EU REGULATIONS: TENDENCIES IN TRANSLATING LEXICAL BUNDLES FROM ENGLISH INTO LITHUANIAN

Mantas Noreika, Inesa Šeškauskienė

Department of Translation and Interpretation Studies
Faculty of Philology
Vilnius University
mantas.noreika23@gmail.com
inesa.seskauskiene@flf.vu.lt

Based on the theoretical framework of lexical bundle research developed by Biber et al. (2004) and on Sinclair's (1991) observations on the compositionality of meaning and delexicalisation of formulaic language, the present study analyses four-word lexical bundles found in English EU legal discourse, focusing on their structural and semantic properties and tendencies of their translation into Lithuanian. The study examines a selection of 43 lexical bundles with a total of 791 unique instances in the corpus, extracted from a self-designed 59,579-word corpus representing EU legal English comprised of five regulations on health claims and food labelling. The results show that lexical bundles can be rendered into Lithuanian as single lexical or function words, phrases, dependent clauses, through the grammatical category of case, or omitted altogether. The length of the translated bundles suggests that there is a tendency to choose a shorter way of rendering them into Lithuanian. High variability of the translations of some lexical bundles points to a lack of their stability in Lithuanian, which might be due to a number of reasons, including a lack of agreement among translators and a much shorter tradition of legal translation into Lithuanian.

1. INTRODUCTION

In formulaic language research there have been several names whose contribution can hardly be disregarded. A number of them are closely linked to the studies of collocations initiated by John Firth several decades ago (1957). Among numerous books and papers dealing with different aspects of formulaic language, seminal publications by Sinclair (1991) and Biber (Biber et al. 1999; Biber et al. 2004) stand out. They initiated a shift in focus from single words to phrases in research on language production. A similar line is maintained in Wray's works, who defines formulaic language from the point of view of its processing: formulaic words and word strings are those "which appear to be processed without recourse to their lowest level of composition" (Wray 2002, 4). Both approaches, language production and language processing, agree on the main notions and adhere to the same principles.

Compositionality and formulaicity are two ends of a continuum. The open-choice principle of compositionality, when language is produced in a linear fashion and the forthcoming lexical element is taken from the lexicon as single, is only one possibility, which language users resort to only occasionally (Sinclair 1991). The other, a much stronger principle, as claimed by Sinclair (*ibid.*, 109–110), is based on formulaicity, when semi-preconstructed phrases act as single words and fit into the required slots.

Challenging the conception of strict language compositionality, Biber et al. argued that "(...) much of our everyday language use is composed of prefabricated expressions" (2004, 372). According to the scholars, research into formulaic language found that multi-word expressions can make up as much as 55 per cent of language that we use, depending on the register and mode of expression. In their previous publication, Biber and his colleagues indicate that recurrent expressions amount to 21 per cent of all language use in academic prose and 28 per cent in conversation (Biber et al. 1999, 993–995), Foster suggests that multi-word expressions constitute approximately 25 to 32 per cent of spoken language of native speakers of English (2001, 85), whereas Erman and Warren report even greater figures, claiming that on average prefabricated expressions comprise 55 per cent of spoken and written language (2000, 37). Despite the fact that figures differ from author to author, it can hardly be questioned that during the last decades formulaic language has been taken out of the periphery and placed to the forefront of linguistic inquiry. Formulaicity has been researched from a variety of perspectives: in reception and production, in spoken and written language, across different genres and different professional discourses. Unsurprisingly, most studies focus on English.

The processing of multi-word units and their representation in the mental lexicon is in the focus of psycholinguistic analysis (Jiang, Nekrasova 2007; Conklin, Schmitt 2008). There is an interesting study attempting to determine the correlation between the density of formulaic language and the readability of a written text (Shrefler 2011). In pedagogy, the compilation of frequently recurring multi-word expression lists paved the way for novel language learning and teaching methods (Simpson-Vlach, Ellis 2010; Martinez, Schmitt 2012), whereas other studies focused on the relationship between the knowledge of formulaic language and linguistic competence of native and non-native speakers (Juknevičienė 2011). Discourse and disciplinary variation of formulaic language have also been addressed, bringing about studies in medical (Grabowski 2013), academic (Biber et al. 2004; Biber, Conrad 2004; Hyland 2008) and legal discourse (Trebits 2009b; Jablonkai 2010).

Legal discourse is notoriously complex, rigid and difficult to understand, despite its aim at precision (Tiersma 2006). The rigidity of the discourse is to a large extent concerned with standardisation, rituals and formulaicity (Biel, Engberg 2013), as pointed out by several researchers. Studies by Trebits (2008, 2009a, 2009b) are worth a closer look. In one of them (Trebits 2008), the scholar was mostly interested in the pedagogical aspects of EU English. She used a self-designed corpus of EU English and concordancing and semantic analysis software to conduct an exploratory study of English lexis in EU legal discourse. The researcher examined the most frequent lexical elements and compared them with the 3,000 most frequent words in the British National Corpus (BNC). The results have shown that almost half of the most frequent lexical elements of EU English did not overlap with the BNC's 3,000 most frequent words and concluded that the vocabulary of EU English could prove challenging for average language learners. Her next study (Trebits 2009a) focused on conjunctive cohesion. She identified the most frequent conjunctions in EU English and compared them with the most frequent conjunctions in the sub-corpus of written English of the BNC. It turned out that causal, clarifying and continuative conjunctions were significantly more frequent in EU English, whereas adversative and hypothetical conjunctions were more frequent in the BNC. Another follow-up study by the same author (Trebits 2009b) focused on the most frequent phrasal verbs in EU English and found that just as academic English, EU English uses considerably fewer phrasal verbs as compared to fiction or news writing or spoken English and showed that polysemous phrasal verbs in EU English were used in significantly fewer senses than in English in general. Moreover, the last two studies provided sample tasks exemplifying how a corpus of EU English could be used to design data-driven pedagogic material.

