

Vilnius University

Rita Juknevičienė

LEXICAL BUNDLES IN NON-NATIVE SPEAKER AND NATIVE SPEAKER  
WRITTEN ENGLISH

Summary of doctoral dissertation

Humanities, Philology (O4 H)

Vilnius, 2011

This doctoral dissertation was written at Vilnius University in the Faculty of Philology in 2006-2010.

Research supervisor:

**Assoc. Prof. Dr. Jonė Grigaliūnienė** (Vilnius University, Humanities, Philology – 04 H)

The dissertation will be defended at the Council of Philology of Vilnius University.

Chair:

**Assoc. Prof. Dr. Meilutė Ramonienė** (Vilnius University, Humanities, Philology – 04 H)

Members:

**Prof. Habil. Dr. Ineta Dabašinskienė** (Vytautas Magnus University, Humanities, Philology – 04 H)

**Assoc. Prof. Dr. Violeta Kalėdaitė** (Vytautas Magnus University, Humanities, Philology – 04 H)

**Assoc. Prof. Dr. Inesa Šeškauskienė** (Vilnius University, Humanities, Philology – 04 H)

**Dr. Loreta Vaicekauskienė** (Institute of the Lithuanian Language, Humanities, Philology – 04 H)

Opponents:

**Prof. Habil. Dr. Aurelija Usonienė** (Vilnius University, Humanities, Philology – 04 H)

**Assoc. Prof. Dr. Jūratė Ruzaitė** (Vytautas Magnus University, Humanities, Philology – 04 H)

The public defence of the dissertation is to be held in the meeting of the Council of Philology of Vilnius University at 11 o'clock on 11 February 2011 in the Faculty of Philology, Vilnius University. Address: Universiteto St. 5, LT-01513 Vilnius.

The summary of the dissertation was sent out to relevant institutions on \_\_\_\_\_ January 2011.

The dissertation is available at the library of Vilnius University.

VILNIAUS UNIVERSITETAS

Rita Juknevičienė

LEKŠINĖS SAMPLAIKOS GIMTAKALBIŲ IR SVETIMKALBIŲ VARTOTOJŲ  
RAŠY TINĖJE ANGLŲ KALBOJE

Daktaro disertacijos santrauka

Humanitariniai mokslai, filologija (04 H)

Vilnius, 2011

Disertacija rengta 2006-2010 metais Vilniaus universiteto Filologijos fakultete.

Mokslinio darbo vadovas:

**doc. dr. Jonė Grigaliūnienė** (Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H)

Disertacija ginama Vilniaus universiteto Filologijos mokslo krypties taryboje.

Pirmininkas:

**doc. dr. Meilutė Ramonienė** (Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H)

Nariai:

**prof. habil. dr. Ineta Dabašinskienė** (Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04 H)

**doc. dr. Violeta Kalėdaitė** (Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04 H)

**doc. dr. Inesa Šeškauskienė** (Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H)

**dr. Loreta Vaicekauskienė** (Lietuvių kalbos institutas, humanitariniai mokslai, filologija – 04 H)

Oponentai:

**prof. habil. dr. Aurelija Usonienė** (Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H)

**doc. dr. Jūratė Ruzaitė** (Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04 H)

Disertacija bus ginama viešame Filologijos mokslo krypties tarybos posėdyje 2011 m. vasario 11 d. 11 val. Vilniaus universiteto Filologijos fakultete. Adresas: Universiteto gt. 5, LT-01513 Vilnius.

Disertacijos santrauka išsiuntinėta 2011 m. sausio \_\_\_\_ d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje.

# LEXICAL BUNDLES IN NON-NATIVE SPEAKER AND NATIVE-SPEAKER WRITTEN ENGLISH

## INTRODUCTION

**The object of the research.** Corpus studies of recurrent word sequences, or lexical bundles (Biber *et al.* 1999), have outlined new directions in ELT/EFL research. The fact that naturally produced English consists of prefabricated multi-word units gave rise to the question of chunkiness in learner language. This research was designed to compare and contrast lexical bundles retrieved from corpora of written English produced by NNSs (non-native speakers) and NSs (native speakers). The corpora of learner language, represented in this research by written English of Lithuanian learners and native speakers, provide valuable empirical data which sheds light on the process of foreign language acquisition and, more specifically, development of lexical competence in writing.

The analysis of lexical bundles, as shown by previous studies (*cf.* Altenberg 1998, Biber 2006, Hyland 2008, Cortes 2008, Chen and Baker 2010), reveals specific discourse features established on the basis of prevailing structural and functional types of lexical bundles. It was thus assumed that research into learner language representing learners at different levels of proficiency may contribute to a better understanding of how the ability to write changes alongside the development of general language proficiency. Furthermore, a contrastive analysis of structural and functional types of lexical bundles in NNS and NS written English also reveals peculiarities of Lithuanian EFL learner vocabulary and writing. Hence findings of this research may have a particular relevance to the practice of teaching English as a foreign language to Lithuanian learners.

**The goals of the dissertation** are to contrast and compare vocabulary of the learners of English at three different levels of proficiency by analysing lexical bundles retrieved from the corpora of written learner language, to carry out structural and functional analyses of the lexical bundles and evaluate their impact on the quality and discourse features of learner writing, as well as to describe phrasal word combinations,

or formulaic sequences, which recur in the lexical bundles. To achieve these goals, the following **research tasks** were set:

1. To compile corpora of written English produced by the Lithuanian EFL learners at two different levels of proficiency.
2. To perform quantitative and frequency analyses of the lexical bundles in the learner corpora.
3. To classify the lexical bundles by their structural features and functional properties.
4. To contrast and compare the learner corpora by the recurrence of phrasal word combinations in the lexical bundles.

**The novelty and relevance of the research.** Research into the acquisition of English as a foreign language has so far been rather scarce in Lithuania while the use of computerized corpora for the investigation of learner language is only currently gaining proper recognition. This study offers many general insights into the development of the lexical competence of Lithuanian EFL learners, which may enhance EFL teaching/learning outcomes, and points out directions for further research in the field, particularly as regards the development of writing skills and vocabulary as well as the processing and currency of phrasal word combinations in learner language. More specifically, this study which is based mainly on the data obtained from the students of one study programme (BA in English Philology) brings forth certain aspects of teaching English for Specific Purposes and reveals that the general academic English vocabulary may be more problematic for the advanced Lithuanian EFL learners than topic-specific lexis.

Furthermore, the study and its findings belong to a body of similar contrastive studies of NNS learner language and contribute to the ongoing research in the field by providing data about the Lithuanian EFL learners. Comparable data obtained from the other corpora of learner language may highlight specific difficulties of the Lithuanian learners and thus contribute to the development of a more efficient and focused teaching practice. Hopefully, the findings of this research could be used for the preparation of learner-corpus informed ELT materials, for example, vocabulary and writing aids.

The study has also revealed some useful insights for language testing and assessment. Since the corpora analysed in this research consist of academic student essays, the research showed how the task rubric and topic may influence the authentic learner production and thus impact the process of test assessment. Finally, the author hopes that the research might be interesting to the general public and contribute to the further development of learner language research.

The following **theses** are to be defended:

1. The analysis of lexical bundles retrieved from the learner corpora allows to establish specific features of the learner language characterizing learners at different levels of proficiency.
2. The structural and functional features of the lexical bundles show that written English produced by the learners of the lower proficiency level bears more similarity to the spoken rather than written variety of the language. Lexical bundles typical of written English become more frequent as the proficiency of the learners increases.
3. Written English produced by the learners of the lower proficiency levels contains more discourse-organizing and stance bundles which are gradually replaced with referential bundles as the level of proficiency increases, *i. e.* the higher language proficiency is directly related to the increase of the propositional density of the learner writing.
4. Differences in the learner language representing the learners of different proficiency arise from the use of the general academic English vocabulary rather than topic-specific lexis.
5. The analysis of the lexical bundles automatically retrieved from a corpus reveals the general tendencies of the use of different lexico-grammatical patterns in the learner language and may thus be seen as an important starting point in the studies of vocabulary.

