Studies in Polish Linguistics
vol. 9, 2014, issue 1, pp. 21–43
doi:10.4467/23005920SPL.14.002.2186
www.ejournals.eu/SPL

Łukasz Grabowski
Opole University, Poland

On Lexical Bundles in Polish Patient Information Leaflets: A Corpus-Driven Study

Abstract
So far little attention has been paid to the corpus analysis of recurrent phraseologies found in Polish texts, in particular texts representing specialists registers of language use. Also, one may note the lack of corpus linguistic studies of lexical bundles (Biber et al. 1999) found in texts originally written in Polish. Conducted from a register perspective (Biber and Conrad 2009), this descriptive and exploratory study is intended as a first step towards a comprehensive corpus-driven description of the use and functions of the most frequent lexical bundles found in patient information leaflets (PILs), one of the most commonly used text types in the healthcare sector in Poland. The research material includes 100 PILs written originally in Polish, extracted from internet websites of ten pharmaceutical companies operating on the Polish market, compiled in a purpose-designed corpus of circa 197,000 words. Based largely on the methodology proposed by Biber, Conrad and Cortes (2003, 2004), Biber (2006), and Goźdź-Roszkowski (2011), which makes possible an analysis of the use and discourse functions of lexical bundles, the present study is primarily meant to provide methodological guidelines for future research on lexical bundles in Polish texts. This appears to be important since so far lexical bundles have been studied predominantly in texts originally written in English. The results of this preliminary research reveal salient links between the frequent occurrence of lexical bundles on the one hand, and situational and functional characteristics of the text variety under scrutiny on the other.

Keywords
corpus linguistics, phraseology, register analysis, corpus-driven approach, lexical bundles, patient information leaflets

Streszczenie
Dotychczas rzadko podejmowano się korpusowych badań nad często powtarzalnymi frazeologizmami w tekstach napisanych w języku polskim, zwłaszcza reprezentujących rejestry specjalistyczne. Rezultatem takiego stanu rzeczy jest m.in. niedobór badań korpusowych nad zbitkami wielowyrazowymi (*lexical bundles*) (Biber et al. 1999) na materiale w języku polskim. Wykonane z perspektywy rejestru języka (Biber i Conrad 2009), niniejsze badanie o charakterze eksploracyjno-opisowym jest pomyślane jako pierwszy krok na drodze do opracowania bardziej wszechstronnego opisu – z wykorzystaniem podejścia sterowanego korpusem – użycia i funkcji dyskursywnych najczęstszych zbitek wielowyrazowych występujących w ulotkach dla pacjentów, które są jednym z najczęściej używanych

typów tekstów w sektorze opieki zdrowotnej w Polsce. Materiał badawczy obejmuje 100 ulotek napisanych w języku polskim, pozyskanych ze stron internetowych dziesięciu firm farmaceutycznych działających na polskim rynku leków, zebranych w specjalnie do tego celu stworzonym korpusie liczącym ok. 197 000 wyrazów tekstowych. Oparte w dużym stopniu na elementach metodologii zaczerpniętej z badań Bibera, Conrad i Cortes (2003, 2004), Bibera (2006) i Goździa-Roszkowskiego (2011), niniejsze badanie ma również na celu przedstawienie propozycji metodologicznych w kontekście przyszłych badań nad zbitkami wielowyrazowymi na materiale polskich tekstów. Takie zamierzenie wydaje się uzasadnione, ponieważ dotychczasowe badania nad zbitkami wielowyrazowymi w przeważającej większości wykonano na materiale tekstów napisanych w języku angielskim. Wyniki niniejszego badania pilotażowego wykazały istotny związek pomiędzy wysoką frekwencją zbitek wielowyrazowych a funkcją komunikacyjną i kontekstem sytuacyjnym, w jakim zazwyczaj używane są ulotki dla pacjentów.

Słowa kluczowe
językoznawstwo korpusowe, frazeologia, analiza rejestru języka, podejście sterowane korpusem, zbitki wielowyrazowe, ulotki dla pacjentów

1. Introduction

Corpus linguistics has enjoyed a period of rapid growth in recent years; newer and newer computational tools and, in consequence, more sophisticated research procedures have been developed in order to facilitate empirical descriptive studies on lexis and phraseology. This has made it easier for researchers to identify and study repeated events in language use, notably the repeated use of multi-word units or strings of word forms typically found in the whole variety of specialist texts (including a pharmaceutical text type, such as Polish patient information leaflets, as it is the case in this study). This makes corpus linguistics approach particularly attractive for analyses of lexis and phraseology of routinized specialist texts since they rely more on limited stocks of prefabricated chunks, linguistic patterns or formulas, notably when compared with more creative literary texts. It is thus in the clichéd specialist texts where language users more often select a large number of prefabricated phrases or phraseologies with single form and meaning, the phenomenon described by Sinclair (1991) as the Idiom Principle.[1] In a similar vein, Wray and Perkins (2000) and Wray (2002) use the term 'formulaicity' with reference to the frequent use of formulaic sequences. Also called formulas, such sequences are characterized as "various types of wordstrings" which appear to be "prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray and Perkins 2000:

[1] These claims accord with the view of restricted languages, represented across multiple text types and genres, each with their own micro-grammar and micro-glossary, originally put forward by the linguist John Firth, as noted in Holtz (2011: 19).

1; Wray 2002: 9). Furthermore, Wray (2002: 75) argues that the use of formulas is related to the demands of the genre or types of activities that require the use of particular genres. This makes language users primed to use certain formulas more often than others depending on social situations; in fact, formulaic sequences are commonplace in language use, constitute a large proportion of discourse and occur in "so many forms that it is presently difficult to develop a comprehensive definition of the phenomenon" (Schmitt and Carter 2004: 1–2). In view of this, it seems highly unrealistic to impose a single research paradigm on the study of formulaicity in language (Read and Nation 2004: 23).

One of many theoretical perspectives or research paradigms applied to the study of formulaicity in language is phraseology, which encompasses a wide spectrum of more or less fixed combinations of words referred to as set phrases, proverbs, fixed expressions up to formulaic sentences and entire texts (Burger 2007). The present study is conducted from a corpus linguistic perspective and what corpus linguists studying phraseology are mostly interested in are frequent and statistically significant multi-word patterns in which particular words occur (Moon 2007: 1046). For many reasons described in the following paragraphs, lexical bundles (LBs), the concept originally proposed by Biber et al. (1999), have been particularly useful in the phraseological research on fixed expressions.

