ISSN 1392–1517. Online ISSN 2029–8315. KALBOTYRA. 2017 • 70
DOI: https://doi.org/10.15388/Klbt.2017.11181

A corpus-driven analysis of structural types of lexical bundles in court judgments in English and their translation into Lithuanian

Donata Berūkštienė
Department of Foreign Language, Literature and Translation Studies
Faculty of Humanities
Vytautas Magnus University
V. Putvinskio g. 23-320, LT-44243 Kaunas, Lithuania
E-mail: donaber@gmail.com

Abstract

Formulaicity is one of the characteristic features of legal discourse, which manifests itself not only at the level of wording, "but also in the content, structure and layout" of legal texts (Ruusila & Londroos 2016, 123). Formulaic language, which includes phrasal and prepositional verbs, idioms, collocations, lexico-grammatical associations, lexical bundles, *etc.*, are building blocks of legal discourse shaping legal text meanings. However, up to now, far too little attention has been paid to the nature of frequently occurring "sequences of three or more words that show a statistical tendency to co-occur" (Biber & Conrad 1999, 183), *i.e.* lexical bundles, in different genres of legal texts. Most studies in the field of lexical bundles in legal texts have only been based on one language (*e.g.* Jablonkai 2009; Goźdź-Roszkowski 2011; Breeze 2013), whereas translation-oriented contrastive studies on lexical bundles are lacking. In respect of the aforementioned gaps, the aim of this pilot study is to analyse structural types of lexical bundles in court judgments of the Court of Justice of the European Union in English and to examine the way these structures are rendered into Lithuanian. To gain insights into the frequency and structure of lexical bundles, the present study uses the methodological guidelines of corpus linguistics. The classification of lexical bundles into structural types is based on the framework suggested by Biber *et al.* (1999, 2004). For the purpose of this study, a parallel corpus of court judgments was compiled comprising approximately 1 million words of original court judgments in the English language and about 8 hundred thousand words of court judgments translated into Lithuanian. Lexical bundles in this research were identified using the corpus analysis toolkit AntConc 3.4.4 (Anthony 2015). A concordance program AntPConc 1.2.0 (Anthony 2017) was employed to find Lithuanian equivalents of the most frequent lexical bundles identified in the English court judgments. The evidence from this study suggests that different structural types of

lexical bundles have more or less regular equivalents in Lithuanian; however, in most cases, these equivalents tend to be shorter.

Keywords: formulaic language, lexical bundles, court judgments, structural types, corpus-driven analysis, translation

1 Introduction

In recent years, there has been an increasing interest in formulaicity in different discourses. Formulaic language, which includes phrasal and prepositional verbs, idioms, collocations, lexico-grammatical associations, lexical bundles, *etc.*, has been an object of analysis in a great variety of linguistic disciplines starting from lexicology, lexicography, discourse analysis, corpus-based and corpus-driven analyses and ending with the first and second language acquisition, foreign language teaching, psycholinguistics, semantics, and stylistics. Such interest in formulaic nature of language was stimulated after it had been established that a great amount of language is not constructed according to word-by-word principle but occurs in sequences of words or strings of linguistic items that are more or less fixed in form (Hunston & Francis 2000, 7). It should be also pointed out that it was the development of computer technology and the introduction of corpus linguistics that made the investigation into longer sequences of words possible.

One of the main categories of formulaic language are lexical bundles, which recently have been studied by many researchers, who have noticed that this type of multi-word expressions constitutes a significant part of spoken as well as written language (Biber & Conrad 1999; Cortes 2004; Biber & Barbieri 2007). The term *lexical bundles* was introduced by Biber *et al.* in the *Longman Grammar of Spoken and Written English* to refer to recurrent expressions or "sequences of word forms that commonly go together in natural discourse" and that "show statistical tendency to co-occur" (*e.g. in the case of the*, *it should be noted that*, *do you want me to* (1999, 989–990). According to the linguists, lexical bundles can be made of three or more words. Chen and Baker, in their turn, expanded the primary definition by adding that lexical bundles "refer to continuous word sequences retrieved by taking a corpus-driven approach with specified frequency and distribution criteria" (2010, 30).

The definitions of lexical bundles above suggest that lexical bundles are identified solely on frequency criterion. To be qualified as a lexical bundle, a sequence of words must meet the criteria of orthographic length, minimum frequency and dispersion range (Biber *et al.* 1999). The first criterion of orthographic length concerns the number of words that constitute a lexical bundle. Depending on the purpose of the study, different authors choose to analyse lexical bundles containing from three to eight words. It has

been observed that shorter lexical bundles are much more common but they are "often incorporated into more than one longer lexical bundle" (Biber *et al.* 1999, 990). The second criterion, namely, minimum frequency refers to the minimum number of times a word sequence must appear in a corpus in order to be counted as a lexical bundle. Previous research on lexical bundles has applied the normalized frequency threshold between 10-40 occurrences per million words (Biber *et al.* 1999; Biber *et al.* 2004; Conrad & Biber 2004; Hyland 2008b; Jablonkai 2009). The higher the frequency cut-off point is set, the shorter the list of lexical bundles is generated. Finally, the third criterion, *i.e.* dispersion range, should be applied in order to avoid individual speaker or writer idiosyncrasies (Biber *et al*. 1999, 992–993). In the vast majority of studies on lexical bundles (*e.g.* Conrad & Biber 2004; Goźdź-Roszkowski 2011; Kopaczyk 2013; Güngör & Uysal 2016; Pan, Reppen & Biber 2016), the dispersion range is set at five, which means that a particular lexical bundle must be used in five or more different texts. However, other studies set the dispersion range at 10% of all texts (*e.g.* Hyland 2008a, 2008b) or even more (*e.g.* Mackiewicz 2017).

It should be mentioned that despite the fact that most lexical bundles represent incomplete structural units, they have been classified according to their structural correlates and their predominant pragmatic functions. Thus, the classification of lexical bundles follows two dimensions: structural and functional. The structural classification is based on the grammatical structure of lexical bundles. Depending on whether lexical bundles incorporate nouns, prepositional phrases, verbs, or clause fragments, they are divided into (1) lexical bundles that incorporate noun phrase fragments (further also referred to as 'nominal lexical bundles') or prepositional phrase fragments (further also referred to as 'prepositional lexical bundles'), *e.g.*, *the end of the*, *those of you who*, *at the end of*; (2) lexical bundles that incorporate verb phrase fragments (further also referred to as 'verbal lexical bundles'), *e.g.*, *I am not going to*, *take a look at*, *are you going to*; (3) lexical bundles that incorporate dependent clause fragments (further also referred to as 'clausal lexical bundles'), *e.g.*, *I want you to*, *if we look at*, *what I want to*; (3) (Biber *et al.* 2004, 382). In addition, each main structural type of lexical bundles may be further divided into smaller subtypes. From the functional perspective, lexical bundles have been classified into stance bundles, which express attitudes, discourse organizers, which show relationships between different parts of a text, and referential bundles, which indicate physical or abstract entities or refer to textual context (Biber *et al.* 2004, 384). Each of the mentioned functional categories include subcategories.