**Review of the earlier research.** The analysis of language corpora has brought to light a considerable amount of new data about languages and contributed to a shift in linguistics from intuition to empirical material (Biber *et al.* 1998, Tognini-Bonelli 2001, Sinclair 2004). The investigation of grammatical structures of a language gave way to the studies of vocabulary which had been hardly possible before the first large

computerized corpora came into being. Gradually, the research into lexical patterns in language established the tradition of lexical grammar, a new approach to the study of language. According to one of its founders, John Sinclair, 'division into grammar and vocabulary obscures a very central area of meaningful organization' (1991: 137) in language and is hardly plausible. Early findings from the corpus research pointed to a new unit of language – *word* became replaced with the *lexical unit*, which is defined as a multi-word unit having a specific meaning (Sinclair 2004). Thus corpus linguistics opened up a new research direction which took into account the syntagmatic relations between words and was focused on the study of word phrases, combinations, collocations, chunks and other multi-word units.

The first extensive account of word combinations established in a corpus of English was presented in *The Longman Grammar of Spoken and Written English* (Biber *et al.* 1999). While *idioms* and to some extent *collocations* are distinguished on the basis of semantic transparency and/or substitutability of their elements, *lexico-grammatical associations* and *lexical bundles* are established on the frequency of co-occurrence, *i. e.* they are sequences of words that are often repeated in the identical form. Irrespective of the rather formal criterion for distinguishing the lexical bundles, these multi-word combinations proved to have certain structural and functional properties which may be useful for the study of different aspects of the texts from which they are retrieved.

Lexical bundles, defined as statistically established word sequences, are also known in literature as recurrent word-combinations (Altenberg 1998), recurrent sequences (De Cock 1998, 2004), chunks (O'Keeffe *et al.* 2007), clusters (Scott 2008, Hyland 2008b). They are non-idiomatic, structurally and semantically incomplete and they recur in the corpus in the identical form, for example, *and then, one of the, the fact that the*. As seen from the examples, the lexical bundles may be combinations of varying length, *i. e.* two-word, three-word *etc.* The longer bundles (three- and four-word) are analysed more often as they carry more content and thus can be associated with particular discourse functions though there are studies in which bundles of varying lengths have been analysed (e. g. O'Keeffe *et al.* 2004, De Cock 2004). The investigation of structural and functional features of the lexical bundles shows that lexical bundles consisting of noun phrases and prepositional phrases, e. g. *the nature of, at the end of*, most often perform the referential function in the text and thus express propositions. In



contrast, the verbal and clausal bundles, e. g. *can be said that, to start with*, usually express authorial stance or contribute to the cohesive features of the text (*cf.* Biber *et al.* 2004, Hyland 2008, Biber 2006). Undoubtedly, these findings influenced the research based on the corpora of learner.

**Multi-word units in the learner language.** One of the earliest contributions in the field belongs to Pawley and Syder (1983), who hypothesized that the differences between the NS and NNS language often arise from the use of word phrases rather than grammaticality of expressions. They observed that in natural speech NSs reproduce prefabricated chunks of words and combine them into meaningful sentences instead of generating every single word sequence by the application of the rules of syntax. In comparison, NNSs try to construct their spontaneous speech by combining individual words. As a consequence, they often produce unnaturally sounding language even though it is grammatically accurate (1983: 215). As it was later suggested by Sinclair (1991), the NNS learners are unaware of the special combinatory properties of the English words which, as he claimed, are governed by the idiom principle. So instead of selecting longer combinations of words as native speakers do, the NNS learners choose separate words: their 'building material is individual blocks rather than prefabricated sections' (Kjellmer 1991: 124). Further work by Nattinger and DeCarrico (1992) and particularly Lewis (1993, 2000) called for a rather radical change in the ELT/EFL teaching and advocated for a much more intense focus on word phrases. Yet the advance of the new ideas was rather inert and perhaps it could be explained by the lack of agreement in linguistics on the treatment of multi-word units which is evidenced by the terminological variation. Wray (2000) reports over forty different terms for the multi-word combinations and proposes a general term, *formulaic sequence*, which is defined as a prefabricated combination of words, 'stored and retrieved whole from memory at the time of use' (2000: 465). Interestingly, Wray observed that the NNS learners fail to realize the non-compositional character of the formulaic sequences and tend to creatively re-produce them, which is particularly evident at the higher levels of language proficiency.

The investigation of multi-word units in the learner language is closely related to the development of corpora representing the authentic learner production. The largest learner language project, ICLE (International Corpus of Learner English), was initiated

by Sylviane Granger in the Catholic University of Louvain-la-Neuve (Belgium). The corpus consists of argumentative essays written by NNSs of different mother tongue backgrounds. A comparable corpus of NS essays (the LOCNESS corpus) was also compiled to enable comparisons across the corpora. This work gave rise to a specific research method, known as the *Contrastive Interlanguage Analysis* (Granger 1996, 1998b, 2003), which has become the main research approach in the studies of learner language. The ICLE and LOCNESS corpora, alongside the spoken corpora LINDSEI (Louvain International Database of Spoken English Interlanguage), continue to be extensively researched worldwide.

One of the first attempts to analyse multi-word units, or more specifically, recurrent word sequences, in the learner language was undertaken by De Cock (1998, 2004), who primarily focused on spoken English produced by the French learners of English. More recently, Chen and Baker (2010) presented a study of lexical bundles in terms of structure and function which provided evidence to the claim that written language competence develops from spoken to written. Collocations, another type of multi-word expressions, was at the focus of Nesselhauf's work (2005) which showed that the German learners of English misused nearly one third of collocations retrieved from their argumentative essays. Waibel (2007) investigated the use of phrasal verbs, which could be also seen as multi-word units, in written English produced by the German and Italian learners. Her data suggests that NNSs indeed may be unnecessarily creative when reproducing fixed and non-compositional verbs which results in inaccurate English expressions. These and some other studies provided evidence to the claims that multi-word units do help differentiate the language produced by NNSs and NSs and should thus be considered an important aspect of teaching EFL.

English produced by the Lithuanian learners of English has not been widely researched though the corpus approach to the study of learner language is certainly encouraging more interest in the field. Problems with the use of English collocations were first described by Aprijaskytė and Pareigyte (1982) who compiled a list of phrases prone to error when used by the Lithuanian learners. The need to teach 'prefabricated sequences' in ESP courses was stressed by Stungienė (1998), who introduced these multi-word units to the Lithuanian readership. The development of the writing teaching tradition at the Department of English Philology in Vilnius University by Assoc. Prof. L.

E. Katkuvienė fostered further interest in various aspects of written learner English (*cf.* Katkuvienė and Šeškauskienė 2005, Burkšaitienė 2008, Burneikaitė 2008, Šeškauskienė 2008, Bikelienė 2009). Hopefully, the currently compiled Lithuanian component of the ICLE corpus (Grigaliūnienė *et al.* 2008) will stimulate further interest in the field.

## DATA AND METHODS

The study was done as a corpus-driven contrastive interlanguage analysis aimed at the comparison of lexical bundles retrieved from three corpora of learner language (Table 1). **Corpus I** represents written English of the first-year students of English Philology (BA) in Vilnius University. It includes examination essays on a variety of topics related to the courses of Contemporary English. **Corpus II** consists of argumentative and literary essays written by the third- and fourth-year students of the same study programme. All the learners, whose language is represented in the two Lithuanian corpora, are native speakers of Lithuanian. **Corpus III** includes a selection of essays from the LOCNESS corpus, compiled within the ICLE project in the Catholic University of Louvain-la-Neuve. The purpose of the selection was to make Corpus III comparable in terms of topic coverage and essay length to the two Lithuanian corpora. The topics of all the essays are given in Annex 1 of the dissertation.

Table 1. Learner corpora used in the study

	<b>Corpus I</b>	<b>Corpus II</b>	<b>Corpus III</b>		
			British English	American English	<b>Total in Corpus III</b>
Number of essays	<b>226</b>	<b>253</b>	102	95	<b>197</b>
Corpus size (in words)	<b>92 050</b>	<b>137 004</b>	93 000	71 687	<b>164 687</b>
Average essay length (in words)	<b>407</b>	<b>519</b>	911	754	<b>835</b>

The corpora were analysed using the WordSmith Tools (v. 5) software (Scott 2008), which automatically generates lists of lexical bundles of the delimited distribution and length. This software package was also used to retrieve the frequency lists of the corpora and concordances of individual words or bundles.

**Research methods** include quantitative and qualitative analyses. Two tests of statistical significance were used in the study: the  $\chi^2$  and log likelihood tests. All the statistical tests were done with raw frequencies and corpus sizes expressed in words. In certain cases, the raw frequencies were normalized to 100,000 words to allow for comparisons across the corpora when reporting the frequency data. The qualitative analysis was largely based on the structural and functional classifications of the lexical bundles proposed by Biber (2006) though these classifications were slightly modified to better accommodate the material of this study. The analysis of phrasal expressions occurring in the lexical bundles was based on the classification of phrasal units presented in the Collins COBUILD English Dictionary for Advanced Learners (2001, 3rd edition).

