Multi-word patterns in the corpus of Information and Communication Technology. Terminological bundles in the genre – 'Textbooks in ICT'

MAREK WEBER
Bialystok University of Technology

KEYWORDS: corpus linguistics, Information and Communication Technology, lexical bundles, functional classification of lexical bundles, terminological bundles, lexical scrutiny

DATA FOR THE STUDY, THE CORPUS OF INFORMATION AND COMMUNICATION TECHNOLOGY

Data for the study consists of an electronic corpus of the broad domain of Information and Communication Technology (ICT) an umbrella term that includes all technologies developed for the purposes of processing and the transfer of data. The texts were selected to represent a cross-section of the field of ICT. The texts were divided into seven categories which can be regarded as genres within the domain of ICT:

Genre 1: Professional articles in ICT;
Genre 2: Academic articles in ICT;
Genre 3: Technical documentation of software applications in ICT;
Genre 4: Technical documentation of ICT hardware;
Genre 5: Textbooks in ICT;
Genre 6: Technical documentation of programming languages and programming environments;
Genre 7: Technical documentation of network technologies.

The corpus contains a collection of **973** texts totaling almost 24 million words. The choice of data for a corpus is one of the most important concerns for the researcher as the corpus should be representative of the analyzed domain of use. As Douglas Biber suggests the quantity of texts is crucial for research in which the scholar concentrates on the text as the primary unit of scrutiny. A sufficient number of texts should be included in each genre to account for variation between categories and authors (Biber, 2006).

Table 1. Composition of the Information and Communication Technology corpus

Genre	Number of texts	Number of tokens (running words)
Professional articles	391	225551
Academic articles	95	1272001
Technical documentation of software applications in ICT	92	2472839
Technical documentation of ICT hardware	100	2471772
Textbooks	99	15315237
Technical documentation of programming languages and programming environments	90	1282175
Technical documentation of network technologies	106	725418
TOTAL	**973**	**23764993**

LEXICAL BUNDLE: THE TERM

Recurrent sequences of words which appear together more frequently than expected by chance are an important part of linguistic output. The term *lexical bundle* was first used by Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan in the now classic *Longman grammar of spoken and written English* to refer to multi-word combinations identified on the basis of the sole criterion of frequency (Biber et al. 1999) while Mike Scott calls them *word clusters* in the manual of the software application *WordSmith Tools* version 4 (1996) and *WordSmith Tools* version 5 (2010). The only consideration in identifying lexical bundles is their frequency. "Clusters are words which are found repeatedly together in each others' company, in sequence" (Scott 2010: 337). Scott points out that word clusters may involve semantic prosody i.e. the tendency for certain lexemes to co-occur with certain other lexemes (e.g. the tendency for *cause* to come with negative effects such as *accident, trouble*, etc.) (Scott 2010: 337).

Biber, Johansson, Leech, Conrad and Finegan define *lexical bundles* as recurrent expressions regardless of their structural properties, i.e. they are sequences of word forms that frequently co-occur in discourse and point out that shorter bundles are often incorporated into more than one longer lexical bundle (Biber et al. 1999). They also observe that in most cases lexical bundles do not form structural units and most of them are

not expressions that language users would recognize as idioms or other fixed lexical expressions. Biber, Johansson, Leech, Conrad and Finegan establish the following cut off thresholds for multi-word sequences to qualify as lexical bundles: they must occur in at least 10 times per million words and at least in five texts (Biber et al. 1999).

Biber notes that a surprising result of a frequency driven approach is that lexical bundles have two unexpected characteristics (Biber 2006: 134). Firstly, most of them are not idiomatic in meaning but the meanings are usually transparent from individual word-forms. Ken Hyland also points to the fact that most bundles are semantically transparent and "formally regular, providing the building blocks of coherent discourse" (Hyland 2008: 6).

Secondly, most bundles are not complete grammatical structures. Biber, Johansson, Leech, Conrad and Finegan observed that approximately 15 % of the lexical bundles in conversation formed complete structural units while only approximately 5 % of the lexical bundles in academic prose could be considered as complete phrases or clauses (Biber et al. 1999: 1000). A valuable finding into the syntactic nature of lexical bundles by Biber is that they are lexical units that frequently cut across grammatical structures, for example they can bridge two phrases or clauses in such a way that the last words of a bundle are the beginning parts of a second syntactic structure (Biber 2006: 135). Hyland also emphasizes that lexical bundles being identified solely on the basis of their frequency usually span structural units (Hyland 2008: 6).

