ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА
THEORY AND PRACTICE OF TRANSLATION

STANCE BUNDLES IN ENGLISH-TO-POLISH TRANSLATION: A CORPUS-INFORMED STUDY

ŁUKASZ GRABOWSKI

University of Opole
11a pl. Kopernika, 45-040 Opole, Poland

Abstract

In this paper, we make an attempt to improve the textual fit of English-to-Polish translation of a peculiar type of multi-word units known in corpus linguistic literature as lexical bundles (Biber et al. 1999). Inspired by a study conducted by Grabar and Lefer (2015), we used the English-Polish parallel corpus Paralela (Pęzik 2016) and the National Corpus of Polish (NKJP) to extract and explore the use — in terms of frequency distributions — of the Polish equivalents of selected English lexical bundles expressing attitudinal and epistemic stance. More precisely, we used the NKJP corpus to check whether the Polish equivalents are typical of contemporary Polish as found in native texts. The results of this corpus-informed study revealed a high number of Polish equivalents, both single- and multi-word units, expressing stance. Also, the results showed that the majority of Polish equivalents are frequently used in native Polish texts and therefore they can potentially help enhance the textual fit of translations. Finally, we discussed limitations of the methods and corpora used in this preliminary study and presented suggestions on how it can be pursued further in the future to better explore the usefulness of lexical bundles for translation teaching and translation practice. To that end, we also presented proposals of in-class translation activities.

Keywords: *corpus linguistics, lexical bundles, English-to-Polish translation, parallel corpus, textual fit*

1. INTRODUCTION

When we read translations, be it literary novels, user manuals, press articles or otherwise, we are sometimes under the impression that the text sounds somewhat unnatural or reads with difficulty. This impression of ours is largely based on the linguistic intuition of native speakers of a given target language, which, in turn, is contingent upon our prior experience (i.e. reading and/or writing) with native non-translational texts. In that respect, the linguistic intuition is largely determined by the memory of contexts, both linguistic and extra-linguistic ones, in which words or expressions were used in the past (Piotrowski 2011: 50). In a similar vein, Hoey (2005, 2007) argues that linguistic intuition of language users represents accumulation of their prior linguistic experience[1].

---

[1] More specifically, Hoey (2005: 13) claims that a "word is primed for use in discourse as a result of the cumulative effects of individual encounters with the word", a statement which is an essence of his lexical priming theory. Accordingly, when we use a particular word, we primed to use it again in the future with its typical collocations, in its typical grammatical function, in similar semantic and pragmatic contexts, in the domains, registers and genres associated with it as well as in familiar social contexts (Hoey 2005: 13; 2007: 8).

A clash between linguistic intuition and linguistic properties of texts may arise in the case of translations, which — by their very nature — are produced under different constraints than native texts, e.g. interference from the language of the original, standardization to the norms of a target language (Toury 1995), to name but a few factors[2]. For example, if a native speaker of Polish is confronted with a choice between two alternative equivalents of the English sentence *I am not at home*, he or she will most probably argue that *Nie ma mnie w domu* sounds more natural in Polish than *Nie jestem w domu*, a calque of the English sentence, which is ungrammatical in Polish. At this point, one may also refer to the concept of text's naturalness, which can be described — capitalizing on the definition proposed by Lewandowska-Tomaszczyk (2012: 34) — as a system of language users' preferences of the use of linguistic items measured by their frequency of occurrence in a particular context. Hence, the use of language corpora providing access to information on the frequency of use of particular linguistic items in a given context and co-text offers a more objective way, notably when compared with linguistic intuition alone, to capture and measure linguistic preferences of language users as well as the text's naturalness. According to Pérez-Paredes (2010: 157), "we can all too easily, maybe too 'easy', make generalizations about language use based on our perceptions or personal experiences, contact with a language or plain introspection." Also, given the fact that native texts and translations are produced under different circumstances, linguistic preferences of translators, in particular those rendering texts into a target language which is not their native one, may not always coincide with the native speakers' preferences.

In view of the above considerations, we can assume that when a translation sounds somewhat unnatural, idiosyncratic or reads with difficulty (due to excessive lexical or syntactic calques, simplification of syntactic structures, overuse or underuse of certain grammatical structures or prefabricated formulas etc.), it most probably does not fit the norms and conventions (grammatical, stylistic, generic etc.) of the target language. Accordingly, such a translation may not closely resemble native texts (i.e. non-translations) originally written in the target language. That being so, one may observe certain linguistic distance or dis/similarity of translated texts to non-translated texts, a hypothesis known in literature on the so-called translation universals as the textual fit hypothesis (Chesterman 2004: 6). It accounts for the relation of acceptability of a text or its fitting into the family of non-translated native texts in the target language, e.g. whether lexical, grammatical or stylistic profile of a translation from source language and culture into target language fits into the corresponding profile of non-translated texts in the target language, which function in the target culture (Chesterman 2004: 6). As argued by Kranich (2016:10), apart from cultural differences and interference from the source language "a tendency to 'say what seems normal or safe' should be also kept in mind as a potential explanation for differing behaviour of translated texts compared to the source and target language originals".

The interest in corpus linguistic research on the textual fit hypothesis has intensified in recent years. For example, Biel (2014a, 2014b) explored the textual fit of EU law

---

[2] Interference and standardization are two laws of translation described by Toury (1995, 2001).

translated into Polish as compared with non-translated Polish law. From a cross-linguistic perspective, Grabowski (2018a) made an attempt to use a custom-designed comparable corpus of English and Polish patient information leaflets (i.e. non-translated English and Polish texts) to extract lexical bundles of similar discourse functions (referential, discourse-organizing and expressing stance), which may help enhance the textual fit of translated texts[3]. In practice, the textual fit hypothesis implies that what is important for the translator to take account of when performing a translation task is adherence to discourse norms and conventions of text production in the target language and culture, which also includes expectations of the receiver of a translated text.

In this preliminary corpus-informed study[4], an attempt is made at enhancing the textual fit of translated texts taking the translation of recurrent multi-word units as a case in point. Also, the study aims to verify whether the results of descriptive research — conducted using corpus linguistic methodology — on the so-called lexical bundles (Biber et al. 1999), a peculiar type of recurrent multi-word units, may be turned into actionable knowledge useful for practitioners of English-to-Polish translation. A similar attempt, which provided motivation to undertake a study like this one, was made by Grabar and Lefer (2015), who focused on English-to-French translation of lexical bundles found in the transcripts of debates held at the European Parliament.