Another author (Jablonkai 2010) focused specifically on lexical bundles in English EU documents and found that they abounded in high frequency lexical bundles. According to the scholar, their prevalence is indicative of high formulaicity of EU English discourse. Furthermore, verb-based lexical bundles appeared more often in the English EU discourse as compared to other written registers; however, most of them incorporated noun phrase and prepositional phrase fragments. Nominal and prepositional phrases, as claimed by the author, are also typical of written technical discourse.

Lexical bundles and their frequency are in the focus of the present study as well. Despite their frequency and relevance in specific texts, such as legal, there seems to be hardly any research done in translating lexical bundles into Lithuanian. The paper

aims at identifying structural and semantic features of some bundles in English and at identifying major tendencies of translating them into Lithuanian by focusing on the retention of semantic content in the target text. Our research also raises the question of stability of some bundles in Lithuanian.

2. THE NOTION OF LEXICAL BUNDLES AND THEIR CLASSIFICATION

Lexical bundles are defined by Biber et al. (1999, 992) as commonly recurring word combinations in a given register. One of the primary criteria for the identification of lexical bundles is frequency. According to another paper by Biber et al. (2004, 376), frequency data not only allows us to isolate and identify lexical bundles, but also indicates the likelihood of any given multi-word expression to be stored and used as an unanalysed prefabricated chunk in the mental lexicon. Jiang and Nekrasova (2007, 433) and Siyanova-Chanturia and Martinez (2015, 549) also identify frequency as an important defining characteristic of formulaic sequences; however, unlike Biber et al. (1999), they do not single out register specificity.

Another feature common to most lexical bundles, according to Biber and his colleagues (Biber et al. 2004, 376), is that their meaning is not idiomatic, i.e. the meaning of most lexical bundles can be derived from the meanings of their constituent words. However, the definitions of other scholars are more specific. For example, Jiang and Nekrasova suggest that the meanings of formulas in general are not idiomatic and can be arrived at by combining the meanings of each component word (2007, 433), whereas Martinez and Schmitt claim the opposite: the meanings of such phrases cannot be "(…) easily discernible by decoding the individual words alone" (2012, 304). As Sinclair suggests, this may be due to progressive delexicalisation when the higher frequency of a word reduces its distinctive contribution to the meaning of a phrase because, according to him, meanings of frequent words are less clear and independent than those of less frequent words, which are easier to explain (1991, 113). In another book Sinclair claims that "the meaning of words chosen together is different from their independent meanings" (2004, 20), i.e. the words are delexicalised. Tognini-Bonelli refers to a similar process as desemantisation "through which a lexical item loses its original lexical value and often acquires other meanings and other functions within a larger unit" (2000, 228–229). Sinclair describes some items of such delexicalised vocabulary as the empty lexicon. He distinguishes between two types of lexicon: one comprised of terminological lexical items and the other comprised of empty lexical items. The

terminological lexical items do not always coincide with terms proper, but they do have meanings that are fixed, autonomous and clearly differentiated, whereas the meanings of the empty lexical items are much more opaque and become more transparent only when placed in context (Sinclair 1996, 112–116, in Marcinkevičienė 2010, 65). In the light of the above considerations, it appears that Biber et al.'s approach (2004) to the issue of compositionality of meaning is most inclusive, because it allows for both idiomatic and non-idiomatic lexical bundles, thus enabling a continuum approach to the study of lexical bundles.

The third aspect to consider is the structural completeness of lexical bundles. Biber et al. claim that most lexical bundles are structurally incomplete because "they begin at a clause or phrase boundary, but the last words of the bundle are the first elements of a second structural unit" (2004, 377), i.e. their structural boundaries overlap and act as bridges to connect two different structural units. Only 15 per cent of lexical bundles in conversation and less that 5 per cent in academic prose are structurally complete units (Biber et al. 1999, 995). However, Biber and his colleagues do not elaborate on the definition of a structurally complete lexical bundle. Their view is similar to Sinclair's, who suggests that there are no clear-cut boundaries between stretches of language constructed on the open-choice principle and the idiom principle, i.e. between the word-based and the phrase-based models of language production, and states that points at which a switch between the two principles occurs are sudden and covert (Sinclair 1991, 114). In other words, there is room for overlap between phrasal boundaries. Arguably, the meanings of structurally incomplete bundles are also incomplete and require a final element to fill the open slot; this element may be a single word or another phrase which acts as a single word. This could also be indicative of some degree of delexicalisation of lexical bundles.

The classification of lexical bundles adopted by Biber et al. (2004, 380–381) is twofold: structural and functional. A lexical bundle may belong to one of three structural types of lexical bundles, depending on what phrasal fragments it contains: bundles incorporating verb phrase fragments, bundles incorporating dependent clause fragments (in addition to simple verb phrase fragments) or bundles incorporating noun phrase and prepositional phrase fragments. The functional classification developed by Biber et al. (2004, 383–384) takes into account functions performed by lexical bundles in a specific discourse. The main functions include stance (expressing attitudes or assessments of certainty), discourse organizing (creating discourse cohesion) and referential (indicating physical or abstract entities). According to the scholars, not only

can these functions be sub-categorized further, but also the same lexical bundle may sometimes be functionally polysemous, serving several functions at once in a single occurrence. The classification has been successfully used by several other scholars working on lexical bundles (Hyland 2008; Jablonkai 2010; Shrefler 2011). It has been adopted in the present study as well.