The vast majority of studies on lexical bundles have focused on the English language, except for several works which have attempted to apply lexical bundles approach to the analysis of different texts in other languages, such as Spanish (Tracy-Ventura *et al.* 2007), Korean (Kim 2009), Polish (Grabowski 2014), and Brazilian Portuguese (Sardinha

et al. 2014). Research on lexical bundles in the English language has been carried out in two directions: the analysis of spoken and written language. Different discourses, from academic discourse to legal discourse, as well as different genres of texts have been the main sources for investigations of frequency, structural and functional features of lexical bundles. Different genres of spoken and written academic discourse have been investigated by Biber *et al.* (1999), Cortes (2004), Biber *et al.* (2004), Simpson (2004), Biber (2006), Hyland (2008b), Simpson-Vlach and Ellis (2010), Salazar (2010), Jalali and Moini (2014), Pan *et al.* (2016), and Yang (2017). Besides different studies focusing on the use of lexical bundles across different disciplines, many of the previous works on lexical bundles in academic discourse have focused on differences and similarities of the use of lexical bundles between L1 and L2 writers (Salazar 1996; De Cock 1998; Schmitt 2005; Salazar 2011; Juknevičienė 2011; Ädel & Erman 2012; Amirian *et al.* 2013; Purificación 2013; Pan *et al.* 2016; Güngör & Uysal 2016; Güngör 2016) as well as between novice and expert writers (Cortes 2004; Chen & Baker 2010).

Besides academic discourse, the lexical bundles approach has been also used to study political (Partington & Morley 2004), religious (Shrefler 2011), literary (Stubbs & Barth 2003), medical (Kopaczyk 2013; Grabowski 2013), and legal (Goźdź-Roszkowski 2004, 2006a, 2006b, 2011; Jablonkai 2009, 2010; Breeze 2013; Kopaczyk 2012, 2013; Tománková 2016) discourse. However, a much greater deal of previous research has been concerned with lexical bundles in academic discourse than in any other discourse. Besides, many of these studies focus on pedagogical perspective of lexical bundles. So far, very little attention has been paid to the role of lexical bundles in translation. Drawing on the idea that combining the fields of phraseology and corpus-based contrastive linguistics is "entering relatively unexplored territory" (Ebeling & Oksefjell Ebeling 2013, 1) and in order to fill the above-mentioned gaps, this pilot study aims to analyse structural types of 4-word lexical bundles prevailing in court judgments of the Court of Justice of the European Union in English and to examine the way they are rendered in the translation of the original court judgments into Lithuanian. For the purpose of this study, the following research questions were formulated:

1. What is the distribution of lexical bundles in English court judgments?
2. What structural types of 4-word lexical bundles prevail in English court judgments?
3. What structures are used to render the prevailing structural types of 4-word lexical bundles in English court judgments into Lithuanian?

2 Lexical bundles in legal discourse and translation

According to Goźdź-Roszkowski and Pontrandolfo, corpus research into legal phraseology has focused on five major areas: (1) research into lexico-syntactic combinations, focusing on specialised collocations; (2) research into routine formulae;

(3) lexicographic studies with a focus on terminology and terminography; (4) contrastive studies of phraseology, including translation; and (5) semantics of legal patterns (2015, 133–134). Investigations of uninterrupted recurrent word sequences, *i.e.* lexical bundles, belong to the second area. It should be noted that, despite the fact that "law itself is an inherently formulaic discipline and formulaicity an integral quality of law" (Ruusila & Lindroos 2016, 126), phraseology in general and lexical bundles in particular have received relatively scant attention in legal language studies for some time. However, recently, there has been a renewed interest in formulaicity in legal discourse. The publication of the special issue of *Fachsprachen*, namely, *Legal Phraseology Today. A corpus-Based view* (2015) and the collection of articles in the book *Phraseology in Legal and Institutional Settings. A Corpus-Based Interdisciplinary Perspective* (2017) have presented the latest developments in the study of formulaicity in legal and institutional discourse focusing on the main areas of research introduced above, including the research into lexical bundles.

So far, there have been several noteworthy studies contributing to the investigation of lexical bundles in legal discourse, namely, works by Goźdź-Roszkowski (2006b, 2011), Jablonkai (2010), Kopaczyk (2013) and Breeze (2013). The studies approach lexical bundles from different perspectives: synchronic (Goźdź-Roszkowski 2006b), variation in legal discourse (Goźdź-Roszkowski 2011; Breeze 2013) and standardisation of early legal discourse (Kopaczyk 2013) (Goźdź-Roszkowski & Pontrandolfo 2015, 133–134). What is more, the investigations focus on different legal genres belonging to four different legal systems, such as in the EU, the US, England, and Scotland. For instance, Goźdź-Roszkowski bases his study on American legal English and analyses such genres of legal texts as legislation, briefs, contracts, court opinions, legal academic journals, legal professional articles, and textbooks. The study suggests that it is legislation and contracts that demonstrate the highest degree of formulaicity. Breeze also focuses on lexical bundles in different English legal genres, including academic law articles, law reports, court opinions, legislation (Companies Acts), and legal documents (*e.g.* contracts, merger agreements and others). Breeze's analysis shows that legislation and documents might be regarded as the most formulaic genres of legal texts. Variation of lexical bundles in a mix of EU genres of legal texts compiled into one corpus against the British National Corpus is analysed in Jablonkai' s work. This study demonstrates that there are many more lexical bundles in EU legal discourse than in different sections of the BNC. Unlike the overviewed studies, the work by Kopaczyk takes a different approach to lexical bundles and analyses them from a historical perspective. The results of her study reveal the highly formulaic nature of Scots legal texts (burgh records, notarial protocols, statutes and other official and administrative texts), which manifests itself through long 7-word and 8-word lexical bundles. In brief, all the above reviewed studies are based on a cross-generic investigation of lexical bundles.

In addition to cross-generic investigations into lexical bundles in legal discourse, one would also expect works analysing lexical bundles from a cross-linguistic, translational perspective. However, there are scarcely any studies approaching the translation of different genres of legal texts from the perspective of lexical bundles. There exist probably one noteworthy work related to the topic of lexical bundles, translation and multilingualism, *i.e.* the study by Biel *The impact of translation process on the patterning of legal language* (2017). In her article, Biel focuses on 4-word lexical bundles in English and Polish. However, instead of applying cross-generic or cross-linguistic perspective towards lexical bundles, the scholar investigates lexical bundles from a new angle, *i.e.* she focuses her analysis on "internal variation of a single genre of legislation – translator-mediate multilingual legislation and domestic legislation" (*i.e.* Polish legislation (Biel 2017, 13). For this purpose, Biel uses Polish Eurolect corpus, which is compared to English Eurolect corpus and the Polish Domestic Law corpus. Biel holds the position that "if we are interested in cross-linguistic comparisons, which lie at the heart of translation, a question should be asked to what extent a 4-gram reflects the same level of formulaicity across languages; in other words, how we can compare 4-grams across languages" (Biel 2017, 16). The scholar presupposes that, as Polish is an inflectional language coding grammatical information morphologically and not marking (in)definiteness through articles, 4-word English lexical bundles may correspond to shorter lexical bundles in Polish. In her opinion, "we are at too early a stage to solve the problem of cross-linguistic comparisons of bundles and more field work is required in this area" (*ibid.*). Biel's study refutes the hypothesis that translations tend to be "less patterned and less formulaic than nontranslations" (Biel 2017, 23). What is more, it is demonstrated that translations share very few bundles with non-translations, which suggests that translations "resort to their own n-grams prompted by source texts rather than prime bundles which are natural and expected in target-language legal texts" (*ibid.*).