2. RECURRENT MULTI-WORD SEQUENCES
AS A PROBLEM IN TRANSLATION

Before the scope and methodology used in this preliminary study is described, it is necessary to justify why the emphasis is put in this paper on translation of recurrent multi-word units (henceforth MWUs). Generally speaking, MWUs pose a plethora of problems in translation, in particular machine translation and computer-assisted translation. As for machine translation, the main problems refer to varying degrees of fixedness, pattern variability, syntactic flexibility (positional and constituency variation), and semantic compositionality of MWUs (Sag et al. 2002; Bouamor et al. 2012; Barreiro & Batista 2016; Skadina 2016). For example, it may happen that the same sequence or combination of words (e.g. *умывать руки*, which literally means 'to wash one's hands') may convey different meanings in different contexts of language use[5], e.g.:

> *Нарезаю лососину на ошметки (...), мажу чуть-чуть васаби с одной стороны (...) и приделываю сверху на бобышки — так, чтобы васаби оказался между лососем и рисом. Получаются суси с лососем.* **Умываю руки.** *Теперь — роллы с лососем. Рис готовится так же, как и для суси.* [Александр Черных. Москва — Токио (2004) // «Хулиган», 2004.08.15]

---

[3] The research conducted by Grabowski (2018a, forth.), which provided another motivation to pursue a study like this one, revealed that genre-specific stylistic conventions also determine the choice of the most natural (acceptable) translation, e.g. it was found that English patient information leaflets are written in a less formal, plain style as compared with their Polish counterparts.

[4] This preliminary descriptive study is described as corpus-informed one (or corpus-illustrated one), as the use of corpora (parallel and monolingual ones) is largely limited to quoting examples of language use (Górski 2012: 292; Lee 2008: 88).

[5] Similar example: *дать прикурить* etc.

ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА

*Петр Антонович пожал плечами. — Если вы настаиваете на своем, — сказал он, — то я **умываю руки**, и слагаю с себя ответственность за возможные последствия* [Ф.К. Сологуб. Турандина (1912)][6].

In the examples presented above, the word combination under scrutiny, namely *умываю руки*, should be translated differently into Polish depending on its sense, which emerges from the context of its use. In the first example, *умываю руки* should be translated into Polish as *myję ręce* (used in the context of washing one's hands, i.e. similar to *умывать лицо* 'wash one's face') while in the second one an acceptable translation should be *umywam ręce* (used to communicate that one accepts no responsibility for something). However, for the reasons described above, machine translation systems often fail to make such sense distinctions, as it is illustrated by the data extracted from Google Translate (as of 14 December 2017), e.g.:

*Kroję łososia na strzępy (...), smaruję odrobinę wasabi po jednej stronie (...) i przymocowuję go na wierzchu łapek — tak, aby wasabi było splecione z ryżem. Zdobyte sushi z łososiem. **Myję ręce**. Teraz — bułki z łososiem. Ryż jest przygotowywany w taki sam sposób, jak w przypadku sushi.* [Alexander Chernykh. Moskwa — Tokio (2004) // "The Hooligan", 2004.08.15]

*Piotr Antonowicz wzruszył ramionami. "Jeśli nalegasz na własną rękę", powiedział, "wtedy **myję ręce** i rezygnuję z odpowiedzialności za możliwe konsekwencje".* [Ф.К. Сологуб. Турандина (1912)]

Also, MWUs pose challenges for computer-assisted translation tools (the so-called CATs), which process texts as sequences of words divided by spaces or punctuation signs. That is why such tools fail to perform text segmentation in a way sequences of words are mapped with particular meanings (senses). In other words, as text segmentation is based on text's orthography or punctuation, a translation unit is usually a sentence or clause rather than a multi-word unit constituting a readily available form-and-meaning mapping[7].

Another closely related problem is described by Piotrowski (1994: 104), who argues that in translation one can hardly speak of a stable translation unit. It is often the case that words or MWUs, which are more or less stable across source-language texts, can be or must be translated using target-language words or expressions at different levels of language organization[8], a change in translation as compared with the original referred to by Catford (1965: 76) as a unit shift, e.g. Eng. *there is no doubt that* vs. Pol. *niewątpliwie*. Also, MWUs may convey different pragmatic meanings depending on the context of language use, e.g. a Polish noun phrase *zły pies* 'bad dog' can be translated into English as *bad dog* if used in a narrative text, or *Beware of the dog!* if used as

---

[6] The examples were extracted from the Russian National Corpus (www.ruscorpora.ru).

[7] More detailed discussion on the limitations of CAT tools can be found in Kornacki (2017).

[8] The original reads: „trudność polega na tym, że w rzeczywistych tłumaczeniach nie istnieje stabilna jednostka przekładu (...) zaś wyrażenia w języku źródłowym, stałe w rozmaitych tekstach w tymże języku, mogą bądź muszą być tłumaczone za pomocą wyrażeń języka docelowego na innym poziomie" (Piotrowski 1994: 104).

a warning nailed to a gate or fence (Grabowski 2018a: 182, forth.). As with any linguistic form, be it a single-word or multi-word unit, its pragmatic meaning emerges from a situation of language use, e.g. from particular speech acts. That being so, the very identification of pragmatic meanings of MWUs largely determines the choice of the most natural and acceptable translation in a given context. Finally, it goes without saying that recurrent MWUs may differ with respect to their length, frequency and distribution in texts produced in typologically different languages (cf. Granger 2014; Grabowski 2014, 2018a).

This paper focuses on the translation of a particular type of recurrent MWUs known in corpus linguistic literature as lexical bundles (Biber et al. 1999), e.g. *I don't think*, *as a result*, *the nature of the*, *when it comes to, it is important to, it is clear that*. In short, lexical bundles (henceforth LBs) are extracted from texts based on their length, frequency and distribution. In essence, they perform specific textual or discourse functions (e.g. referential, discourse-organizing, expressing stance) across the whole variety of text types, genres or specialist domains of language use (Biber et al. 2004; Hyland 2008; Biber 2006, 2009; Goźdź-Roszkowski 2011; Breeze 2013; Salazar 2014; Grabowski 2015, 2018b; Fuster-Marquez 2017; McVeigh 2018). In short, those studies provide evidence that the number, distribution, structure and functions of LBs vary across spoken and written registers according to many factors related to situational contexts and communicative functions, such as topic, setting, participants, relations among participants, production circumstances, communicative purposes etc. (Biber & Conrad 2009: 37—47). However, most research studies on LBs have been conducted using English-language material and they are largely descriptive. An overarching aim of those studies, which are predominantly targeted at teaching English in various academic contexts, is to describe and later isolate those MWUs which are potentially the most pedagogically useful (e.g. Simpson-Vlach & Ellis 2010; Martinez & Schmitt 2012; Salazar 2014). One may also note the scarcity of cross-linguistic studies focusing on recurrent n-grams or LBs, with the notable exceptions of Forchini and Murphy (2008), Granger (2014), Oksefjell Ebeling and Ebeling (2016), Biel (2017), Berūkštienė (2017), Grabowski (2018a) or Grabar and Lefer (2015). Approaching those peculiar MWUs from the perspective of translation, the last-mentioned study is targeted at identification of LBs in English and French EU parliamentary debates in order to develop bilingual lexicons to be further used in computer-assisted translation tools or machine translation tools. In a similar vein, Berūkštienė (2017) explored how different structural types of LBs found in English court judgments were rendered into Lithuanian. The rationale behind those cross-linguistic studies is the assumption that LBs, which represent recurrent and reproducible MWUs in a given source language, should have more or less regular equivalents in other languages (Jukneviciene 2017: 63).