## RESEARCH FINDINGS

**The quantitative analysis** showed that the three corpora are very different in terms of the frequency of repetitive sequences of words, or lexical bundles. The differences are statistically significant for the lexical bundles of all lengths yet the greatest difference was established for the frequencies of three- and four-word bundles (Table 2).

Table 2. Frequencies of lexical bundles in the three corpora

Lexical bundles	Raw frequencies of the bundles of delimited frequency and distribution (4 occurrences in the essays of 4 different authors)			Normalized frequencies per 100 000 words		
	Corpus I	Corpus II	Corpus III	Corpus I	Corpus II	Corpus III
2-word	3200	3868	3458	3478	2823	2108
3-word	1319	1382	900	1434	1011	549
4-word	386	259	133	419	189	81
5-word	109	43	25	118	31	15
6-word	28	19	5	30	14	3

The raw frequencies reflect the sizes of the corpora – the longer the corpus, the larger the number of the lexical bundles. As evident from Table 2, the differences in the frequency of the lexical bundles become more obvious after the normalization of the raw statistics. The results of the quantitative analysis were taken into account for the further investigation of the corpora. It was also decided that the four-word bundles should be

chosen for the structural and functional analyses as they have a fuller structure and meaning than the shorter lexical bundles. The sample of the four-word lexical bundles was then increased by setting the frequency of three occurrences in the texts of three different authors.

Furthermore, the analysis of the automatically generated lists of lexical bundles highlighted certain aspects which had to be resolved before the following stages of the study could be undertaken. It was observed that the most frequent lexical bundles in all the corpora consist of topic-specific words, e. g. *capital punishment is* (Corpus I), *the English language* (Corpus II), *the people of Argos* (Corpus III). It was established that such topic-specific lexical bundles are often verbatim repetitions of the input material (task rubrics or quotations of famous people) presented to the students as essay topics. Clearly, such expressions hardly reflect the authentic learner language. To minimize their impact on the data it was decided to exclude from the sample such topic-related bundles which cover fiction titles, references to writers, characters *etc.* and verbatim repetitions of the essay tasks and rubrics.

**The structural analysis** of the lexical bundles draws on the previous work by Biber *et al.* (2004) and Biber (2006). The sample of four-word lexical bundles for this analysis included 1135 lexical bundles: 581 from Corpus I, 363 from Corpus II and 184 from Corpus III, which corresponds to the general frequency tendencies in the three corpora. Table 3 presents the structural types of lexical bundles established in the three corpora.

Table 3. Structural types of lexical bundles

Structural types and subtypes	Examples
<b>1. Verbal lexical bundles that incorporate verb phrases and/or their fragments</b>	
1.1. (connector) + 1st person pronoun + verb phrase (fragment)	<i>we live in the</i> (Corpus I) <i>I do not think</i> (Corpus II) <i>I would like to</i> (Corpus III)
1.2. (connector) + it/there + verb phrase (fragment)	<i>it is obvious that</i> (Corpus I) <i>there are a lot</i> (Corpus II) <i>it is true that</i> (Corpus III)
1.3. (connector) + noun phrase + verb phrase (fragment)	<i>the English language is</i> (Corpus I) <i>the same situation is</i> (Corpus II) <i>a single Europe would</i> (Corpus III)
1.4. Verb phrase (fragment) with a verb in the active voice	<i>is one of the</i> (Corpus I) <i>do not think that</i> (Corpus II)

	<i>have the right to</i> (Corpus III)
1.5. Verb phrase (fragment) with a verb in the passive voice	<i>are written in English</i> (Corpus I) <i>is considered to be</i> (Corpus II) <i>is shown to be</i> (Corpus III)
1.6. Interrogative statement	<i>but is it really</i> (Corpus I) <i>is it really so</i> (Corpus II)
<b>2. Clausal lexical bundles that incorporate dependent clauses and/or their fragments</b>	
2.1. WH-clause (fragment)	<i>what is more the</i> (Corpus I) <i>when it comes to</i> (Corpus II) <i>what to do with</i> (Corpus III)
2.2. (connector) + <i>to</i> -clause (fragment)	<i>to live under one</i> (Corpus I) <i>in order to be</i> (Corpus II) <i>to be able to</i> (Corpus III)
2.3. <i>That</i> -clause (fragment)	<i>that there is no</i> (Corpus I) <i>that language is a</i> (Corpus II) <i>that it is a</i> (Corpus III)
2.4. <i>-ing</i> form clause (fragment)	<i>living under one roof</i> (Corpus I) <i>being a member of</i> (Corpus II) <i>not being able to</i> (Corpus III)
2.5. <i>If</i> -clause (fragment)	<i>if you want to</i> (Corpus I)
<b>Nominal lexical bundles that incorporate noun phrases and prepositional phrases and/or their fragments</b>	
3.1. (connector) + noun phrase with <i>of</i> -phrase fragment	<i>the invention of the</i> (Corpus I) <i>one of the most</i> (Corpus II) <i>the end of the</i> (Corpus III)
3.2. Noun phrase with other post-modifier fragments	<i>people who are against</i> (Corpus I) <i>the changes in the</i> (Corpus II) <i>the way in which</i> (Corpus III)
3.3. Other noun phrase expression or its fragment	<i>his or her life</i> (Corpus I) <i>more and more people</i> (Corpus II) <i>United States of America</i> (Corpus III)
3.4. Prepositional phrase or its fragment	<i>in front of the</i> (Corpus I) <i>all over the world</i> (Corpus II) <i>at the beginning of</i> (Corpus III)

Previous studies (Biber 2006, Hyland 2008a, 2008b) have shown that spoken and written English differ by the prevailing structural types of lexical bundles. While the verbal and clausal bundles tend to be more frequent in the spoken language, the nominal lexical bundles are more characteristic of the written variety of English. This study provided interesting data on the discourse features of the learner writing.

The distribution of the lexical bundles across the structural types in the analysed corpora gradually changes as the level of the learners increases. While Corpus I, which represents the written language of the intermediate Lithuanian learners (first-year

university students), contains the largest proportion of the verbal lexical bundles, Corpus III, which consists of NS essays, yields the largest proportion of the nominal and prepositional bundles (see Figure 1). Corpus II, representing the advanced Lithuanian EFL learners (third- and fourth-year university students) in this respect occupies an intermediary position between the other two corpora while the proportions of the verbal and nominal/prepositional bundles in this corpus are nearly equal (43 and 46 per cent).

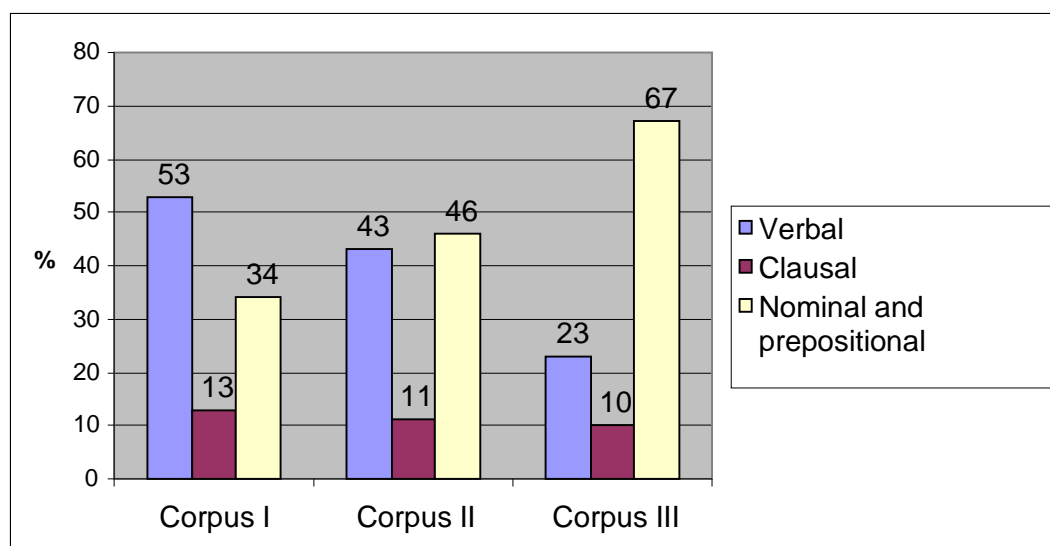


Figure 1. Distribution of the structural types of lexical bundles in the corpora

The analysis provides evidence to the claim that the written language of the lower-level learners contains a larger proportion of lexical expressions that are more frequently found in spoken English. As shown in Figure 1, the proportion of 'spoken' bundle types, i.e. verbal and to a certain extent clausal bundles, becomes smaller as the level of the learners increases. Obviously, the less proficient writers tend to rely on the language that they use in speech while the more advanced learners, as the data shows, resort to lexis which is more typical of written English.