A range of corpus studies have been devoted to the analysis of recurrent strings of uninterrupted word-forms and they demonstrate how important they are in various types of discourse as well as they show considerable variation of lexical bundles in different genres and registers (e.g. Biber 2006; Biber, Conrad, Cortes 2004; Hyland 2008; Scott, Tribble 2006; Gozdz-Roszkowski 2011).

METHODOLOGY USED IN THE STUDY

The decision was made to concentrate on the analysis of **4-word bundles** because on the one hand the frequencies of 4-word bundles are substantially higher than 5-word sequences and on the other hand they frequently contain 3-word strings and enable to identify more apparent patters and structures than 3-word expressions. Lists of 4-word bundles in each of the

seven ICT genres were generated using the *WordSmith Tools* v. 5.0 – the software package for searching patterns in corpora.

The following two cut-off points were set in this study: a minimum frequency of occurrence of 20 times per million words and another criterion that a bundle should occur in at least five texts. The process of applying the cut-off criteria required appropriate calculations in each of the seven files containing the list of bundles in a given genre. The calculations are described in the following steps:

1. Creating an additional column named "Frequency per million" in the spreadsheet obtained from the *WordSmith Tools* software.
2. Inserting the running words value into a free Excel spreadsheet cell within a given category.
3. Filling in the first cell of the "Frequency per million" column with the following formula:

$$\text{Freqency per million words} = \frac{\text{Frequency x 1 000 000}}{\text{Running words}}$$

4. Copying the formula to other cells by clicking on the right bottom corner of the filled-in cell and dragging in downwards to other cells.
5. Highlighting the "Number of texts" column.
6. Choosing the "Data" menu and selecting the field "Sort &Filter" in the sorting tool.
7. Choosing the descending sorting order.
8. Deleting all rows of the table, for which the values of the "Number of texts" field is lower than 5.
9. Highlighting the "Frequency per million" column.
10. Choosing the "Data" menu and selecting the field "Sort & Filter" in the sorting tool.
11. Choosing the descending sorting order.
12. Deleting all rows of the table, for which the values in the "Frequency per million" field are lower than 20.

1886 different bundles were found in the corpus after applying the above **cut-off criteria.** The total number of bundles identified in the entire data amounted to 197 553.

Table 2. Distribution of lexical bundles in ICT genres

Genre	The total number of bundles after applying cut-off criteria	Number of different bundles after applying cut-off criteria	% of running words in bundles
Professional articles	1057	**131**	1,8
Academic articles	5548	**120**	1,7
Software applications	44669	**403**	7,2
Hardware	55296	**672**	8,9
Textbooks	72483	**119**	1,9
Programming languages and programming environments	10984	**180**	3,4
Network technologies	7516	**261**	4,1
Total	197553	**1886**	

The percentage of running words in bundles was obtained by multiplying the number of total cases for a given genre by 4 (4-word bundles are analyzed) and dividing by the number of running words in a given genre and then multiplying by 100 % according to the formula:

$$\text{\% of total words in bundles} = \frac{\text{Number of total cases x 4}}{\text{Running words}} \text{ x 100\%}$$

In our view two parameters: the range of different bundles and the percentage of running words in bundles can be regarded as indicators of the degree to which a given genre is formulaic and repetitive in comparison to other genres. In other words the degree of formulaicity and repetitiveness of a genre can be measured by the range of different bundles employed in a genre and the percentage of running words in bundles.

The ICT genres can be divided into three groups according to the criterion of **formulaicity and repetitiveness**.

Genre 3 – 'technical documentation of software applications in ICT' and genre 4 – 'technical documentation of ICT hardware' are marked by high degrees of formulaicity and repetitiveness as they display comparable patterns by employing the biggest scope of different bundles (403 and 672 respectively) and the highest percentage of running words in bundles (7,2 % and 8,9 % respectively).