Despite the fact, that there remains a paucity of evidence on the role of lexical bundles in legal translation, several cross-linguistic studies have been published on other categories of formulaic language in legal discourse. These include the article by Biel (2015), who compares translated and non-translated English and Polish legal language focusing on complex prepositions; a translation-oriented lexicographical study conducted by Castro and Faber (2015), whose work analyses the phraseological units with the head word *trial* in four legal English-Spanish histories, and the investigation of different phraseological units (*e.g.* complex nominal phrases, extended lexical collocations, *etc.*) in the language of contracts in English and Croatian by Basaneže (2015). The most recent research on formulaic language in legal discourse includes the work by Nebot (2017), who focuses on the structure of binomials and multinomials in the International Bill of Human Rights in English, French, and Spanish; and the comparative analysis of noun binomials in UK and Scottish legislation by Kopaczyk (2017). In addition, the study by Salkie (2017) focuses on a contrastive analysis of the English phrase *the fact that* and its German

counterparts; whereas Goźdź-Roszkowski (2017) investigates the same phrase *the fact that* in American and Polish judicial discourse revealing a general similarity between the usage of the mentioned phrase in both languages. All these works give valuable insights not only into different legal languages but also into different legal cultures.

In the light of the overview of cross-linguistic and translational research on lexical bundles, it should be mentioned that, in general, most of works on the translation of legal discourse have focused on terminology. For this reason, there exists a number of bilingual lexical resources that are mainly restricted to single words and compounds, whereas there is a general lack of bilingual terminological resources including multi-word units, which play a significant role in the process of translation. As Parra and ten Hacken state (2008, 1), "[w]hereas multi-word units are linguistically heterogeneous, in translation they raise a very similar set of problems. In order to translate them, they first have to be recognised as belonging together". Thus, the translation of multi-word units poses challenges not only to human translators but also to Machine Translation, as explained by Monti *et al.* (2013, 8):

> *In spite of the recent positive developments in translation technologies, multi-word units still present unexpected obstacles to Machine Translation and translation technologies in general, because of intrinsic ambiguities, structural and lexical asymmetries between languages, and cultural differences. Multi-word unit identification and translation problems are far from being solved and there is still considerable room for improvement.*

Referring to the above-mentioned problems, Grabar and Lefer (2015) offer a lexical bundle approach to the identification of multi-word units by applying this method for the extraction of lexical bundles from English and French comparable and parallel corpora. They focus on 3-word lexical bundles and longer lexical bundles containing them. The methodology offered by Grabar and Lefer involve extraction of lexical bundles, partial lemmatization of French lexical bundles, manual selection of structurally complete bundles, automatic extraction of target language equivalents, and manual validation of target language equivalents. It should be noted that this work approaches lexical bundles in two languages from the perspective of functional taxonomy of lexical bundles suggested by Biber *et al*. (2004). Altogether, 400 identified English discourse organizers and stance expressions were analysed in respect to their French equivalents. By limiting the scope of the analysis to French equivalents with minimum frequency of two and discarding hapaxes, the final list of over 4000 of French equivalents was generated. Only 32 out of 400 English lexical bundles were found to have no equivalents in the French language. This study highlights that it is polyfunctional bundles and categorial changes when translating them from one language into another that pose challenges for human translators, Computer Assisted Translation and Machine Translation. Although

this research focuses more on political than on legal discourse, its insights can be successfully applied to the analysis of lexical bundles in legal language from a cross-linguistic perspective.

3 Data and methods

In the course of the investigation of structural properties of lexical bundles in English and Lithuanian, methodological guidelines of corpus linguistics have been followed. For the purpose of this study, a parallel corpus of judgments delivered by the Court of Justice of the European Union has been compiled. The corpus contains 104 different court judgments in English delivered between June 2015–November 2016 together with their translations into Lithuanian. The size of the corpus is 1 million words of the original texts and 730,000 words of the translated texts. Obviously, the translations of the court judgments are shorter than their English counterparts.

A list of English lexical bundles was extracted with the software program AntConc, version 3.4.4 (Anthony 2015), using the N-Grams tool. In this study, the frequency cut-off point was set at 50 times per million words to make the quantity sample of lexical bundles manageable for the analysis. The distribution threshold was set as at 10, meaning that, in order to be considered as a lexical bundle, the multi-word sequence had to appear in at least 10 different English court judgments. To follow the prevailing choice in other studies, only 4-word lexical bundles were considered.

The extracted list of English lexical bundles was further refined manually. During the process of the manual identification and selection of lexical bundles to be included in the final list, all lexical bundles that contained titles of cases (*e.g. and others v council*), numbers (*e.g. regulation no. 207 2009*), meaningless fragments (*e.g. p eer eu e*), or overlaps (*e.g. the meaning of article* and *within the meaning of* is a part of 5-word bundle *within the meaning of article*) were excluded from the count. In total, 164 cases of 4-word lexical bundles incorporated in 5-word lexical bundles were deleted. As a result, the total number of lexical bundles after the refinement of the initial list decreased from 590 to 245 lexical bundles. All 245 lexical bundles were further grouped into structural types according to their grammatical structure. The structural analysis of English lexical bundles drew on the previous work by Biber *at al.* (1999) and Biber *et al.* (2004).

In order to analyse how different structural types of English lexical bundles are rendered into Lithuanian, the software program AntPConc 1.2.0 (Anthony 2017) was used, and translational equivalents of the identified structural types of English lexical bundles were manually retrieved from the Lithuanian corpus. It should be pointed out that the investigation of translational equivalents of lexical bundles focusing on their structure was limited to 10 most frequent English lexical bundles of each structural subtype.

Besides, to make the analysis manageable, only 10 translational cases of each lexical bundle distributed evenly throughout all the texts were taken into consideration. Based on the translational equivalents in the Lithuanian court judgments, the patterns of structural correspondences were established.

4 Research findings

The extraction of lexical bundles from a one-million-word corpus of English court judgments revealed that even with a relatively high frequency cut-off point and distribution threshold set, the list of 4-word lexical bundles is rather long. In addition, many of the extracted bundles occur much more frequently than 50 times per one million words. Actually, there are lexical bundles which are very common and repeated in the corpus more than 500 times. These findings support the view that formulaicity is a characteristic feature of legal language (Kopaczyk 2013; Ruusila & Lindroos 2016; Trklja 2017). Table 1 below shows the first 20 most frequent lexical bundles in the corpus:

Rank	Lexical bundles	Frequency
1.	*of the basic regulation*	558
2.	*of the contested decision*	568
3.	*in the present case*	420
4.	*the board of appeal*	358
5.	*the fact that the*	312
6.	*in accordance with the*	312
7.	*in accordance with article*	257
8.	*on the basis of*	257
9.	*in the context of*	222
10.	*for the purpose of*	211
11.	*for the purposes of*	211
12.	*with regard to the*	200
13.	*in relation to the*	195
14.	*laid down in article*	190
15.	*in the contested decision*	178
16.	*see to that effect*	175
17.	*that the general court*	175
18.	*in respect of the*	173
19.	*referred to in article*	172
20.	*in so far as*	171

Table 1. The first 20 most frequent English lexical bundles in the corpus

The list of 20 most frequent lexical bundles in the corpus suggests that most of the bundles belong to shared or unmarked vocabulary, to use Varo and Hughes terms (2002, 16) (*e.g. in respect of the, in relation to the, in accordance with the*). The other group of lexical bundles in the list include lexical bundles related to legal domain (*e.g. of the basic regulation*, *laid down in article*, *of the contested decision*). Besides, this short list reveals the prevalent structural type of lexical bundles in court judgments. It is prepositional lexical bundles that constitute the major part of all the bundles in the list. They are used to express abstract, logical relations (*e.g. on the basis of*, *in relation to the*), or specify legal context (*e.g. of the basic regulation*, *in respect of the*).