3. MATERIALS AND METHODOLOGICAL FRAMEWORK

The present study is based on the analysis of a self-designed corpus of written EU legal texts in English collected from the EUR-Lex database of EU law (EUR-Lex 2016). The corpus is comprised of five regulations, 59,579 words in total, and focuses on nutritional claims and food additives. The corpus includes two regulations, two regulation proposals and one regulation marked as a non-legislative act. The choice of this specific type of EU legal act was primarily governed by the fact that regulations do not need to be transposed into the legal systems of each member state individually and remain the primary sources of reference in the interpretation of EU law. Although regulations are translated into all languages of the EU and all language versions have equal legal status, the fact that they need not be transposed into national legislation means that the linguistic elements comprising these legal acts, lexical bundles in this case, do not undergo any changes that may take place during transposition, which makes the analysis of authentic EU legal language possible.

Moreover, all regulations focus on one thematic area. This seems crucial for the results of our study, since the thematic area is directly linked to the choice of lexical items.

Before conducting the analysis of lexical bundle translation tendencies, lexical bundles had to be extracted and analysed. The bundles were extracted using AntConc (Laurence 2014) software. The N-Grams tool returned 266 four-word lexical bundles based on the following criteria: the raw minimum frequency was set to five and the same type of lexical bundles had to occur in at least three different texts. Both criteria were set to allow for relatively infrequent yet evenly distributed lexical bundles. Afterwards, lexical bundle frequency was normalised per one million words, which showed that the relative frequency of the least frequent lexical bundles was 88 per million words. For comparison, this figure is well above the frequency cut-off value per one million words used by Biber et al. (1999) and Biber et al. (2004), i.e. 10 and 40, respectively. A further note to be made is that numerals and symbols were not considered to be constituent parts of lexical bundles. If a strong pattern emerged that a certain lexical

bundle had a numeral or a symbol in its internal structure, these elements were inserted into the lexical bundle, for example, ***more than # % by volume***[1]. However, such lexical bundles were still considered to be four-word lexical bundles.

The procedure of the analysis was as follows. First, the extracted lexical bundles were analysed with regard to their structural and functional type. The structural analysis mainly included the identification of structural bundle initial and final completeness and structural dependence. Then, of the 266 lexical bundles, 43 were chosen for translation analysis. To qualify, lexical bundles had to be structurally complete in at least one position, i.e. initial, final or both, and express EU legal discourse-specific procedural reference, i.e. to refer to general legal procedures rather than topic-specific terms or names of institutions or bodies. Topic-specific bundles and bundles referring to institutions or bodies were disregarded, because such bundles have a higher likelihood to overlap with terms which, in turn, would increase the probability of them being included in glossaries or terminology databases. This enabled the study to focus on bundles that language users would be less likely to find in dictionaries.

Finally, to examine the translations of the 43 selected lexical bundles, AntPConc (Laurence 2013) software was used to compile and analyse a parallel corpus of EU legal discourse in English and Lithuanian. However, in cases when the texts in the parallel corpus did not line up, the translations which are also freely accessible in the EUR-Lex database (2016) were checked manually in regulation files in the Printable Document Format (PDF) in both languages.

4. LEXICAL BUNDLES AND THEIR TRANSLATION

Of 43 lexical bundles, 21 were noun phrase (NP) and prepositional phrase (PP) bundles, 14 were verb phrase (VP) bundles, six were dependent clause (DC) bundles and two were adjective phrase (AP) bundles. The size of the corpus was 791 tokens. The distribution of types and tokens of the analysed bundles is given in Table 1.

Table 1. Type and token distribution of lexical bundles in English

Structural type	NP and PP	VP	DC	AP	Total
Types	21 (48.84 %)	14 (32.56 %)	6 (13.95 %)	2 (4.65 %)	43 (100 %)
Tokens	281 (35.53 %)	269 (34.01 %)	231 (29.2 %)	10 (1.26 %)	791 (100 %)

---

[1] For ease of reference all lexical bundles analysed in the present study and, where relevant, their translations into Lithuanian are highlighted in bold type.

For comparison, Table 2 gives a general overview of the distribution of the analysed lexical bundle translations in terms of structure.

Table 2. Structural types of lexical bundle translations into Lithuanian

Structural type	Noun- and preposition-based	Verb-based	Conjunction-based (including DCs)	Adjective-based	Adverb-based	Token total
Single word and phrasal translation	196 (26.56 %)	323 (43.77 %)	204 (27.64 %)	5 (0.68 %)	10 (1.35 %)	738 (100 %)
Omissions	--	--	--	--	--	52
Unfit for analysis	--	--	--	--	--	1
					Total	791

As demonstrated in Tables 1 and 2, the most striking difference was identified in the proportion between noun- and preposition-based and verb-based tokens. Whereas English features an almost equal number of NP and PP bundle tokens and VP bundle tokens, 35.53 per cent and 34.01 per cent respectively (see Table 1), Lithuanian uses significantly more verb-based tokens (43.77 per cent, Table 2) and fewer noun- and preposition-based tokens (26.56 per cent, Table 2). This supports Pažūsis' claim that English relies more on nominal means of expression, whereas Lithuanian favours verbal expressions (2014, 380). Huddleston and Pullum suggest that nouns constitute "(…) about 37 per cent of the words in almost any text" (2007, 16) in English. This figure closely corresponds to the 35.53 per cent of NP and PP bundle tokens. However, it should be noted that although Huddleston and Pullum refer to the percentage of words in a text, the figure may still be comparable with the percentage of NP and PP lexical bundle token distribution, because prepositions are closely linked to nouns in terms of grammar and every lexical bundle of this type has at least one noun in its structure.

The distribution of DC lexical bundles and conjunction-based translations were relatively similar, 29.2 per cent (Table 1) and 27.64 per cent (Table 2), respectively. AP lexical bundles occurred ten times, whereas adjective-based translations were half as frequent. One of the reasons for fewer adjective-based translations is the fact that adjectival constructions can be translated into Lithuanian using adjectives or

participles, the latter of which in Lithuanian have features characteristic of both verbs and adjectives. Moreover, Lithuanian is extremely rich in various types of participles employed in different functions (Arkadiev 2012; Ambrazas et al. 1997), which makes them more likely to appear in the translated text.