Figure 1: Genres with high degrees of formulaicity and repetitiveness

In contrast, genre 1 – 'professional articles in ICT', genre 2 – 'academic articles in ICT' and genre 5 – 'textbooks in ICT' are characterized by relatively low degrees of formulaicity and repetitiveness as they employ the lowest scope of different bundles (131, 120 and 119 respectively) and the lowest percentage of running words in bundles (1,8 %, 1,7 % and 1,9 % respectively).

The remaining two genres: genre 6 – 'technical documentation of programming languages and programming environments' and genre 7 – 'technical documentation of network technologies' form the third group with the values of both parameters in the middle of the range (numbers of different bundles: 180 and 261 respectively; percentage of running words in bundles 3,4 % and 4,1 % respectively).

However, it is worth bearing in mind that as Stanislaw Gozdz-Roszkowski rightly points out direct comparisons can only be safely made between genres with similar word counts. In order to account for that consideration the parameter *percentage of running words in bundles* is computed for each genre in such a way as to reflect the word counts of each of the subcorpora representing respective genres (Gozdz-Roszkowski 2011: 111).

FUNCTIONAL CLASSIFICATION OF BUNDLES

A framework for the functional analysis of the bundles obtained in this corpus was established from Biber's (Biber 2006; Biber et al. 2004), Hyland's (2008) and Gozdz-Roszkowski's (2011) taxonomies.

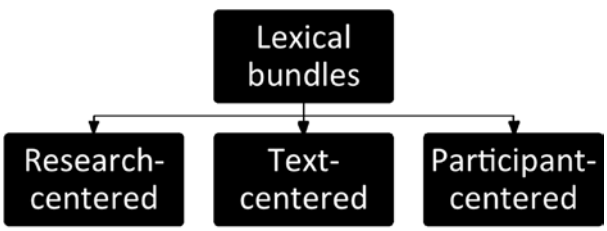
Biber's (2006) classification resulted from the scrutiny of a broad corpus of spoken and written registers which covered among others such types of discourse as: casual conversations, class sessions tapes, classroom teaching, office hours, study groups, on-campus service encounters, textbooks, course packs, institutional texts (e.g. university catalogs, brochures).

The corpus in Hyland's (2008) study (size 3,5 million words) comprises research articles, PhD dissertations and MA/MSc theses from four disciplines: electrical engineering and microbiology from the applied and pure sciences, and business studies and applied linguistics from the social sciences.

Gozdz-Roszkowski (2011) analyzed a corpus of American Law containing over 5,5 million words and representing seven genres within American legal culture and education: academic articles, briefs, contracts, legislation, opinions, professional articles and textbooks.

Drawing upon the abovementioned taxonomies lexical bundles in ICT were functionally divided into three broad categories with respect to their meanings in the texts: **research-centered**, **text-centered** and **participant-centered**. Bundles grouped in the first category help writers to organize their activities and experiences in the domain of ICT. **Text-centered bundles** are employed to indicate the organization of the text and its meaning. Finally **participant-centered bundles** are used to signal different attitudes or assessments and they are focused on the writer or reader of the text.

Figure 2: Functional taxonomy of lexical bundles



Each of the three major functional categories of lexical bundles was further subdivided into a number of subcategories.

Research-centered bundles include the following subcategories:

Quantity bundles (e.g. *one of the biggest; a large number of; one of the following; the number of elements*);

Time reference bundles (e.g. *at the time of; end of the year; the next few weeks; in the coming weeks*);

Place / direction reference bundles (e.g. *in the United States; of the main window; bottom of the window; in the status bar*);

Procedure bundles – used to describe diverse functions pertaining to ICT (e.g. *the use of the; the use of a*);

Topic indicator bundles – pertaining to the area of research (e.g. *proceedings of the IEEE; programming languages and systems; the drop down menu; the following configuration options*);

Multi-functional reference bundles – bundles which can be used as time / place / text reference (e.g. *the end of the; the start of the; the left of the; at the beginning of*);

Description bundles – specify characteristics of the following noun (e.g. *the complexity of the; the surface of the; the scope of the; the height of the*).