Another observation from the structural analysis suggests that the Lithuanian learners have in their lexical repertoire a certain number of fixed sentence stems which they often use in their essays, for example, *some people claim that, there are many different, it is important to* etc. In this respect, the Lithuanian learners are comparable to EFL learners of the other mother tongue backgrounds, for example, Norwegian or French. Some of the repetitively used sentence stems are directly related to the essay

topics and thus may be seen as non-authentic learner language, especially those that rephrase statements from the task rubrics. Yet the functional analysis revealed that quite often such stems acquire a specific discourse function in the essays of the Lithuanian students, which influences the quality of their written language in terms of its propositional density.

**The functional analysis** based on the classifications proposed in Biber *et al.* (2004), Biber (2006) and Hyland (2008a) was undertaken to compare lexical bundles performing different discourse functions in learner writing. The functional grammar distinguishes three major discourse functions: ideational, interpersonal and textual (*cf.* Lock 1996). Thus the use of every lexical bundle was analysed in the corpora to categorize them under three major functional types which correspond to the discourse functions, namely, referential, stance and discourse-organizing lexical bundles. Table 4 presents a more detailed subdivision of the major functional types into subtypes.

Table 4. Functional types and subtypes of lexical bundles

<b>1. Referential subtypes of lexical bundles</b>	<b>Examples</b>
Topic-specific lexical bundles	<i>the importance of English</i> (Corpus I) <i>educational system of Lithuania</i> (Corpus II) <i>ethnic American literature</i> (Corpus III)
Lexical bundles expressing abstract notions	<i>a right to live</i> (Corpus I) <i>on the basis of</i> (Corpus II) <i>in the case of</i> (Corpus III)
Lexical bundles expressing concrete notions	<i>in the form of</i> (Corpus III)
Lexical bundles specifying number/amount	<i>a lot of people</i> (Corpus I) <i>the number of students</i> (Corpus II) <i>a great deal of</i> (Corpus III)
Lexical bundles specifying time/duration	<i>for a long time</i> (Corpus I and II) <i>the start of the</i> (Corpus III)
Lexical bundles specifying place	<i>all over the world</i> (Corpus I and II) <i>in the modern world</i> (Corpus II) <i>in the United States</i> (Corpus III)
Multifunctional lexical bundles	<i>the end of the</i> (Corpus I) <i>at the end of</i> (Corpus II) <i>at the beginning</i> (Corpus III)
Lexical bundles expressing actions	<i>to take care of</i> (Corpus I) <i>to communicate with other</i> (Corpus II)
Lexical bundles expressing states	<i>to know the English</i> (Corpus I) <i>being a part of</i> (Corpus II) <i>do not have a</i> (Corpus III)



Lexical bundles specifying people	<i>people who are against</i> (Corpus I) <i>he or she is</i> (Corpus II)
Lexical bundles expressing imprecision	<i>to a certain extent</i> (all Corpora)
Lexical bundles expressing text deixis	<i>as it was mentioned</i> (Corpus II)
Other lexical bundles	<i>on their own and</i> (Corpus I) <i>it is up to</i> (Corpus II and III)
<b>2. Stance bundles</b>	
Lexical bundles expressing dynamic modality	<i>to be able to</i> (Corpus I) <i>people are able to</i> (Corpus II) <i>not being able to</i> (Corpus III)
Lexical bundles expressing deontic modality	<i>there is no need</i> (Corpus I) <i>it is necessary to</i> (Corpus II) <i>it should not be</i> (Corpus III)
Lexical bundles expressing epistemic modality	<i>it is known that</i> (Corpus I) <i>it is possible to</i> (Corpus I) <i>it is obvious that</i> (Corpus I) <i>the fact that the</i> (Corpus II) <i>have a possibility to</i> (Corpus II) <i>it is clear that</i> (Corpus II) <i>it is true that</i> (Corpus III) <i>should be able to</i> (Corpus III) <i>is shown to be</i> (Corpus III)
Lexical bundles expressing evaluation	<i>is a controversial issue</i> (Corpus I) <i>the most important reason</i> (Corpus I) <i>one of the most</i> (Corpus I) <i>in my opinion this</i> (Corpus II) <i>it is important to</i> (Corpus II) <i>one of the best</i> (Corpus II) <i>no use to the</i> (Corpus III) <i>one of the greatest</i> (Corpus III)
Other stance bundles	<i>but is it really</i> (Corpus I and II) <i>what to do with</i> (Corpus III)
<b>3. Discourse-organizing lexical bundles</b>	
Lexical bundles that introduce the topic	<i>to begin with the</i> (Corpus I) <i>there are a lot of</i> (Corpus II) <i>when it comes to</i> (Corpus III)
Lexical bundles that develop the topic	<i>however there are people</i> (Corpus I) <i>at the same time</i> (Corpus I) <i>on the contrary</i> (Corpus II) <i>that it is not</i> (Corpus II) <i>this is not the</i> (Corpus III) <i>so that they can</i> (Corpus III)
Lexical bundles that conclude the topic	<i>taking everything into account</i> (Corpus I) <i>to conclude language is</i> (Corpus II)

The classification of the lexical bundles by their discourse functions once again revealed a gradual shift in the distribution across the major types and subtypes in the three

corpora. While the referential lexical bundles account for the largest proportion of the functional subtypes in all the corpora (see Figure 2), their proportional weight increases alongside the proficiency level of the learners: the higher the proficiency level, the larger the proportion of referential bundles.

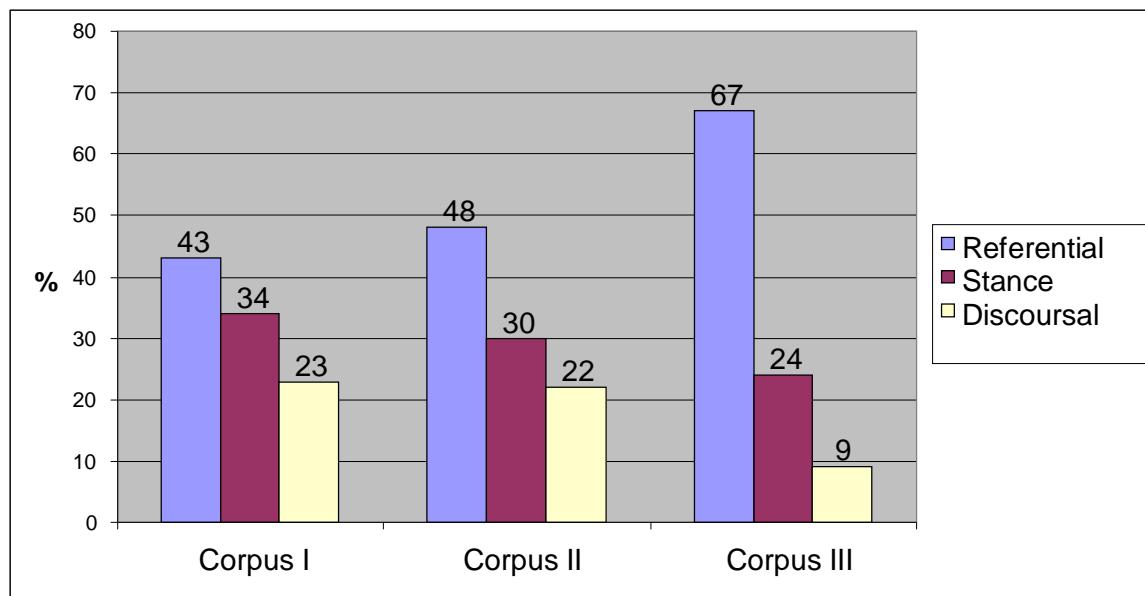


Figure 2. Distribution of functional types of lexical bundles in the corpora

In comparison to the Lithuanian corpora (I and II), Corpus III yields a significantly larger proportion of the referential bundles though it has fewer discourse-organizing bundles which account for more than one fifth of all the bundles in Corpora I and II (23 and 22 per cent respectively). In general, the data suggests that the language of NSs is more propositionally dense than the language of NNSs, which is evidenced by the number and frequency of referential expressions repeated in the lexical bundles. The Lithuanian learners seem to be compensating the propositional density of their essays by a rather intense use of discourse-organizing expressions and stance expressions.