Finally, the study identified ten occurrences of adverb-based translation which does not have a corresponding structural type of lexical bundles. In one instance, ***as the case may*** *be* was rendered as *atitinkamai* 'respectively' and ***hereinafter referred to as*** in all nine occurrences was rendered as *toliau* 'further' followed by an en dash indicating an ellipsis which stands for the *referred to as* portion of the bundle. It appears that adverbs, as an open-class of words, may be an efficient and economical means of rendering strings of formulaic language into Lithuanian.

The present study also looked at the number of lexical bundle tokens translated as single words and as phrases. A quarter, 188 cases, of bundles was rendered as single words and three quarters, 550 cases, as phrases (one case unfit for analysis and 52 cases of omission were not included in the count).

4.1. Single word translations of lexical bundles

Of the 188 cases, the most frequent single word translations of lexical bundles were gerunds (52 tokens or 27.66 per cent), prepositions (41 tokens or 21.81 per cent) and nouns (31 tokens or 16.49 per cent). Let us take a closer look at four-word bundles rendered as gerunds in Lithuanian:

(1) ***In accordance with the*** principle of proportionality (...) [Reg_EN5][2]
Laikantis (...) *proporcingumo principo* (...) [Reg_LT5]

(2) *Health claims authorised **on the basis of** proprietary data* (...) [Reg_EN4]
*Teiginiai apie sveikatingumą, kuriems vartoti leidimas suteiktas **remiantis** nuosavybės teise saugomais duomenimis* (...) [Reg_LT4]

(3) *This Regulation shall apply **without prejudice to the*** (...) [Reg_EN4]
*Šis reglamentas taikomas **nepažeidžiant*** (...) [Reg_LT4]

As seen in the examples, gerunds are convenient in rendering prepositional phrases. Some of them, such as *in accordance with* or *without prejudice to* are part of EU legal discourse, mostly occurring in documents. At the same time, some of such

---

[2] Indexes provided with each example refer to the respective regulations from which they were extracted. The full list of regulations is given at the end of the paper.

phrases are also rendered as prepositions, mainly as the preposition *pagal* 'along; in line with; according to'*,* for example:

(4)   (...) **in accordance with the** *manufacturer's instructions.* [Reg_EN4]
       (...) **pagal** *gamintojo nurodymus* (...) [Reg_LT4]

(5)   (...) *indicated* **on the basis of** *their proportion by weight* (...) [Reg_EN2]
       (...) *nurodomas* **pagal** *jų svorio* (...) *santykį.* [Reg_LT2]

These examples show that Lithuanian translations are still not entirely stable. Despite an attempt at precision, a requirement for legal documents, the same phrase is rendered in Lithuanian in several ways. Some particularly interesting cases are given below; in examples (6), (7) and (8) the lexical bundle **in the case of** is translated as a preposition; however, in each case a different preposition is employed:

(6)   (...) **in the case of** *milk and milk products* (...) [Reg_EN2]
       Dėl *pieno ir pieno produktų* (...) [Reg_LT2]

(7)   **In the case of** *packaging or containers* (...) *only the particulars listed in Article 9(1) (a), (b),(c), (e) and (f) shall be mandatory on the package or on the label.* [Reg_EN2]
       Ant *pakuočių ar taros* (...) *įpakavimo ar etiketėse privaloma nurodyti tik 9 straipsnio 1 dalies a, c, e ir f* [*punktuose*]³ *nurodytus duomenis.* [Reg_LT2]

(8)   **In the case of** *prepacked food, mandatory food information shall appear on* (...) [Reg_EN2]
       Privalomoji informacija **apie** *fasuotus maisto produktus nurodoma ant* (...) [Reg_LT2]

Different prepositions chosen to translate the same phrase are indicative of slightly different interpretation of the whole phrase, presumably, relying on different meanings of prepositions, capable of expressing a number of spatial and non-spatial relations. Further examples provide more rendering varieties of the lexical bundle **in the case of**. In examples (9) and (10) it is translated as a noun:

(9)   **In the case of** *food supplements* (...) [Reg_EN4]
       Maisto papildų **atveju** (...) [Reg_LT4]

(10)   (...) **in the case of** *alcoholic drinks* (...) [Reg_EN2]
       (...) *alkoholinių gėrimų* **srityje** (...) [Reg_LT2]

---

³   Translation error: it appears that the translator in this case omitted the head word *punktas* 'point' that is modified by the letters indicating the points in a paragraph.

In example (9) the noun *atveju* (case-INS.SG) is used and thus is closer to the original. In example (10) the noun *srityje* (area-LOC.SG) is employed. Example (9) is closer to the original, whereas in (10) the translator has chosen a word of somewhat broader semantics.

Two more cases given below show that the information in English encoded in prepositions and longer phrases tend to be "wrapped" in a single Lithuanian word. Interestingly, in (11) the choice of *derinimas* (*lit.* 'adjustment') shows a slightly different interpretation of the basic meaning of the main lexical word *approximation* 'a thing that is similar to something else but not exactly the same' (OD 2017). In (12), the preposition *into* seems to be encoded in the Lithuanian prefix *į-*; again, the English metaphorical *entry into force* is rendered as *įsigaliojimas* 'becoming valid':

(11) (…) *in Directive 2000/13/EC* (…) *on **the approximation of the** laws of the Member States* (…) [Reg_EN5]
(…) *direktyvoje 2000/13/EB dėl valstybių narių įstatymų* (…) ***derinimo**.* [Reg_LT5]

