**Grammar patterns as an exploratory tool for studying formulaicity in English-to-Polish translation: a corpus-based study**

Łukasz Grabowski1, Nicholas Groom2

1University of Opole, 2University of Birmingham

**Abstract.** In this chapter, we explore the use – and argue for the usefulness – of the concept of grammar patterns (Francis et al 19966, 1998; Hunston & Francis 1999) in descriptive research on the English-to-Polish translation of formulaic language. Specifically, we use the *Paralela* English-Polish parallel corpus (Pęzik 2016) to explore – largely in terms of frequency distributions – the use of the Polish equivalents of selected English multi-word items, which are textual manifestations of grammar patterns. As a proof-of-concept, we will focus on a pre-selected grammar pattern (‘*it* v-link ADJ *to*-inf’), where a given word (e.g. the adjective *possible*) may convey different senses depending on the pattern in which it occurs (Groom 2005). We aim to check whether and to what extent somewhat similar lexico-grammatical patterns emerge from the Polish language data (i.e. from the Polish translations under scrutiny). Using this method, we are also able to investigate whether the Polish equivalents of that English grammar pattern (i.e. from its textual manifestations) are realized with the same level of regularity in translation. Early findings revealed the English pattern ‘*it* v-link ADJ *to*-inf’, when filled by adjectives conveying the sense of ‘difficulty’, corresponds to a set of Polish syntagmatic patterns, such as ‘ADJ v-link, *aby*’, ‘ADV v-link’ or ‘ADV’ (the last two ones followed by verbs in the infinitive form), which contribute to the formulaicity of Polish texts. It is argued not only that the study’s findings indicate that grammar patterns are a useful starting point for the exploration of formulaicity in translation, but also that they help us explain some more general differences in terms of semantics, pragmatics and usage in source texts and their translations.

1. Introduction.
   1. General remarks on Pattern Grammar

In the last three decades, corpus linguists have proposed a number of novel research methods and procedures for capturing and exploring recurrent patterns of language use. One of them is Pattern Grammar (Hunston & Francis 2000), a corpus-based grammar of the English language that describes the lexico-syntactic environments of individual lexical items. These ‘grammar patterns’ (henceforth GPs) are defined as “a phraseology frequently associated with (a sense of) a word, particularly in terms of the prepositions, groups, and clauses that follow the word” (Hunston & Francis 2000: 3) or as “all the words and structures which are regularly associated with the word and which contribute to its meaning” (ibid. 37). As argued by Roemer (2009: 142), GPs are neither single words nor empty grammatical structures; they constitute abstract representations of frequent lexico-grammatical patterns. For example, Hunston and Francis (2000: 29) claim that in the patterns ***it* v-link ADJ that** (e.g. *it is interesting/clear that*) and ***it* v-link ADJ *to*-inf** (e.g. *it is sensible/possible to*), the adjectives belong to similar meaningfully-related groups, e.g. expressing a range of different concepts such as likelihood, importance, desirability and obviousness.

Whilst Pattern Grammar is principally an empirical approach to linguistic description, it also puts forward two major theoretical claims. The first of these is that the different meanings (i.e. senses) of polysemous words can be distinguished on the basis of their typical occurrence in different patterns. Consider, by way of example, the following two (invented) sentences:

1. ***It’s possible that*** *she sent a message.*
2. ***It’s possible to*** *send a message.*

It should be immediately clear that, although the GPs highlighted in bold in these sentences both feature the same adjective, ***possible***, the meanings that they make are very different. In example (1), the highlighted words can be replaced with the single adverb ***maybe***, whereas in example (2) the highlighted part can be replaced with the two words ***you can***. From a traditional perspective, it might be said that this shows that the adjective ***possible*** has two different meanings: an epistemic modal meaning in example (1), and a dynamic modal meaning in example (2). From the point of view of Pattern Grammar, however, the epistemic and dynamic modal meanings made by these two sentences reside not in any of the individual words, nor in the morphosyntactic configurations into which these words fall, but in the interaction between them. That is, the epistemic ‘maybe’ meaning belongs to the whole sequence ‘*it is* *possible* + that-clause’, and the dynamic ‘you can’ meaning is made by the whole sequence ‘*it is* *possible* + to-infinitive clause’. This insight is of major significance because, if it is true for a language as a whole, it means that semantic ambiguity is virtually non-existent in naturally-occurring language. Each of the different meanings of a polysemous word will be associated with a different structural pattern, and language users will interpret which meaning is being made in each case on the basis of this patterning. In other words, Pattern Grammar proposes that disambiguation is far more a matter of attending to linguistic co-text than it is a matter of attending to extralinguistic context, as is generally assumed in traditional theories of semantics and pragmatics.

The second major theoretical claim put forward by Pattern Grammar is that “words which share a given pattern tend also to share an aspect of meaning” (Hunston & Francis 2000: 3). To return to examples (1) and (2) above, ***possible*** is just one member of a class of adjectives that have a broadly epistemic modal meaning when they occur in the pattern ***it* v-link ADJ *that*** (e.g. *certain, definite, demonstrable, doubtful, feasible, implausible, incredible, irrefutable, likely, true, uncertain, unthinkable,* etc.), and just one of many adjectives that have a broadly dynamic modal meaning when they occur in the pattern ***it* v-link ADJ *to*-inf** (e.g. *difficult, easy, feasible, hard, impossible, impractical., simple, tough, tricky*, etc.). In Pattern Grammar, these classes of words that share a particular pattern/meaning association are referred to as ‘meaning groups’. It is important to note that each grammatical pattern is not restricted to just one meaning group; on the contrary, many if not most patterns have several different meaning groups associated with them. For example, Francis et al (1998) identify eight distinct meaning groups for the pattern ***it* v-link ADJ *that***, and nine different meaning groups for the pattern ***it* v-link ADJ to-*inf***.

