www.volsu.ru

ENGLISH ACADEMIC DISCOURSE IN TRANSLINGUISTIC CONTEXT: CORPUS-BASED STUDY OF LEXICAL MARKERS

Larisa A. Kochetova

Volgograd State University, Volgograd, Russia

Elena Yu. Ilyinova

Volgograd State University, Volgograd, Russia

Abstract. The paper applies the Matrix method to an investigation into translinguistic features of English academic discourse that is serving worldwide a means of cross-cultural exchange between researchers with translingual skills. Based on the corpus approach to the comparison of the two corpora that comprise samples of professional academic writing in various fields of study (Art and Humanity, Natural and Social sciences), the paper seeks to identify both quantitatively and qualitatively correlations in repertoire and frequencies of recurrent linguistic expressions between the native English-language and non-native (Russian) academic discourse performers. The corpora were investigated along with the use of lexical bundles, re-occurring lexical units, which were grouped into noun-based and preposition-based phrases with post-modifier fragments, verb-based phrases with any form of verb components. The data comparison points to a code-mixing trend at the syntagmatic layer, which is a translingual fusion of English words in accord with a mixture of syntagmatic relations typical of English and Russian variations of academic discourse. It was found that non-native writing does not reveal as much lexical flexibility as native writers do and to a large extent relies on formulaic expressions, most of which are not conventional for expert native academic writing. Native Russian writers use excessively noun-based phrases with abstract nouns and underuse noun phrases without prepositions. Verb phrases with *that*- and *to*-clauses are mainly characteristic of native professional writing whereas non-native writing employs patterns with active verbs and passive constructions. It was found that non-native writing lacks quantifying phrases and hedging expressions that mitigate the proposition.

Key words: academic discourse, corpus, corpus methodology, lexical bundles, English language, translingualism.

АНГЛОЯЗЫЧНЫЙ АКАДЕМИЧЕСКИЙ ДИСКУРС В ТРАНСЛИНГВАЛЬНОЙ СИТУАЦИИ: КОРПУСНО-ОРИЕНТИРОВАННОЕ ИЗУЧЕНИЕ ЛЕКСИЧЕСКИХ МАРКЕРОВ

Лариса Анатольевна Кочетова

Волгоградский государственный университет, г. Волгоград, Россия

Елена Юрьевна Ильинова

Волгоградский государственный университет, г. Волгоград, Россия

Аннотация. В статье представлены результаты, полученные при применении матричного подхода к изучению признаков транслингвальности англоязычного академического дискурса, ставшего каналом кросс-культурного обмена научной информацией между учеными, обладающими трансъязычными средствами коммуникации. Корпусные методики сравнительного анализа двух корпусов научно-ориентированных текстов по различным сферам знаний (гуманитарные, социальные, естественные науки) позволили выявить количественные и функциональные корреляции в лексико-синтагматическом репертуаре единиц дискурса на основе критерия частотности употребления типичных конструкций носителями английского языка и русскоговорящими авторами научных статей. В качестве лексических маркеров английского академического дискурса были выбраны лексические коллокации, высокочастотные словосочетания трех синтагматических групп: именные и предложные словосочетания с атрибутивными компонентами, глагольные словосочетания, включающие любые предикативные формы, в том числе *to-clause*, безличные *it-patterns* модели. Сравнение корпусных данных указывает на наличие кодовых смешений на синтагматическом уровне, употребление английских слов, являющихся транслингвальными кальками, для выражения синтагматических отношений, характерных для русскоязычного академического дискурса. Показано, что русскоязычные авторы, пишущие на английском языке, часто используют формульные клише, не характерные для стандартного англоязычного академического текста, именные конструкции с абстрактными существительными и беспредложные номинативные конструкции. Глагольные конструкции с придаточными, вводимыми союзом *that* и частицей *to*, типичны для профессионального дискурса носителей английского языка, в то время как русскоязычные авторы употребляют преимущественно конструкции с глаголами в действительном залоге и синтагматические клише, содержащие глаголы определенной семантики в страдательном залоге. Установлено, что в академическом дискурсе русскоязычных авторов отсутствуют дискурсивные формулы с семантикой количества и выражения, смягчающие категоричность высказывания.

Ключевые слова: научный дискурс, корпус, корпусная методология, лексические сочетания, английский язык, транслингвизм.

Introduction

Translingualism has become increasingly popular as a theoretical construct that points to a current trend – a dynamic fusion between permeability and convergence of languages and cultures in human brains; it is a characteristic of modern translingual communication in the globalizing world. Translangualism results from an increase in the number of people who apart from their mother tongue have proficiency in one or several foreign languages, and while communicating information they might refer to translingual signs to convey their ideas through language code-switching and messing. It results in translanguaging, a kind of communication process whereby signs and rules of various languages are integratively used in the discursive practice without being merged or submitted.