(12) (…) *from the date of **entry into force of** this Regulation* (…) [Reg_EN4]
(…) *nuo šiame reglamente nurodytos **įsigaliojimo** datos* (…) [Reg_LT4]

Several further examples illustrate how the four-word bundles, especially the phrase ***in the case of**,* are translated as relative adverbs or conjunctions. This, again, shows that the phrase has several equivalents in Lithuanian and its translation has not been completely agreed upon:

(13) (…) ***in the case of** a liquid medium* (…) [Reg_EN2]
(…) ***jei** tai skysta terpė* (…) [Reg_LT2]

(14) (…) ***in the case of** beverages containing more than 1,2 % by volume of alcohol.* [Reg_EN2]
(…) ***kai** gėrimo sudėtyje yra daugiau kaip 1,2 % alkoholio, išreikšto tūrio procentais.* [Reg_LT2]

As seen in the above examples, the overwhelming majority of bundles that can be translated as single words are comprised of one lexical word and the surrounding functional words. The translator mostly relies on the lexical word; however, in some cases, as in (1), it is difficult to identify a single lexical word, since the whole phrase *in accordance with* is an idiom and completely non-compositional. Another interesting case is the bundle *in the case of* translated using an unusually vast array of single words: nouns, prepositions and conjunctions. The reduction of a whole phrase to a single word at first sight is indicative of the loss of its semantic content. However, as pointed out by Martinez and Schmitt (2012, 304), the semantic content of a phrase cannot be the sum

of its components due to what Sinclair called progressive delexicalisation, i.e. the indirect correlation between the frequency of a word and the transparency and independence of its meaning (1991, 113). When treated as a process, delexicalisation is similar to semantic bleaching (Murphy 2010, 97), which in the course of language development, goes through different stages with the word's lexical meaning eventually "washed away".

4.2. Phrasal translations of lexical bundles

Three quarters, or 550 cases, of analysed lexical bundles were translated as phrases. Of these, the three largest phrasal groups in terms of frequency were DCs (196 tokens or 35.64 per cent), PPs (157 tokens or 28.55 per cent) and NPs (111 tokens or 20.18 per cent). Let us consider the dependent clause lexical bundle with the word *referred* rendered in several ways: as a dependent clause, a participial phrase and as a nominal phrase:

(15) (…) ***as referred to in Article 13(1)*** (…) [Reg_EN3]
      (…) ***kaip nurodyta*** (…) ***13 straipsnio 1 dalyje***. [Reg_LT3]

(16) (…) *in accordance with the procedure* ***referred to in Article 24(2)***. [Reg_EN4]
      (…) ***24 straipsnio 2 dalyje nurodyta*** *tvarka*. [Reg_LT4]

(17) (…) *in accordance with the* ***procedure referred to in*** *Article 14(2).* [Reg_EN5]
      (…) *pagal 14 straipsnio 2 dalyje* ***numatytą procedūrą***. [Reg_LT5]

Interestingly, the word *referred* is translated in two ways relying on the verb *nurodyti* 'to point out' and the verb *numatyti* 'to foresee'. This variability is not surprising since in English the idea of referring to different parts of the documents is also rendered in several ways, e.g.:

(18) *The period* ***laid down in Article 5(6)*** (…) [Reg_EN4]
      (…) ***5 straipsnio 6 dalyje nustatytas*** *laikotarpis* (…) [Reg_LT4]

(19) *Acting in accordance with the* ***procedure laid down in*** *Article 251 of the Treaty* (…) [Reg_EN2]
      (…) *laikydamiesi Sutarties 251 straipsnyje* ***nustatytos tvarkos*** (…) [Reg_LT2]

(20) (…) *the derogation* ***provided for in Article 4(b)***. [Reg_EN5]
      (…) ***4 straipsnio b punkte nurodyta*** *nukrypti leidžianti nuostata*. [Reg_LT5]

(21) (…) *in accordance with the principle of subsidiarity as* ***set out in Article 5*** *of the Treaty.* [Reg_EN4]
      (…) *pagal Sutarties* ***5 straipsnyje nustatytą*** *subsidiarumo principą*. [Reg_LT4]

Apparently, the above bundles employing the phrasal verbs *lay down, provide for* and *set out* are synonymous and so far have not caused either legal misinterpretations or, more importantly, disagreement among linguists or lawyers. The Lithuanian translators' choice in this case is understandable.

Alongside variability, there is an interesting tendency in the stability of some bundles. For example, the bundle **by way of derogation** is always translated as *nukrypstant.* Bundles, which make part of the titles of documents, have shown hardly any variability either, e.g.

(22) (...) *on the **approximation of the laws** of the Member States* (...) [Reg_EN4]
      (...) *dėl valstybių narių **įstatymų*** (...) ***derinimo**.* [Reg_LT4]

Presumably, the titles have been rather carefully edited and finally, been standardised by linguists. Another example of a bundle which is uniformly rendered throughout the whole corpus is **binding in its entirety** translated as *privalomas visas.* The full formula is given below:

(23) *This Regulation shall be **binding in its entirety** and directly applicable in all Member States*. [Reg_EN1]
      *Šis reglamentas yra **privalomas visas** ir tiesiogiai taikomas visose valstybėse narėse.* [Reg_LT1]

So far we have discussed cases when the Lithuanian translation was shorter than the original English text. However, there have been several four-word bundles resulting in lengthier translations. One rather typical case is concerned with the translation of phrases containing alphanumeric article codes, which, following the Lithuanian legal tradition, are rendered in much lengthier phrases, employing full words rather than codes and abbreviations, e.g.:

(24) (...) ***provided for in Article 17(1)(b)*** (...) [Reg_EN5]
      (...) ***pagal 17 straipsnio 1 dalies b punktą*** (...) [Reg_LT5]