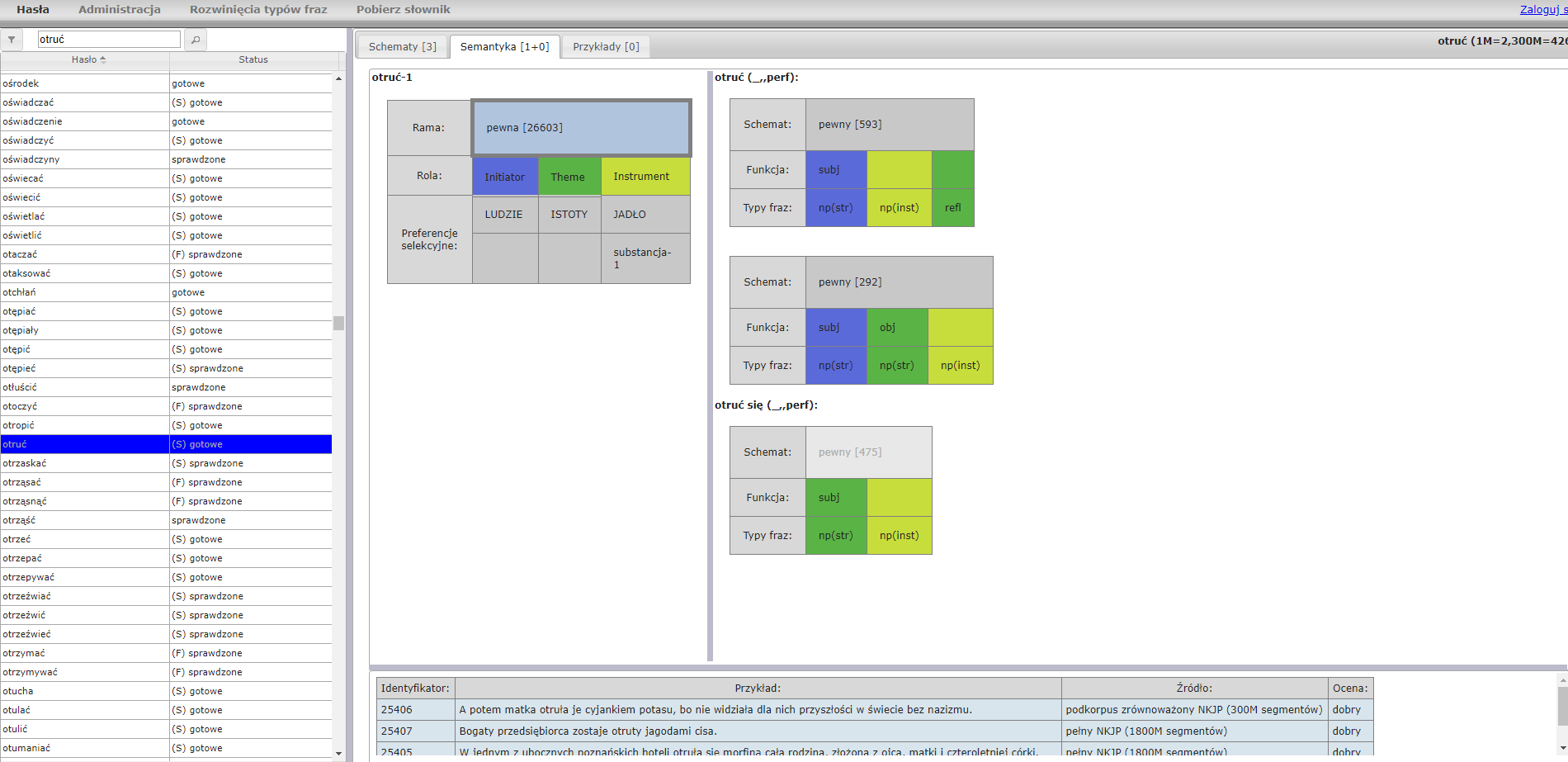
Although Pattern Grammar is unquestionably a corpus-based approach to grammatical description, it is nevertheless highly reliant on the manual qualitative analysis of concordance lines, and thus on the application of human judgement in the identifying of patterns and the meaning groups associated with them. This is because the identification of GPs in attested language corpora essentially involves perceiving a similarity between linguistic items which may not be identical on the surface, but which have some underlying regularity of form and meaning, which at the current time can only be reliably identified by a human analyst.[[1]](#footnote-1) For example, the sequence of words *it is ironic that* is a representative of a general pattern beginning with it, ending with a *that*-clause (with optional *that*) and containing a linking verb (e.g. *be, seem, look*), followed by an adjective expressing evaluation of a given situation, e.g. *it is not surprising that, it looks very unlikely that, it seems very peculiar that* (Hunston & Francis 2000: 154). Such judgements are often highly nuanced and subtle, and it is sometimes problematic even for human researchers, let alone computer software, to distinguish between sequences of words which are formally the same yet differ in terms of their patterns, e.g. ‘annoyance + relative clause’ vs. ‘annoyance + appositive that-clause’ (Hunston & Francis 2000: 67).

Despite its labour intensive nature, Pattern Grammar has been widely and successfully applied to the description and analysis of English over the last two decades or so. Perhaps the best known fruits of this research are the two monumental *COBUILD Grammar Patterns* reference works (Francis et al. 1996, 1998), one covering 700 patterns across 9,000 different verb senses, and the other covering 100 further patterns across 10,000 nouns and adjectives. In addition to these general reference works, corpus-based studies have also provided empirical evidence that GPs differ in systematic ways across different discourses and genres, with academic English being a particular focus of interest (e.g. Groom 2005; Charles 2006, 2007; Larsson 2016; Su & Hunston 2019). The COBUILD Grammar Patterns reference works have also been used as an empirical basis for psycholinguistic investigations into first and second language speaker knowledge of verb argument constructions (e.g. Ellis et al. 2014; Römer et al. 2014, 2015).

These latter inter-varietal research perspectives suggest that Pattern Grammar might also be used to carry out empirical research in cross-linguistic contexts. Surprisingly, however, no work along such lines has been done to date; in fact, although there is some comparative cross-linguistic research on grammar patterns in the specialized genre of judicial decisions (Pontrandolfo & Gozdz-Roszkowski 2014) focusing on the discoursal function of evaluation in English and Italian, PG has as yet never been systematically applied to any other language than English (Hunston, personal communication).

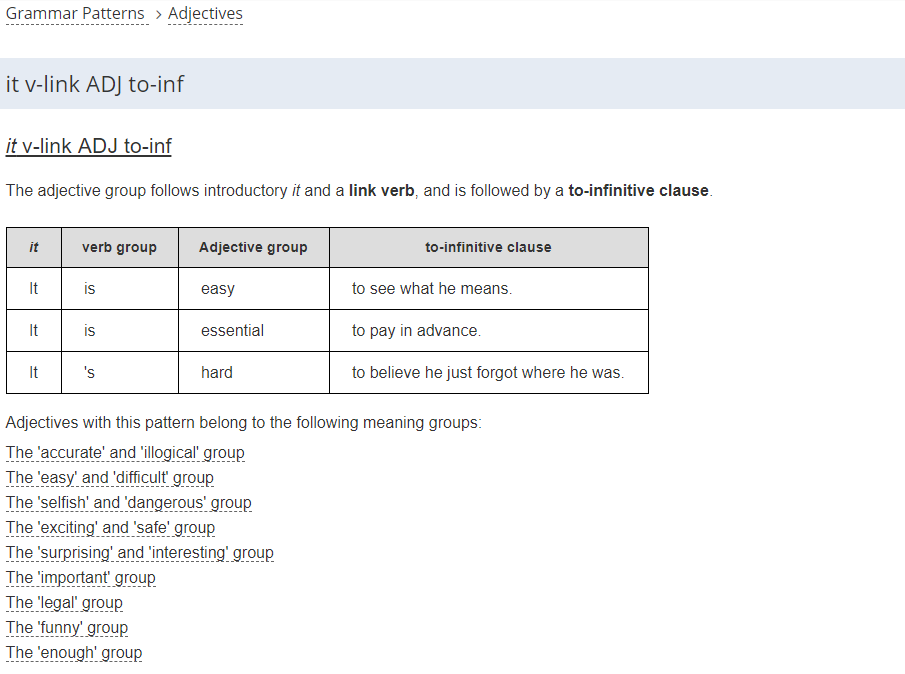
* 1. Towards Pattern Grammar for languages other than English