Text-centered bundles include the following subcategories:

Elaboration bundles – further elaborate on the analyzed topic and clarify it (e.g. *at the same time, as well as the, this means that the, in this case the*);

Transition bundles – provide additive or contrastive connections between portions of texts (e.g. *on the other hand, as opposed to the, in addition to the, in contrast to the*);

Framing attributes bundles – "situate arguments by specifying limiting conditions" (Hyland 2008: 14) for making claims or arguments (e.g. *in terms of the, in the context of, in the presence of, the contents of the*);

Conditions bundles – express conditions (e.g. *if you want to, if you do not, if you have a, if you have an*);

Results bundles – indicate logical links between elements in terms of cause and result relationships (e.g. *so that you can, as a result of, as a result the, as a function of*);

Structure markers bundles – are used to point to other parts of the text (e.g. *as shown in figure, as discussed in section, is shown in ex ample, later in this chapter*).

Participant-centered bundles include the following subcategories:

Engagement bundles – address readers directly; "actively address readers as participants in the unfolding discourse" (Hyland 2008: 18) (e.g. *do one of the, select the type of, is recommended that you, select one of the*);

Stance bundles – convey emotions, attitudes, value judgments and assessments; "provide a frame for the interpretation of the following proposition" (Biber, 2006: 139) (e.g. *it is necessary to, is no guarantee that, it is important to, it is not possible, it is possible to*);

Modality bundles – express that something is probable, permissible or necessary (e.g. *must be fulfilled a, can be used to, as can be seen, in which you can, may not be displayed, may be reproduced or, must accept any interference, interference that may cause, that may cause undesired*);

Prediction bundles – express the writer's prediction of some future action (e.g. *is expected to be, you will be prompted, will be displayed in, this will open the, will be asked to*).

Table 3 shows the numbers of bundles belonging to the three major functional categories in each genre.

Table 3: Distribution of lexical bundles across functional categories

	Research-centered	Text-centered	Participant-centered	Others
Professional articles	45	23	13	45
Academic articles	48	35	9	28
Software applications	123	55	122	83
Hardware	207	51	138	187
Textbooks	32	35	18	20
Programming languages and programming environments	39	39	26	62
Network technologies	91	24	11	91

Figure 3: Classification of research-centered bundles

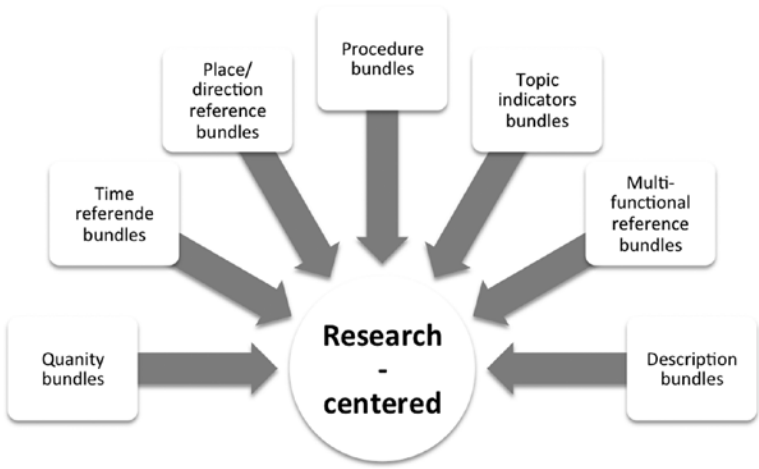


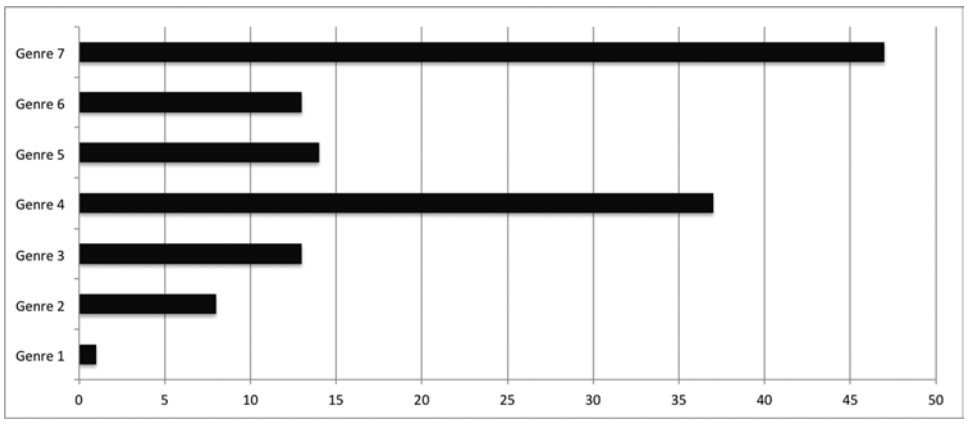
Table 4: Distribution of research-centered bundles across subcategories