An observation made by Grabar and Lefer (2015), who argue that terminological databases used by translators rarely, if ever, include MWUs that express writer's stance or structure texts, provided motivation to undertake a study like this one. The following section describes the methodology, research material and goals of this study.

ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА

3. METHODS

Likewise in Grabar and Lefer (2015), the general aim of this preliminary study is to verify the usefulness of LBs for translation purposes. More precisely, following selected elements of the methodology used by Grabar and Lefer (2015), we aim to explore whether LBs may be used to improve naturalness — in this study operationalized as the textual fit — of English-to-Polish translation of selected LBs expressing stance and found in the EU parliamentary debates. As mentioned earlier, a unit of analysis used in this paper are LBs expressing attitudinal stance, i.e. the speaker's subjective feelings, emotions, attitudes, value judgments or assessments of the following proposition, and epistemic stance, i.e. the speaker's expression of certainty, doubts, reliability or limitations of the following proposition (Biber et al. 1999: 966; Biber 2006: 139; Mindt 2011: 74; Gray & Biber 2013)[9]. Capitalizing on the results of the study conducted by Grabar and Lefer (2015), who identified a high number of stance LBs in EU parliamentary debates[10] in English and aligned them with their French equivalents, in this paper we want to explore, first, how four stance bundles[11] (*it is not surprising that, it would be wrong to, there is no doubt that, it may well be*) were translated into Polish and, second, whether the Polish equivalents are at the same time typical of the Polish language (i.e. whether they are the ones that enhance the textual fit of translations as compared with native texts produced originally in Polish). Employed to strengthen or weaken the force of the following proposition, epistemic stance LBs (*there is no doubt that*, *it may well be*) can be said to pragmatically function as boosters or hedgers[12]. As for attitudinal stance LBs (*it is not surprising that, it would be wrong to*), which are used to subjectively evaluate or assess the content of the following proposition, they may help persuade someone into accepting the speaker's interpretation of information conveyed in the text or his/her point of view. Hence, the study results may also offer cursory insight into pragmatic preferences in English and Polish as regards the linguistic expression of stance.

As a research material, we will use two corpora: a parallel one and a monolingual one. More specifically, in order to identify Polish equivalents of the four aforementioned English stance bundles, we will use Paralela corpus (Pęzik 2016), an English-Polish and Polish-English parallel corpus. Currently, the corpus includes 262 million words in 10,877,000 translation segments found predominantly in legal texts (European Union legislation, proceedings of the European Parliament etc.), press releases, medical texts (provided by the European Medicine Agency) as well as film subtitles (Pęzik 2016: 68). The English and Polish translation segments are aligned at the sentence level (Pęzik 2016: 70), with 5.3% of the segments aligned manually. The size of the sub-corpus of the European Parliament proceedings (EPP) is 13,026,414 words stored in 693,139

---

[9] The division between attitudinal and epistemic stance is also described by Jalali (2017: 31).

[10] The transcripts of EU parliamentary debates were extracted from the Europarl corpus (Koehn 2005).

[11] The LBs under scrutiny were randomly selected from the ones identified in the English Europarl corpus by Grabar and Lefer (2015).

[12] In a similar vein, Kranich (2016: 95) argues that expressions of epistemic modality can perform the function of hedgers.

translation segments. Recorded on 11th, 12th and 23rd October 2006, the debates were originally translated from English into Polish.

Having identified Polish target language equivalents of the English LBs under scrutiny, the monolingual corpus of Polish will be used to verify the status — in terms of the frequency of use — of the equivalents as they are used in native texts originally written in Polish. The selection of the reference corpus is not devoid of methodological problems. Ideally, one should employ a corpus representing the same genre, e.g. a collection of debates held in the Polish parliament. However, such a corpus is not readily available to researchers. That is why a decision has been made to use a balanced sub-corpus of the National Corpus of Polish (NKJP), which includes 240,192,461 words found in texts published after the year 1945 and represents the whole variety of text types and genres, both written and spoken. In fact, 10% of the texts represent spoken language, including parliamentary debates held in the Polish parliament (Pęzik 2012: 39). However, throughout validation of the target-language equivalents the frequencies obtained from the spoken language component of NKJP were found to be too low to arrive at any definite conclusions. That is why we decided to use the entire balanced sub-corpus of NKJP, also in view of the fact that in terms of their use both source- and target-language equivalents are not restricted to spoken texts[13]. Another limitation of the procedure adopted in this study is that the target language equivalents were searched for in their exact form, which follows that any variation within MWUs was ignored.

In the following section, the results of the quantitative and qualitative analysis will be presented. They will provide an insight into the equivalent Polish lexical items, be it single-words or MWUs, expressing attitudinal and epistemic stance. It is believed that LBs, which are recurrent MWUs typical of specialist discourses, text types or registers and which perform specific discourse functions, can be used as a starting point in the search for target language equivalents. The search will be conducted through close reading of parallel concordances and manual identification of equivalent pairs of translation units. Later, in order to identify the most salient equivalents, i.e. the most typical of contemporary Polish, the frequency of the Polish translation units will be verified against native texts collected in the NKJP corpus.

4. RESULTS

The first attitudinal stance bundle under scrutiny, namely *it is not surprising that*, occurs only 9 times in the EPP sub-corpus of Paralela. One may find there the following Polish equivalents: *nie może zaskakiwać, że; nie zaskakuje* [*(propozycja), by*]*; nie jest zaskoczeniem* [*(propozycja), by*]*; nie zaskakuje, że; nie jest zaskakujące, że*; *nic dziwnego, że; nie należy się dziwić, że; nie dziwi (fakt), że; nie dziwi (to), że*. In this particular case, there was no unit shift in the translation, i.e. the MWUs in the original were translated using MWUs in the target language. However, the manual verification of the Polish equivalents revealed that two items, namely *nie może zaskakiwać, że*

---

[13] According to Pęzik (2012: 39), transcripts of pre-written speeches are referred to in specialist literature as 'to-be-spoken' texts.

(0 occurrences *in* NKJP), *nie jest zaskakujące, że* (1 occurrence in NKJP) are very rare (or not used at all) in the National Corpus of Polish, e.g.:

(1) *Biorąc pod uwagę te niepewności, **nie jest zaskakujące, że** wielu z nich ma opory co do inwestowania, jak też zatrudniania nowych pracowników.* (IJP-PAN_p00009600946).