In fact, there has been no usage-based description similar to Pattern Grammar capturing recurrent lexico-grammatical patterns of language use in Polish, i.e. based on the relationship between meanings or discourse functions on the one hand and structural patterns on the other. However, similar to valency dictionaries developed for English (e.g. *Erlangen Valency Patternbank*[[2]](#footnote-2) or *Pattern Dictionary of English Verbs*[[3]](#footnote-3)), there are dictionaries of Polish that combine syntactic and semantic information on particular lexical items in ways which bear some comparison with the Pattern Grammar approach. One of them is *Walenty[[4]](#footnote-4),* a comprehensive valency dictionary of Polish (Przepiórkowski et al. 2017a, 2017b) that specifies arguments of predicates for verbs (and – to a lesser extent – for nouns, adjectives and adverbs), which consist of comprehensive syntactic and semantic formalisms, the latter ones not fully implemented yet (Przepiórkowski et al. 2017a: 10). Although utilized by two parsers of Polish (*Świgra* and POLFIE), the dictionary has been designed primarily for computer processing of natural language texts as well as for researchers (linguists) and lexicographers (ibid.: 9). For example (and unlike in the case of Pattern Grammar), verbs recorded in *Walenty* are described using the valency frames, each with a set of argument specifications (ibid.), e.g. the verb *otruć* (‘to poison’) has three syntactic frames and one semantic frame, all accompanied by examples of use (Figure 1).

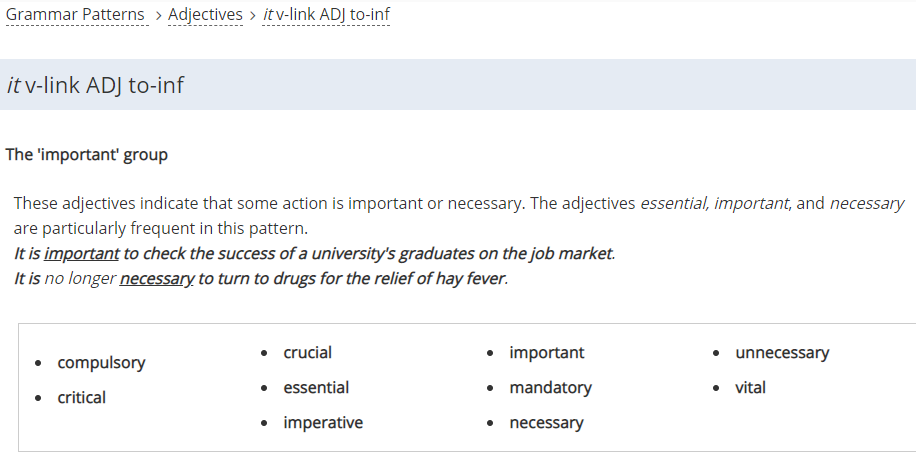


*Figure 1. The verb otruć ‘to poison’ in Walenty (retrieved from: http://walenty.clarin-pl.eu/)*

Although *Walenty* is similar to Pattern Grammar in that it presents syntactic environments of lexical items, it does not allow the researcher to browse through larger lexico-grammatical patterns (e.g. verbs, nouns or adjectives and the words or syntagmatic frames that follow or precede the words) or to classify the slot-fillers in those patterns into specific meaning or functional groups, which is precisely what the Pattern Grammar approach does. For example, the so-called ‘introductory *it*’ pattern followed by a link-verb, adjective and to-infinitive clause (***it* v-link ADJ *to*-inf**)[[5]](#footnote-5) reveals a number of meaning groups for adjectives filling in the pattern (Figure 2) as well as specific examples of their use (Figure 3).



*Figure 2. The pattern ‘it v-link ADJ to-inf’*



*Figure 3. The pattern ‘it v-link ADJ to-inf’ filled with adjectives conveying the sense of importance.*

As can be seen in Figure 3, the sense of importance (or attitudinal stance) of the following proposition is conveyed by the entire grammar pattern ***it* v-link ADJ *to*-inf** rather than by individual adjectives filling in the pattern. Consequently, Pattern Grammar offers an inventory of lexico-grammatical constructions which constitute pairings of form with semantic or discoursal function.[[6]](#footnote-6) Such a resource describing lexico-grammatical patterns extracted from corpora in a bottom-up fashion has not been developed for the Polish language so far.

The main goal of this study is therefore to explore whether GPs may be employed as a useful exploratory tool for cross-linguistic studies. In other words, we aim to identify and describe lexico-grammatical patterns that emerge from the English-to-Polish translations – extracted from Paralela corpus (Pęzik 2016) – of a pre-selected English GP performing specific discoursal functions. More precisely, we will focus on one English ‘introductory it’ pattern with *to*-clause complementation (***it*** *v-link* ***ADJ to-inf***), which will constitute the starting point for our analysis.[[7]](#footnote-7) However, the *tertium comparationis* are the functional categories or discoursal functions (necessity, importance, obviousness etc.) contingent on individual words filling in the patterns or, in other words, worked out on the basis of intuitive understandings of words in the English GPs. This will allow us to investigate whether we may find corresponding generalizable lexico-grammatical patterns in Polish that perform the same discoursal functions as the English GP under scrutiny. The underlying assumption is that in theory the Polish translations should convey the same information as their English source texts. We will also attempt to verify whether GPs are useful as a discovery tool for detecting translation patterns. We believe that a study like this one may pave the way to a wider application of the Pattern Grammar approach for the description of languages other than English (in this case, Polish) and/or for cross-linguistic comparisons or translation-oriented research, which have not been attempted hitherto.

The chapter is structured as follows. In Section 2, the research material (i.e. the study corpus), units of analysis and methodology will be described. Next, we describe and discuss the empirical results exemplifying translation patterns emerging from textual realizations of GPs in the sample of English source-texts and their Polish translations. The concluding section discusses the study limitations and offers suggestions on how this research may be developed further in the future.