4.1 Structural classification of English lexical bundles

After the refinement of the extracted initial list of lexical bundles, all 245 bundles were categorized structurally. It should be noted that even though the majority of the selected lexical bundles are not complete structural units, they have certain structural associations, which serve as a basis of the classification. The distribution of the four major structural categories of lexical bundles is illustrated in Figure 1:

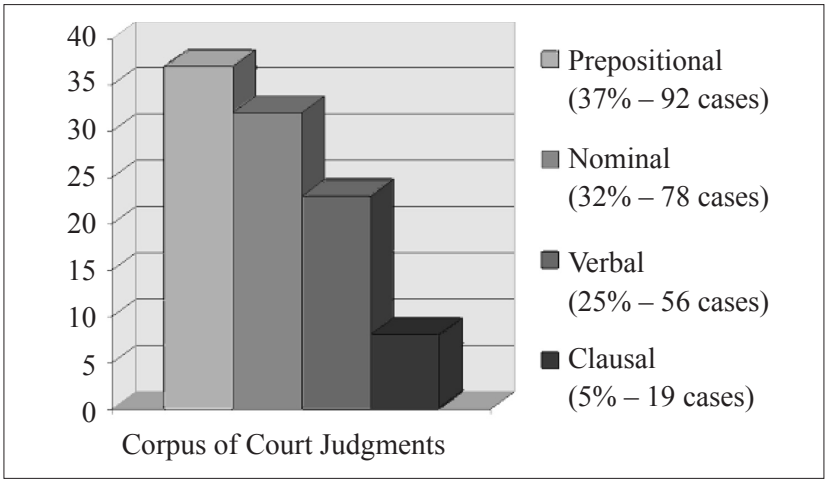

Figure 1. Distribution of structural types of lexical bundles in English court judgments

As can be seen in Figure 1 above, lexical bundles in court judgments are primarily prepositional. Lexical bundles that incorporate prepositional phrase and noun phrase fragments account for almost 70% of all the bundles, whereas verbal and clausal lexical bundles tend to be infrequent. These results are in line with those of the previous studies conducted on lexical bundles in legal discourse by Goźdź-Roszkowski (2011), Breeze (2013) and Tománková (2016). Similar tendencies were observed by Conrad and Biber (2004) in academic prose. According to Conrad and Biber, lexical bundles that are parts of noun phrases and prepositional phrases account for about 60% of all lexical bundles

in academic prose. This is in contrast to the findings in conversation, where even 90% of all lexical bundles incorporate verb phrases (Conrad & Biber 2004, 63).

Several subtypes were identified in the major four categories of lexical bundles. The structural classification and distribution of the lexical bundles across the structural subtypes are presented in Table 2. For the purpose of this study, lexical bundles that incorporate noun phrases are taken separately from the ones that incorporate prepositional phrase fragments, although they are included under one category in Biber *et al.*'s (2004) classification, since it is presumed that different structures might be used in Lithuanian to render prepositional lexical bundles and nominal lexical bundles.

Structural types	**Subtypes**	**Examples**	**Percentage – Cases**
1. Lexical bundles that incorporate prepositional phrase fragments	**1.1** *in* + noun phrase fragment	*in the present case* *in the light of* *in accordance with article*	37% – 34
	1.2 *of* + noun phrase fragment	*of the contested decision* *of the basic regulation* *of the principle of*	30% – 28
	1.3 Other prepositions (*on, for, to, at, according to, under, by, from, with*) + noun phrase fragment	*on the basis of* *for the purposes of* *with regard to the*	33% – 30
2. Lexical bundles that incorporate noun phrase fragments	**2.1** Noun phrase + *of*-phrase fragment	*the board of appeal* *the amount of the* *the scope of the* *the application of the*	72% – 56
	2.2 Noun phrase + other post-modifier fragment	*the fact that the* *the contested decision in* *background to the dispute*	28% – 22
3. Lexical bundles that incorporate verb phrase fragments	**3.1** Verb phrase fragment with a verb in the active or passive voice	*must be rejected as* *be taken into account* *gives the following judgment*	34% – 19
	3.2 Noun phrase + verb phrase fragment	*the applicant claims that* *the commission did not* *the applicant submits tha*t	23% – 13
	3.3 Past participle + prepositional phrase fragment	*referred to in article* *laid down in article* *lodged at the court*	23% – 13
	3.4 *It* + verb phrase fragment	*it is apparent that* *it should be noted* *it follows that the*	20% – 11

4. Lexical bundles that incorporate dependent clause fragments	**4.1** *That*-clause fragment	*that the court should* *that the board of* *that the general court*	69% – 13
	4.2 *To*-clause fragment	*to pay the costs* *to take into account* *to rule on the*	26% – 5
	4.3 *-ing* form clause fragment	*having regard to the*	5% – 1

Table 2. Structural classification and distribution of lexical bundles in English court judgments

It is apparent from Table 2 that, in total, twelve different subtypes of lexical bundles have been identified. The table above shows that lexical bundles that incorporate prepositional phrase fragments have been found to contain the following three structural subtypes: *in* + noun phrase fragment, *of* + noun phrase fragment, and other prepositions (*on, for, to, at, according to, under, by, from, with*) + noun phrase fragment. *In* + noun phrase fragment is, undoubtedly, the most numerous category of lexical bundles incorporating prepositional phrase fragments; it makes 37% of all prepositional lexical bundles. Many of these bundles are used to interpret the prior or forthcoming text, to indicate relations in the text (*e.g. in accordance with article*, *in the preamble of*), or abstract relations (*in relation to the*, *in the event of the*, *in the form of*). Lexical bundles with the construction o*f* + noun phrase fragment account for 30% of all lexical bundles of this type. These bundles help to specify a particular text, legal context, values, and principles (*e.g. of the basic regulation*, *of the principle of*, *of the value of*). 33% of all prepositional lexical bundles are covered by lexical bundles that start with other prepositions.

It can be seen from the data in Table 2 that lexical bundles that incorporate noun phrase fragments can be divided into two subtypes. The first group of nominal lexical bundles, which accounts for 72% of all lexical bundles of this type, include lexical bundles that consist of a noun phrase followed by a post-modifying *of*-phrase fragment. The second group of nominal lexical bundles is made of noun phrases with other post-modifier fragments. This group covers 28% of all nominal lexical bundles. Together these subtypes are used to specify particular aspects of information in court judgments (*e.g. an infringement of the*, *the nature of the*, *the general court in*).