The most frequent Polish equivalent found in NKJP is *nic dziwnego, że* (2,744 occurrences), followed by *nie dziwi, że* (184 occurrences) and *nie należy się dziwić, że* (52 occurrences). That is why these target language equivalents can be considered to be more typical of contemporary Polish and hence they may help enhance the textual fit of translated texts. The examples of their use in the EPP sub-corpus of Paralela are presented below.

(2) ***It is not surprising that***, *at the end of this period, we have actually created the greatest productive power and the greatest degree of clarity in this period.*
***Nic dziwnego, że** pod koniec tego okresu rzeczywiście stworzyliśmy największy produktywny potencjał i przejrzystość najwyższego stopnia w tym czasie.*

(3) *(...) **it is not surprising that** the first full impact on the real economy of the crisis in the financial markets has hit the car market.*
***Nie dziwi to, że** oddziaływanie kryzysu finansowego na gospodarkę realną jest w pierwszej kolejności odczuwalne na rynku samochodowym.*

(4) *I should also say that Shen Yun promotes the philosophy of truthfulness, tolerance and compassion so **it is not surprising that** the Chinese Government and Communist Party fear that contrary ideology.*
*Trzeba też powiedzieć, że Shen Yun promuje filozofię prawdy, tolerancji i współczucia, więc **nie należy się dziwić, że** chiński rząd i partia komunistyczna obawiają się tej obcej sobie ideologii.*

The remaining Polish equivalents do not enhance the degree of textual fit to the same extent. The reason for that is that they occur in NKJP with considerably lower frequencies and in different lexical and grammatical contexts, e.g. *nie zaskakuje* [*(propozycja), by*] and *nie jest zaskoczeniem* [*(propozycja), by*] do not occur in the said corpus in the form of constructions such as '*nie zaskakuje/nie jest zaskoczeniem* + noun + *by*'. Interestingly, the expression *nie jest zaskoczeniem* occurs in NKJP 121 times, in most cases either in sentence-final position (e.g. *Opór ludowców **nie jest zaskoczeniem**.* (PELCRA_1303919931001)) or followed by conjunctions, such as *bo* or *gdyż* introducing explanations to information introduced earlier in the text (e.g. *Brak w tym gronie Unibaksu **nie jest zaskoczeniem, bo** żużlowcy jako spółka akcyjna dostaną wsparcie z funduszu promocji* (IJPPANp0006300176)). Finally, the expression *nie jest zaskoczeniem, że/iż* is used in NKJP 12 times only.

The next bundle subjected to the analysis, *it would be wrong to*, is found 17 times in the EPP sub-corpus of Paralela, and its two Polish equivalents, namely *błędem byłoby* and *byłoby błędem*, are the most frequent ones (10 occurrences in total), e.g.:

(5) *At the same time **it would be wrong to** compare the African Union with the European Union, because they are different types of Unions and we should not try to compare them one to one.*
*Równocześnie **błędem byłoby** porównywanie Unii Afrykańskiej z Unią Europejską, ponieważ są one różnymi rodzajami unii; nie powinniśmy więc porównywać ich ze sobą.*

(6)   In both cases, however, I think **it would be wrong to** break off the talks.
       *Sądzę jednak, że w obydwu przypadkach zrywanie rozmów **byłoby błędem**.*

Other Polish equivalents include *niewłaściwe byłoby* (1), *byłoby złym* (np. *posunięciem*) (1), *niesłuszne/niesłusznym byłoby* (2 occurrences), *byłoby niestosowne* (1 occurrence) *nie byłoby dobre* (1 occurrence) or *nie można* (2 occurrences), e.g.:

(7)   That is why **it would be wrong to** agree with him in this instance.
       *Dlatego też **nie można** zgodzić się z nim w tym względzie.*

(8)   It would be wrong to deny that.
       ***Niesłuszne byłoby** zaprzeczanie temu.*

(9)   In the rapporteur 's view, **it would be wrong to** miss this opportunity to ensure that this directive does more than supply a set of definitions.
       *W opinii sprawozdawczyni **nie byłoby dobre** przeoczenie możliwości zapewnienia przez tę dyrektywę czegoś więcej niż tylko zbioru definicji.*

However, the data found in the National Corpus of Polish show that the most frequent equivalents in Paralela (*błędem byłoby* and *byłoby błędem*) are at the same time the most typical of contemporary Polish (156 occurrences in NKJP). Other equivalents occur in the corpus with lower frequencies (*niewłaściwe byłoby* — 6 occurrences; *byłoby złym* — 11 occurrences; *niesłuszne/niesłusznym byłoby* — 12 occurrences; *nie byłoby dobre* (followed by gerunds — 4 occurrences). As for the impersonal construction with *nie można* followed by the infinitive, it occurs 7,283 times in NKJP in the whole variety of contexts ('must not', 'one cannot', 'it is not permitted to' etc.), i.e. not limited to *it would be wrong to* followed by the infinitive, as it is the case in the English original.

The third lexical bundle analyzed in this paper, *there is no doubt that,* is used in Paralela 181 times and its most frequent Polish equivalent is *nie ma wątpliwości, że* (75 occurrences in Paralela), e.g.:

(10)   **There is no doubt that** the damage to the Fukushima nuclear power plant is a disaster, but the final death toll will not be counted in thousands or hundreds, and perhaps not even in tens.
       ***Nie ma wątpliwości, że** szkody w elektrowni jądrowej w Fukushimie to katastrofa, ale ostatecznie ofiary nie będą liczone w tysiącach czy setkach, a być może nawet nie w dziesiątkach.*

(11)   **There is no doubt that** the US is a superpower, and its views, proposals and requests cannot be swept off the table just like that.
       ***Nie ma wątpliwości, że** Stany Zjednoczone to supermocarstwo oraz że poglądów, propozycji i żądań tego kraju nie można tak po prostu ignorować.*

Among other equivalents, one may find both MWUs and single-word units. The former ones include *nie ma wątpliwości co do tego, że* (10 occurrences), *nie ma żadnych wątpliwości, że* (1 occurrence), *nie ulega wątpliwości, że* (21 occurrences), *bez wątpienia* (28 occurrences), *co oczywiste* (1 occurrence), *z całą pewnością* (3 occurrences), *nie podlega wątpliwości* (1 occurrence), *brak wątpliwości co do tego, że* (1 occurrence), e.g.:

(12)   For example, **there is no doubt that** the Court of Justice, in particular, would use the accession to once again extend the EU 's powers.

> *Przykładowo **nie ma wątpliwości co do tego, że** w szczególności Trybunał Sprawiedliwości może wykorzystać przystąpienie do kolejnego rozszerzenia uprawnień UE.*

(13) **There is no doubt that** this is an EP own-initiative report that is highly relevant and topical.
Bez wątpienia, przedmiotowe sprawozdanie PE z inicjatywy własnej jest w wysokim stopniu trafne i rzeczowe.