RESEARCH - CENTERED							
No of genre	Quantity	Time reference	Place/direction reference	Procedure	Topic indicator	Multifunctional reference	Description
1	4	7	4	0	27	3	0
2	7	0	3	1	34	0	3
3	4	1	63	0	40	8	7
4	3	11	33	2	154	0	4
5	4	0	4	1	16	3	4
6	6	0	2	1	15	3	12
7	6	0	5	2	73	2	3

TERMINOLOGICAL BUNDLES

Terminological bundles have been indentified as a subcategory of topic indicator bundles and defined as multi-word combinations which contain terms.

Figure 4: Distribution of terminological bundles across 7 ICT genres



As Table 5 shows genre 7 – 'technical documentation of network technologies' contains the largest percentage (18 %) as well as the largest number of terminological bundles – 47. Such terminological density could be linked to the highly specialized domain specific nature of texts in genre 7. Disseminating highly specialized knowledge requires the recurrent usage of fixed and specialized multi-word combinations.

Table 5: Numbers and proportions of terminological bundles in ICT genres

Genre number	No of bundles in a genre	No of terminological bundles in a genre	Percentage of terminological bundles in a genre	Percentage of terminological bundles in the category RESEARCH - CENTERED
1	131	1	0.7	2.2
2	120	8	6.6	16.6
3	403	13	3.2	10.5
4	672	37	5.5	17.8
5	119	14	11.7	66.6
6	180	13	7.2	38.2
7	261	47	18.0	50.0

The second largest number of bundles is found in genre 4 – 'technical documentation of ICT hardware' amounting to 37. However, the proportion of all bundles belonging to the category of terminological bundles for genre 4 is relatively low – only 5,5 %. But it should be born in mind that genre 4 has by far the largest number of different bundles after applying cut-off criteria – 672 and the highest percentage of running words in bundles – 8,9 %.

The genre 5 – 'textbooks in ICT' has a relatively large proportion of terminological bundles – 11,7 %. Table 6 below contains terminological bundles identified in genre 5 – 'textbooks in ICT'.

Table 6: Terminological bundles identified in genre 5 – 'textbooks in ICT'

GENRE 5			
Bundle	Freq.	Number of texts	Freq. per Mil.
the **java io package**	345	12	22.53
the **java lang package**	322	9	21.02
import j**ava awt** import	336	15	21.94
public static final int	1306	21	85.27
static void main string	967	39	63.14
public static void main	952	40	62.16
public final static int	861	5	56.22
void main string args	846	36	55.24

GENRE 5			
int x int y	600	14	39.18
public instance methods public	539	5	35.19
import **java io** import	447	22	29.19
the number of **bytes**	433	38	28.27
the **java lang package**	322	9	21.02
x int y int	316	12	20.63

A substantial proportion of textbooks (52 %) concern various topics related to programming in a number of programming languages such as Java, C++, C#, Objective-C, PHP and others. This is reflected in the considerable number of terminological bundles in genre 5 which contain frequently occurring lines of code. For example, terminological bundles such as *STATIC VOID MAIN STRING* (967 occurrences in 39 texts; frequency per million 63,14), *PUBLIC STATIC VOID MAIN* (952 occurrences in 40 texts; frequency per million 62,16) and *VOID MAIN STRING ARGS* (846 occurrences in 36 texts; frequency per million 55,24) are all parts of programming code including the keyword *main*. The keyword *main* refers to the **main function** responsible for the high-level organization of the program's functionality. Functions, also referred to as methods, are sets of instructions used to operate on input data, e.g. mathematical calculations. In many programming languages, the main function is where a program starts execution. For example, Java programs start executing at the main function which has the following method heading: public static void main (String[] args). The main function is responsible for the high-level organization of the program's functionality. The keyword *void* means that the method has no return value. If the method returned an integer value then the keyword *int* would be used instead of *void*.