The analysis of the functional subtypes of lexical bundles pointed out certain peculiarities of written English produced by the Lithuanian learners. As regards the referential bundles, the Lithuanian students often resort to very general and broadly understood references. For example, many lexical bundles of this functional type contain the words *people* and *world* (cf. *a lot of people, all over the world, peole who say that*) which diminish the referential definiteness of the text and make it inappropriately vague. Another peculiarity of the Lithuanian learner writing is evident from the analysis of the

stance bundles. In comparison to the data of Corpus III, the Lithuanian students are significantly overusing expressions of deontic and epistemic modality, for example, *there is no need, it is necessary to, it is known that, there is no doubt* etc. They are particularly fond of evaluative expressions, e. g. *it is better/hard/easier/difficult to* etc. Such findings may be explained by the specificity of the essay topics in which the students had to argue for and against certain ideas which, as a consequence, called for a more explicit expression of the authorial stance. Furthermore, the significant overuse of discourse-organizing bundles reflects the students' attempts to observe the requirements of cohesion in writing which, obviously, is seen as an attribute of good writing. Yet the overuse of certain expressions hardly contributes to the quality of writing and, most importantly, reflects the limited lexical expression of the Lithuanian learners.

The functional and structural analysis pointed out yet another research direction. It was observed that part of the lexical expressions which occur in the automatically retrieved lists of lexical bundles are phrasal, or formulaic. It was thus decided to compare the corpora on the basis of phrasal expressions which are repeated in the lexical bundles.

For this purpose, the sample of lexical bundles was enlarged to include two-, three-, four-, five- and six-word bundles that are repeated at least three times in the essays of three different authors. The automatically generated lists were manually edited to eliminate lexical bundles of varying lengths representing identical expressions, for example:

*to be able* (11 occurrences)  
→ *to be able to* (11)  
*to be able to* (11)  
 (Corpus I)

The longer bundle was included in the further analysis while the shorter was eliminated from the sample. The other revision strategy concerned topic-specific bundles all of which were excluded from the sample, irrespectively of whether they were identical repetitions of the essays rubrics or they merely contained one topic-specific word. So the total number of lexical bundles for this analysis included 6488 lexical bundles (1683 from Corpus I, 2330 from Corpus II and 2475 from Corpus III). The Collins COBUILD

English Dictionary for Advanced Learners (2001, 3rd edition) was used to check every individual phrase occurring in the lexical bundles for its phraseological status.

**The analysis of phrasal expressions** showed that lexical bundles contain four types of phrasal expressions:

- 1 - phrasal conjunctions, e. g. *as if, as soon as, in order to*;
- 2 – phrasal prepositions, e. g. *due to, in front of, in the course of*;
- 3 – phrasal verbs, e. g. *to go through, to sum up, to give birth*;
- 4 – phrases, e. g. *in addition to, in conclusion, take the easy way out*.

The first three types of phrasal expressions correspond to particular word classes while the category *phrases* covers expressions which are non-compositional semantically and which may belong to different parts of speech. Figure 3 summarizes the quantitative data of the phrasal expressions occurring in the lexical bundles.

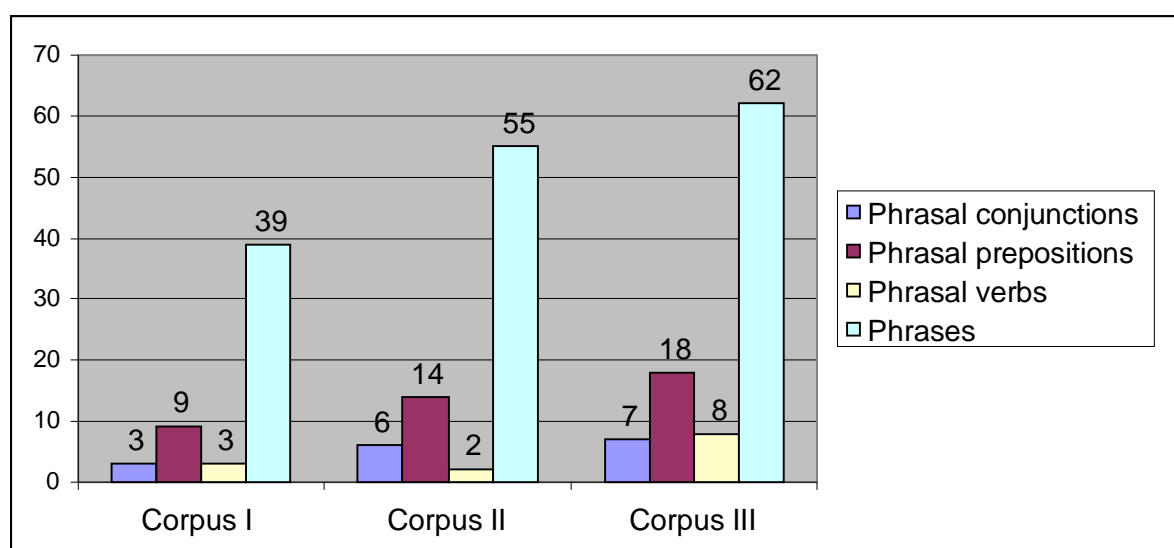


Figure 3. Number of phrasal expressions occurring in the lexical bundles

As shown in Figure 3, the number of different phrasal expressions occurring in the lexical bundles is directly related to the general proficiency of the learners – the largest number of such expressions was established in the data retrieved from Corpus III.

In order to get a better view of the differences in the use of the phrasal expressions, several items were chosen for a more detailed investigation. They include the phrasal preposition *in order*, the phrasal conjunction *because of*, the phrasal verb *to sum up*, and the phrases *first of all, on the other hand, at the same time, the fact that*, and

*in my opinion*. In general, the analyses showed that the Lithuanian learners tend to overuse very significantly those expressions that are either related to discourse organization, e. g. *first of all, in order*, or are very well-known to the learners, e. g. *all over, to sum up* or *as well*. Yet the qualitative analysis of the selected phrasal expressions provides data which shows that these expressions are often reproduced in identical lexical contexts, for example, the preposition *all over* occurs exclusively with the noun phrase *the world* as in the phrase *all over the world*. It proves that the learners of this level of proficiency still do not fully understand the meaning of the phrasal preposition *all over* and thus use it in the only context they are familiar with.

Furthermore, quite often phrasal expressions seem to be used inappropriately in the Lithuanian learner writing. The corpus data suggests that these expressions are sometimes misunderstood by the learners even though the learners do not have problems in reproducing the expressions accurately in writing. The misunderstanding may sometimes be explained by the mother tongue interference yet certain difficulty in the usage could be related to the developmental problems of the learners' written competence. The qualitative analysis allows to conclude that the two Lithuanian corpora, *i. e.* Corpus I and Corpus II, bear more similarity as regards the use of phrasal expressions while Corpus III in this respect stands out as an example of a much more varied phraseology.

## CONCLUSIONS

The analysis of lexical bundles retrieved from the three corpora of written English produced by the learners at different levels of proficiency provides useful insights into the development of lexical competence in writing. Since the study involves a contrastive analysis of Lithuanian EFL learners and native speakers of English, it also sheds light on the specific problems that the Lithuanian learners of English encounter and the impact of these problems on **the quality of Lithuanian learner writing**.

The study showed that the development of written competence progresses from spoken to written, *i. e.* when writing the less proficient learners tend to rely on their spoken lexical resources. Gradually, as the level of proficiency increases, lexical expressions characteristic of spoken English give way to the written English vocabulary. This conclusion first of all draws on the findings of the structural analysis of lexical

bundles which shows that the lexical bundles incorporating noun and prepositional phrases, typical of written English, are less prominent in the data retrieved from Corpus I and Corpus II and account for the largest proportion of the structural bundles types in Corpus III. Hence the process of teaching English as a foreign language could include a more direct teaching of register-specific vocabulary and in this way increase the learners' awareness of register differences.