The verification of the findings in the National Corpus of Polish revealed that the most frequent equivalent in the EPP subcorpus of Paralela is not necessarily the most typical one of contemporary Polish. More precisely, the most frequent expression in the NKJP is *nie ulega wątpliwości, że* (1,289 occurrences), followed by *nie ma wątpliwości, że* (953 occurrences), *nie ma żadnych wątpliwości, że* (53 occurrences), *nie ma wątpliwości co do tego, że* (48 occurrences), *nie podlega wątpliwości* (19 occurrences). The equivalent *brak wątpliwości co do tego, że* is not found in NKJP. Other equivalents, namely *bez wątpienia* 'without doubt' (3,866 occurrences in NKJP), *co oczywiste* 'obviously' (262 occurrences in NKJP)*, z całą pewnością* 'certainly' (3,463 occurrences in NKJP), represent interesting translational choices yet they can be also used as equivalents of other words or expressions.

As for the single-word items, adverbials such as *niewątpliwie* ('undoubtedly', 'doubtless') with 18 occurrences in Paralela (and 10,891 in NKJP), *oczywiście* ('of course', 'obviously') with 2 occurrences in Paralela and 86,424 in NKJP) and *niezaprzeczalnie* 'undeniably' (1 occurrence in Paralela and 103 in NKJP) account for all the three equivalents of *there is no doubt that*, e.g.:

(12) **There is no doubt that** cluster munitions are very cruel weapons systems which cause great suffering to civilians.
Niewątpliwie amunicja kasetowa należy do bardzo okrutnych typów broni, który powoduje ogromne cierpienia wśród ludności cywilnej.

(13) **There is no doubt that** the Commission is telling us that this will mean a reduction in bureaucracy.
Komisja oczywiście zapewnia nas, że zabieg ten ograniczy biurokrację.

(14) *This is a pity, because **there is no doubt that** science allows us to assess what influence economic changes have on the environment in the region.*
Szkoda, bo niezaprzeczalnie to nauka pozwala nam ocenić, jaki wpływ w tym rejonie wywierają zmiany ekonomiczne na środowisko.

A relatively high frequency of *niewątpliwie* ('undoubtedly') in both the EPP sub-corpus of Paralela and NKJP shows that it may also be treated as an acceptable translation equivalent of a MWU *there is no doubt that*, which is another example of the so-called unit shift (Catford 1965: 76).

Finally, the bundle *it may well be* occurs in the EPP subcorpus of Paralela 12 times with the following equivalents: *być może* (2 occurrences); *bardzo możliwe, że* (1 occurrence); *jest możliwe* (1 occurrence) *niewykluczone, że* (1 occurrence); *może się okazać, że* (1 occurrence); *równie dobrze* (2 occurrences); *może* (4 occurrences), e.g.:

(15) **It may well be** that I will then be among them.
Być może będę wtedy jedną z nich.

(16) *However, **it may well be** the case that tools such as XBRL tagging can develop that.*
***Może się** jednak **okazać**, że umożliwią to takie narzędzia jak format elektroniczny XBRL.*

(17) *We need to adopt a completely different approach to dismantling and, in my opinion, **it may well be** possible to induce the shipowners to do so, especially given all the negative publicity on this issue in recent years.*
*Musimy zająć zupełnie inne stanowisko wobec demontażu statków i moim zdaniem **równie dobrze** można nakłonić właścicieli statków do tego samego, zwłaszcza biorąc pod uwagę wszelkie negatywne materiały, jakie zostały wydane w ciągu ostatnich kilku lat.*

(18) *Indeed, **it may well be** the case that liberalisation fuels liberalisation.*
*W istocie, liberalizacja w jednym miejscu **może** przyśpieszać liberalizację w drugim.*

(19) *As for your agreement with Australia, **it may well be** a cut above other agreements, for example with the United States.*
*Jeżeli chodzi o umowę z Australią, **niewykluczone**, że jest lepsza od innych umów, na przykład tej ze Stanami Zjednoczonymi.*

The manual verification of the Polish equivalents in the National Corpus of Polish revealed that all equivalents occur there with high frequencies, e.g. *być może* (35,247 occurrences), *bardzo możliwe, że* (242 occurrences), *niewykluczone, że* (2,353 occurrences), *może się okazać, że* (1,091 occurrences), *równie dobrze* (2,430 occurrences) and *może* (395,510 occurrences). On the one hand, these high frequencies show that all the equivalents are typical of contemporary Polish. On the other hand, one may expect that they occur in the whole variety of contexts that require the expression of epistemic stance. For example, an impersonal construction starting with *niewykluczone, że* could as well mean 'it is possible that' or 'there may be', likewise 'it may well be'; *być może* could as well mean 'perhaps', 'maybe', 'possibly', 'might be', 'could be' etc.

5. DISCUSSION

Based on the selected examples of English-to-Polish translations under scrutiny, the results of this study revealed that the translator may use, at least in theory, an infinite number of linguistic means as suitable equivalents that express writer's or speaker's attitudinal or epistemic stance. In practice, by creating adequate contexts of language use — taking into consideration both the original text as well as similar native texts in the target language — the translator is restricted neither to those linguistic items which have already occurred in the target language nor to those which are frequent in the target language (Piotrowski 2011: 48), which has been often the case in the examples presented throughout this study (cf. example 1). Also, the translator may attach the expression of stance to a text fragment in the translation which does not correspond to a text fragment expressing stance in the original (cf. example 13). Hence, it is often the case that a MWU in the source language is translated as a single-word unit in the target language. In such a situation, the actual verification of the target language equivalents — in terms of their frequency and potential textual fit — in monolingual reference corpora such as NKJP poses particular challenges. Since monolingual general

language corpora (e.g. NKJP, BNC), by their very nature, contain the whole variety of text types and genres, the target language equivalents subject to verification may occur in various contexts of language use. Moreover, the number and distribution of stance bundles may vary across written and spoken registers according to communicative purposes implied by their co-text and context. That is why it is recommended in the future to replicate this study by using a relatively large target language corpus with native texts representing the same text type, namely transcripts of parliament debates originally conducted in Polish.

6. CONCLUSIONS

The aim of this preliminary study was to verify whether (and if so, then how) lexical bundles may be used to enhance the naturalness — in this paper operationalized as the textual fit (Chesterman 2004: 6) — of English-to-Polish translation of EU Parliament debates. Inspired by the study conducted by Grabar and Lefer (2015), we used the European Parliament sub-corpus (EPP) of Paralela (Pęzik 2016), an English-Polish and Polish-English parallel corpus, as well as the National Corpus of Polish (NKJP), a general language corpus, to explore how four attitudinal and epistemic stance bundles (*it is not surprising that, it would be wrong to, there is no doubt that, it may well be*) are translated into Polish and, second, whether the Polish equivalents are at the same time typical of contemporary Polish language in terms of their frequency of use.