Even though, as illustrated in Figure 1, verbal lexical bundles constitute only 23% of all lexical bundles in court judgments, they are more varied in structural subtypes. Overall, four subtypes of this category have been distinguished. The largest subtype, *i.e.* 34% of all verbal lexical bundles, is made of lexical bundles that incorporate a verb phrase fragment with a verb in the active or passive voice, which are used to denote some kind

of an action (*e.g. must be rejected as, see to that effect*). The table above shows that each of the other two subtypes, namely, lexical bundles that incorporate a noun phrase with a verb phrase fragment and a past participle with a prepositional phrase fragment covers 23% of all verbal lexical bundles. Lexical bundles that incorporate a noun phrase with a verb phrase fragment are mainly used to refer to actions of the parties to the case (*e.g. the applicant claims that*), whereas lexical bundles representing the subtype of lexical bundles made of past participles with prepositional phrase fragments are used to specify a particular part of the legal text or an institution (*e.g. laid down in article*, *provided for in article*). 20% of verbal lexical bundles are covered by the subtype that incorporates *it* with verb phrase fragments, which serves to report the stance of the author such as necessity (*e.g. it should be noted*), logical consequence (*e.g. it follows that the*, *it is clear that*), or appropriateness (*e.g. it is appropriate to*).

The last, rather infrequent type of lexical bundles in court judgments, *i.e.* clausal lexical bundles, can be divided into three subtypes, as shown in the table above. Lexical bundles that incorporate *that*-clause fragment are the most common structure of clausal lexical bundles as it accounts for almost 70% of all lexical bundles of the clausal type. These multi-word sequences appear mainly either with the parties to the case as the subject (*e.g. that the general court*, *that the applicant had*, *that the commission was*), with an extraposed clause embedded in *that*-clause, with *it* as the subject and the copula *is/was* as the verb (*e.g. that it is not*, *that it was not*), or *that* with prepositional phrase fragments (*e.g. that in order to*, *that in accordance to*). Lexical bundles with the *to*-clause fragment covers 26% of all clausal lexical bundles and are used to refer to particular actions (*e.g. to pay the costs*, *to ensure that the*, *to determine whether the*). Only one case of the lexical bundle with *-ing* form clause fragment (*having regard to the*) was found in the corpus, which covers only 5% of all clausal bundles.

Taking into consideration the analysis of structural types of lexical bundles in court judgments it can be noted that lexical bundles mainly incorporate prepositional and nominal phrases. As explained by Goźdź-Roszkowski, "such distribution of structural types contributes to the perception of legal genres as linguistically highly nominal" (2011, 117). However, court judgments have a small proportion of lexical bundles incorporating verb phrase and dependent clause fragments, which are more characteristic of spoken language. All four main structural types of lexical bundles are expressed by different structural subtypes, which help to shape the meanings in court judgments. The following subsection of the paper focuses on the way the discussed structures of lexical bundles in court judgments are rendered into Lithuanian.

4.2 Rendering of the structure of English lexical bundles into Lithuanian

The main goal of the analysis of the translation of English lexical bundles into Lithuanian is to examine what happens with the structure of 4-word English lexical bundles in translation and to establish the translational patterns of different structural types of lexical bundles.

First of all, the largest group, *i.e.* prepositional lexical bundles, was taken into consideration. The contrastive analysis of English prepositional lexical bundles and their Lithuanian equivalents revealed that prepositional lexical bundles incorporating *in* + noun phrase fragments are rendered into Lithuanian by five different structures, as illustrated in Table 3.

Structural patterns	**English lexical bundles**	**Translation into Lithuanian**
EN *in* + NP > LT PREP	*in accordance with the*	*pagal* 'according to'
	in accordance with article	*pagal <...> straipsnio* 'according to article <…>'
	in relation to the	*dėl* 'due to' *apie* 'about'
	in respect of the	*dėl* 'due to'
EN *in* + NP > LT PTCP	*in accordance with the*	*remiantis* 'based on'
	in support of the	*grindžiamas* 'based on'
	in the context of (takes a verb after of, *e.g.* 'application')	*atlikdama* 'accomplishing' *vykdydama* 'performing'
EN *in* + NP > LT CONJ	*in so far as*	*kadangi* 'since'
EN *in* + NP > LT ADV	*in the first place*	*pirma, pirmiausia* 'first of all'
EN *in* + NP > LT NP	*in the present case*	*nagrinėjamu atveju* 'in the present case'
	in the contested decision	*ginčijamo sprendimo* 'of the contested decision'
	in the present case	*šioje byloje* 'in this case'

Table 3. Rendering of lexical bundles that incorporate *in* + NP into Lithuanian

The table above shows that the same structural pattern in English may be translated using prepositions, participles, conjunctions, adverbs, or noun phrases. Although all prepositional lexical bundles with the preposition *in* incorporate a noun (except for *in so far as*) in English, the noun element is not rendered in most of the translational

equivalents. This pattern is used when translating lexical bundles that incorporate *in* + a noun referring to an abstract entity (*e.g. in accordance to*, *in support to*), whereas if this structural subtype of lexical bundles incorporates noun phrases referring to legal domain (*e.g. in the contested decision*, *in the present case*), the noun element of these lexical bundles is transferred in the translation. What is more, all lexical bundles of this structural subtype are translated into Lithuanian by one to two words, since functional words such as *the*, *of*, *to*, *with* are not transferred.

The table below illustrates some of the main structures used to render English lexical bundles incorporating *of* + noun phrase fragments into Lithuanian. Besides 2-word noun phrases, these lexical bundles are also translated simply by single nouns or dependant clause fragments, as shown in Table 4.

Structural patterns	**English lexical bundles**	**Translation into Lithuanian**
EN *of* + NP > LT NP	*of the basic regulation*	*pagrindinio reglamento* 'of the basic regulation'
	of the contested decision	*ginčijamo sprendimo* 'of the contested decision'
	of the member states	*valstybių narių* 'of the member states'
	of the first plea	*pirmojo ieškinio pagrindo* 'of the first plea'
	of the case law	*teismų praktiką* 'of the case law'
	of the present judgment	*šio sprendimo* 'of this decision'
EN *of* + NP > LT N	*of the value of the*	*vertės* 'of the value of'
	of the judgment in	*sprendimo* 'of the judgment'
	of the principle of	*principo* 'of the principle of'
EN *of* + NP > LT dependant clause fragment	*of the fact that*	*į tai, kad*

Table 4. Rendering of lexical bundles that incorporate *of* + NP into Lithuanian

In most cases, noun phrase fragments present in this structural subtype of lexical bundles are retained in translational equivalents, only these nouns are used in the Genitive case in the Lithuanian language. What is more, the majority of English lexical bundles incorporating *of* + NP is translated directly into Lithuanian; thus, this structural subtype of lexical bundles should be least challenging to translators.

Table 5 shows that other prepositional lexical bundles starting with such prepositions as *on*, *for*, *with*, *before*, and *to* are rendered into Lithuanian by participles, conjunctions, noun phrases, prepositions, or prepositional constructions.