The phenomenon of translingualism is being studied worldwide in social and psychological perspectives. Linguists focus more on the trans-crossing of grammatical and lexical features of various languages in the context of translingual communication, in particular, some theoretical points are set on the bilingual corpus data and by employing a Matrix methodology groups of researchers investigate aspects of code-switching in written or oral speech [Chauncey, Grainger, Holcomb, 2008; Deuchar, 2006; Fricke, 2016; Gabdreeva, Ageeva, Timirgaleeva, 2013; Gardner-Chloros, 2009; Gevers, 2018; Gullberg, Indefrrey, Muysken, 2009; Maksimova, 2016; Marinova, 2008; Molina, 2011; Moradi, Chen, 2019; Myers-Scotton, 2006; etc.]. While recognizing translanguaging as a hybrid cognitive strategy of communication using mixing language patterns and units from native and second languages, we would like to suggest for consideration another issue on translingualism – discursive aspect of academic writing in English by Russian professionals due to the necessity to exchange scientific facts and discoveries in the global academic discourse. This paper presents an analysis of corpus-based data to support a hypothesis on translingual code-mixing in scientific articles authored by Russian professional scientists that is manifested in some

lexical markers of Russian academic discourse incorporated into papers written in English.

Terminology review

The transcultural and translingual trend was distinguished in speech practices in the late 1990s by specialists in applied linguistics and language teaching [Zamel, 1996; 1997]. While analyzing speech behavior of bilinguals and translinguals, they noted that due to their polylanguage repertoire they could produce translinguistic discursive unities (based on a mix of language units and rules) which pointed to their ability to code-switching and code-mixing. Having chosen a syntactic (grammar) basis of one of the languages, they could layer morphemes from various languages with some evident deviations in syntagmatic patterns from the standard grammar and other rules of the basic language [Canagarajah, 2002; 2013; 2016; Pennycook, 2008; Zamel, 1996].

The trend on translanguaging is viewed in modern globalized communities worldwide, especially in prominent economically advanced, and cultural centers where the mixture of languages is an objective reality. In many locations a trend of expansive multilingualism with the English language as a basis of national-wide communicative behavior has been noticed; however, there has established some novel style of globalized communication that is based on English as *lingua franca,* used by professionally-oriented communities with their national mother tongues to exchange information on science and humanities, technical achievements, social issues.

The trend pronounced contradicts strict requirements in prescriptive grammar to bilingualism that means fluency in two languages (practical skills of efficient communication in a certain linguistic and cultural environment); however, the term *translingualism* states the fact of dynamics in speech behavior in times of transcultural globalization [Crystal, 2003; Tlostanova, 2008]. If there exists a trend than it is a task of linguists to start monitoring such type of discourse realization and with the tools of corpus linguistics collect data for further observation on social functions and linguistic patterns of mixed-speech discourse and pave the way to better understanding translanguage abilities represented in cognitive strategies of code-switching and code-mixing.

Thus, the term *translingualism* specifies some phenomenon in modern communicative practice; it is implemented in translanguage abilities that are used for constructing a kind of translanguage hybrid text through mixing units and grammar rules of various languages. In translingual situations people demonstrate cognitive abilities of code-switching and code-mixing due to which there appears some degree of syntagmatic fusion without one language expansion; by using non-native English the language users preserve their mother tongue identity, which is revealed in some language equivalents and represent some native-like cognitive strategy of discourse production. The mixture of language units and rules of their combinability while the native-like elements are installed into hybrid discourse structures point to the notion of *translingualism,* the way how non-native English users communicate in professionally oriented communication in Global English.

Methodological basis

Accepting the fact that *translingualism* is a dynamic feature of modern globalized communication it is necessary to lay theoretical background and methodological basis for linguistic analysis of translanguaging as a visual representation of the phenomenon under study. A critical review of publication by European, American, Asian, and Russian research groups tells about the corpus-assisted approach to translanguage markers studies which attempt to explain some cognitive strategies of this type of verbal behavior (code-switching and code-mixing, in particular) and offer some methods to single out some language markers.

Translingualism is supposed to present some novel ways of communication in oral and written forms. Translingual situations are specified, firstly, by participants – they are non-native English users with various cultural and linguistic bases; secondly, it is due to their goals of pragmatic efficiency of communication that they refer to hybrid strategies of discourse production; thirdly, in their translanguaging discursive performance, people refer to a fusion of language codes.