As expected, the results obtained from the EPP sub-corpus of Paralela revealed a high number of Polish equivalents, both single- and multi-word units, expressing stance, which means that the translators use the whole variety of translation techniques when selecting the equivalents. Notably, we reported a high number of unit shifts, where a MWU in the original was translated using a single-word item in the translation. It was also reported that occasionally entirely different sentence fragments in the original and in the translation conveyed attitudinal and epistemic stance.

Next, the results obtained from NKJP corpus revealed a number of Polish equivalents (e.g. *nie może zaskakiwać, że; nie jest zaskakujące, że*)[14] which are very rare or do not occur — in their exact form — at all in the National Corpus of Polish. As a result, it may be argued that they fail to enhance the textual fit of Polish translations. On the other hand, the majority of the Polish equivalents (e.g. *nic dziwnego, że*) are frequently used in native Polish texts and therefore they can potentially help enhance the textual fit of translations.

As for verification of the Polish equivalents in the entire NKJP corpus, we encountered a number of problems. Most importantly, since NKJP includes a plethora of text types and genres[15], the equivalents occur in the whole variety of contexts that require expression of attitudinal and epistemic stance. Hence, the verification of the equivalent in a given context, e.g. in a parliament debate, requires that a custom-designed

---

[14] These are the translations of the English-original lexical bundle *it is no surprise that*.

[15] See Przepiórkowski et al. (2012) for a more detailed description of NKJP.

collection of transcripts of parliament debates originally conducted in Polish be used in the future to further verify the obtained results. Such collections of native texts in the source language and in the target language, i.e. non-translations, are referred to as bilingual comparable corpora (Laviosa 2002: 101). Obviously enough, in this study we largely focused on English-to-Polish translation of selected stance bundles originally found — by Grabar and Lefer (2015) — in EU Parliament debates, yet it is possible to replicate the procedures described in this paper using other text types or genres.

Both research procedures described in this paper, that is, using a parallel (Paralela) and monolingual reference corpus (NKJP) to extract and verify the use of translation equivalents, constitute the skills that enhance translation competence: using language corpora is nowadays recommended when designing translation training programmes at universities (Biel 2011: 165—169). Importantly, unlike the extraction of LBs from texts, following the methodology proposed by Biber et al. (1999)[16], the use of parallel and comparable corpora is a realistic scenario in the translator's work, which offers repeated exposure to authentic linguistic data. All in all, the use of monolingual, parallel and comparable corpora may help eliminate interference from the source language, identify formulaic expressions and collocations, adapt translations to stylistic conventions of the target language, among others (Biel 2011: 168—169). That is why practical exercises, e.g. focusing on stylistics, aimed at extraction and validation of the use of MWUs in translation and native texts — conducted using monolingual, parallel and comparable corpora as well as online multilingual resources (e.g. Linguee[17]) — should be encouraged in the translation classroom. Capitalizing on the proposals put forward by Jukneviciene (2017: 62—64) and Salazar (2011: 189)[18], the translation tasks may involve, for example, identifying recurrent n-grams or LBs and their functions in source texts and then searching for their equivalents in target texts; comparing the use of LBs (or other types of MWUs) across text samples in L1 and L2, e.g. by focusing on translation of particular MWUs expressing stance or performing text-organizing functions, e.g. cause-and-effect, connectives. For the sake of illustration, Appendix 1 presents a proposal of two translation tasks.

Since the use and distribution of LBs and other types of recurrent word combinations varies across proficiency levels of language learners (Jukneviciene 2009; Staples et al. 2013; Appel & Wood 2016), it may be expected that the frequency and distribution of LBs will also vary between trainee and professional translators. For example, Novita and Kwary (2018), who studied English-to-Indonesian translation of

---

[16] Extraction of lexical bundles is a task for corpus linguists rather than for professional translators. An attempt at extracting lexical bundles from a custom-designed comparable corpus of English and Polish patient information leaflets, aimed at developing bilingual glossaries (in the form of functionally-aligned lexical bundles in English and Polish) to be implemented into translation memories used in CATs, is described in Grabowski (2018a, forth.).

[17] https://www.linguee.com/.

[18] The research conducted by Salazar (2011, 2014) was focused, among others, on teaching LBs in ESP contexts.

ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА

literary texts using 600-word samples of short stories, showed that professional translators produce more LBs, which also occur with higher frequencies, as compared with trainee translators. Hence, similar future studies conducted from the perspective of English-to-Polish translation may provide valuable pedagogical insights into the use of recurrent phraseologies by trainee translators, notably if compared with translations produced by professionals as well as with native texts originally produced in Polish. The results of such studies may also potentially help improve the textual fit of translations.

Summing up, it is hoped that the results of this preliminary research, likewise the results of the study conducted by Grabar and Lefer (2015), showed that the findings from descriptive studies on LBs, most of which were conducted using English language materials, can also be potentially useful for practitioners of translation.

REFERENCES

Appel, R. & Wood, D. (2016). "Recurrent Word Combinations in EAP Test-Taker Writing: Differences between High- and Low-Proficiency Levels". *Language Assessment Quarterly*, 13 (1): 55—71.

Barreiro, A. & Batista, F. (2016). "Machine Translation of Non-Contiguous Multiword Units". Proceedings of DiscoNLP 2016, 22—30. Available: http://www.inesc-id.pt/publications/12162/pdf (accessed in October 2017).

Berūkštienė, D. (2017). "A corpus-driven analysis of structural types of lexical bundles in court judgments in English and their translation into Lithuanian" *Kalbotyra*, 70: 7—31.

Biber, D. (2006). *University Language. A corpus-based study of spoken and written registers.* Amsterdam: John Benjamins.

Biber, D. (2009). "A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing". *International Journal of Corpus Linguistics*, 14(3): 275—311.

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

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. (2009). *Register, genre and style*. Cambridge: Cambridge University Press.

Biel, Ł. (2011). "Professional Realism in the Legal Translation Classroom: Translation Competence and Translator Competence". *Meta*, 56(1), 162—178.

Biel, Ł. (2014a). "The textual fit of translated EU law: a corpus-based study of deontic modality". *The Translator*, 20 (3): 332—355.

Biel, Ł. (2014b). *Lost in the Eurofog. The Textual Fit of Translated Law.* Frankfurt: Peter Lang.

Biel, Ł. (2017). "Lexical bundles in EU law: the impact of translation process on the patterning of legal language". In" S. Goźdź-Roszkowski & G. Pontrandolfo (Eds), *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective*. London/New York: Routledge, 10—26.

Bouamor, D., Semmar, N., Zweigenbaum, P. (2012). "Identifying Bilingual Multi-Word Expressions for Statistical Machine Translation". In: *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC12), 674—679. Available at: http://www.lrec-conf.org/proceedings/lrec2012/pdf/886_Paper.pdf (accessed in December 2017).