The following two terminological bundles *PUBLIC STATIC FINAL INT* (1306 occurrences in 21 texts; frequency per million 85,27) and *PUBLIC FINAL STATIC INT* (861 occurrences in 5 texts; frequency per million 56,22) are parts of lines of code in programming languages, e.g. Java. They define the type of functions to be applied. The keyword *public* means that the method is visible and can be called from objects of other types. The keyword *static* means that the method is associated with the class, not a specific instance (object) of that class. This means that it is possible to call a static method without creating an object of the class.

The bundle *INT X INT Y* occurs 600 times in 14 texts with the frequency per million words 39,18. The bundle represents a declaration of variables, the type of variables in this case is integer, or *int* in short. The set of integers is formed by the natural numbers (0, 1, 2, 3, ...) together with the negative natural numbers (–1, –2, –3, ...).

The frequent keyword *Java*, the name of a programming language, occurs in 6 bundles: *JAVA ENTERPRISE IN A* (653 occurrences in 5 texts; frequency per million 42,64); *IMPORT JAVA IO IMPORT* (447 occurrences in 22 texts; frequency per million 29,19); *THE JAVA IO PACKAGE* (345 occurrences in 12 texts; frequency per million 22,53); *IMPORT JAVA AWT IMPORT* (336 occurrences in 15 texts; frequency per million 21,94); *JAVA IO IMPORT JAVA* (328 occurrences in 20 texts; frequency per million 21,42) and *THE JAVA LANG PACKAGE* (322 occurrences in 9 texts; frequency per million 21,02).

Java gained popularity because of such characteristics as ease of use, cross-platform capabilities and security features. By using Java, one program can be run on many different platforms. This means that there is no need to put efforts on developing a different version of software for each platform. Java was designed to be easy to use and it is regarded as easier to write, compile, debug, run and learn than some other programming languages.

The abbreviation *AWT*, appearing in the bundle *IMPORT JAVA AWT IMPORT* stands for The Abstract Window Toolkit which is a platform-independent windowing, graphics, and user-interface toolkit.

The full name of *JAVA ENTERPRISE*, the sequence being part of the frequent bundle *JAVA ENTERPRISE IN A*, is Java Platform, Enterprise Edition. It denotes Oracle's enterprise Java computing platform which provides runtime environment for developing and running enterprise software, including network and web services, and other large-scale, scalable and secure network applications. The bundle is part of a longer 5-word formulaic expression (653 occurrences in 5 texts; frequency per million 42,64): *Java Enterprise in a Nutshell* as illustrated in the following example:

> **Sample 1**: Sun makes life even easier by supplying providers for many naming services. So while the service provider gets to spend lots of time implementing the javax.naming.spi interfaces, the client merely needs to provide a URL for the provider and the context factory class, and then use the naming service. For a lot more on JNDI, as well as related topics, you can pick up **Java Enterprise in a Nutshell**. [Genre 5 – 'textbooks in ICT']

The sequence *JAVA EDITOR*, part of the bundle *IN THE JAVA EDITOR* (75 occurrences in 5 texts; frequency per million 58,49) denotes a software tool designed to edit and compile programs written in Java as well as to execute them. The usage of the bundle is illustrated in the following example:

> **Sample 2:** In order to implement content assist, your editor's source viewer configuration must be configured to define accountant assistant. This is done in the **Java editor** example inside the JavaSourceViewerConfiguration.
>
> Spelling errors are displayed in the **Java editor** and corresponding Quick Fixes are available: You can make the dictionary also available to the content assist.
>
> Create your own code assist proposals similar to the ones proposed in the **Java editor**. Instantiate CompletionProposalCollector to get the same proposals as the Java editor, or subclass it to mix in your own proposals. [Genre 6 – 'technical documentation of programming languages and programming environments']

CONCLUSIONS

The results summarized in this work reveal substantial differences in the frequency of forms and functions of lexical bundles across respective genres of the language of Information and Communication Technology. The variations are reported in quantitative as well as qualitative terms. Quantitatively, Genre 3 – 'technical documentation of software applications in ICT' and genre 4 – 'technical documentation of ICT hardware' are most formulaic and repetitive. The other ICT genres employ considerably lower numbers of lexical bundles and lexical bundles account for markedly lower proportions of the total number of words. All 14 terminological bundles occurring in the genre textbooks which are the subject of scrutiny in the paper relate to various aspects of programming. This can be explained by the fact that a considerable percentage of textbooks (52 %) concern various topics in programming.