LBs are described as sequences of three or more words that occur frequently in natural discourse and constitute lexical building blocks used frequently by language users in different situational and communicative contexts, e.g. *I don't think*, *as a result*, *the nature of the, as well as* etc. (Biber et al. 1999: 990–991). Typically, LBs are not idiomatic in meaning and not perceptually salient; on the contrary, the meaning of an LB is transparent from the individual words contained in it (Biber 2006: 134). According to Biber (2006: 174), "the functions and meanings expressed by these lexical bundles differ dramatically across registers and academic disciplines, depending on the typical purposes of each". Stubbs and Barth (2003: 81) argue that recurrent LBs, which they refer to as 'chains', are "not necessarily linguistic or psycholinguistic units": some LBs are not complete syntactic units yet they may contain one; some strongly predict a complete syntactic unit; some are not necessarily pre-constructed. Also Kopaczyk (2012: 5) underlines that the lexical bundles approach is not limited to exploration of phrasal constituents; conversely, the uninterrupted sequences of words extracted from corpora are often either smaller or larger than a phrase. According to Goźdź-Roszkowski (2011: 44), the lexical bundles methodology has refined traditional phraseological research in that it is based on the analysis of frequencies of fixed sequences of word forms without taking into consideration their grammatical structure, or – to put it simply – their form.

Although operationalized in different ways and only implicitly grounded in corpus linguistics, the concepts similar to LBs have also appeared in Polish phraseology. For example, a phraseme (*frazem*) is defined by Chlebda (2003: 52) as a linguistic sign (irrespective of its semantic status and form) that is a ready-made chunk of language reproduced by language users as a verbalizer of their propositions, notions, intentions etc. Consequently, the discipline of phrasematics (*frazematyka*) studies various forms of linguistic signs with the main common denominator, namely their reproducibility and recurrence in specific contexts of language use (Chlebda 2003: 49–52). In a similar vein, Chlebda (2009, 2010) proposes – chiefly for lexicographic purposes – the concept of a reproduct (*reprodukt*), where one can find a mention of using some research methods typical of corpus linguistics. Defined as linguistic units (components of the language system of an ethnic language) isolated from texts following verification of their regular repeated occurrence, reproducts play the role of verbalizers of specific content, e.g. a notion, proposition, intention, emotion etc. (Chlebda 2010: 15–16, 140). Thus, a single-word or multi-word unit is a reproduct only if its form is regularly associated with invariable and identifiable discourse functions and as such it is frequently and repeatedly used in texts. Although Chlebda does not operationalize any frequency threshold, considering this to be a purely arbitrary matter, he argues that the higher the frequency of a given reproduct and the higher the number of texts in which it occurs, the more probably it is the unit of language system and hence will be recorded in a dictionary (Chlebda 2009: 20).

The assumption at the heart of this paper is that LBs can be treated as a subcategory of phrasemes and/or reproducts. In fact, the frequent occurrence of LBs is due to their fulfilling specific discourse functions in texts. According to Kilgarriff (2005: 263–264), one has now access to ample empirical evidence revealing that "language users never choose words randomly, and language is essentially non-random" because people tend to speak or write with purposes in mind. Therefore, the aim of this paper is to identify the most frequent LBs in the corpus of Polish patient information leaflets (PILs) and then to explore their discourse functions, capitalizing on the functional typology originally proposed by Biber, Conrad and Cortes (2004). This will make it possible to determine the links between situational, linguistic and functional features of PILs, the approach that falls within the framework of register analysis described by Biber and Conrad (2009: 51–81). Prior to that, however, a short characteristics of the PILs genre is presented in the following section, largely based on the framework for description of genres and registers proposed by Biber and Conrad (2009: 37–47).

2. Patient information leaflets: a brief characteristics

In this paper, the emphasis is put on the description of the most frequent lexical bundles in patient information leaflets (PILs) written originally in Polish. PILs are found in sales packages of medicines and they are written in the language of the country where the medicines are sold, in this case – the Polish language. Nowadays, PILs are produced (or computer-generated) by pharmaceutical companies in accordance with guidelines issued by regulatory authorities (e.g. *Urząd Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych* 'The Office for Registration of Medicinal Products, Medical Devices and Biocidal Products' in Poland (URPL)), as well as with more specific guidelines, pursuant to Article 65 of Directive 2001/83/EC, issued by the European Medicines Agency (*Europejska Agencja Leków* (EMA)) applicable to pharmaceutical companies selling their products on the territory of the European Union (Montalt Resurrecio and Gonzalez Davies 2007: 68).

PILs have a specific institutional addressor (a pharmaceutical company) and a singular addressee (a patient/consumer of a medicine or medicinal product, pharmacist, nurse, general practitioner etc.). However, this pharmaceutical text variety also has intermediate users such as regulatory authorities. The main communicative purpose of PILs is to provide specific information on proper and safe use and administration of medicines. Needless to say, PILs contain information about medical conditions, doses and side effects associated with the use of medicines (Montalt Resurrecio and Gonzalez Davies 2007: 69). Since PILs are primarily targeted at consumers of medicines, they should be written in a plain and user-friendly style.[2] Intended to facilitate this goal, specific linguistic guidelines are stipulated in the Polish Health Minister's Regulation of 20 February 2009 regarding requirements for labeling packages of medicinal products and contents of patient information leaflet.[3] It is often the case that different manufacturers of medicines give different titles to PILs' sections in order to make them easier to read and comprehend; also, technical terms are frequently accompanied by or substituted with explanations or deleted altogether (Montalt Resurrecio and Gonzalez Davies 2007: 71; Cacchiani 2006: 33).

---

[2] Article 63(2) of Directive 2001/83/EC reads as follows: "The package leaflet must be written in clear and understandable terms for the users and be clearly legible in the official language or languages of the Member State where the medicinal product is placed on the market."

[3] Rozporządzenie Ministra Zdrowia z dnia 20 lutego w sprawie wymagań dotyczących oznakowania opakowań produktu leczniczego i treści ulotki (Dz. U. Nr 39 poz. 321). Attachment 2 Section IV Point 3 of the said regulation reads as follows: "One shall use active voice and verbs in infinitive form. If possible, one shall avoid long sentences (i.e. longer than 20 words). It is recommended that a line has no more than 70 letters. One shall avoid complex sentences (…). One shall avoid abbreviations. Name of a medicinal product may be substituted with a pronoun (e.g. *on* 'he') as long as its referent is clear from the context" (translated by ŁG).