Breeze, R. (2013). "Lexical bundles across four legal genres". *International Journal of Corpus Linguistics* 18 (2): 229—253.

Catford, J. (1965). *A Linguistic Theory of Translation*. London: Oxford University Press.

Chesterman, A. (2004). Hypothesis about translation universals. In: G. Hansen, K. Malmkjaer & D. Gile (Eds), *Claims, Changes and Challenges in Translation* Studies. Amsterdam: John Benjamins, 1—13.

Forchini, P. & Murphy, A. (2008). N-grams in comparable specialized corpora. Perspectives on phraseology, translation and pedagogy. *International Journal of Corpus Linguistics*, *13*(3), 351—367.

Fuster-Marquez, M. (2017). "The Discourse of US Hotel Websites: Variation through the Interruptibility of Lexical Bundles". In: M. Gotti, S. Maci and M. Sala (Eds), Ways of Seeing, Ways of Being: Representing the Voices of Tourism. Frankfurt am Main: Peter Lang Verlag, 400—420.

Goźdź-Roszkowski, S. (2011). Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study. Frankfurt am Main: Peter Lang Verlag.

Górski, R. (2012). "Zastosowanie korpusów w badaniu gramatyki". In: A. Przepiórkowski, M. Bańko, R. Górski & B. Lewandowska-Tomaszczyk (Eds), *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN, 291—300.

Grabar, N. & Lefer, M-A. (2015). "Building a lexical bundle resource for CAT and MT". Presentation delivered at Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT2015) of EUROPHRAS 2015. 29 Jun — 1 Jul 2015, Malaga, Spain. Available at: http://natalia.grabar.free.fr/publications/grabar-MUMTTT2015.pdf (accessed in October 2017).

Grabowski, Ł. (2014). "On Lexical Bundles in Polish Patient Information Leaflets: A Corpus-Driven Study". *Studies in Polish Linguistics*, 19 (1): 21—43.

Grabowski, Ł. (2015). "Keywords and lexical bundles within English pharmaceutical discourse: a corpus-driven description". *English for Specific Purposes*, 38: 23—33.

Grabowski, Ł. (2018a, forth.). "On identification of bilingual lexical bundles for translation purposes. The case of an English-Polish comparable corpus of patient information leaflets". In: R. Mitkov, J. Monti, G. Corpas Pastor and V. Seretan (Eds), *Multiword Units in Machine Translation and Translation Technology* [*Current Issues in Linguistic Theory* 341], Amsterdam: John Benjamins, pp. 181—200.

Grabowski, Ł. (2018b). "Fine-tuning lexical bundles: A methodological reflection in the context of describing drug-drug interactions". In: J. Kopaczyk & J. Tyrkkö (Eds), *Applications of Pattern-driven Methods in Corpus Linguistics*. Amsterdam: John Benjamins, pp. 57—80.

Granger, S. (2014). A lexical bundle approach to comparing languages. Stems in English and French. In: M-A. Lefer & S. Vogeleer (Eds.), *Genre- and register-related discourse features in contrast.* Special issue of *Languages in Contrast*, *14*(1), 58—72.

Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, *18*(1), 109—135.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language.* London: Routledge.

Hoey, M. (2007). Lexical priming and literary creativity. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.), Text, Discourse and Corpora. London: Continuum, 7—30.

Hyland, K. (2008). "As can be seen: Lexical bundles and disciplinary variation". *English for Specific Purposes* 27: 4—21.

Jalali, H. (2017). "Reflection of stance through it bundles in applied linguistics". *Ampersand* 4, 30—39. https://doi.org/10.1016/j.amper.2017.06.001 (accessed in October 2017).

Jukneviciene, R. (2009). "Lexical bundles in learner language: Lithuanian learners vs. native speakers". *Kalbotyra* 61 (3): 61—72.

Jukneviciene, R. (2017). *English phraseology and corpora: An introduction to corpus-based and corpus driven phraseology.* Vilnius: Vilniaus Univesiteto Leidykla.

Koehn, P. (2005). "Europarl: A Parallel Corpus for Statistical Machine Translation". In: *Conference Proceedings: the tenth Machine Translation Summit*, Phuket, Thailand: AAMT, 79—86.

Kornacki, M. (2017). *Computer-assisted translation (CAT) tools in the translator training process.* Unpublished PhD dissertation. University of Łódź.

Kranich, S. (2016). *Contrastive Pragmatics and Translation: Evaluation, epistemic modality and communicative styles in English and German*. Amsterdam: John Benjamins.

Laviosa, S. (2002). *Corpus-based translation studies: theory, findings, applications.* Amsterdam: Rodopi.

Lee, D. (2008). "Corpora and discourse analysis". In V. Bhatia, J. Flowerdew & R. Jones (Eds.), *Advances in Discourse Studies*. London: Routledge, 86—99.

Lewandowska-Tomaszczyk, B. (2012). "Cognitive Corpus Studies: A New Qualitative & Quantitative Agenda for Contrasting Languages". *MFU Connexion: Journal of Humanities and Social Sciences*, 1 (1): 26—64. Available at: http://doi.nrct.go.th/ListDoi/Download/252098/f2601120e1c60067d1328094376a8c8d?Resolve_DOI=10.14456/connexion.2012.2 (accessed in February 2018).

Martinez, R. & Schmitt, N. (2012). "A Phrasal Expression List". *Applied Linguistics*, 33(3): 299—320.

McVeigh, J. (2018). "Join us for this. Lexical bundles and repetition in email marketing texts". In: J. Kopaczyk & J. Tyrkko (Eds), *Applications of Pattern-driven Methods in Corpus Linguistics*. Amsterdam: John Benjamins, 213—250.

Mindt, I. (2011). *Adjective Complementation: An Empirical Analysis of Adjectives Followed by That-clauses.* Amsterdam: John Benjamins.

Novita, H. & Kwary, D. (2018). "Comparing the use of lexical bundles in Indonesian-English translation by student translators and professional translators". *Translation & Interpreting* 10 (1): 53—74.

Oksefjell Ebeling, S. & Ebeling, J. (2017). "A Cross-Linguistic Comparison of recurrent word combinations in a comparable corpus of English and Norwegian Fiction". In: M. Janebova, E. Lapshinova-Koltunski & M. Martinkova (Eds), *Contrasting English and Other Language through Corpora*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2—31.

Pérez-Paredes, P. (2010). "The death of the adverb revisited: attested uses of adverbs in native and non-native comparable corpora of spoken English". In: M. Moreno Jaén, F. Serrano Valverde, & M. Calzada Pérez (Eds), *Exploring new paths in language pedagogy. Lexis and corpus-based language teaching*. London: Equinox, 157—172.