REFERENCES

Biber D. 2006: *University language: A corpus-based study of spoken and written registers*, Philadelphia/Amsterdam: John Benjamins Publishing Company.

Biber D., Conrad S., Cortes V. 2004: If you look at ...: Lexical bundles in university teaching and textboks, *Applied Linguistics*, 25L, 371–405.

Biber D., Conrad S., Cortes V. 2003: Lexical bundles in speech and writing: an initial taxonomy. – *Corpus Linguistics by the Lune: A festschrift for Geoffrey Leech*, Wilson A., Rayson P. and McEnery T. (eds.), Frankfurt/Main: Peter Lang.

Biber D., Johansson S., Leech G., Conrad S. and Finegan E. (1999): *Longman grammar of spoken and written English*, Harlow: Pearson.

Hyland K. 2008: As can be seen: Lexical bundles and disciplinary variation. – *English for Specific Purposes* 27 (2008), 4–21.
Gozdz-Roszkowski S. 2011: *Patterns of linguistic variation in American legal English*, Frankfurt/Main: Peter Lang.
Gozdz-Roszkowski S. (ed.) 2011: *Explorations across languages and corpora*, PALC 2009, Frankfurt/Main: Peter Lang.
Scott M. 1996: *WordSmith Tools 4 manual*, Oxford University Press.
Scott M. 2010: *WordSmith Tools 5 manual*, Oxford University Press.
Scott M., Tribble C. 2006: *Textual patterns*, Philadelphia/Amsterdam: John Benjamins Publishing Company.

DAUGIAŽODŽIAI MODELIAI INFORMACINIŲ IR KOMUNIKACINIŲ TECHNOLOGIJŲ TEKSTYNE. TERMINOLOGINĖS SAMPLAIKOS IKT VADOVĖLIUOSE

Pagrindinis straipsnio tikslas – informacinių ir komunikacinių technologijų (IKT) kalboje pasikartojančių tam tikrų lingvistinių modelių tyrimas pasinaudojant tekstynų lingvistikos naujovėmis.

Straipsnyje nagrinėjamos keturžodės samplaikos iš 24 milijonų žodžių IKT srities tekstyno, suskirstyto į 7 kategorijas (žanrus): profesiniai straipsniai, moksliniai straipsniai, programinės įrangos dokumentacija, aparatinės įrangos dokumentacija, vadovėliai, programavimo kalbų ir aplinkos dokumentacija ir tinklo technologijų dokumentacija.

Atliekant tyrimą buvo nutarta apsiriboti keturžodėmis samplaikomis. Naudojant modelių paieškos tekstynuose programų paketą *WordSmith Tools* v. 5.0 sudaryti kiekvieno iš septynių IKT srities tekstų žanrų keturžodžių samplaikų sąrašai. Pasirinktos dvejopos ribos: mažiausias samplaikos dažnumas turi būti 20 kartų milijonui žodžių ir samplaika turi būti rasta mažiausiai 5 tekstuose.

Pritaikius tokius apribojimus tekstyne rastos 1886 skirtingos samplaikos. Bendras duomenų rinkinyje nustatytų samplaikų skaičius – 197 553.

Pagal reikšmę tekste samplaikos suskirstytos į tris plačias funkcines kategorijas: orientuotas į tyrimus, į tekstą ir į dalyvį. Pirmai kategorijai priskirtos samplaikos toliau išskirstytos į keletą subkategorijų: kiekybės, laiko, vietos / krypties, procesų, temos, daugiafunkcių nuorodų, apibūdinimo. Terminologinės samplaikos buvo priskirtos temos rodiklio subkategorijai ir apibrėžtos kaip keliažodžiai junginiai, kuriuose yra terminų.

Straipsnyje nagrinėjamos terminologinės samplaikos, rastos IKT vadovėliuose. Didelė vadovėlių dalis (52 %) skirta įvairiems programavimo klausimams. Tai paaiškina faktą, kad visos 14 šio žanro terminologinių samplaikų susijusios su įvairiais programavimo aspektais.

Gauta 2012-10-30

Marek Weber
Bialystok University of Technology
ul. Wiejska 45A, Bialystok, Poland
E-mail 1234marek@interia.pl