The functional analysis of lexical bundles confirmed that the development of language proficiency and, particularly, written competence, is related to the learners's ability to express in writing specific discourse functions. The corpora representing more proficient and advanced learners, *i. e.* Corpus II and Corpus III, contain more lexical bundles that perform the referential function. These lexical bundles incorporate phrases, naming abstract and concrete objects, they also contain topic-specific words, time and place references *etc.* Significantly smaller proportions of lexical bundles in these two corpora have been categorized as incorporating stance and discourse-organizing expressions. In contrast, the distribution of lexical bundles across the functional types in Corpus I is more even.

The Lithuanian learners, whose written English is represented in the study by Corpus I and Corpus II, are much more focused on the appropriate and extensive use of discourse-organizing expressions and in this respect they certainly demonstrate their awareness of text cohesion. The frequencies and numbers of the lexical bundles incorporating such expressions are significantly higher in Corpus I and Corpus II than in Corpus III. This could be explained both by the teaching foci of the academic writing courses in the students' study programmes, where much attention is given to the formal attributes of the academic essay, and the more varied lexical expression of the NS students who create cohesion by other lexical means than the overt discourse organizers.

Another distinctive feature of the Lithuanian student writing was observed in the analysis of the lexical bundles incorporating stance expressions, which are significantly overused in the Lithuanian corpora in comparison to the Corpus III data. This overuse partly reflects the younger age of the learners whose language is represented by Corpus I and who tend to be more categorical and uncompromising in their views and opinions. On the other hand, the essay topics encouraged the authors to express their opinions which they did by using a limited number of stance expressions. The language of Corpus III,

where certain topics also required to express a personal opinion, is more lexically varied and this explains why this corpus yields fewer repetitive sequences. The data suggests that the Lithuanian learners, similarly to the tendency observed in the use of discoursal expressions, rely on a limited set of stance expressions, for example, phrases with the adjective *important* or the modal verb *need*, which explains the considerable differences in the frequencies of lexical bundles incorporating stance expressions among the three corpora.

The analysis of phrasal expressions occurring in the lexical bundles allows to conclude that the phraseological competence of the Lithuanian learners does not develop very intensely during the two-three years of their study in the university. They seem to be overusing such phrasal expressions which are easily transferrable from their mother tongue and which are associated with a clear discourse function, especially discourse organization. Yet a more in-depth analysis of certain expressions proves that the learners do not always fully understand the meaning of the phrasal expressions which they overuse and, as a consequence, they fail to use them appropriately.

The study carries diverse **implications for further research**. Firstly, it revealed the impact of topic-specific vocabulary on the essay language. Since the essay is a popular test format in many testing contexts, test developers have to bear in mind the influence of task rubrics and other input material given to the testees for essay writing. The study showed that students make an extensive use of all linguistic resources available to them in the stressful test situation and in this way try to compensate for their limited vocabulary. It may be acceptable in some testing situations but it certainly impoverishes the authenticity of learner language. Secondly, the study also raises the question of the assessment criteria of the essay. The data showed that the less proficient learners are fully aware of certain formal attributes of essay writing, for example, connecting devices, and they seem to be very particular about their use. One of the reasons behind it is to be found in the essay assessment scales where text organization is one of the criteria. Unfortunately, it sometimes unduly diminishes the importance attached by the students to the content of the essay and expression of arguments, which is particularly evident in the language of intermediate and partly advanced EFL learners. They seem to be pre-occupied with the matters of text organization and forget about the ideas expressed in their texts.

**Limitations of the study** mainly arise from the size of the learner corpora. It is assumed that the larger corpora would provide more evidence to certain tendencies which could not be further investigated due to the scarcity of material. Moreover, Corpus I and Corpus II represent the language produced by the students of one study programme so the findings of the study are undoubtedly limited and applicable to this particular group of learners. To allow for broader generalizations, the study should be extended to a larger corpus of learner language representing learners from a variety of educational contexts.

Another limitation of the study may be ascribed to the research approach which involved the analysis of automatically retrieved recurring sequences, or lexical bundles. Admittedly, this research approach leaves out many interesting lexical units that do not recur in the corpora in the identical form and, as a consequence, are not represented in the lexical bundles. Yet it could be argued that the lexical bundles quite accurately reflect frequency data derived from the analysis of individual lexical items. Moreover, the lexical bundles offer additional information about the development of word patterns in learner language and the learners' ability to reproduce phrasal expressions.

Finally, the study confirmed that research into learner language is hardly possible within one discipline. This study was undertaken as corpus linguistics research in the tradition of functional and lexical grammar yet it would have also benefited from a more direct reference to the insights of the theory of second language acquisition. As observed by Granger (2009), the tradition to approach learner language research as interdisciplinary is not yet very strong so, hopefully, further studies involving other disciplines will contribute to a better understanding of the development of learner language.



# LEKŠINĖS SAMPLAIKOS SVETIMKALBIŲ IR GIMTAKALBIŲ VARTOTOJŲ RAŠYTINĖJE ANGLŲ KALBOJE

## Reziumė

**Tyrimo objektas.** Pasikartojančių žodžių sekų, t. y. leksinių samplaikų (Biber *et al.* 1999), tyrimai tekstynų lingvistikoje parodė naujų anglų kaip svetimšios kalbos (ASK) tyrimo galimybių. Duomenys, patvirtinantys, kad autentišką anglų kalbą sudaro keliažodžiai leksiniai vienetai, paskatino tyrėjus atkreipti dėmesį į tokių junginių vartojimą svetimkalbių mokinių kalboje. Šis darbas skirtas gretinamajai leksinių samplaikų, pasikartojančių rašytinėje svetimkalbių ir gimtakalbių mokinių (angl. *learner*) anglų kalboje, analizei. Darbe analizuojama iki šiol išsamiau netirta lietuvių gimtosios kalbos vartotojų rašytinė anglų kalba, sukaupta elektroniniuose tekstynuose. Tikimasi, jog naujausių kalbotyros mokslo pasiekimų ir tyrimo metodų taikymas pateiks vertingų įžvalgų apie lietuvių ASK mokinių kalbinę kompetenciją, leis nustatyti specifinius jų patiriamus sunkumus besimokant anglų kalbos ir prisidėti prie tolesnės ASK mokymo proceso pažangos.

Kaip rodo ankstesni kitų autorių darbai (plg. Altenberg 1998, Biber 2006, Hyland 2008, Cortes 2008, Chen ir Baker 2010), leksinių samplaikų analizė atskleidžia rašytinių tekstų ypatumus, nustatomus tiriant vyraujančius struktūrinius ir funkcinius leksinių samplaikų tipus. Tikėtina, jog skirtingų mokėjimo lygių mokinių kalbą reprezentuojančių tekstynų analizė leis nustatyti, kaip kylant bendrajam kalbos mokėjimo lygiui kinta rašytinės kalbos kompetencija. Be to, struktūrinių ir funkcinių leksinių samplaikų tipų tyrimas gali parodyti ir lietuvių ASK mokinių rašytinės anglų kalbos ypatumus. Taigi šio tyrimo įžvalgos gali būti itin naudingos ASK mokymo praktikai Lietuvoje.

Disertacijoje aprašomas darbas sumanytas kaip tekstynų inspiruotas tyrimas, kuriame iš anksto formuluoti tik tyrimo klausimai ir tikslai, nekeliant jokių pirminių hipotezių. Būtent už tokį mokinių kalbos tyrimų kelią pasisako ir šiuolaikinės tekstynų lingvistikos atstovai (Sinclair 2004, Tognini-Bonelli 2001), ir mokinių kalbos tyrinėtojai (Granger 1998a). **Tyrimo tikslai** yra palyginti skirtingų mokėjimo lygių anglų kalbos vartotojų rašytinės kalbos leksiką, tiriant pasikartojančias žodžių sekas, t. y. leksines samplaikas; aprašyti vyraujančius struktūrinius ir funkcinius leksinių samplaikų tipus ir įvertinti jų įtaką rašytinės kalbos kokybei; aprašyti į leksines samplaikas patenkančius

sustabarėjusius (pastoviuosius) junginius. Šių tikslų siekiama sprendžiant tokius **tyrimo uždavinius**:

1. Sukaupti skirtingų lygių lietuvių gimtosios kalbos ASK mokinių rašytinės anglų kalbos tekstynus.
2. Ištirti leksinių samplaikų kiekybinius ir dažnumo parametrus mokinių kalbos tekstynuose.
3. Išanalizuoti leksinių samplaikų struktūrinius ir funkcinius ypatumus tiriamuosiuose tekstynuose.
4. Palyginti tekstynus pagal sustabarėjusių (pastoviųjų) junginių realizavimą leksinėse samplaikose.