Such bundles were excluded from the count because alphanumeric article codes were not considered lexical elements within the structure of a four-word lexical bundle in the present study. In our corpus, only five bundles occurring in 26 utterances were rendered in Lithuanian as four-word bundles. Most of them employed very explicit clauses or phrases, e.g.:

(25) ***In order to ensure*** *harmonised scientific assessment of these claims* (...) [Reg_EN4]
      Tam, kad būtų užtikrintas *suderintas mokslinis šių teiginių vertinimas* (...) [Reg_LT4]

(26) ***Directive as last amended*** *by the 2003 Act of Accession* (...) [Reg_EN4]
Direktyva su paskutiniais pakeitimais*, padarytais 2003 m. Stojimo aktu* (...)
[Reg_LT4]

(27) *Directive as last **amended by Commission Directive** 2003/120/EC* (...) [Reg_EN2]
*Direktyva su paskutiniais **pakeitimais, padarytais Komisijos direktyva** 2003/120/EB*
(...) [Reg_LT2]

(28) (...) ***in particular Article 95 thereof*** (...) [Reg_EN1]
(...) ***ypač į jos 95 straipsnį*** (...) [Reg_LT1]

(29) (...) *having regard to the **provisions of Article 8 thereof**.* [Reg_EN2]
(...) *atsižvelgiant į **minėto sprendimo 8 straipsnio nuostatas**.* [Reg_LT2]

The above bundles are rather well-established in the EU documents. The length of some of them, for example, (28) and (29), are determined by the structure of Lithuanian, since the word *thereof* does not have a single word equivalent and has to be translated considering the reference words in context. They are usually restored in the form of nouns or pronouns. The length of the phrases in (25) and (26) can be accounted for by the genre and register of the text. The word *amend* in documents is usually rendered with the help of the phrase *daryti pakeitimus* 'to make amendments' rather than the verb *pakeisti* 'to amend, change, modify', which is much broader in its semantics. Its usage in this type of text might result in ambiguity, which in a legal document should be avoided.

Phrasal translations of four-word lexical bundles in Lithuanian have resulted in most cases in participial and nominal phrases and dependent clauses, often shorter than the original bundle. The translated bundles have demonstrated both extensive variability and stringent formulaicity in rendering this type of lexical bundles into Lithuanian.

4.3. Omission

The study identified five lexical bundles that in 52 cases were not rendered at all. The lexical bundles in examples (30)–(34) are given in ascending order of omission ratio: ***in accordance with the*** (omission ratio: 23 out of 83, or 27.71 %), ***as part of the*** (omission ratio: 4 out of 9, or 44.44 %), ***in the case of*** (omission ratio: 14 out of 30, or 46.67 %), ***in the context of*** (omission ratio: 5 out of 10, or 50 %) and ***for the purposes of*** (omission ratio: 6 out of 7, or 85.71 %). Consider the following examples:

(30)   *In accordance with the procedure referred to in Article 24(2), a decision shall be taken* (...) [Reg_EN4]
Sprendimas priimamas 24 straipsnio 2 dalyje nurodyta tvarka [order-INSTR.SG] (...) [Reg_LT4]

(31)   (...) *per quantified portion **as part of the** meal.* [Reg_EN3]
(...) *kiekybiškai išreikštoje porcijoje* (...) [Reg_LT3]

(32)   ***In the case of** non-prepacked food, the provisions of Article 41 shall apply.* [Reg_EN2]
Nefasuotiems maisto produktams taikomos 41 straipsnio nuostatos. [Reg_LT2]

Interestingly, the lexical bundle ***in accordance with the*** (30) has been translated in a number of ways including prepositions and nouns. The strategy of omission added to numerous other strategies gives a clue to its high variability. The same applies to the lexical bundle ***in the case of*** (32), discussed in previous sections of the paper. Another example (33) might again be an indication of delexicalisation and semantic bleaching, since the loss of the lexical bundle does not seem to affect the quality of the translated text:

(33)   (...) *individual products may have an important role **in the context of** an overall diet.* [Reg_EN4]
(...) *atskiri maisto produktai gali vaidinti svarbų vaidmenį bendroje mityboje.* [Reg_LT4]

Arguably, the bundle ***for the purposes of*** in example (34) is slightly different. The rendering is simplified and some content is lost. The phrase seems to be employed for the purpose of emphasis and should have been preserved:

(34)   ***For the purposes of** this Regulation* (...) [Reg_EN2]
Šiame reglamente [lit. 'in this regulation'] (...) [Reg_LT2]

The omission evidenced by the above examples might have been due to different reasons, including a decision of a translator. At least some of the five cases above demonstrate several degrees of delexicalisation. The bundle ***in the context of*** could be called empty and the bundle ***in the case of*** in this respect is very similar despite the fact that its translation still demonstrates high variability. The semantic contribution to the overall text of ***in accordance with the***, ***as part of the*** and ***for the purposes of*** seems to be rather varied. It could have influenced the choice of translation strategy. Interestingly, the omission of ***in accordance with the*** only occurred when the lexical bundle was used before one of two other lexical bundles, namely ***procedure referred to in*** or ***procedure laid down in***. The translation in (30) could also be interpreted

as preserved in the instrumental case of the word *tvarka* 'order, procedure', which expresses a broadly understood idea of instrumentality when Article 24(2) is used to make a decision.

5. CONCLUSION

The aim of the paper was to identify translation tendencies of non-terminological four-word lexical bundles from English into Lithuanian. The selected phrases have a lower likelihood to appear in terminological databases and, therefore, are much less likely to be located by professional translators. The study revealed some structural and semantic aspects of bundle translation.

The four-word lexical bundles tend to be translated as single lexical words, mainly as gerunds, prepositions and nouns, also as phrases, dependent clauses or omitted altogether. The overall tendency is to choose a shorter way of expression in terms of the number of words, which may be due to an entirely different structure of the target language. One exception in this respect is the translation of phrases containing alphanumeric article codes, which in Lithuanian are rendered in full words. Another tendency is concerned with a much higher percentage of verb-based expressions in Lithuanian (rather than noun-based in English), which confirms a previously established tendency.

