Papers on Linguistics*. Oxford University Press.
- Halliday, M.A.K. et al. 1964. *The Linguistic Sciences and Language Teaching*. London. Longman.
- Halliday, M. A. K. 1988. On the language of physical science. In Ghades, M. *Registers of written English: situational factors and linguistic features* (pp. 161-178). London: Frances Pinter.
- Hyland, K. 1998. *Hedging in Scientific Research Articles*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Nattinger, R.J., and S.J. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford University Press.
- Pawley, S., and Syder, F. 1983. *Two Puzzles for Linguistic theory: nativelike selection and nativelike fluency*. Longman.
- Yorio, C. et al. 1979. *On TESOL '79*. Washington D.C.: TESOL.