**Tyrimo taikomoji vertė ir naujumas.** Iki šiol lietuvių gimtosios kalbos vartotojų anglų kalbos kompetencija tyrinėta fragmentiškai, todėl mokslo literatūroje nėra išsamių darbų apie tai, kokių specifinių sunkumų patiria lietuvių ASK mokiniai ir kaip kinta jų žodynas kylant bendrajam kalbinės kompetencijos lygiui. Kitas naujas tyrimo aspektas yra susijęs su tyrimo metodika. Šiame tyrime pirmą kartą lietuvių ASK mokinių leksinės kompetencijos tyrimui naudojami du skirtingų mokėjimo lygių lietuvių ASK mokinių kalbą reprezentuojantys elektroniniai rašytinės kalbos tekstynai.

Aprašomas tyrimas gali turėti nemažos taikomosios naudos. Tyrimo duomenų pagrindu gali būti rengiamos žodyno plėtrai skirtos mokymo priemonės vidurinių mokyklų mokiniams ir aukštųjų mokyklų studentams. Kitos tyrimo išdavų taikymo galimybės sietinos su kalbos mokėjimo lygių aprašais ir vertinimu bei testavimu. Tyrimo duomenys leidžia manyti, jog vertinant rašymo užduotis verta atsižvelgti į mokinių gebėjimą vartoti sustabarėjusius junginius, kurie iš dalies lemia ir rašomo teksto kokybę bei priartina mokinių kalbą prie gimtakalbių vartosenos.

Tyrimui naudoti tekstynai turi ir išliekamosios vertės. Specialiai šiam tyrimui sukauptas pirmakursių rašinių tekstynas, kaip ir kiti du tyrime naudoti tekstynai, turi daug įvairių taikymo paskirčių. Be savo tiesioginės naudos mokslo reikmėms, šie tekstynai, kaip rodo daugelio užsienio universitetų patirtis, gali būti intensyviai naudojami ASK mokinių mokymui. Tolesnis šių tekstynų kaupimas suteiktų studentams ir tekstynų kaupimo praktikos, ir tekstynų analizės pagrindų.

Be to, šis tyrimas vertintinas ir kaip vienas iš daugelio šiuo metu pasaulyje atliekamų gretinamųjų mokinių kalbos tyrimų, papildantis juos duomenimis apie lietuvių ASK mokinių rašytinės anglų kalbos ypatumus. Autorė taip pat tikisi, jog tyrimas gali sudominti ir platesnę skaitytojų auditoriją bei prisidėti prie tolesnės mokinių kalbos tyrimų plėtros.

### **Ginamieji teiginiai:**

1. Kiekybinis ir kokybinis tiriamuosiuose tekstynuose nustatytų leksinių samplaikų gretinimas rodo, jog tekstynai reprezentuoja skirtingų mokėjimo lygių mokinių kalbą. Vadinasi, leksinių samplaikų analizė leidžia nustatyti specifinius skirtingų mokėjimo lygių anglų kalbos vartotojų rašytinės kalbos ypatumus.
2. Pagal leksinių samplaikų struktūrą ir funkciją trys tirtieji mokinių rašytinės anglų kalbos tekstynai reikšmingai skiriasi, o nustatytieji skirtumai rodo, jog žemesnio mokėjimo lygio mokinių kalba turi daugiau sakytinės kalbos bruožų. Kylant mokėjimo lygiui mokinių kalboje randasi daugiau autentiškai rašytinei anglų kalbai būdingų ypatumų.
3. Žemesnio kalbos mokėjimo lygio mokinių rašytinėje anglų kalboje esama daugiau diskurso organizavimo ir autoriaus pozicijos raiškos priemonių, kurias kylant bendrajam kalbinės kompetencijos lygiui laipsniškai keičia referentinė leksika, t. y. intensyvėja teksto propozinio turinio raiška.
4. Skirtumus tarp skirtingų mokėjimo lygių mokinių rašytinės kalbos lemia ne tiek tematiškai specifinės leksikos vartojimas, kiek tematiškai neutrali ir rašytinei anglų kalbai būdinga leksika.
5. Leksinių samplaikų tyrimas atskleidžia bendrąsias leksinių dėsningumų ir leksinių-gramatinių konstrukcijų vartojimo tendencijas mokinių kalbos tekstynuose, todėl leksinių samplaikų tyrimas suteikia vertingų įžvalgų tolesnėms tyrimo kryptims.

**Tiriamoji medžiaga ir tyrimo metodai.** Tyrimas atliktas kaip tekstynų inspiruotas gretinamasis mokinių kalbos tyrimas. Jo duomenys rinkti iš trijų mokinių kalbos tekstynų. I tekstynas reprezentuoja Vilniaus universiteto pirmojo kurso Anglų filologijos bakalauro pakopos studentų rašytinę anglų kalbą. Jį sudaro egzaminų rašiniai įvairiomis su studijų programa susijusiomis temomis. II tekstynas sudarytas iš tos pačios

studijų programos trečiojo ir ketvirtojo kursų studentų samprotaujamųjų rašinių. Visų I ir II tekstynų rašinių autorių gimtoji kalba – lietuvių. III tekstyną sudaro iš LOCNESS tekstyno paimti gimtakalbių (britų ir amerikiečių) studentų rašiniai. Tekstynų dydžiai: I tekstynas – 92050 žodžių, II tekstynas – 137 004 žodžiai ir III tekstynas – 164 687 žodžiai.

Tekstynų analizei naudota programa „WordSmith Tools“ (5 v.). Šia programa buvo generuojami leksinių samplaikų sąrašai, sudaromi dažniniai žodžių sąrašai bei individualių žodžių konkordansai. Tekstynų duomenys tirti taikant kiekybinius ir kokybinius tyrimo metodus: atlikti statistinio reikšmingumo testai ( $\chi^2$  ir logaritminės tikimybės testai), kokybiniai metodai taikyti analizuojant leksinių samplaikų struktūrą ir funkcijas bei frazinių junginių, realizuojamų leksinėse samplaikose, vartojimą.

**Darbo struktūra.** Disertaciją sudaro įvadas, trys pagrindinės dalys, išvados, literatūros sąrašas ir priedai. Įvade aptariamas darbo aktualumas, pristatomas tyrimo objektas, pateikiami tikslai ir uždaviniai, formuluojami ginamieji teiginiai. Pirmojoje dalyje apžvelgiami svarbiausi žodžių junginių tyrimai tekstynų lingvistikoje, aptariama jų įtaka mokinių kalbos tyrimams bei pristatomi svarbiausi žodžių junginių tyrimai ASK mokinių kalboje. Antroji disertacijos dalis skirta tyrimo medžiagai ir metodams. Trečiojoje dalyje pateikiama duomenų analizė, suskirsčius ją į tokias skyrius: kiekybiniai duomenys, leksinių samplaikų struktūra, leksinių samplaikų funkcijos, į leksines samplaikas patenkantys fraziniai junginiai. Disertacija baigiama išvadomis. Darbo pabaigoje pateikiami literatūros sąrašas ir septyni priedai.

**Išvados.** Tyrimas parodė, kad leksinių samplaikų analizė gali pateikti vertingų duomenų apie mokinių kalbos raidą. Kadangi tyrime buvo gretinami lietuvių ASK mokinių ir gimtakalbių vartotojų kalbą reprezentuojančių tekstynų duomenys, tyrimas leidžia daryti išvadas apie lietuvių mokinių rašytinės anglų kalbos ypatumus.

Leksinių samplaikų struktūros tyrimas parodė, kad I ir II tekstynai, reprezentuojantys skirtingų mokėjimo lygių lietuvių ASK mokinių rašytinę anglų kalbą, turi daugiau sakytinei anglų kalbai būdingos struktūros samplaikų, t. y. veiksmazodinių samplaikų ir samplaikų su predikatiniais sakinių dėmenimis. III tekстыne, reprezentuojančiame gimtakalbių anglų kalbos vartotojų kalbą, tokios struktūros samplaikų skaičius kone dvigubai mažesnis. Taigi galima konstatuoti, kad rašytinės kalbos kompetencijos raida apima sakytinei kalbai būdingos leksikos perkėlimą į

rašytinę kalbą, tad ugdant ASK mokinių rašytinę kompetenciją daugiau dėmesio verta skirti sakinio ir rašytinio anglų kalbos diskursų ypatumams.