As regards their macro-structure, PILs are highly conventionalized in that they follow a standard form for every medicinal product and provide the same types of information; however, different pharmaceutical companies present this information in different order. Montalt Resurrecio and Gonzalez Davies (2007: 70–71) describe the scope of information provided in PILs, not limited to posology and method of administration. It also includes specification of the name of a product, specification of its pharmaceutical form, specification of the marketing authorization holder, description of the composition of a medicine, indications, contraindications etc.

Although PILs constitute one of the most commonly used specialist text type by doctors, pharmacists, nurses and patients in Poland, one may note the scarcity of studies devoted to description of the use and functions of the most frequent phraseologies in this pharmaceutical text variety. In fact, there have been many corpus linguistic studies exploring linguistic variation in PILs originally written in English and other languages (e.g. Italian), conducted by Paiva (2000), Cacchiani (2006), Clerehan, Hirsh and Buchbinder (2009) and Grabowski (2013), among others. This observation provided motivation to undertake a preliminary corpus-driven study like this one using Polish language material.[4] More specifically, the aim of this paper is the identification of the most frequent LBs found in a sample of PILs originally produced in the Polish language. In order to explain how to achieve this goal, the research material and methodology are described in greater detail in the following section.

3. Research material and methodology

This study is conducted from a register perspective (Biber and Conrad 2009: 51–81), which means that apart from the identification of the most frequent LBs, the research is aimed to explain their frequent use in PILs. To that end, the empirical corpus-driven identification of LBs and further concordance-based analyses of their use provide primary evidence for the description of phraseologies in PILs. The research material encompasses a corpus of 100 PILs (full-texts) produced by 10 pharmaceutical companies operating on the Polish market (AstraZeneca Polska, Boehringer Ingelheim, Egis, Farmapol, Krka, Lundbeck, Polfa Warszawa, Sanofi, Takeda, and Teva).[5] The only information

[4] This paper has been inspired by the ongoing project "Lexical and Phraseological Variation Across English Pharmaceutical Texts: A Corpus Linguistics Study" conducted by the author and financed by the National Science Centre (*Narodowe Centrum Nauki*) pursuant to a decision no. DEC- 2013/09/D/HS2/00543.

[5] The research material was retrieved on 18–24 April 2014 from the following websites: http://www.astrazeneca.pl/zdrowie-i-choroby/leki-az; http://www.boehringer-ingelheim.pl/produkty/leki_na_recepte.html; http://www.egis.pl/24/nasze-leki; http://www.farmapol.pl/produkty/;

deleted from the PILs were the addresses of the companies' headquarters in those European countries where medicines are available for sale. All in all, the corpus' size is 196,757 word tokens (12,754 word types) and each company is represented by 10 PILs in the corpus. As mentioned earlier, a radical corpus-driven approach is used in this study. This means that the data contained in the corpus are not adjusted to fit any predefined categories or theories of language description. On the contrary, the empirical bottom-up corpus analysis of frequency distributions of recurrent linguistic items, exemplified by LBs, provides evidence for the description of phraseologies in PILs. Such an approach enables one to pinpoint the most frequent phraseologies in a more objective way as compared with the intuition-based approach, which prioritizes unusual and rare linguistic items. The study is conducted in a number of stages discussed below.

Firstly, the LBs in PILs are generated with the help of the purpose-designed software for text analysis Wordsmith Tools 4.0 (Scott 2007). However, in order to obtain an analyzable and representative sample of LBs, it is required that specific selection criteria be determined. The most important criteria and their parameters discussed by Biber et al. (1999); Biber, Conrad and Cortes (2003, 2004) and Biber (2006) are the following: determination of a length of an LBs (to determine the scope of the analysis); a frequency cut-off point (to limit the number of LBs subjected to the analyses), which is a normalized frequency of occurrence of an LB per 1 million words; a number of texts in which an LB must occur (to remove idiosyncratic LBs from the analyses). Consequently, the methods used to derive LBs are quantitative (frequency-driven) ones and, as it is shown by a number of studies (e.g. Biber et al. 1999, 2003, 2004; Biber 2006; Hyland 2008; Chen and Baker 2010; Goźdź-Roszkowski 2011; Adel and Erman 2012, to name but a few), the parameters used to identify lexical bundles are arbitrarily set. As for the length of LBs, this study focuses on 5-word and 6-word LBs only. This is due to the fact that PILs constitute a highly patterned and formulaic text variety and therefore one may expect to find a high number of prefabricated chunks or strings of words. Assuming that the longer the string, the lower its frequency of occurrence, it is believed that focusing on 5-word and 6-word LBs will yield more manageable dataset for the analysis, as compared with 3-word or 4-word LBs that are bound to be more numerous and more frequent in texts. Regarding the criterion of a frequency cut-off point, the study focuses on high-frequency LBs only and hence the frequency threshold is set at 200 occurrences per 1 million words for 5-word LBs and 150 occurrences per million words for 6-word LBs. Finally, as regards the criterion

---

http://www.krkapolska.pl/pl/; http://www.lundbeck.com/pl/dla-pacjentow-i-rodzin/ulotki-dla-pacjentow; http://www.polfawar.com.pl/katalog/kategoria/bez-recepty; http://www.sanofi.pl/l/pl/pl/layout.jsp?scat=1C6DC549-B4AE-4C04-B375-41127592C70E; http://www.takeda.com.pl/products/healthcare-professionals/; http://teva.pl/u235/navi/31243.

of distribution range (i.e. a minimum number of texts in which a given contiguous sequence of words must occur in order to be considered an LB), only those LBs which occur in at least 25% of PILs (i.e. in 25 PILs) are analyzed in this study. Importantly, those sequences of five or six words which go over sentence boundaries (i.e. are divided by full-stops or semi-colons) or contain numbers are not analyzed. In practice, application of the said parameters to the selection criteria resulted in a relatively high number of 5-word and 6-word LBs (278 and 295, respectively), a sample not amenable to detailed qualitative investigation. Consequently, additional exclusion criteria, based on the ones applied by Chen and Baker (2010: 33) and Salazar (2011: 48–50), were applied in order to further limit the number of LBs. More specifically, those LBs which occurred on the clause or phrase boundaries (e.g. *tę ulotkę aby w* 'this leaflet in order to', *leku należy poradzić się* 'medicine one should consult', *są takie same jeśli* 'are the same if' etc.) were deleted from the list. Likewise, those LBs which are fragments of other, typically longer LBs, have been eliminated from the analyses (e.g. *się z treścią ulotki* 'self with the contents of the leaflet', *w jakim celu się go* 'what is the purpose of' etc.). Also, all LBs ending in prepositions or conjunctions were dropped. Finally, the LBs with acronyms or abbreviations were also deleted from the list (e.g. *Sp. z o. o.*, an abbreviation for a private limited company). After this stage of filtering out the data, nineteen 5-word LBs (Table 1) and twenty four 6-word LBs (Table 2) have been eventually selected for the analyses (i.e. 43 LBs altogether).