In terms of translation variability, the analysed lexical bundles in translation have demonstrated two tendencies. One of them is concerned with a high degree of variability, sometimes resulting in four or five different versions of rendering one and the same phrase, such as ***in accordance with the***, ***in the case of***. It might have been due to a lack of agreement between the translators and/or lack of stability of the phrase in Lithuanian, also, presumably, to a much shorter tradition of legal translation into Lithuanian. The other tendency is to keep to one stable equivalent. The latter signals an apparent agreement among the translators, which might be due to tendencies of language standardisation.

It should be noted, however, that some variability might have been due to the choice of an individual translator, which was not investigated in the present study. A psycholinguistic experiment might be instrumental in uncovering strategies chosen by individual translators and the constraints imposed by the target language. Further research could focus on some extremely variable phrases and investigate their frequency and semantics in a larger and thematically less restricted corpus.

Data sources

EUR-Lex. 2016. [Online.] Available from: http://eurlex.europa.eu/homepage.html, [accessed on 21 November 2016].

Reg_EN1: European Commission. 2007. *Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EC) No 1924/2006 on nutrition and health claims made on foods*, available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52007PC0368&qid=1495529325456&from=EN [accessed on 21 November 2016].

Reg_EN2: European Commission. 2008. *Proposal for a Regulation of the European Parliament and of the Council on the provision of food information to consumers {SEC(2008) 92} {SEC(2008) 93} {SEC(2008) 94} {SEC(2008) 95}*, available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52008PC0040&qid=1495529770659&from=EN [accessed on 21 November 2016].

Reg_EN3: European Commission. 2012. *Commission Regulation (EU) No 432/2012 of 16 May 2012 establishing a list of permitted health claims made on foods, other than those referring to the reduction of disease risk and to children's development and health Text with EEA relevance*, available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012R0432&qid=1495530036382&from=EN [accessed on 21 November 2016].

Reg_EN4: European Parliament and Council of the European Union. 2006. *Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods*, available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R1924&from=EN [accessed on 21 November 2016].

Reg_EN5: European Parliament and Council of the European Union. 2006. *Regulation (EC) No 1925/2006 of the European Parliament and of the Council of 20 December 2006 on the addition of vitamins and minerals and of certain other substances to foods*, available from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R1925&from=EN [accessed on 29 November 2016].

Reg_LT1: Europos Komisija. 2007. *Pasiūlymas Europos parlamento ir Tarybos reglamentas iš dalies keičiantis Reglamentą (EB) Nr. 1924/2006 dėl teiginių apie maisto produktų maistingumą ir sveikatingumą*, available from: http://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:52007PC0368&qid=1495529325456&from=EN [accessed on 21 November 2016]

Reg_LT2: Europos Komisija. 2008. *Pasiūlymas Europos parlamento ir Tarybos reglamentas dėl informacijos apie maistą teikimo vartotojams {SEK(2008) 92} {SEK(2008) 93} {SEK(2008) 94} {SEK(2008) 95}*, available from: http://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:52008PC0040&qid=1495529770659&from=EN [accessed on 21 November 2016].

Reg_LT3: Europos Komisija. 2012. *2012 m. gegužės 16 d. Komisijos reglamentas (ES) Nr. 432/2012 dėl tam tikrų leidžiamų vartoti teiginių apie maisto produktų sveikumą, išskyrus teiginius apie susirgimo rizikos mažinimą, vaikų vystymąsi ir sveikatą, sąrašo sudarymo Tekstas svarbus EEE*, available from: http://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:32012R0432&qid=1495530036382&from=EN [accessed on 21 November 2016].

Reg_LT4: Europos Parlamentas, Europos Sąjungos Taryba. 2006. *2006 m. gruodžio 20 d. Europos Parlamento ir Tarybos reglamentą (EB) Nr. 1924/2006 dėl teiginių apie maisto produktų maistingumą ir sveikatingumą*, available from: http://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:32006R1924&from=EN [accessed on 21 November 2016].

Reg_LT5: Europos Parlamentas, Europos Sąjungos Taryba. 2006. *2006 m. gruodžio 20 d. Europos Parlamento ir Tarybos reglamentas (EB) Nr. 1925/2006 dėl maisto produktų papildymo vitaminais ir mineralais bei tam tikromis kitomis medžiagomis*, available from: http://eur-lex.europa.eu/legal-content/LT/TXT/PDF/?uri=CELEX:32006R1925&from=LT [accessed on 29 November 2016].

References

Ambrazas V., Geniušienė E., Girdenis A., Sližienė N., Tekorienė D., Valeckienė A., Valiulytė E. 1997. *Lithuanian Grammar*. Vilnius: Baltos lankos.

Arkadiev P. 2012. *"External" Verbal Prefixes in Lithuanian*. Conference presentation at *Scuola Normale Superiore di Pisa*.

Biber D., Conrad S., Cortes V. 2004. If You Look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25 (3), 371–405.

Biber D., Conrad S. 2004. The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica* 20, 56–71.

Biber D., Johansson S., Leech G., Conrad S., Finegan E. 1999. *The Longman Grammar of Spoken and Written English*. Harlow (Essex): Longman.

Biel L., Engberg J. 2013. Research Models and Methods in Legal Translation. *Linguistica Antverpensia, New Series – Themes in Translation Studies* 12, 1–11, available from: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/316/225 [accessed on 7 February 2017].

Conklin K., Schmitt N. 2008. Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics* 29 (1), 72–89.

Erman B., Warren B. 2000. The Idiom Principle and the Open Choice Principle. *Text* 20 (1), 29–62.