1. Methodology
   1. Research material

For the purposes of the current study, we will use the English-Polish parallel corpus *Paralela* (Pęzik 2016), which is available at: http://paralela.clarin-pl.eu. The parallel corpus under scrutiny includes a little more than 262 million word tokens in 10,877,000 translation segments from various text types and genres, including written, spoken and to-be spoken texts, e.g. legal documents from various European Union institutions (legislation, transcripts of proceedings of the European Parliament etc.), press releases, medical texts, film subtitles, popular science texts, literary classics, transcripts of European Parliament proceedings etc. As a rule, a parallel corpus contains source texts aligned with their translations in the target language. In this study, we will analyze a pre-selected GP (‘*it* v-link ADJ *to*-inf’) found in a single genre of English source texts, namely European Parliament proceedings (henceforth EPP), and we will try to align its textual realizations with their Polish equivalents as found in translation segments in the corresponding Polish sub-corpus (with more than 13 million word tokens in almost 700,000 translation segments)[[8]](#footnote-8) Then, we will attempt to quantitatively and qualitatively analyze the target language equivalents and see whether any regular patterns, i.e. lexico-syntactic associations similar to GPs, emerge from the Polish language data.

* 1. Units of analysis

At least in theory, parallel corpora include texts (originals and translations) that express the same meanings and perform the same discoursal functions, which allows one to search for correspondences between linguistic items in source and target texts (Johansson 2007: 9, cited in Marco 2019: 43). As mentioned earlier, in this study we will focus on a pre-selected GP, which means that the lexical items under scrutiny will not be extracted from texts in a bottom-up approach. This is mainly because we do not have access to full texts collected in *Paralela*. Hence, the use of the study corpora will be limited to the analysis of bilingual concordances illustrating particular translation patterns, i.e. frequency and distribution of the Polish equivalents of the English GP, which provide a starting point for our investigation.

We will capitalize on the results of the study conducted by Groom (2005), who analyzed, among others, two GPs, namely ***it* v-link *ADJ that*** and ***it* v-link *ADJ to-*inf**. Groom (2005: 259-260) noticed that adjectives which convey the meaning (i.e. sense) that can be generalized as “validity” (e.g. *clear, inconceivable, obvious*) tend to fall into the pattern ***it* v-link *ADJ that***while the adjectives conveying the sense of “difficulty” (e.g. *difficult, easy, hard*) tend to fall into the pattern ***it* v-link *ADJ to-*inf***[[9]](#footnote-9)*. Also, it was found that depending on the pattern, one and the same word (e.g. the adjective *possible*) may convey different senses, that is either “difficulty” or “validity” (Groom 2005: 259). In summary, the findings of the study conducted by Groom (2005) provide strong evidence of the relationship between particular sense conveyed by particular words and the structural patterns in which those words tend to occur.

We will search for a single pre-selected pattern using the SlopeQ query syntax implemented in *Paralela* (Pęzik 2016) as well as morphosyntactic tags (e.g. *<tag=j.\*>* stands for adjectives). Thus, the pattern ***it* v-link *ADJ to-*inf** will be searched for using the following query:

* *it <tag=v.\*> <tag=j.\*> to* (34,047 occurrences in Paralela; 3,067 occurrences in EPP).

In view of a high number of occurrences, it is necessary to filter out the results to facilitate the qualitative analysis of concordance lines. To this end, we will apply systematic sampling by recording translation pairs of every 30th concordance. This means that we will ultimately focus on a sample of 100 translation pairs (i.e. bilingual concordances illustrating Polish translations of specific textual realizations of the English GP) extracted from the EPP sub-corpus of *Paralela*.

* 1. Research questions and hypotheses

The main problem addressed in this study concerns whether the corresponding items (translation equivalents) found in target texts can also be described in terms of recurrent lexico-syntactic associations similar to the GP identified in the English source texts. Hence, in this exploratory paper we aim to provide answers to the following research questions:

1) What are the Polish equivalents of the English source-language multi-word items emerging from GPs?

2) Can the Polish equivalents be generalized into a more abstract set of lexico-grammatical patterns similar to GPs?

3) Can the Pattern Grammar approach be used for a description of Polish?

4) Can GPs be applied as a unit of analysis in cross-linguistic contexts?

* 1. Research questions and hypotheses

This study will be conducted in a number of stages. First, we will preselect one GP (described earlier) and develop an inventory of their textual realizations in English source-texts, i.e. in the EPP sub-corpus of *Paralela*. Then, we will generalize the results by means of grouping overlapping textual realizations into a list of n-grams performing specific discourse functions. After identification of the Polish equivalent, or translation variant (be it a single-word or a multi-word unit), we will investigate whether the observed patterns of Polish translations can be generalized to a more abstract phraseological, syntagmatic or lexico-grammatical units, similar to GPs.

1. Preliminary results: grammar patterns in contrast

In this exploratory study, we used the GP ***it* v-link *ADJ to-*inf** as a unit of analysis and a tool for discovering potential translation patterns. All in all, the said GP occurred in the EPP sub-corpus of Paralela (Pęzik 2016) 3,067 times. In order to limit the amount of data for manual analysis of bilingual concordances, we used systematic sampling and selected every 30th concordance, which resulted in the set of 100 English-Polish translation segments to be analyzed. Depending on the adjectives filing in the slot, we used the procedure put forward by Groom (2005) and – by conducting manual analysis of English source-text fragments – we classified the textual variants of the GP found in the sample into semantic/functional categories (discoursal functions) corresponding to the senses conveyed by the adjectives. This way, the linguistic data were classified into the following categories: ‘importance’, ‘validity’, ‘desirability’ and ‘difficulty’.

First, we present the results of the analysis of the textual instantiations of the GP ***it* v-link *ADJ to-*inf** as filled with adjectives conveying the sense of ‘difficulty’ in the English-original texts in the EPP sub-corpus (8 occurrences) and its equivalents in the Polish translations.

|  |  |  |  |
| --- | --- | --- | --- |
| **Textual realizations in the English-original** | **Discoursal function** | **Polish equivalents** | **Generalized pattern of Polish translations** |
| *it is* ***difficult*** *to* (5) | DIFFICULTY | ***trudno*** *jest* (5) | ADV v-link v-inf |
| *it is* ***hard*** *to* (2) | DIFFICULTY | ***trudno****, aby* (1)  ***trudno*** (1) | ADV, *aby*  ADV v-inf |
| *it is* ***easier*** *to* (1) | DIFFICULTY | ***łatwiej*** (1) | ADV v-inf |

*Table 1. Textual realizations of the GP* ***it v-link ADJ to-inf*** *in English source-texts and their Polish translations: discoursal function of difficulty*

The findings (Table 1) show that in the sample under scrutiny there are three different instantiations of the English GP ‘*it* v-link ADJ *to*-inf’ when filled with adjectives conveying the sense of ‘difficulty’, namely *it is difficult to*, *it is hard to* and *it is easier to*, with the total frequency of 8. Those multi-word items have the following Polish equivalents in Paralela corpus (Pęzik 2016), such as *trudno jest* (used 5 times as an equivalent of *it is difficult to*), *trudno, aby* and *trudno* (used 1 each as equivalents of *it is hard to*) and *łatwiej* (used once as an equivalent of *it is easier to*). Apart from insights into certain translational choices, it has been possible to reconstruct abstract lexico-grammatical patterns (or syntagmatic frames) based on the Polish equivalents and conveying the sense of difficulty with respect to the following proposition, e.g. ‘ADV v-link v-inf’ (*trudno jest*) or ‘ADV v-inf’, both followed by a verb in infinitive form (*trudno jest udowodnić ‘*it is difficult to prove’*, trudno uwierzyć* ‘it is difficult to believe’) or ‘ADV, *aby*’ (*trudno, aby* etc.) followed by a complement clause, e.g.:,