Table 1. 5-word LBs in PILs

No.	5-word LBs in PILs	Frequency	Frequency pmw	*n* texts
1	*należy zwrócić się do lekarza* 'should turn to/contact a doctor'	249	1265	99
2	*informacje ważne przed zastosowaniem leku* 'information important before taking a medicine'	164	833	82
3	*należy powiedzieć o tym lekarzowi* 'should tell a doctor about it'	139	706	77
4	*należy skontaktować się z lekarzem* 'should contact a doctor'	128	650	65
5	*objawy niepożądane niewymienione w ulotce* 'undesirable side effects not listed in the leaflet'	115	584	75
6	*zawartość opakowania i inne informacje* 'contents of the pack and other information'	91	462	46

No.	5-word LBs in PILs	Frequency	Frequency pmw	*n* texts
7	*nie należy wyrzucać do kanalizacji* 'should not be thrown into sewage system'	77	391	77
8	*poradzić się lekarza lub farmaceuty* 'consult a doctor or pharmacist'	73	371	61
9	*prowadzenie pojazdów i obsługiwanie maszyn* 'driving vehicles and operating machines'	71	361	69
10	*takie postępowanie pomoże chronić środowisko* 'such conduct will help protect the environment'	70	356	70
11	*nie należy stosować dawki podwójnej* 'should not take a double dose'	66	335	66
12	*w celu uzupełnienia pominiętej dawki* 'in order to make up for a missed dose'	64	325	64
13	*jeśli wystąpią jakiekolwiek objawy niepożądane* 'if any undesirable side effects occur'	63	320	43
14	*lek może zaszkodzić innej osobie* 'a medicine may do harm to another person'	62	315	62
15	*należy powiedzieć lekarzowi lub farmaceucie* 'should tell a doctor or pharmacist'	58	295	56
16	*zachować szczególną ostrożność stosując lek* 'act with great caution while taking a medicine'	58	295	47
17	*niezwłocznie skontaktować się z lekarzem* 'immediately contact a doctor'	33	168	25
18	*jeśli pacjentka jest w ciąży* 'if a patient is pregnant'	31	157	28
19	*należy przechowywać w miejscu niewidocznym* 'should be kept in a hidden place'	31	157	31

Table 2. 6-word LBs in PILs

No.	6-word LBs in PILs	Frequency	Frequency pmw	n texts
1	*w jakim celu się go stosuje* 'what is the purpose of taking it'	199	1011	100
2	*zwrócić się do lekarza lub farmaceuty* 'turn to/contact a doctor or pharmacist'	188	955	91

No.	6-word LBs in PILs	Frequency	Frequency pmw	n texts
3	*jeśli nasili się którykolwiek z objawów* 'if any symptoms intensify'	105	533	61
4	*powiedzieć o tym lekarzowi lub farmaceucie* 'tell a doctor or pharmacist about it'	100	508	65
5	*chociaż nie u każdego one wystąpią* 'although they do not occur in everyone'	92	467	92
6	*zastosowanie większej niż zalecana dawki leku* 'taking a larger dose of medicine than recommended'	64	325	64
7	*w celu uzyskania bardziej szczegółowych informacji* 'in order to obtain more detailed information'	61	310	61
8	*lek ten przepisano ściśle określonej osobie* 'this medicine has been prescribed for a specific person'	59	300	59
9	*przechowywać w miejscu niedostępnym i niewidocznym* 'keep in a safe and hidden place'	57	290	57
10	*skontaktować się z lekarzem lub farmaceutą* 'contact a doctor or pharmacist'	56	285	47
11	*zawiera ona informacje ważne dla pacjenta* 'contains information important to a patient'	56	285	56
12	*należy natychmiast skontaktować się z lekarzem* 'should immediately contact a doctor'	54	274	32
13	*zawsze stosować zgodnie z zaleceniami lekarza* 'always take in accordance with doctor's guidelines'	53	269	53
14	*należy zapoznać się z treścią ulotki* 'should familiarize oneself with the contents of the leaflet'	51	259	51
15	*kiedy zachować szczególną ostrożność stosując lek* 'when to act with great caution while taking a medicine'	44	223	38
16	*przechowywać w miejscu niedostępnym i niewidocznym* 'keep in a safe and hidden place'	42	213	42

No.	6-word LBs in PILs	Frequency	Frequency pmw	n texts
17	*lek ten może powodować działania niepożądane* 'this medicine may cause undesirable side-effects'	41	208	41
18	*w tym wszelkie możliwe objawy niepożądane* 'including any possible undesirable side effects'	36	183	28
19	*zwrócić się do przedstawiciela podmiotu odpowiedzialnego* 'turn to/contact a representative of a responsible subject/marketing authorization holder'	36	183	36
20	*ważne informacje o niektórych składnikach leku* 'important information about certain ingredients of a medicine'	35	178	31
21	*uważnie zapoznać się z treścią ulotki* 'carefully familiarize oneself with the contents of the leaflet'	33	168	33
22	*lek należy przechowywać w miejscu niewidocznym* 'medicine should be kept in a hidden place'	30	152	30
23	*pacjent powinien skontaktować się z lekarzem* 'patient should contact a doctor'	30	152	29
24	*na stronie internetowej Europejskiej Agencji Leków* 'on the website of European Medicines Agency'	27	137	27

Finally, in order to illustrate the relationship between situational contexts of use of PILS on the one hand, and the frequent use of LBs on the other, forty three top-frequency LBs, presented in Table 1 and 2 above, are further subjected to the functional interpretation. More specifically, the functional typology largely based on the one originally proposed by Biber, Conrad and Cortes (2004) and Biber (2006) is applied so that the LBs are divided into three inclusive functional categories, namely *referential*, *discoursal* and *expressing stance*.[6] Also, they are further subdivided into more specific subcategories to accommodate more fine-grained functional distinctions specific to Polish PILs. However, as Wray and Perkins (2000: 8) note, any specific functional tax-

[6] The typology in question was later used – with minor modifications of labels applied to more fine-grained functional distinctions – by, among others, Hyland (2008); Chen and Baker (2010); Jablonkai (2010); Goźdź-Roszkowski (2011); Adel and Erman (2012), in explorations of LBs used across academic or legal text varieties.

onomy for recurrent multi-word units is bound to suffer from proliferation of types and subtypes, a situation not conducive to distilling the data into a compact functional model applicable across corpora representing various domains of language use. Importantly, the functional labels assigned by the author to the LBs represent tendencies or approximations foregrounding their primary functions fulfilled in the majority of contexts in which the LBs occur. In view of this, it is justified that the functional categories originally proposed by Biber, Conrad and Cortes (2004) be treated flexibly rather than strictly and that the insights from other functional taxonomies (Hyland 2008; Goźdź-Roszkowski 2011) be used in the course of the analyses, notably when applied to linguistic data produced in a language other than English.