Structural patterns	**English lexical bundles**	**Translation into Lithuanian**
EN *on, for* + NP > LT PTCP	*on the basis of*	*remdamasi, remiantis* 'on the basis of'
	for the purpose of	*siekiant, siekdamas* 'seeking'
EN *with, on* + NP > LT CONJ	*with the result that*	*taigi* 'thus' *todėl* 'therefore'
	on the other hand	*tačiau* 'however' *be to* 'besides'
EN *before, to* + NP > LT NP	*before the general court*	*bendrajame teisme* 'in the general court'
	to the contested decision	*ginčijamo sprendimo* 'of the contested decision'
EN *on* + NP > LT PREP construction	*on the one hand*	*iš tikrųjų, iš tiesų* 'really'
EN *with* + NP > LT PREP	*with regard to the*	*dėl* 'concerning'

Table 5. Rendering of lexical bundles that incorporate other prepositions + NP into Lithuanian

The translational profile of lexical bundles incorporating other prepositions followed by noun phrase fragments shows that they are rendered into Lithuanian by the structures of one or two words omitting functional words of the source language. Besides, the same translational pattern identified in the case of the translation of the structural subtype *in* + noun phrase is retained when translating lexical bundles which incorporate different prepositions + noun phrases. To be more specific, lexical bundles which are made of prepositions and nouns referring to abstract entities (*e.g. on the other hand*, *with regard to the*) are translated by different parts of speech, except for nouns, whereas lexical bundles incorporating prepositions and noun phrases related to legal domain (*e.g. to the contested decision*, *before the general court*) are rendered into Lithuanian by noun phrases.

The analysis of translational equivalents of nominal lexical bundles has revealed that this type of lexical bundles is mainly rendered into Lithuanian by noun phrases or simply by nouns. Lexical bundles with noun phrase + *of*-phrase fragment tend to be translated by noun phrases either in Nominative or Genitive case. In addition, this structural subtype of lexical bundles is rendered into Lithuanian by participles, which refer to the actions expressed by nouns in the English language (*e.g. adoption* – 'the act of adopting').

Structural patterns	English lexical bundles	Translation into Lithuanian
EN NP + *of*-phrase fragments > LT NP	*the board of appeal*	*apeliacinė taryba* 'the board of appeal'
	the court of justice	*teisingumo teismas* 'the court of justice'
EN NP + *of*-phrase fragments > LT N	*the statement of reasons*	*motyvavimas, motyvai* 'motivation, motives'
	an infrigement of the	*pažeidimo* 'an infrigement of'
	the application of the	*taikymo, taikymas* 'the application of'
	the scope of the	*apimties, apimtimi* 'the scope of'
	the existance of the	*buvimas* 'the existance of'
EN NP + *of*-phrase fragments > LT PTCP	*the adoption of the*	*priimant, priimdamas* 'adopting'
	the calculation of the	*apskaičiuojant* 'calculating'
EN NP + post-modifier fragments > LT NP	*the member state concerned*	*atitinkama valstybė narė* 'respective member state'
	the decision at issue	*ginčijamas sprendimas* 'contested decision'
	background to the dispute	*bylos, ginčo aplinkybės* 'facts of the case, dispute'
EN NP + post-modifier fragments > LT N + dependant clause fragment	*the fact that the*	*aplinkybė, kad* 'the fact that'
	the question whether the	*klausimas, ar* 'the question whether'

Table 6. Rendering of lexical bundles that incorporate nominal lexical bundles into Lithuanian

It can be seen from the data in Table 6 that lexical bundles incorporating noun phrases with other post-modifier fragments are rendered into Lithuanian either by noun phrases or nouns + dependant clause fragments. In general, in most cases of the translation of nominal lexical bundles, the noun element is retained in Lithuanian.

Table 7 provides the summary of translational equivalents of verbal lexical bundles. It has been observed that lexical bundles incorporating only verb phrases in an active or passive voice are rendered into Lithuanian by infinitive phrases. Further, it has been found that the structural pattern of a noun phrase + a verb phrase is retained in the translation, as all of these lexical bundles are translated by the structure made of a noun and a verb.

Structural patterns	English lexical bundles	Translation into Lithuanian
EN VP (active or passive voice) > LT INFP	*see to that effect*	*šiuo klausimu žr.* 'on this subject see'
	see by analogy judgment	*pagal analogiją žr.* 'see by analogy'
	as is apparent from	*kaip matyti iš* 'as is aparent from'
	be taken into consideration; be taken into account	*atsižvelgti į* 'to take into consideration'
	call into question the	*paneigti* 'to deny'
	must be rejected as	*reikia atmesti kaip* 'must be rejected as'
EN NP + VP > LT N + V	*the applicant claims that*	*ieškovė tvirtina, teigia, nurodo kad* 'the applicant confirms, states, indicates that'
	the applicant submits that	*ieškovė teigia, pažymi, kad* 'the applicant states, notes'
	the general court held that	*bendrasis teismas nusprendė, padarė išvadą* 'the general court decided, concluded'
	the gound of appeal must	*apeliacinio skundo pagrindą reikia* 'the ground of appeal must'
	the commision submits that	*komisija tvirtina kad* 'the commission states that'
EN PPTCP + PP > LT N + PPTCPP	*laid down in article*	*straipsnyje numatytas* 'provided for in article'
	referred to in paragraph	*punkte minėtus* 'mentioned in paragraph'
	provided for in article	*straipsnyje numatytą, numatytas straipsnyje* 'provided for in article'
	set out in paragraphs	*punktuose nurodytų, išdėstytų* 'referred to in paragraphs'
EN *it* + VP > LT V or VP	*it is necessary to*	*reikia* 'it is necessary'
	it is appropriate to	*reikia* 'it is necessary'
	it should be noted	*reikia pažymėti, nurodyti, konstatuoti* 'it should be noted, indicted, stated'
	it is clear that	*reikia pažymėti, nurodyti, konstatuoti, kad* 'it should be noted, indicated, stated that'

Table 7. Rendering of verbal lexical bundles

The investigation of the parallel corpus has also revealed that past participle + prepositional phrase constructions are rendered into Lithuanian using past participle phrases; thus, the main elements of the original structure are retained. The last subtype of verbal lexical bundles, *it* + verb phrases, are translated by single verbs or verb phrases with infinitives, as illustrated in the table above. The data in the table also indicate that many synonyms are used to render the structure of English verbal lexical bundles into Lithuanian. What is more, the translation of some lexical bundles (*e.g. see to that effect*, *see by analogy judgment*) by equivalents incorporating abbreviations (*e.g.* šiuo *klausimu* žr.) reveals some of the stylistic conventions in the Lithuanian language.

The data in the table below gives an overview of the translational correspondences of clausal lexical bundles in the Lithuanian language.

Structural patterns	**English lexical bundles**	**Translation into Lithuanian**
EN *that*-clause fragment > LT *that*-clause fragment	*that the contested decision*	*kad ginčijamas sprendimas* 'that the contested decision'
	that the fact that	*kad tai jog; jog dėl to kad* 'that the fact that'
	that in order to	*tam kad; kad* 'that in order to'
	that in accordance with	*kad pagal* 'that in accordance to'
	that in order to	*tam kad; kad* 'that in order to'
	that according to the	*kad pagal* 'that according to'
EN *to*-clause fragment > LT INF	*to take into account*	*atsižvelgti į* 'take into account'
	to rule on the	*nagrinėti* 'to investigate' *pareikšti* 'to state'
	to ensure that the	*užtikrinti kad* 'to ensure that'

Table 8. Rendering of clausal lexical bundles

The table provides clear evidence that, when translating both lexical bundles incorporating *that*-clause fragments and lexical bundles incorporating *to*-clause fragments, the original structural patterns are retained, although translational equivalents are shorter, which may be explained by differences between the two languages.