***It is hard to*** *believe when reading it.*

***Trudno*** *uwierzyć w te słowa, kiedy się je czyta.* [EVOeRj]

*This is because* ***it is difficult to*** *prove that the service rendered was of poor quality.*

*Dzieje się tak dlatego, że* ***trudno jest*** *udowodnić, że świadczone usługi były złej jakości.* [qoavWA]

*Simplification of the common agricultural policy is a beautiful idea, and* ***it is hard to*** *imagine that someone would oppose it.*

*Uproszczenie wspólnej polityki rolnej to piękna idea i* ***trudno, aby*** *ktoś był jej przeciwny.* [ea2Eoq]

As can be seen, the English GP ‘*it* v-link ADJ *to*-inf’, when filled by adjectives conveying the sense of ‘difficulty’, corresponds to a set of Polish syntagmatic patterns, such as ‘ADV, *aby +* complement clause’, ‘ADV v-link v-infinitive’ or ‘ADV v-infinitive’.

The following example under scrutiny refers to the GP ‘*it* v-link ADJ *to*-inf’ as filled with adjectives (*possible, clear, impossible, true*) conveying the sense of ‘validity’ in the English-source texts (Table 2).

|  |  |  |  |
| --- | --- | --- | --- |
| **Textual realizations in the English-original** | **Discoursal function** | **Polish equivalents** | **Generalized pattern of Polish translations** |
| *(will) it be* ***possible*** *to* (1) | VALIDITY | ***możliwe*** *(będzie)* (1) | ADJ v-link |
| *it is* ***clear*** *to (us/me)* (2) | VALIDITY | *dla (nas)* ***jasne*** *jest, że* (1)  *jest dla (mnie)* ***jasne****, że* (1) | ADJ v-link, *że* ‘that’  v-link ADJ, że ‘that’ |
| *it is* ***impossible*** *to* (1) | VALIDITY | ***niemożliwością*** *jest* (1) ‘impossibility is’  *nie można* (1) ‘ | NN v-link  NEGprt + MOD v |
| *it is* ***possible*** *to* (4) | VALIDITY | ***możliwe*** *jest* (2) ‘possible that’  *można* (1) ‘may/can’  *możliwość[[10]](#footnote-10)* (1) ‘possibility’ | ADJ v-link  MODv v-inf  TRANSFORMATION |
| *it is* ***true*** *to* (1) say | VALIDITY | ***prawdą*** *jest, że* (1) ‘truth is that’ | NN v-link, *że* ‘that’ |

*Table 2. Textual realizations of the GP ‘it v-link ADJ to-inf’ in English source-texts and their Polish translations: discoursal function of validity*

The findings show that there is more variety among Polish lexico-grammatical patterns that convey the sense of ‘validity’ in the Polish translations as compared with the sense of ‘difficulty’. The most frequent one (4 occurrences) is the pattern ‘ADJ v-link’ (with positional variation) and ‘ADJ v-link, *że*’, which is realized with the following Polish equivalents, namely *możliwe* (*będzie*); *możliwe jest;* (…) *jasne jest, że*; *jest* (…) *jasne, że*. Other patterns are centred around nouns (‘NN’, ‘NN v-link’ and ‘NN v-link, *że*’), which include the following words and phrases: *możliwość*, *niemożliwością jest* and *prawdą jest, że* respectively, e.g.:

*I would like to say that it is becoming clear in this discussion that* ***it is possible to*** *have a 'two-speed ' Europe.*

*Chciałbym powiedzieć, że w tej dyskusji oczywista staje się* ***możliwość*** *istnienia Europy "dwóch szybkości[[11]](#footnote-11)”.*[Drag45]

There are also two patterns centred on the modal verb *można* ‘can/may’ (‘MODv’, ‘NEG PRT MODv’), namely *można* and *nie można*, followed by the infinitive form of the verb, e.g.:

(…) *but the one thing I have learned is that* ***it is impossible to*** *book a ticket on the Eurostar when you are travelling*.

(…) *lecz nauczyłem się jednego, a mianowicie, że* ***nie można*** *zarezerwować biletu na pociąg*. [BNzRG4]

A large group of multi-word items in the English originals conveys a sense of ‘importance’ (e.g. *it is important to, it is crucial to, it is essential that*) and there is a high variety among their Polish equivalents, which can be grouped into a number of lexico-grammatical patterns. The most prominent ones include the pattern ‘ADJ v-link’ and ‘ADJ v-link, *aby/by’* (e.g. *istotne jest*; *ważne jest, aby*; *niezbędne jest; konieczne jest*), with adjectives occasionally modified by adverbs (e.g. *bardzo ważne jest*, *niezwykle istotne jest*). It is particularly noticeable that the pattern ‘ADJ v-link’ is followed by a verbal noun functioning as a direct object, while the pattern ‘ADJ v-link, *aby/*by’ or ‘ADJ, *by*’) is followed by the verb in the infinitive form, e.g.:

*Finally, I would like to stress that* ***it is important to*** *have full transparency regarding the founding of the initiative and sources of financial support for the organisers.*

*Wreszcie, pragnę podkreślić, że* ***ważne jest*** *zapewnienie pełnej przejrzystości w odniesieniu do finansowania inicjatywy oraz źródeł wsparcia finansowego dla jej organizatorów* [RDwmvo]

***It is important to*** *overcome the current problems that characterise the sector: the lack of competition, the regulatory over-dependence on ratings, and the low reliability of notes.*

***Ważne jest, aby*** *rozwiązać problemy, które obecnie spotyka się w tej branży, a mianowicie przesadne zawierzanie ratingom i małą wiarygodność ocen.* [APaN8r]