In the case of LBs, their use and functions are commented upon in greater detail since the functional analysis is supplemented by concordances illustrating the actual uses of these linguistic items in the PILs under scrutiny. All in all, it is hypothesized in this study that the frequent use of LBs in the PILs originally written in Polish is due to discipline-specific practices as well as situational contexts of use and communicative functions of this text type. Therefore, this paper aims to determine whether there are any LBs repeatedly used in PILs, what discourse functions of the most frequent LBs in PILs are, and whether the discourse functions of those LBs are linked with situational contexts and communicative purposes of the text type. The results are presented in the following empirical section.

4. Empirical part: the use and functions of LBs in PILs

As mentioned earlier in the paper, only a sample of forty three 5-word and 6-word LBs is further interpreted functionally and divided into three general categories, namely *referential LBs*, *discoursal LBs* and *stance LBs* (Table 3). The following sections explicate their discourse roles.

Table 3. General functional classification of LBs (43) in PILs

Functional categories	PILs
Referential LBs	6
Discoursal LBs	8
Stance LBs	29

4.1. Referential bundles in PILs

Six referential LBs identified in PILs are used to refer to abstract or physical entities or to identify an attribute of an entity as particularly important (Biber et al. 2004: 384), such as information about side-effects, contents of the packaging, information on indications or counter-indications etc. In this group, three functional subcategories are distinguished in Polish PILs, namely identification/focus LBs (4), temporal LB (1) and location LBs (1).

The identification/focus LBs (4), also referred to by Hyland (2008: 13) as topic-bundles, are typically extended noun phrases that identify key aspects of the use and properties of medicines or key information contained in the PILs, e.g. *objawy niepożądane niewymienione w ulotce* 'undesirable side effects not listed in the leaflet', *zawartość opakowania i inne informacje* 'contents of the pack and other information', *prowadzenie pojazdów i obsługiwanie maszyn* 'driving vehicles and operating machines', *zastosowanie większej niż zalecana dawki leku* [trade name of a medicine] 'taking a larger dose of medicine than recommended', as in the following:

1. *Jeśli wystąpią jakiekolwiek objawy niepożądane w tym wszelkie możliwe* ***objawy niepożądane niewymienione w ulotce****, należy zwrócić się do lekarza lub farmaceuty.* (PIL Escitil, 5/10/15/20mg)
   'If any undesirable side-effects occur, including any possible undesirable side-effects not listed in the leaflet, one should turn to a doctor or pharmacist'
2. ***Zastosowanie większej niż zalecana dawki leku*** *Forxiga* (PIL Forxiga_5/10mg)
   'Taking a larger dose of Forxiga than recommended'

Next, the analysis revealed one temporal LB, namely *kiedy zachować szczególną ostrożność stosując lek* [trade name of a medicine] 'when to act with great caution while taking the medicine'. Its main function, however, is not limited to instructing patients on when to properly use or administer medicines, but it is primarily to alert consumers of medicines to any precautions before taking medicines, as illustrated below:

3. ***Kiedy zachować szczególną ostrożność stosując lek*** *Zincas* (PIL Zincas)
   'When to act with great caution while taking Zincas medicine'

Finally, a location LB *na stronie internetowej Europejskiej Agencji Leków* 'on the website of European Medicines Agency' refers to a place on the world wide web where one may find more specific information on medicines described in PILs. In fact, pharmaceutical companies are required to submit to the European Medicines Agency (EMA) documents known as Summaries of Product

Characteristics (*Skrócona charakterystyka produktu leczniczego*), describing pharmacological, chemical, pharmaceutical, toxicological and other properties of medicines. This is necessary in order to commercialize any medicine on the territory of the European Union. Since this information is publicly available on the EMA website, the LB in question is often found (137 occurrences pmw in 27 texts) in PILs, as in:

4. *Szczegółowa informacja o tym leku jest dostępna* ***na stronie internetowej Europejskiej Agencji Leków*** *http://www.ema.europa.eu/* (PIL Abilify_7.5mg/ml)
   'Detailed information about this medicine is available on the website of European Medicines Agency *http://www.ema.europa.eu/*'

4.2. Discoursal bundles in PILs

The category of discoursal LBs is represented by eight items concerned with text organization (Biber et al. 2004: 391), also referred to by Hyland (2008: 13) as text-oriented bundles. They express several different textual functions, such as elaborating on a given topic, expressing emphasis, introducing conditions, increasing or decreasing writer's distance from information conveyed in texts etc. In this group, four functional subcategories are distinguished, namely condition LBs (3), purpose bundles (3), focus LBs (1) and concession LBs (1).

As regards condition LBs (*jeśli wystąpią jakiekolwiek objawy niepożądane* 'if any undesirable side effects occur', *jeśli pacjentka jest w ciąży* 'if a patient is pregnant', *jeśli nasili się którykolwiek z objawów* 'if any symptoms intensify'), these are usually conditional clauses, with the conjunction *jeśli* 'if' used in the sentence-initial position, followed by the main clause with verbs expressing directives. The LBs found in this group introduce various conditions that may arise while using medicines, such as patients' questions or doubts, occurrence of any side effects etc. Also, condition LBs help convey special warnings and precautions, encourage patients to consult a doctor or pharmacist etc., e.g.:

5. ***Jeśli wystąpią jakiekolwiek objawy niepożądane****, w tym wszelkie objawy niepożądane niewymienione w ulotce, należy powiedzieć o tym lekarzowi, lub farmaceucie.* (PIL Casodex_50mg)
   'If any undesirable side effects occur, including those not listed in the leaflet, one should tell a doctor or pharmacist about it'
   Jeśli pacjentka jest w ciąży *lub myśli, że może być w ciąży, powinna poinformować o tym lekarza prowadzącego.* (PIL Clopixol_10/25mg)
   'If a patient is pregnant or thinks that may be pregnant, she should tell a doctor in charge of the case about it'