Pęzik, P. (2012). "Język mówiony w NKJP. (Spoken Language in NKJP)" In: A. Przepiórkowski, M. Bańko, R. Górski & B. Lewandowska-Tomaszczyk (Eds), *Narodowy Korpus Języka Polskiego* (National Corpus of Polish), Warszawa: Wydawnictwo Naukowe PWN, 37—47.

Pęzik, P. (2016). "Exploring phraseological equivalence with Paralela". In: E. Gruszczyńska & A. Leńko-Szymańska (Eds), *Polish-Language Parallel Corpora*. Warszawa: Instytut Lingwistyki Stosowanej UW, 67—81.

Piotrowski, T. (1994). *Z zagadnień leksykografii* (Problems in lexicography). Warszawa: Wydawnictwo Naukowe PWN.

Piotrowski, T. (2011). "Ekwiwalencja w słownikach dwujęzycznych (Equivalence in bilingual dictionaries)". In: W. Chlebda (Ed.), *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych* (Searching for translation equivalents). Opole: Wydawnictwo Uniwersytetu Opolskiego, 89—114.

Przepiórkowski, A., Bańko, M., Górski, R. & Lewandowska-Tomaszczyk, B. (Eds) (2012). *Narodowy Korpus Języka Polskiego* (National Corpus of Polish). Warszawa: Wydawnictwo Naukowe PWN.

Sag, I., Baldwin, T., Bond, F., Copestake, A., & Flickinger D. (2002). Multiword Expressions: A Pain in the Neck for NLP. *Computational Linguistics and Intelligent Text Processing: Third International Conference* (CICLing 2002), 1—15. Available at: http://lingo.stanford.edu/pubs/WP-2001-03.pdf (accessed May 2013).

Salazar, D. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing.* Unpublished PhD dissertation. University of Barcelona.

Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing.* Amsterdam: John Benjamins.

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

Skadina, I. (2016). "Multi-Word Expressions in English-Latvian SMT: Problems and Solutions. In: I. Skadina & R. Rozis (Eds), *Human Language Technologies — The Baltic Perspective: Proceedings of the Seventh International Conference Baltic HLT 2016.* Amsterdam: IOS Press. 97—106.

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

Toury, G. (1995). *Descriptive Translation Studies and Beyond.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Toury, G. (2001). ─The Nature and Role of Norms in Translation. In: L. Venuti (ed.), *The Translation Studies Reader*, London: Routledge, 198—211.

APPENDIX 1. EXAMPLES OF TRANSLATION TASKS

Task 1. Since lexical bundles constitute building blocks of specialist discourses (genres, text types and registers), they should have readily-available equivalents across languages. Identify Polish and/or Russian equivalents of the following stance bundles found in a collection of English patient information leaflets (Grabowski 2018a, forth.). To that end, you may compile a custom-designed corpus of patient information leaflets or use multilingual online resources.

please ask your doctor or pharmacist
tell your doctor or pharmacist
tell your doctor immediately
talk to your doctor
check with your doctor
please read this leaflet carefully
what you should know about
as directed by your doctor
never give it to someone else

ТЕОРИЯ И ПРАКТИКА ПЕРЕВОДА

stop taking the tablets
you may need to
you should not take

Task 2. Look at the following list of the top-20 lexical bundles — by frequency — found in a sample of specialist texts describing drug-drug interactions (Grabowski 2018b: 71). Some of them express stance (<u>underlined</u>). Can you find their Polish and/or Russian equivalents?

did not affect the
a single dose of
on the pharmacokinetics of
the concomitant use of
<u>*it is recommended that*</u>
the concomitant administration of
drug laboratory test interactions
in the presence of
<u>*the patient should be*</u>
has not been studied
<u>*caution should be exercised*</u>
had no effect on
<u>*caution should be used*</u>
has been reported to
have been reported in
no effect on the
affect the pharmacokinetics of
<u>*should be observed closely*</u>
the clinical significance of
did not alter the

Bionote:
ŁUKASZ GRABOWSKI is Associate Professor at the Institute of English, University of Opole (Poland). His research interests include corpus linguistics, phraseology, formulaic language, translation studies and lexicography. He is also interested in computer-assisted methods of text analysis. He has published research articles and book chapters internationally in *International Journal of Corpus Linguistics* and *English for Specific Purposes* as well as with John Benjamins and Emerald, among others. He is also Managing Editor of the journal *Explorations: A Journal of Language and Literature*. *Contact information:* e-mail: lukasz@uni.opole.pl

ЛЕКСИЧЕСКИЕ СВЯЗКИ
СО ЗНАЧЕНИЕМ ОЦЕНКИ И ОТНОШЕНИЯ В ПЕРЕВОДЕ
С АНГЛИЙСКОГО ЯЗЫКА НА ПОЛЬСКИЙ:
КОРПУСНОЕ ИССЛЕДОВАНИЕ

ЛУКАШ ГРАБОВСКИЙ

Опольский университет
45-040, Ополе, Польша пл. Коперника 11а

Целью исследования является поиск более точных текстуальных соответствий в переводе с английского языка на польский специфического типа многословных блоков, известных в литературе по корпусной лингвистике как лексические связки (Biber et al. 1999). Под влиянием исследования М.-А. Грабарь и Н. Лефер (Grabar, Lefer 2015) мы используем англо-польский параллельный корпус Paralela (Pęzik 2016) и Национальный корпус польского языка (NKJP) для выделения и верификации статуса (с точки зрения частотности) польских эквивалентов выборочных английских лексических связок, выражающих оценку и отношение. Точнее, Национальный корпус польского языка использовался для проверки того, характерны ли польские эквиваленты для аутентичных текстов современного польского языка. В результате проведенного корпусного исследования было выявлено значительное количество польских эквивалентов как одного, так и многословных блоков, выражающих оценку и отношение. Кроме того, было установлено, что большая часть польских эквивалентов часто встречается в аутентичных польских текстах и, следовательно, они потенциально могут способствовать повышению уровня текстуального соответствия переводов. И, наконец, мы обсуждаем методические и корпусные ограничения данного исследования и намечаем перспективы его продолжения с целью дальнейшего изучения роли лексических связок в преподавании перевода и в переводческой практике. С этой целью мы также даем примеры заданий по переводу для работы в классе.

Ключевые слова: *корпусная лингвистика, лексические связки, перевод с английского на польский, параллельный корпус, текстуальное соответствие*

Сведения об авторе
ЛУКАШ ГРАБОВСКИЙ — доцент Института английского языка Опольского университета, Польша. Его научные интересы включают корпусную лингвистику, фразеологию, лексикографию и теорию и практику перевода. Он также интересуется методами анализа текста с привлечением компьютерных технологий. Среди его публикаций — статьи и главы книг, опубликованные в международных журналах, в частности, в *International Journal of Corpus Linguistics*, *English for Specific Purposes* и др. Он является редактором журнала *Explorations: A Journal of Language and Literature. Контактная информация:* e-mail: lukasz@uni.opole.pl