Other frequent patterns include ‘MODv v-inf’ (*trzeba, musimy/musi, należy* followed by verbs in the infinitive form) or ‘NN ADJ v-link’ (e.g. *sprawą podstawową jest*; *ważną rzeczą jest/są*) or ‘V ADJ NN’ (e.g. *ma fundamentalne znaczenie*), cf. Table 3 below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Textual realizations in the English-original** | **Discoursal function** | **Polish equivalents** | **Generalized pattern in Polish translations** |
| *it is* ***important*** *to* (1) | IMPORTANCE | ***ważne*** *jest, aby* (1) | ADJ v-link, *aby* |
| *it is* ***crucial*** *to* (2) | IMPORTANCE | ***istotne*** *jest* (1)  ***kluczowe*** *znaczenie ma* (1) | ADJ v-link  ADJ NN V |
| *it is* ***essential*** *to* (10) | IMPORTANCE | ***konieczne*** *jest* (2)  ***kluczowe*** *znaczenie ma* (2)  ***niezbędne*** *jest* (2)  *trzeba* (1)  *ma* ***istotne znaczenie***(1)  ***niezwykle******istotne*** *jest* (1)  *należy* (1) | ADJ v-link  ADJ NN V  ADJ v-link  MODv v-inf  V ADJ NN  (ADV) ADJ v-link  MODv v-inf |
| *it is* ***fundamental*** *to* (1) | IMPORTANCE | *ma* ***fundamentalne znaczenie*** (1) | V ADJ NN |
| *it is* ***imperative*** *to* (1) | IMPORTANCE | ***bardzo ważne*** *jest* (1) | (ADV) ADJ v-link |
| *it is* ***important*** *to* (24) | IMPORTANCE | *należy* (6)  ***ważne*** *jest* (6)  *musimy* (1)  *należy* ***koniecznie*** (3)  ***ważne*** *jest, aby* (2)  *trzeba* (2)  ***bardzo ważne*** *jest* (1)  ***konieczne*** *jest* (1)  ***ważną rzeczą*** *jest/są* (1)  ***ważne****, by* (1)  ***najważniejsze*** *jest* (1) | MODv v-inf  ADJ v-link  MODv v-inf  MODv ADV v-inf  ADJ v-link, *aby*  MODv v-inf  ADJ v-link  ADJ v-link  ADJ N v-link  ADJ, *by*  ADJ v-link |
| *it is* ***necessary*** *to* (15) | IMPORTANCE | *należy* (4)  ***konieczne*** *jest* (3)  *wymaga[[12]](#footnote-12)* (2)  *trzeba* (2)  *trzeba* ***koniecznie*** (1)  *rzeczą* ***konieczną*** *jest* (1)  *musi* (1) | MODv v-inf  ADJ v-link  V  MODv v-inf  MODv ADV v-inf  NN ADJ v-link  MODv v-inf |
| *it seems* ***fundamental*** *to* (1) me | IMPORTANCE | ***sprawą podstawową*** *jest* (1) | NN ADJ v-link |
| *it was* ***essential*** *to* (1) | IMPORTANCE | *było (jest)* ***niezwykle istotne*** (1) | v-link ADV ADJ |

*Table 3. Textual realizations of the GP ‘it v-link ADJ to-inf’ in English source-texts and their Polish translations: discoursal function of importance*

Finally, similar correspondences can be observed in the case of structures conveying the sense of ‘desirability’ in the English originals and their Polish equivalents (Table 4). The most frequent patterns – with positional variation due to free word order – in the Polish translations were found to be ‘ADJ v-link’ (e.g. *konieczne jest, jest niestosowne, nieodpowiedzialne jest, słuszne jest*), ‘ADV v-link’ (e.g. *dobrze jest*), ‘MODv v-inf’ (e.g. *warto, należy*). The comparison of certain GPs provides interesting insights into cross-linguistic correspondences on the syntactic level. For example, the English pattern ‘*it* v-link ADJ v-inf’ may correspond, among others, to the Polish GP ‘MODv v-inf’, where the position of direct and indirect object with respect to the patterns under scrutiny changes, e.g.

***It is essential to*** *give people* [Indirect object] *access* [Direct object] *to healthcare, drinking water and sanitation.*

*Ludziom* [Indirect object] ***należy*** *zapewnić dostęp* [Direct object][[13]](#footnote-13) *do opieki zdrowotnej, wody pitnej i kanalizacji.* [DrZG0o]

We also identified ready-made formulas corresponding to specific speech acts, (e.g. the polite phrase used after a greeting *Cieszymy się, że jest Pan tu z name* versus *it is nice to have you with us*) or transformations resulting from stylistic changes in the translation as compared to the original. Both fixed formulas and transformations do not fall into GPs that we attempted to identify among the Polish equivalents.

|  |  |  |  |
| --- | --- | --- | --- |
| **Textual realizations in the English-original** | **Discoursal function** | **Polish equivalents** | **Generalized pattern in Polish translations** |
| *it is* ***necessary*** *to* (1) | DESIRABILITY | ***konieczne*** *jest* (1) | ADJ v-link |
| *it is* ***advisable*** *to* (1) | DESIRABILITY | OMISSION[[14]](#footnote-14) (1) | OMISSION |
| *it is* ***appropriate*** *to* (1) | DESIRABILITY | ***warto*** (1) ‘it is worth’ | MODv v-inf |
| *it is* ***better*** *to* (1) | DESIRABILITY | ***lepiej*** (1) ‘better’ | ADV V |
| *it is* ***fair*** *to* (1) (*say that…*) | DESIRABILITY | ***uczciwie*** (1) ‘fairly’ | TRANSFOR-MATION[[15]](#footnote-15) |
| *it is* ***good*** *to* (1)know | DESIRABILITY | ***dobrze*** *jest* (1) | ADV v-link |
| *it is* ***inappropriate*** *to* (1) | DESIRABILITY | *jest* ***niestosowne*** (1) | v-link ADJ |
| *it is* ***irresponsible*** *to* (1) | DESIRABILITY | ***nieodpowiedzialne*** *jest* (1) | ADJ v-link |
| *it is* ***nice*** *to* (1) (*have you with us*) | DESIRABILITY | *Cieszymy się, że* (1) *jest Pan tu z nami* ‘we are happy to have you here with us’ | FIXED FORMULA |
| *it is* ***pointless*** *to* (1) | DESIRABILITY | *nie ma sensu* (1) | NEGprt V NN |
| *it is* ***profitable*** *to* (1) | DESIRABILITY | *przynosi zysk* | V NN |
| *it is* ***right*** *to* (3) | DESIRABILITY | *należy* (1)  *jest* ***słuszne*** (1)  ***słuszne*** *jest* (1) | MODv v-inf  v-link ADJ  ADJ v-link |
| *it is* ***sensible*** *to* (1) | DESIRABILITY | *(jest)* ***rozsądne*** (1) | v-link ADJ |
| *it is* ***unacceptable*** *to* (1) | DESIRABILITY | *nie do przyjęcia jest* (1) | NEGprt PREP NN v-link |
| *it is* ***uplifting*** *to* (1) | DESIRABILITY | *świadomość* (…) *podnosi na duchu* (1) ‘awareness of (…) raises sb’s spirits' | TRANSFOR-MATION |
| *it is* ***worthwhile*** *to* (1) | DESIRABILITY | ***warto*** (1) | MOD v v-inf |
| *it was* ***agreeable*** *to* (1) | DESIRABILITY | *miło było* (1) ‘(it) was nice to’ | ADV v-link v-inf |

*Table 4. Textual realizations of the GP ‘it v-link ADJ to-inf’ in English source-texts and their Polish translations: discoursal function of desirability*