Firth J. R. 1957. *Studies in Linguistic Analysis*. Oxford: Blackwell.

Foster P. 2001. Rules and Routines: A Consideration of Their Role in the Task Based Language Production of Native and Non-native Speakers. Bygate M., Skehan P., Swain M. (Eds.). *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Harlow: Longman, 75–93.

Grabowski L. 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* 25, 391–401.

Huddleston R., Pullum G. K. 2007. *A Student's Introduction to English Grammar*. New York: Cambridge University Press.

Hyland K. 2008. As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes* 27 (1), 4–21.

Jablonkai R. 2010. English in the Context of European Integration: A Corpus-driven Analysis of Lexical Bundles in English EU Documents. *English for Specific Purposes* 29 (4), 253–267.

Jiang N., Nekrasova T. M. 2007. The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal* 91 (3), 433–445.

Juknevičienė R. 2011. *Leksinės samplaikos svetimkalbių ir gimtakalbių vartotojų rašytinėje anglų kalboje*. Daktaro disertacija. Vilnius: Vilniaus universitetas.

Laurence A. 2013. *AntPConc (Version 1.0.2)*. [Computer software]. Tokyo: Waseda University, available from: http://www.laurenceanthony.net/software/antpconc/ [accessed on 23 November 2016].

Laurence A. 2014. *AntConc (Version 3.4.1w)*. [Computer software]. Tokyo: Waseda University, available from: http://www.laurenceanthony.net/software/antconc/ [accessed on 23 November 2016].

Marcinkevičienė R. 2010. *Lietuvių kalbos kolokacijos*. Kaunas: Vytauto Didžiojo universiteto leidykla.

Martinez R., Schmitt N. 2012. A Phrasal Expressions List. *Applied Linguistics* 33 (3), 299–320.

Murphy M. L. 2010. *Lexical Meaning*. Cambridge: Cambridge University Press.

OD (Oxford Dictionaries). 2017. [Dictionary online]. Oxford: Oxford University Press, available from: https://en.oxforddictionaries.com/ [accessed on 22 February 2017].

Pažūsis L. 2014. *Kalba ir vertimas*. Vilnius: Vilnius universiteto leidykla.

Shrefler N. 2011. Lexical Bundles and German Bibles. *Literary and Linguistic Computing* 26 (1), 89–106.

Simpson-Vlach R., Ellis N. C. 2010. An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics* 31 (4), 487–512.

Sinclair J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair J. 2004. *Trust the Text: Language, Corpus and Discourse.* London and New York: Routledge.

Siyanova-Chanturia A., Martinez R. 2015. The Idiom Principle Revisited. *Applied Linguistics* 36 (5), 549–569.

Tiersma P. 2006. Some Myths about Legal Language. *Journal of Law, Culture and Humanities* 2 (1): 29–50.

Tognini-Bonelli E. 2000. Corpus Classroom Currency. *Darbai ir dienos* 24, 205–243.

Trebits A. 2008. English Lexis in the Documents of the European Union – A Corpus-based Exploratory Study. *Working Papers in Language Pedagogy* 2, 38–54, available from: http://langped.elte.hu/WoPaLParticles/W2Trebits.pdf [accessed on 20 January 2016].

Trebits A. 2009a. The Most Frequent Phrasal Verbs in English Language EU Documents – A Corpus-based Analysis and Its implications. *System* 37, 470–481.

Trebits A. 2009b. Conjunctive Cohesion in English Language EU Documents – A Corpus-based Analysis and Its Implications. *English for Specific Purposes* 28, 199–210.

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

ES REGLAMENTAI: LEKSINIŲ SAMPLAIKŲ VERTIMO IŠ ANGLŲ Į LIETUVIŲ KALBĄ TENDENCIJOS

Mantas Noreika, Inesa Šeškauskienė

Santrauka

Remiantis Biber et al. (2004) leksinių samplaikų tyrimo įžvalgomis ir Sinclair (1991) pastebėjimais, susijusiais su sustabarėjusios kalbos (angl. *formulaic language*) semantiniu nedalomumu (angl. *compositionality of meaning*) ir kai kurių vienetų deleksikalizacija, šiame tyrime nagrinėjamos keturių žodžių ilgio leksinės samplaikos, randamos ES teisės diskurse anglų kalba, daug dėmesio skiriant leksinių samplaikų semantinių ir struktūrinių ypatybių išlaikymui verčiant į lietuvių kalbą. Tyrime nagrinėjamos 43 leksinės samplaikos, išskirtos iš 59 579 žodžių apimties tyrimui sudaryto ES teisės anglų kalbos tekstyno, susidedančio iš penkių reglamentų, susijusių su maisto produktų ženklinimu ir ant maisto produktų leidžiamais vartoti teiginiais, tekstų. Tiriant nustatyta, kad leksinės samplaikos gali būti verčiamos į lietuvių kalbą pasitelkiant frazes ir šalutinius sakinius, pavienius leksinius žodžius, pavienius funkcinius žodžius, gramatinę linksnio kategoriją arba visai praleidžiamos vertime. Verčiant leksines samplaikas linkstama neilginti teksto. Dažniau pasitaiko trumpesni atitikmenys. Kai kurių samplaikų vertimas itin įvairus – jos verčiamos vienu daiktavardžiu, prielinksniu, fraze arba praleidžiamos. Labiau linkstama rinktis veiksmažodines, o ne daiktavardines frazes. Kelios samplaikos turi lietuvių kalboje labai stabilius atitikmenis. Šios tendencijos susijusios su santykinai neilga teisės kalbos vertimo į lietuvių kalbą tradicija, samplaikų variantiškumu originalo kalboje, ryškia kai kurių iš jų deleksikalizacija ir, manytina, individualiu vertėjo pasirinkimu.