Pagal tam tikrus struktūrinius leksinių samplaikų ypatumus, pavyzdžiui, sakinio kamieną realizuojančių samplaikų dažnį, lietuvių ASK mokinių rašytinė anglų kalba lygintina su kitų gimtųjų kalbų vartotojų (pvz., prancūzų, švedų) rašytine anglų kalba. Tačiau, apskritai, visuose trijuose tirtuosiuose tekstynuose didžiausią samplaikų dalį sudaro vardažodinės ir prielinksninės samplaikos, kurios ir yra dažnesnės rašytinėje anglų kalboje.

Funkcinė analizė parodė, kad kylant bendrajam kalbos mokėjimo lygiui nuosekliai kinta skirtingų kalbos funkcijų raiškos intensyvumas. I ir II tekstyne daugiau samplaikų vartojamos autoriaus požiūriui reikšti ir susieti atskiras teksto dalis, o III tekstyne sumažėja autoriaus požiūrio raiškos ir ypač teksto junglumo funkcijai skirtų samplaikų, tačiau čia nustatyta kur kas daugiau referentinių samplaikų. Vadinas, žemesnio lygio mokinių kalboje pasitaiko daugiau nuomonių, vertinamųjų pasakymų ir teksto junglumo raiškos priemonių, tačiau mažiau propozicinį turinį perteikiančios leksikos.

Frazinių junginių tyrimas patvirtino, kad gebėjimas rašant vartoti keliažodžius vientisos reikšmės junginius tiesiogiai susijęs su bendruoju kalbos mokėjimo lygiu, nors pagal atskirų junginių dažnį žemesnio mokėjimo lygio mokiniai kartais net pranoksta gimtakalbius vartotojus. Lietuviai ASK mokiniai itin dažnai rašydami atkuria tokius pasakymus, kurie kone pažodžiui gali būti perkelti iš gimtosios kalbos arba kurių reikšmę mokiniams nesunku suprasti. Deja, tyrimo duomenys rodo, jog mokiniai ne visuomet teisingai suvokia vieno ar kito pasakymo paskirtį tekste, kartais juos vartoja neapgalvotai ar bando skaidyti sustabarėjusius pasakymus, o tai menkina jų rašytinės kalbos kokybę.

Atliktas tyrimas pateikė keletą išvalgų apie ASK gebėjimų tikrinimą ir vertinimą. Galima daryti išvadą, kad kalbos testuose ypatingą svarbą įgyja užduočių pateikčių leksika, ypač jei pateiktys formuluojamos ilgesnėmis žymių autorių citatomis ar ištraukomis iš jų darbų. Rašydami egzamino rašinį mokiniai intensyviai išnaudoja visą jiems prieinamą kalbinę medžiagą, todėl jų rašiniuose nuolat kartojami iš pateikčių pasiskolinti žodžiai, ir tai neabejotinai mažina mokinio kalbos autentiškumą.

Kita testavimo požiūriu įdomi tyrimo įžvalga susijusi su samprotaujamojo rašinio žanriniais ypatumais ir vertinimo kriterijais. Šis testo formatas yra itin dažnas įvairiuose akademinuose mokymo ir vertinimo kontekstuose, nors tyrimo duomenys iškėlė įvairių klausimų apie tokio pobūdžio teksto žanrinius ypatumus. Iš tiesų neaišku, kas yra svarbiau samprotaujamajame rašinyje – autoriaus pozicijos raiška, turinys ir dalykiškumas ar formalieji diskurso organizavimo dalykai? Lietuvių studentų rašiniuose dalykiškumą kartais gali užgožti perdėm formalizuoti pasakymai, kuriais kuriamas teksto junglumas ir grindžiama autoriaus pozicija, tačiau palyginti su gimtakalbių studentų rašiniais juose mažiau dėmesio skiriama turinio ir argumentacijos raiškai.

## LIST OF PUBLICATIONS ON THE SUBJECT OF THE DISSERTATION

1. Juknevičienė R. 2007. Analyzing Topic-Specific Vocabulary in EFL Student Writing, in *Research Papers of Nordic-Baltic FIPLV Conference "Innovation in Language Teaching and Learning in the Multicultural Context"*. Riga: Izglitibas soli. 161–168.
2. Juknevičienė, R. 2008. Collocations with high-frequency verbs in learner English: Lithuanian learners vs. native speakers. *Kalbotyra* 59 (3): 119–127.
3. Grigaliūnienė J., Bikelienė, L. and R. Juknevičienė. 2008. The Lithuanian Component of the International Corpus of Learner English (LICLE): a resource for English language learning, teaching and research at Lithuanian institutions of higher learning. *Žmogus ir žodis* 10 (III): 62–65.
4. Juknevičienė, R. 2009. Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra* 61 (3): 61–72.

## CONFERENCE PRESENTATIONS ON THE SUBJECT OF THE DISSERTATION

1. *Analyzing topic-specific vocabulary in EFL student writing*, FIPLV (Fédération Internationale des Professeurs de Langues Vivantes) Conference in Riga (Latvia), 15–16 June, 2007
2. *The use of collocations in argumentative essays of Lithuanian EFL learners* (poster presentation), ALTE (Association of Language Testers in Europe) 3<sup>rd</sup> International Conference, Cambridge University, Cambridge (UK), 10–12 April, 2008
3. *Anglų kaip svetimosios kalbos produktyviojo žodyno vertinimo parametrai* (Vocabulary assessment in the tests of writing), Conference of PhD students, Vilnius University (Lithuania), 24 April, 2008
4. *"Verb-Noun" Collocations in Learner Language: Lithuanian learners of English vs. Native Speakers*, International Conference on English and German Studies, Vilnius University (Lithuania), 18–20 September, 2008
5. *Insights from the Lithuanian Learner Corpus of English* (a co-authored poster presentation with Lina Bikelienė), International Seminar on Corpus Linguistics in honour of John Sinclair, Granada (Spain), 22–24 September, 2008
6. *Lexical bundles in learner language: Lithuanian learners of English vs. native speakers*, International conference on language teaching, Institute of Foreign languages, Vilnius University (Lithuania), 8–9 October 2009
7. *Lexical bundles in learner language: Lithuanian learners of English vs. native speakers*, FLaRN (Formulaic Language Research Network) International Conference in Paderborn (Germany), 23–26 March, 2010
8. *Patterns of lexis in learner language: Lithuanian learners of English vs. native speakers*, SLE (Societas Linguistica Europaea) 43rd Annual Meeting in Vilnius (Lithuania), 2–5 September 2010

Rita Juknevičienė (g. 1968) 1991 m. baigė Anglų filologijos specialybės studijas Vilniaus universiteto Filologijos fakultete. Nuo 1996 m. dirba Vilniaus universiteto Filologijos fakulteto Anglų filologijos katedroje, yra dėščiusi įvairių anglų kaip svetimosios kalbos (gramatikos, rašymo, žodyno) ir kitų dalykų (kalbos testų rengimo, bendrosios kalbotyros, leksikografijos, kursinio darbo rašymo ir kt.) kursų. 2003 ir 2004 m. Lankasterio universitete mokėsi kalbos gebėjimų testavimo ir statistinės analizės kursuose.

2006 m. įstojo į Vilniaus universiteto Filologijos mokslo krypties doktorantūrą. Doktorantūros studijų metais dalyvavo tarptautinėse mokslinėse konferencijose, vykdė mokinių kalbos tekstyno rengimo projektą (ICLE), paskelbė keturis mokslinius straipsnius. 2009 m. gavusi VU doktorantų mobilumo fondo paramą buvo išvykusi trumpalaikiai stažuotei į Liuveno (Louvain-la-Neuve) katalikiškojo universiteto (Belgija) Anglų tekstynų lingvistikos centrą.

Rita Juknevičienė (b. 1968) graduated from Vilnius University and received a diploma in English Philology (equivalent to an MA degree). Since 1996 she has been working at the Department of English Philology of the Faculty of Philology in Vilnius University where she has taught various EFL courses (grammar, academic writing, vocabulary) and other subjects, namely, language testing, introduction to general linguistics, lexicography, research paper writing etc. In 2003 and 2004 she attended the courses in language testing and test statistics in the University of Lancaster (UK).

In 2006 she started her PhD studies in Vilnius University. During the studies she took part in a number of international conferences in Lithuania and abroad, published four research articles, was engaged in the compilation of the Lithuanian component of the ICLE corpus. In 2009 she was on a short library visit in the Catholic University of Louvain-la-Neuve in Belgium.