The next category are purpose bundles (3). Also referred to as causative-resultative bundles, they are typically used to signal links in terms of cause and effect between arguments in a text (Goźdź-Roszkowski 2011: 130). Centred around the connective *w celu* 'in order to', the purpose clauses in the Polish PILs are often related to various conditions that may arise while using medicines, to administration of medicines to patients or to commencement of a particular drug therapy. The LBs in this group include items such as *w celu uzupełnienia pominiętej dawki* 'in order to make up for a missed dose', *w jakim celu się go stosuje* 'what is the purpose of taking it', *w celu uzyskania bardziej szczegółowych informacji* 'in order to obtain more detailed information', as in:

7. *Nie należy stosować dawki podwójnej* ***w celu uzupełnienia pominiętej dawki****.* (PIL Dasselta_5mg)
   'One should not take a double-dose in order to make up for a missed dose'
8. *Co to jest lek Nolvadex D i* ***w jakim celu się go stosuje*** (PIL Nolvadex D_20mg)
   'What is Nolvadex D and what is the purpose of taking it'
9. ***W celu uzyskania bardziej szczegółowych informacji*** *należy zwrócić się do miejscowego przedstawiciela podmiotu odpowiedzialnego.* (PIL Ebixa_20mg, PIL Fasturtec_1.5mg/ml)
   'In order to obtain more detailed information, one should contact a representative of a responsible subject/marketing authorization holder'

Next, a focus LB (*w tym wszelkie możliwe objawy niepożądane* 'including any possible undesirable side effects') attempts to directly engage readers in the new problem signalled in the text, typically revolving around potential side-effects, as in:

10. *Jeśli wystąpią jakiekolwiek objawy niepożądane,* ***w tym wszelkie możliwe objawy niepożądane*** *niewymienione w ulotce, należy zwrócić się do lekarza, farmaceuty lub pielęgniarki.* (PIL Kwetax_25mg, PIL)
    'If any undesirable side effects occur, including any possible undesirable side-effects not listed in the leaflet, one should contact doctor, a pharmacist, or a nurse'

Finally, a concession LB (*chociaż nie u każdego wystąpią* 'although they do not occur in everyone'), a clause starting with the adversative conjunction *chociaż* 'although', is used in PILs to increase and emphasize the distance of the writer from a specific piece of information, in most cases referring to side-effects that may arise as a result of using a particular medicine, as in:

11. *Jak każdy lek, lek ten może powodować działania niepożądane,* ***chociaż nie u każdego one wystąpią****.* (PIL Lantus_100 units/ml)

'As with any other medicine, this medicine may cause undesirable side-effects, although they do not occur in everyone'

4.3. Stance bundles in PILs

Among forty three 5-word and 6-word top-frequency LBs, there are twenty nine stance LBs in the Polish PILs, making them the largest functional group in this pharmaceutical register. These LBs are typically used as a means of expressing PILs writers' attitudes, value judgments or assessments, providing a frame for interpretation of the following proposition (Biber et al. 1999: 966; Biber, Conrad and Cortes 2004: 384). In this category, one may distinguish between epistemic stance LBs (4), attitudinal/evaluative LBs (3), desire LBs (1) and obligation/directive LBs (21).

In general terms, epistemic stance LBs signal writers' certainty or uncertainty about the truth or knowledge status of the following or preceding proposition.[7] In PILs, epistemic stance LBs (*lek może zaszkodzić innej osobie* 'a medicine may do harm to another person', *lek ten może powodować działania niepożądane* 'this medicine may cause undesirable side-effects', *lek ten przepisano ściśle określonej osobie* 'this medicine has been prescribed for a specific person', *takie postępowanie pomoże ochronić środowisko* 'such conduct will help protect the environment') signal manufacturers' opinions about the activity of particular medicines or about the consequences of improper use of the medicines. Oftentimes, these stance LBs contain a modal verb *może* 'may', as in:

12. ***Lek ten przepisano ściśle określonej osobie****. Nie należy go przekazywać innym.* ***Lek może zaszkodzić innej osobie****, nawet jeśli objawy jej choroby są takie same.* (PIL Pradaxa_150mg, PIL Trajenta_5mg)
    'This medicine has been prescribed for a specific person. One should not give it to other people. This medicine may do harm to another person, even if they have the same symptoms'
13. *Należy zapytać farmaceutę, jak usunąć leki, których się już nie używa.* ***Takie postępowanie pomoże chronić środowisko.*** (PIL Trajenta_5mg, PIL Wolarex_25/50mg)
    'One should ask a pharmacist how to dispose of medicines no longer used/required. Such conduct will help protect the environment'

The next subcategory, namely attitudinal/evaluative LBs (3), emphasizes the importance of the contents of the following proposition, e.g. *zawiera ona informacje ważne dla pacjenta* 'contains information important to a patient', *informacje ważne przed zastosowaniem leku* [trade name of a medicine] 'information important before taking the medicine', *ważne informacje o niektórych*

[7] Goźdź-Roszkowski 2011: 138.

składnikach leku [trade name of a medicine] 'important information about certain ingredients of the medicine'. Hence, centred around an adjective *ważne* 'important', these LBs help raise awareness of patients about the importance of following doctors' instructions and/or manufacturers' guidelines on how to properly take medicines. Also, the LBs in this group help one ensure that patients read PILs carefully. Typically used in headings or in the sentence-initial position, these items are conspicuous and perceptually salient in the Polish PILs under scrutiny, as in:

14. *Należy przeczytać uważnie całą ulotkę, ponieważ* ***zawiera ona informacje ważne dla pacjenta***. (PIL Zincas)
    'One should read carefully the entire leaflet as it contains information important to a patient'
15. ***Informacje ważne przed zastosowaniem leku*** *Zoladex* (PIL Zoladex_3.6mg)
    'Information important before taking Zoladex medicine'
16. ***Ważne informacje o niektórych składnikach*** *leku Arava* (PIL Arava_10mg)
    'Important information about certain ingredients of Arava medicine'

As for desire LBs (1), represented by a single item *pacjent powinien skontaktować się z lekarzem* 'patient should contact a doctor', centred around the modal verb *powinien* 'should' expressing epistemic modality, it is used to express manufacturers' wishes as to a desirable course of action undertaken by patients in the event of any problems arising from the use of medicines, as in:

17. *Jeśli u pacjenta stwierdzono wcześniej nietolerancję niektórych cukrów, przed rozpoczęciem przyjmowania leku* ***pacjent powinien skontaktować się z lekarzem***. (PIL Gliclada_30mg)
    'If a patient had been diagnosed with intolerance to certain sugars, the patient should contact a doctor before starting on a medicine'