1. Discussion and conclusions

The findings of our exploratory study indicate that the Pattern Grammar approach holds unexplored potential for cross-linguistic research, of both text-oriented and system-oriented kinds. The approach showcased in our study used a set of GPs as a starting point to identify recurrent English multi-word items with specific meanings (senses) and discoursal functions, which were then aligned with their Polish equivalents. We obtained further empirical evidence against a widely held misconception about translation whereby one has to translate a given fixed phrase from L1 (English) into a corresponding fixed phrase in L2 (Polish). Using this methodology, we were able to obtain a set of lexico-grammatical frames emerging from the English-to-Polish translation patterns in the corpus under scrutiny (EPP). As well as yielding insights into language in use in English-to-Polish translation in the EPP corpus, our methodology revealed a number of novel and valuable cross-linguistic correspondences between formulaic structures in English and Polish. As our study was based on a specific text type, it remains to be seen to what extent those Polish lexico-grammatical patterns can be generalized to other text types or genres (if we adopt a textual perspective), and to the Polish language system (if we adopt a cross-linguistic systemic perspective). Nevertheless, our preliminary results are certainly encouraging.

The findings of the current study also show that the Pattern Grammar approach holds unexplored potential not only for cross-linguistic studies, but also for the description of recurrent lexico-syntactic constructions found in other languages, such as Polish. Clearly, such a research agenda needs to be pursued with caution in view of the many typological differences between English and Polish (cf. Fisiak 1978; Willim & Mańczak-Wohlfeld 1997). In practice, what this means is that a “Polish version” of Pattern Grammar will need to be mindful of formal descriptions of Polish, and that we should expect to find non-formally corresponding grammar patterns that convey the same meanings or discourse functions in English and Polish, as was the case with the English GP ‘*it* v-link ADJ *to*-inf’ and the lexico-syntactic patterns that emerged from the Polish translations, rather than exact equivalents for each given English GP. Indeed, we may find that there are one-to-one, one-to-many, many-to-many and many-to-one relationships between GPs conveying the same meaning or discourse function in English and Polish. Since adjective complementation patterns have received considerable attention (e.g. Su & Hunston 2019), it would be useful for future research to conduct analyses that involve other patterns (e.g. *there* v-link *sth/anything/nothing* ADJ *about/in* NP). Since this study concerns translation, it would be also interesting for future research to examine whether the observed translation patterns are triggered by English (due to interference), and if so then to what extent, or whether they are natural and equally formulaic in native Polish texts. Since this preliminary study is exploratory in nature, no frequency threshold was set to verify when an equivalent is regarded as a pattern; however, setting such a threshold will be essential in the future.

Finally, Pattern Grammar-based descriptions of Polish may also feed into the future development of more comprehensive lexicographic resources, the so-called ‘dictionaries of constructions’ or ‘constructicons’ which are presently being compiled for several European languages, including German (Boas & Ziem 2018), Swedish (Lyngfelt et al. 2018b) and Russian (Janda et al. 2018) [[16]](#footnote-16), to name but a few, and which are designed to model entire languages as inventories of constructions at all levels, i.e. from morpheme to discourse. A promising example of this synergy of approaches is provided by Perek & Patten (2019), who are currently building an English constructicon by semi-automatically combining GPs (as represented by the COBUILD reference works) with the semantic frames and valency relations found in the FrameNet database. Since constructicons may serve lexicographers, language learners as well as NLP applications, it is postulated that such a resource, which could utilize information already available in valency dictionaries (e.g. *Walenty*), should be also developed in the future for the Polish language[[17]](#footnote-17).

Acknowledgements

This research has been funded by the Polish National Agency for Academic Exchange (NAWA) under the agreement no: PPN/BEK/2018/1/00081/, and conducted during a research stay of Łukasz Grabowski at the University of Birmingham (UK).

1. References

Boas, H. & Ziem, A. (2018). “Constructing a constructicon for German: Empirical, theoretical, and methodological issues”. In: B. Lyngfelt, L. Borin, K. Ohara and T. Timponi Torrent (Eds), Constructicography: Constructicon development across languages. Amsterdam: John Benjamins, 183-228.

Charles, M. (2006). ‘Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines.’ English for Specific Purposes 25: 310–331.

Charles, M. (2007). 'Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction.' English for Specific Purposes 26: 203–218.

Ellis, N.C., O'Donnell, M.B. & Römer, U. (2014). ‘Second language processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality.’ Linguistic Approaches to Bilingualism 4(4): 405-431.

Fisiak, J., Lipińska Grzegorek, M. & Zabrocki, L. (1978). An introductory English-Polish contrastive grammar. Warszawa: PWN.

Fillmore, Ch. (1982). ‘Frame semantics’. In: The Linguistic Society of Korea (eds), Linguistics in the Morning Calm. Seoul: Hanshin, 111-137.

Francis, G., Manning, E. & Hunston, S. (1996). Collins COBUILD grammar patterns 1: Verbs. London: HarperCollins.

Francis, G., Manning, E. & Hunston, S. (1998). Collins COBUILD grammar patterns 2: Nouns and adjectives. London: HarperCollins.

Goldberg, A. (2006). Constructions at Work. The Nature of Generalization in Language. Oxford: Oxford University Press.

Groom, N. (2005). “Pattern and meaning across genres and disciplines: An exploratory study.” A Journal of English for Academic Purposes 4: 257-277.

Hanks, P. (2008a). "Mapping meaning onto use: a Pattern Dictionary of English Verbs". AACL 2008, Utah.

Hanks, P. (2008b). “Lexical Patterns: from Hornby to Hunston and beyond”. Afrilex 2008, Stellenbosch.

Hanks, P. (2013). Lexical Analysis: Norms and Exploitations. Cambridge. MIT Press.

Herbst, T., Heath, D., Götz, D., & Roe, I. (2004). A Valency Dictionary of English. Berlin/New York: De Gruyter.

Hunston, S., Manning, E., Francis, G. (1998). Grammar Patterns II: Nouns and Adjectives. London: Collins Cobuild.

Hunston, S. & Francis, G. (2000). Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English. Amsterdam: John Benjamins.

Janda, L., Lyashevskaya, O., Nesset, T., Rakhilina, E. & Tyers, F. (2018). “A constructicon for Russian”. In: B. Lyngfelt, L. Borin, K. Ohara and T. Timponi Torrent (Eds), Constructicography: Constructicon development across languages. Amsterdam: John Benjamins, 165-182.

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

Larsson, T. (2016). “The introductory it pattern: Variability explored in learner and expert writing. Journal of English for Academic Purposes, 22, 64-79.

Lyngfelt, B., Borin, L., Ohara, K. & Timponi Torrent, T. (Eds) (2018a). Constructicography. Constructicon development across languages. Amsterdam: John Benjamins.

Lyngfelt, B., Bäckström, L., Borin, L., Ehrlemark, A. & Rydstedt, R. (2018b). “Constructicography at work”. In: B. Lyngfelt, L. Borin, K. Ohara and T. Timponi Torrent (Eds), Constructicography: Constructicon development across languages. Amsterdam: John Benjamins, 41-106.