5 Conclusions

The present pilot study was designed to investigate lexical bundles from structural and translational perspectives. In contrast to the majority of the previous studies on lexical bundles, which have been based on the English language, this is the first study to investigate lexical bundles in legal discourse by taking into consideration two languages, *i.e.* English and Lithuanian, and conduct a contrastive analysis. The findings of this research suggest that lexical bundle approach could be adopted not only to language learning and teaching but also to translation.

This study has shown that the analysis of lexical bundles provides insights into formulaic nature of legal discourse. Taken together, the results of the paper have revealed that four different structural categories, which are characteristic of this particular genre of legal texts, prevail in English court judgments. The great frequency of pre-formulated multi-word sequences in English court judgements reflects a high level of formality of legal texts, which leaves little room for creativity. The most obvious findings to emerge from the analysis are that prepositional and nominal lexical bundles are the most frequent multi-word expressions in court judgments, whereas verbal and clausal bundles are not prevailing in court judgments. This tendency confirms the observation made by many researchers that verbal lexical bundles are more characteristic of spoken language. The four major types of lexical bundles were further classified into smaller structural groups of lexical bundles, identifying 12 different structural subtypes of lexical bundles found in English court judgments.

The analysis of the translation of the identified subtypes of English lexical bundles showed that prepositional lexical bundles are rendered from English into Lithuanian by a great variety of different structures, *i.e.* conjunctions, adverbs, participles, prepositions, and noun phrases. Nominal lexical bundles are rendered from English in Lithuanian by noun phrases, nouns, participles, nouns + dependant clause fragments. Verbal lexical bundles are translated from English into Lithuanian mainly by infinitive phrases, nouns + verbs, nouns + past participles, verbs, and verb phrases. Both in the translation of nominal and verbal lexical bundles, the noun element and verb element are retained respectively. Finally, clausal lexical bundles are rendered from English into Lithuanian by *that*-clause fragments and infinites. As the paper demonstrates, a lexical bundle can be successfully treated as a single cognitive unit for the purposes of establishing a translational equivalent. The identification of translational equivalents in the Lithuanian corpus is in line with the fact that English and Lithuanian are typologically different languages.

There are several noteworthy limitations of this pilot study. First of all, this pilot study is limited to the analysis of only 4-word lexical bundles. What is more, this paper does not provide a complete picture of translation tendencies followed when translating different English lexical bundles into Lithuanian since not all 245 cases of the identified lexical bundles were taken into consideration in respect of translational equivalents. Moreover, even structurally incomplete lexical bundles were taken into consideration when identifying translational equivalents in the parallel corpus. Finally, the translation of different structural types of lexical bundles from English into Lithuanian was approached only from the structural perspective. Despite the limitations, this study is considered to contribute to the existing knowledge of lexical bundles in legal discourse in general and the application of lexical bundle approach to legal translation in particular.

There is still room for further research to provide better understanding of the application of this approach to the translation of different genres of legal texts. Firstly, the investigation into longer lexical bundles may provide better insights into the formulaic nature of legal discourse. What is more, a research focusing on Lithuanian lexical bundles in translated court judgments and court judgments originally written in Lithuanian would offer some important insights into the phenomenon of formulaicity in legal discourse in general and the role of lexical bundles in both languages in particular. Finally, the contrastive study of translational equivalents of lexical bundles based on the functional perspective could reveal more promising results.

List of abbreviations

LT – Lithuanian
N – noun
NP – noun phrase
PP – prepositional phrase
PPTCP – past participle
PPTCPP – past participle phrase
PREP – preposition
PTCP – participle
V – verb
VP – verb phrase

Data sources

InfoCuria. *Case-law of the Court of Justice*. Available at: http://curia.europa.eu/juris/recherche.jsf? cid=92048

References

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

Anthony, Laurence. 2015. *AntConc (Version 3.4.4)* [Computer Software]. Tokyo, Japan: Waseda University. Available at: http://www.laurenceanthony.net/.

Anthony, Laurence. 2017. *AntPConc (Version 1.2.0)* [Computer Software]. Tokyo, Japan: Waseda University. Available at: http://www.laurenceanthony.net/.

Amirian, Zahra, Somaye Ketabi & Harmed Eshaghi. 2013. The use of lexical bundles in native and non-native post-graduate writing: the case of applied linguistics MA theses. *Journal of English Language Teaching and Learning* 11, 1–29.

Basaneže, Katja Dobrić. 2015. Investigating "concgrams" in the language of contracts and legal agreements. In *Legal phraseology today: corpus-based view*.

Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. Special Issue of *Fachsprache* 3-4, 176–192.

Biber, Douglas & Federica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26, 263–286.

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

Biber, Douglas, Susan Conrad & Vivian Cortes. 2004. If you look at…: lexical bundles in university teaching and textbooks. *Applied Linguistics* 25 (3), 371–405.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Biber, Douglas & Susan Conrad. 1999. Lexical bundles in conversation and academic prose. In *Out of corpora. Studies in honour of Stig Johansson*. Hilde Hasselgård & Signe Oksefjell, eds. Amsterdam: Rodopi. 181–191.

Biel, Łucja. 2017. Lexical bundles in EU law: the impact of translation process on the patterning of legal language. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. London, New York: Routledge. 10–26.

Breeze, Ruth. 2013. Lexical bundles across four legal texts. *International Journal of Corpus Linguistics* 18 (2), 229–253.

Castro, Miriam Buendía & Pamela Faber. 2015. Phraseological units in English-Spanish legal dictionaries: a comparative study. In *Legal phraseology today: corpus-based view*. Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. Special Issue of *Fachsprache* 3-4, 161–175.

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

Conrad, Susan & Douglas Biber. 2004. The frequency and use of lexical bundles in conversation and academic prose. *Applied Linguistics* 20, 56–71.

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

De Cock, Sylvie. 1998. A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics* 3 (1), 59–80.

Ebeling, Jarle & Signe Oksefjell Ebeling. 2013. *Patterns in contrast*. Amsterdam: John Benjamins.

Goźdź-Roszkowski, Stanisław. 2004. Multi-word lexical units in legal genres – the pedagogic perspective. In *Practical applications in language and computers*. Barbara Lewandowska-Tomaszczyk, ed. Frankfurt am Main: Peter Lang. 401–416.

Goźdź-Roszkowski, Stanisław. 2006a. Frequent phraseology in contractual instruments: a corpus-based study. In *New trends in specialized discourse analysis*. Maurizio Gotti & Davide S. Giannoni, eds. Bern: Peter Lang. 147–162.

Goźdź-Roszkowski, Stanisław. 2006b. Recurrent word combinations in judicial argumentation. A corpus-based study. In *Langue, droit, société/Language, law, society*. Danuta Bartol, Anna Duszak, Hubert Izdebski & Jean-Marie Pierrel, eds. Nancy: Presses universitaires de Nancy. 139–152.

Goźdź-Roszkowski, Stanisław. 2011. *Patterns of linguistic variation in American legal language: a corpus-based study*. Frankfurt am Main: Peter Lang.