Finally, the remaining and at the same time the biggest subcategory of stance LBs in PILs, are obligation/directive LBs (21) starting with either the modal verb *należy* 'should' followed by action verbs or with action verbs in their infinitive form (*zwrócić się* 'turn to/contact', *powiedzieć* 'tell', *skontaktować się* 'contact', *zapoznać się* 'familiarize oneself', *przechowywać* 'keep', *stosować* 'use/take' etc.). Oftentimes, for greater emphasis, these constructions include adverbs of time or manner, such as *zawsze* 'always', *niezwłocznie* 'immediately' or *uważnie* 'carefully'. The main function of these LBs is to direct patients to carry out specific actions that manufacturers of medicines want them to complete in the event of any problems with the use of medicines (e.g. directing patients to read PILs care-

fully, requesting them to consult doctors or pharmacists, instructing patients on recommended methods of disposal of medicines etc.), as in:

18. *Leków **nie należy wyrzucać do kanalizacji** lub domowych pojemników na odpadki.* (PIL Gliclada_30mg)
'Medicines should not be thrown into sewage system or waste containers at home'
19. *Lek Hydroxyzinum Teva należy **zawsze stosować zgodnie z zaleceniami lekarza**. W przypadku wątpliwości należy ponownie **skontaktować się z lekarzem lub farmaceutą**.* (PIL Hydroxizinum Teva_50mg/ml)
'Hydroxyzinum Teva should always be taken in accordance with doctor's guidelines. In case of doubt, contact a doctor or pharmacist again'
20. *Jeśli u pacjenta kiedykolwiek wystąpią myśli samobójcze lub myśli o samookaleczeniu, należy **niezwłocznie skontaktować się z lekarzem** lub zgłosić się do szpitala.* (PIL Kwetax_25mg)
'If at any time a patient has suicidal thoughts or is thinking about self-mutilation, one should immediately contact a doctor or report to the hospital'
21. *Należy **uważnie zapoznać się z treścią ulotki** przed zastosowaniem leku, ponieważ zawiera ona informacje ważne dla pacjenta.* (PIL Jevtana_60mg)
'One should carefully familiarize oneself with the contents of the leaflet as it contains information important to a patient'

5. Conclusions and implications for the future

This descriptive and exploratory study interfaces certain theoretical concepts from the field of register analysis and phraseology with selected elements of corpus linguistic methodology, namely the corpus-driven approach. The research therefore falls within the scope of frequency-driven or distributional phraseology (Granger and Meunier 2008; Pęzik 2013), or corpus linguistics phraseology. This research paradigm may be described as a bottom-up, inductive, empirical study of the use, meaning and function of various types of contiguous or non-contiguous recurrent combinations of words, which are typically, but not limited to, incomplete grammatical units with compositional meaning, retrieved automatically or semi-automatically from texts or corpora with the help of purpose-designed computer software and by using quantitative frequency-driven methods. Corpus linguistics phraseology can help one discover and test new methods of identification of multi-word units as well as obtain new facts about the use, distribution and function of recurrent co-occurring sequences of words, with emphasis on regularities in their

use and distribution rather than irregularities or rare uses. Although in recent years researchers have gained access to more and more language corpora with Polish texts (e.g. the National Corpus of Polish) and to various tools and resources developed alongside the corpora (Przepiórkowski et al. 2012), it has to be noted, regrettably, that corpus linguistic research on phraseology in texts originally written in Polish is still scarce. This preliminary paper is therefore intended to be a small step to change this state of affairs as well as to inspire further research in the future.

The main goal of this empirical study was to provide a comprehensive corpus-driven description of the use and function of the most frequent LBs found in a purpose-designed corpus of 100 Polish patient information leaflets, extracted from internet websites of ten pharmaceutical companies operating on the Polish market. Also, the study aimed to test the methodology used for extraction of LBs from highly patterned and formulaic texts produced in the Polish language. Finally, the study aimed to qualitatively examine discourse functions fulfilled by the LBs identified in the course of the analyses. The results provided an outline of a phraseological profile of PILs and revealed salient links between the communicative purpose and situational contexts of the use of PILs on the one hand, and the use and functions of the most frequent LBs on the other. Finally, it was found that the LBs explored in this study help authors of PILs fulfil the main communicative functions of this text variety. Hence, the LBs identified in this study shall be treated as important building blocks of this register.

All in all, forty three LBs were qualitatively explored in the course of this exploratory and descriptive study and the results revealed that the PILs are dominated by stance LBs (29) describing attitudes towards the actions and events described in the following proposition, such as obligation/directive LBs (21), epistemic stance LBs (4), attitudinal evaluative LBs (3) and desire LBs (1). Such a high number of stance LBs in PILs results from the communicative function of this text variety, namely of alerting patients to potential side-effects as well as of instructing them on how to properly use medicines or what course of action to pursue in the event of any problems arising from the use of medicines. These functions are also fulfilled by condition LBs and purpose LBs, representing the subcategories of discoursal LBs. The results are discussed in greater detail in the empirical part of this paper.

Importantly, however, the results shall be interpreted with caution as they apply only to the texts analyzed in this research, rather than to Polish patient information leaflets in their totality. Hence, if an LB has not been found in the study corpus or identified in the course of the analyses, it does not automatically mean that it is unacceptable or not used in other PILs than the ones analyzed in this study. Similar problems with interpreting linguistic data extracted from corpora are briefly discussed by Piotrowski and Grabowski (2013).

There are many possible ways in which this study can be pursued further in the future. Firstly, it is possible to explore in greater detail text-internal distribution and dispersion of LBs within a macrostructure of the PILs genre. This would provide information on exact positions of these phraseologies, also within the theme–rheme structure in the text. Secondly, in order to extend the scope of description of phraseologies in PILs, it is possible to additionally investigate phrase frames (Fletcher 2007) or concgrams (Cheng et al. 2006; Greaves 2009), two recently proposed approaches to the analysis of non-contiguous multi-word units, providing further generalizations of phraseological patterns found in texts. In fact, both phrase frames and concgrams may constitute an attractive starting point for more detailed explorations of specific co-occurrence patterns, including collocations with specialist terms. The studies conducted by Forchini and Murphy (2008); Biber (2009); Cheng and Leung (2012); Gray and Biber (2013); Fuster-Marquez (2014), among others, show how to successfully apply and operationalize these concepts in research on phraseological variation across a wide variety of registers in the English language. It is particularly tempting to use and test these concepts, including lexical bundles, in further explorations of phraseological patterns across text types or genres written originally in Polish and other than patient information leaflets.