Mason, O. & Hunston, S. (2004). ‘The automatic recognition of verb patterns: a feasibility study.’ International Journal of Corpus Linguistics 9: 253-270.

Perek, F. & Patten, A. (2019). “Towards an English Constructicon using patterns and frames”. International Journal of Corpus Linguistics, 24(3): 356-386.

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.

Pontrandolfo, G., & Goźdź-Roszkowski, S. (2014). ‘Exploring the Local Grammar of Evaluation: The Case of Adjectival Patterns in American and Italian Judicial Discourse’. Research in Language 12(1): 71-91.

Przepiórkowski, A., Hajic, J., Hajnicz, E. & Uresova, Z. (2017a). „Phraseology in two Slavic Valency Dictionaries: Limitations and Perspectives”. International Journal of Lexicography, 30(1), 1-38. [10.1093/ijl/ecv048]

Przepiórkowski, A., Hajnicz, E., Andrzejczuk, A., Patejuk, A. & Woliński, M. (2017b). „Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego”. Język Polski, XCVII(1): 30–47.

Römer, U., O'Donnell, M.B., & Ellis, N.C. (2014). ‘Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology.’ The Modern Language Journal 98(4): 952-975.

Römer, U., O'Donnell, M.B., & Ellis, N.C. (2015). ‘Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions: Exploring corpus data and speaker knowledge.’ In N. Groom, M. Charles & S. John (eds.). Corpora, Grammar and Discourse: In Honour of Susan Hunston. Amsterdam: John Benjamins, 43-71.

Su, H., & Hunston, S. (2019). ‘Language patterns and attitude revisited Adjective patterns, Attitude and Appraisal’. Functions of Language 26(3): 343-371.

Wierzbicka-Piotrowska, E. (2019). “Od synonimii składniowej do polisemii konstrukcyjnej (rozważania metodologiczne na przykładzie analizy polskich konstrukcji wyrażających relację przyczynowo-skutkową)”. In: Book of Abstracts. Polish Cognitive Linguistics Association Conference 2019 (September 26-28, 2019, Białystok), 95-96. Retrieved on 21 Nov 2019 from: <http://ptjk2019.uwb.edu.pl/wp-content/uploads/2019/09/PCLA2019-Book-of-Abstracts.pdf>

Willim, E. & Mańczak-Wohlfeld, E. (1997). A contrastive approach to problems with English. Warszawa: PWN.

1. The automatic identification of GPs has been demonstrated as feasible in principle (Mason & Hunston 2004), but has not as yet been achieved at scale on open text. [↑](#footnote-ref-1)
2. EVP is available at: <http://www.patternbank.uni-erlangen.de/cgi-bin/patternbank.cgi>; it provides a list of valency patterns for 511 verbs, 544 adjectives and 274 nouns. See Herbst et al. (2004) for more details. [↑](#footnote-ref-2)
3. PDEV is available at: <http://pdev.org.uk/#about_cpa>; it provides systematic description of meaning and use of verb patterns of 1451 verbs (as of 9 July 2019). See Hanks (2008a, 2013) for more details. [↑](#footnote-ref-3)
4. Primarily a dictionary of Polish subcategorization frames, Walenty can be directly compared to FrameNet (http://framenet. icsi.berkeley.edu) grounded in the theory of frame semantics (Fillmore 1982). [↑](#footnote-ref-4)
5. The example is available at: https://grammar.collinsdictionary.com/grammar-pattern/it-v-link-adj-to-inf\_1 [↑](#footnote-ref-5)
6. In practice, identification of GPs is possible by analyzing concordance lines, which later involves grouping the patterns into notional categories (e.g. topical or functional ones) on the basis of different types of meanings conveyed in contexts. Although individual words can help in determination of those groupings, a qualitative analysis of a wider context of their occurrence is necessary to form appropriate groups and thus identify the patterns (Hunston & Francis 2000: 162). [↑](#footnote-ref-6)
7. Both patterns were studied by Groom (2005) in terms of their variability across different academic registers. [↑](#footnote-ref-7)
8. The transcripts were originally extracted from Europarl corpus (Koehn 2005) and included in Paralela (Pęzik 2016). The debates were recorded on 11-12th and 23rd October 2006, and translated from English into Polish. [↑](#footnote-ref-8)
9. However, Groom (2005) found that distributions and more fine-grained rhetorical functions of the adjectives in those grammar patterns vary across corpora representing different language varieties. [↑](#footnote-ref-9)
10. (EN) *I would like to say that* ***it is becoming clear*** *in this discussion that it is possible to have a 'two-speed' Europe.* (PL) *Chciałbym powiedzieć, że w tej dyskusji* ***oczywista staje się możliwość*** *istnienia Europy "dwóch szybkości”.* [Drag45] [↑](#footnote-ref-10)
11. The phrase *Europa dwóch szybkości* ‘two-speed Europe’ has not been adopted in the Polish press discourse. Instead, the phrase *Europa dwóch prędkości* has become commonly used. [↑](#footnote-ref-11)
12. (EN) *There is also a need to measure quality of life in societies, because in order to ensure and sustain quality of life,* ***it is necessary to*** *take into account important, consensual factors such as health, education, culture, employment, housing, environmental conditions etc.* (PL) *Istnieje także potrzeba pomiaru jakości życia w społeczeństwach, ponieważ osiągnięcie oraz utrzymanie odpowiedniej jakości życia* ***wymaga*** *(‘requires’) uwzględnienia istotnych i jednomyślnie uznanych czynników, takich jak zdrowie, edukacja, kultura, zatrudnienie oraz warunki mieszkaniowe, środowiskowe itp.* [6nv4bb] [↑](#footnote-ref-12)
13. Lit. ‘people should be given access’ [↑](#footnote-ref-13)
14. (EN) *We shall check what you have said, and you can rest assured that, within the context of measures available to the Council and the work on implementing the Pact on Immigration and Asylum, we shall look once more at whether* ***it is advisable to*** *reinforce this point within the framework of the Schengen area.* (PL) *Sprawdzimy to, o czym pan mówił, i może pan mieć pewność, że w kontekście środków, jakimi dysponuje Rada, oraz prac nad wdrożeniem paktu o imigracji i azylu, ponownie przemyślimy kwestię udoskonalenia tego elementu w ramach strefy Schengen.* [AP01Ae] [↑](#footnote-ref-14)
15. (PL) (…) *będzie uczciwie, jeżeli powiem, że (…)* [rBwJxM] [↑](#footnote-ref-15)
16. According to Janda et al. (2018: 165), the Russian constructicon prioritizes multi-word constructions. [↑](#footnote-ref-16)
17. The idea of developing a construction grammar of Polish has also been briefly mentioned recently by Wierzbicka-Piotrowska (2019). [↑](#footnote-ref-17)