Goźdź-Roszkowski, Stanisław & Gianluca Pontrandolfo. 2015. Legal phraseology today: corpus-based applications across legal languages and genres. Special Issue of *Fachsprache* 3-4, 130–138.

Goźdź-Roszkowski, Stanislaw & Gianluca Pontrandolfo, eds. 2015. Legal phraseology today. A corpus-based view. Special Issue of *Fachsprache* 3-4.

Goźdź-Roszkowski, Stanislaw & Gianluca Pontrandolfo, eds. 2017. *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). London, New York: Routledge.

Goźdź-Roszkowski, Stanislaw. 2017. Facts in law: a comparative study of *fact that* and its phraseologies in American and Polish judicial discourse. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. London, New York: Routledge. 143–159.

Grabar, Natalia & Marie-Aude Lefer. 2015. Building a lexical bundle resource for CAT and MT. MUMTTT workshop (*Multi-word units in machine translation and translation technology*), University of Málaga, 1–2 July 2015.

Grabowski, Łukasz. 2013. Register variation across English pharmaceutical texts: a corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics. *Procedia: Social and Behavioral Sciences* 95, 391–401.

Grabowski, Łukasz. 2014. On lexical bundles in Polish patient information leaflets: a corpus-driven analysis. *Studies in Polish Linguistics* 9 (1), 21–43.

Güngör, Fatih. 2016. *Crosslinguistic analysis of lexical bundles in L1 English, L2 English and L1 Turkish research articles*. PhD dissertation. Gazi University, Gaduate school of educational science.

Güngör, Fatih & Hacer H. Uysal. 2016. A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching* 9 (6), 176–188.

Hunston, Susan & Gill Francis. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hyland, Ken. 2008a. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18, 41–62.

Hyland, Ken. 2008b. As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes* 27, 4–21.

Jablonkai, Réka. 2009. "In the light of": a corpus-based analysis of lexical bundles in two EU-related registers. *VoPaLp* 3, 1–27.

Jablonkai, Réka. 2010. *A corpus-linguistic investigation into the lexis of written English EU discourse: an ESP pedagogic perspective.* PhD dissertation. Faculty of Pedagogy and Psychology, Eötvös Loránd University, Budapest.

Jalali, Zahra Sadat & Mohammad Raouf Moini. 2014. Structure of lexical bundles in introduction section of medical research articles. *Procedia Social and Behavioral Sciences* 98, 719–726.

Juknevičienė, Rita. 2011. *Leksinės samplaikos svetimkalbių ir gimtakalbių vartotojų rašytinėje anglų kalboje* [Lexical bundles in non-native speaker and native speaker written English]. Daktaro disertacija [PhD dissertation]. Vilnius: Vilniaus universitetas.

Kim, YouJin. 2009. Korean lexical bundles in conversation and academic texts. *Corpora* 4 (2), 135–165.

Kopaczyk, Joanna. 2012. Applications of the lexical bundles method in historical corpus research. In *Corpus data across languages and disciplines*. Piotr Pęzik, ed. Frankfurt am Main: Peter Lang. 83–95.

Kopaczyk, Joanna. 2013. *The legal language of Scottish burghs: standartization and lexical bundles*. Oxford: Oxford University Press.

Kopaczyk, Joanna. 2017. Terms and conditions**.** A comparative study of noun binomials in UK and Scottish legislation. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. London, New York: Routledge. 160–185.

Mackiewicz, Jo. 2017. *The aboutness of writing center talk. A corpus-driven and discourse analysis*. New York: Routledge.

Monti, Johanna, Ruslan Mitkov, Gloria Corpas Pastor & Violeta Seretan, eds. 2013. *Multi-word units in machine translation and translation technologies workshop proceedings*. Machine Translation Summit XIV. Allschwil: The European Association for Machine Translation.

Nebot, Esther Monzó. 2017. The out-grouping society: phrasemes othering underprivileged groups in the International Bill of Human Rights (English-French-Spanish). In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. London, New York: Routledge. 109–125.

Pan, Fan, Randi Reppen & Douglas Biber. 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.

Parra, María Fernández & Pius ten Hacken. 2008. Beyond terms: multi-word units in multiterm extract. *Proceedings of the 30th international conference on translating and the computers*. London: ASLIB. Available at: http://www.mt-archive.info/Aslib-2008-Fernandez-Parra.pdf. Accessed: 13 October 2017.

Partington, Alan & John Morley. 2004. From frequency to ideology: investigating word and cluster/bundle frequency in political debate. In *Practical applications in*

language and computers. Barbara Lewandowska-Tomaszczyk, ed. Frankfurt a. Main: Peter Lang. 179–192.

Purificación, Sánchez Hernández. 2013. Lexical bundles in three oral corpora of university students. *Nordic Journal of English Studies* 13 (1), 187–209.

Ruusila, Anna & Emilia Londroos. 2016. *Conditio sine qua non:* on phraseology in legal language and its translation. *Language and Law* 3 (1), 120–140.

Salazar, Danica. 2011. *Lexical bundles in scientific English: a corpus-driven study of native and non-native writing*. PhD dissertation. Barcelona: University of Barcelona.

Salazar, Danica. 2010. Lexical bundles in Philippine and British scientific English. *Philippine Journal of Linguistics* 41, 94–109.

Salazar, Danica. 1996. *Lexical bundles in native and non-native writing*. Amsterdam: John Benjamins.

Salkie, Raphael. 2017. Legal phraseology in contrast: *the fact that* and its German counterparts. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective* (e-book). Stanislaw Goźdź-Roszkowski & Gianluca Pontrandolfo, eds. London, New York: Routledge. 126–142.

Sardinha, Tony Berber, Rosana Teixeira & Telma Ferreira. 2014. Lexical bundles in Brazilian Portuguese. In *Working with Portuguese corpora*. Tony B. Sardinha & Telma Ferreira, eds. Bloomsbury: Bloomsburg Publishing Plc. 33–68.

Schmitt, Norbert. 2005. Formulaic language: fixed and varied. *Elia* 6 (6), 13–39.

Shrefler, Nathan. 2011. Lexical bundles and German bibles. *Literary and Linguistic Computing* 26 (1), 89–106.

Simpson-Vlach, Rita & Nick Ellis. 2010. An academic formulas list: new methods in phraseology research. *Applied Linguistics* 31 (4), 487–512.

Simpson, Rita. 2004. Stylistic features of academic speech: the role of formulaic expressions. In *Discourse in the professions: perspectives from corpus linguistics*. Thomas A. Upton & Ulla Connor, eds. Amsterdam: John Benjamins. 37–64.

Stubbs, Michael & Isabel Barth. 2003. Using recurrent phrases and text-type discriminators: a quantitative method and some findings. *Functions of Language* 10 (1), 61–104.

Tománková, Veronika. 2016. Lexical bundles in legal texts corpora – selection, classification and pedagogical implications. *Discourse and interaction* 9 (2)*,* 75–94.

Tracy-Ventura, Nicole, Viviana Cortes & Douglas Biber. 2007. Lexical bundles in Spanish speech and writing. In *Working with Spanish corpora*. Giovanni Parodi, ed. London: Continuum. 354–375.

Yang, Yanfeng. 2017. Lexical bundles in argumentative and narrative writings by Chinese EFL learners. *International Journal of English Linguistics* 7 (3), 58–69.

Submitted July 30, 2017