References

Adel Annelie, Erman Britt (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical LBs approach. *English for Specific Purposes* 31, 81–92.

Biber Douglas (2006). *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

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

Biber Douglas, Conrad Susan (2009). *Register, Genre and Style*. Cambridge: Cambridge University Press.

Biber Douglas, Conrad Susan, Cortes Viviana (2003). Lexical LBs in speech and writing: An initial taxonomy. In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Andrew Wilson, Paul Rayson, Tony McEnery (eds.), 71–92. Frankfurt am Main: Peter Lang.

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

Biber Douglas, Johansson Stig, Leech Geoffrey, Conrad Susan, Finegan Edward (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Burger Harald (ed.) (2007). *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung,* Vol. 2. Berlin: Walter de Gruyter.

Cacchiani Silvia (2006). Dis/similiarities between Patient Information Leaflets in Britain and Italy: Implications for the Translator. *New Voices in Translation Studies* 2, 28–43.

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

Cheng Winnie, Leung Maggie (2012). Exploring phraseological variations by concgramming: The realization of complete patterns of variations. *Linguistic Research*, 29(3), 617–638.

Cheng Winnie, Greaves Chris, Warren Martin (2006). From n-gram to skipgram to concgrams. *International Journal of Corpus Linguistics* 11, 411–433.

Chlebda Wojciech (2003). *Elementy frazematyki: wprowadzenie do frazeologii nadawcy*. Łask: Leksem.

Chlebda Wojciech (2009). Idiomatykon 4: gdzie jesteśmy, dokąd zmierzamy (i parę zdań o tym, skąd przychodzimy). In *Podręczny idiomatykon polsko-rosyjski 4*. Wojciech Chlebda (ed.), 9–38. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Chlebda Wojciech (2010). Nieautomatyczne drogi dochodzenia do reproduktów wielowyrazowych". In *Na tropach reproduktów: w poszukiwaniu wielowyrazowych jednostek języka*. Wojciech Chlebda (ed.), 15–35. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Clerehan Rosemary, Hirsch Di, Buchbinder Rachelle (2009). Medication information leaflets for patients: the further validation of an analytic linguistic framework. *Communication & Medicine* 6(2), 117–128.

Fletcher William (2007). KfNgram. Annapolis: USNA. Retrieved from: http://www.kwicfinder.com/kfNgram/kfNgramHelp.html.

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

Fuster-Marquez Miguel (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84–121.

Goźdź-Roszkowski Stanisław (2011). *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Frankfurt: Peter Lang.

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

Granger Sylviane, Meunier Fanny (2008). Introduction: The many faces of phraseology. In *Phraseology: An Interdisciplinary Perspective*. Sylviane Granger, Fanny Meunier (eds.), xix–xxx. Amsterdam: John Benjamins.

Gray Betany, Biber Douglas (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), 109–135.

Greaves Chris (2009). *ConcGram 1.0: A Phraseological Search Engine*. Amsterdam: John Benjamins.

Holtz Monica (2011). Lexico-grammatical properties of abstracts and research articles. A corpus-based study of scientific discourse from multiple disciplines. Unpublished PhD dissertation. Technische Universitaet Darmstadt. [URL: http://

tuprints.ulb.tu-darmstadt.de/2638/1/PhD-Thesis-Monica-Holtz.pdf, accessed: October 23, 2013].

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

Jablonkai Reka (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes* 29, 253–267.

Kilgarriff Adam (2005). Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* 1(2), 263–276.

Kopaczyk Joanna (2012). Long lexical bundles and standardisation in historical legal texts. *Studia Anglica Posnaniensia* 47(2–3), 3–25.

Montalt Resurrecio Vicent, Gonzales Davies Maria (2007). *Medical Translation Step by Step. Translation Practices explained*. Manchester: St. Jerome.

Moon Rosamund (2007). Corpus linguistic aspects of phraseology. In Burger (ed.), 1045–1059.

Paiva Daniel (2000). Investigating style in a corpus of pharmaceutical leaflets: results of a factor analysis. In *Proceedings of the Student Workshop of the 38th Annual Meeting of the ACL*, Hong Kong, China, 1–8 Oct 2000, 52–59. [URL: http://www.itri.brighton.ac.uk/~Daniel.Paiva/acl2000student.finalversion.pdf; accessed: October 08, 2011].

Pęzik Piotr (2013). Paradygmat dystrybucyjny w badaniach frazeologicznych. Powtarzalność, reprodukcja i idiomatyzacja. In *Metodologie Językoznawstwa. Ewolucja Języka, Ewolucja Teorii Językoznawczych*. Piotr Stalmaszczyk (ed.), 143–160. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Piotrowski Tadeusz, Grabowski Łukasz (2013). Interpretacja danych frekwencyjnych z korpusów językowych: opis pewnych problemów (na kilku przykładach z życia wziętych). In *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*. Wojciech Chlebda (ed.), 59–71. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Przepiórkowski Adam, Bańko Mirosław, Górski Rafał, Lewandowska-Tomaszczyk Barbara (eds.) (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: PWN.

Read John, Nation Paul (2004). Measurement of formulaic sequences. In *Formulaic Sequences: Acquisition, Processing and Use*. Norbert Schmitt (ed.), 23–35. Amsterdam: John Benjamins.

Salazar Danica (2011). *Lexical Bundles in Scientific English: A Corpus-based Study of Native and Non-native Writing*. Unpublished PhD dissertation. University of Barcelona. [URL: http://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf; accessed: March 26, 2013].

Schmitt Norbert, Carter Ronald (2004). Formulaic sequences in action: An introduction. In *Formulaic Sequences: Acquisition, Processing and Use*. Norbert Schmitt (ed.), 1–22. Amsterdam: John Benjamins.

Scott Mike (2007). *WordSmith Tools 4.0*. Liverpool: Lexical Analysis Software.

Sinclair John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stubbs Michael, Barth Isabel (2003). Using recurrent phrases as text-type discriminators: a quantitative method and some findings. *Functions of Language* 10, 65–108.

Wray Alison (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray Alison (2009). Identifying formulaic language. Persistent challenges and new opportunities. In *Formulaic Language. Vol. 1. Distribution and historical change*. Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, Kathleen Wheatley (eds.), 27–51. Amsterdam: John Benjamins.

Wray Alison, Perkins Michael (2000). The functions of formulaic language: an integrated model. *Language and Communication* 20, 1–28.

Instytut Filologii Angielskiej
Uniwersytet Opolski
Pl. Kopernika 11
45-040 Opole
Polska
[lukasz@uni.opole.pl]