The data obtained through corpus analyses point to some translingual strategies that consist in switching and meshing language codes, cross-cultural borrowings, simplifications, and hybridization of syntactic layers of speech

utterances [Caragarajah, 2002; Gardner-Chloros, 2009; Lefringhausen, Marchall, 2016; Leonard, Nowacek, 2016; Proshina, 2016; Spencer-Oatey, 2018; etc.]. The methodology of linguistic analysis is based on a discrete approach with language layers distinction and interpretative comparison of real-life (natural) samples of translanguaging to the standard norms of languages that are mixed (as identified languages). To date, linguists have published many reports about dialectal translingual studies (for ex., Welsh-English, Persian-English, Spanish-English, French-English-Cyprian, Russian-English, etc. [Chauncey, Grainger, Holcomb, 2008; Gabdreeva, Ageeva, Timirgaleeva, 2013; Gardner-Chloros, 2009; Deuchar, 2006; Marinova, 2008; Moradi, Chen, 2019; Myers-Scotton, 2006; etc.]). In their works there dominates an opinion about the Matrix language basis for cases of translingualism, with a grammatical (syntactic) structure as some definite core of the utterances (referring to one language standard, usually standard English) and mixed-language lexical nominations inserted (words, phrases, set expressions from different languages). However, their explanations are mainly devoted to social and psycholinguistic aspects of translingualism, enumeration of cases when the transmixture of words are noted as cases of code-switching, identification of cases with deviations from standard norms of syntactic structure [Gardner-Chloros, 2009; Moradi, Chen, 2019; Myers-Scotton, 2006; etc.], the aspect of discourse production by non-native professionals in the translingual environment as the mechanism of code-mixing has not been by now in the focus of linguists' interest.

Considering blurring differences between code-switching and code-mixing, we assume that these mechanisms are described through various methodological approaches with more interest to code-switching, none of them on its own can provide a complete overview. The mechanism of code-switching is described as the process of alternation of units of two or more languages in an individual speech utterance. It is studied with a discrete method, translinguistic cases are described separately as leapfrogging at certain language layers (phonomorphological, grammatical, syntactical), and then multilayered data are used for making a general conclusion about the degree of social integration or acculturation of a person, besides they might formulate grammatical tendencies rather than general syntactic rules of translingual language usage [Gardner-Chloros, 2009; Spencer-Oatey, 2018; Lefringhausen, Marchall, 2016]. However, code-mixing as a strategy of speech production remains practically unclear. Code-mixing can be described as the process of language blending, which presupposes mingled co-occurrence of signs of different languages and norms of their realization in speech practice produced by non-native users. It is not an individual switch from one language marker to another, it is a kind of switch-mode characteristic of a group of users. The conversational efforts of a certain group of non-natives (let's say in English) should demonstrate some common frequency of other language marker occurrences (mainly, mother tongue cognitive strategies). The veracity (reliability) of such data requires a corpus-oriented matrix methodology grounded on a multilayer search. The multilayered realization of code-mixing implies more technical efforts to result in distinguishing translingual cognitive patterns that require references to discourse oriented material for analysis. This research was based on two collections of academic papers written in English by Russian scientists who are thought to be proficient in written academic discourse and by native British scientists; their lexical matrix comparison resulted in distinguishing some syntagmatic and lexical markers of code-mixing.

The English language has undoubtedly become the preferred medium of communication in the academic context on a global scale. Non-native speakers of English in the academic professional community have been encouraged to use English without constant recourse to the native-speaker ideal. As a result, the English used in globalized professional academic writing lacks uniform standard but displays a variety in the use of language structures that is due to an array of factors among which the most important is a cognitive lingual background that determines how the non-native speakers verbalize their ideas in English. To date, a large number of papers that compare and contrast academic texts produced by native and non-native speakers have been published. They concentrated on the discrete-layered method of analysis, outlining lexical and grammatical features of the writings,

accentuating the use of linking adverbials, modal verbs, conjunctions, collocations [Altenberg, Granger, 2001; Chen, Baker, 2010; Gao, 2016; Martin, 2003; Siyanova, Schmitt, 2008; Yang, 2018]. These studies, most of which employed the corpus-based approach to counting and comparing the use of linguistic features, have found some ways in which professional academic writing of English natives and non-natives differ. However, markers of code-mixing in non-native academic writing have not been fully investigated and the lexical layer, which, in particular, is regarded as most prominent in revealing markers of code-mixing, has not received proper attention so far. Recent research in English academic discourse [Biber et al., 1999; Biber, 2006; Biber, Barbieri, 2007; Cortes, 2004; Hyland, 2008; Simpson-Vlach, Ellis, 2010; Wray, 2000] has evidenced the importance of discursive lexical bundles (recurrent sequences of words) as a major component in providing coherent discourse and an essential aspect of the shared knowledge of a professional discourse community on academic discourse genres. The focus of attention has been given to the identification and structural implementation of these discourse bundles as well as their functions and discipline variations. Having discovered some definite markers of academic discourse in native English writings, the researchers continued investigation by analyzing the use of lexical bundles in non-native speakers' writing and stated some deviation in the language means mainly based on implied references to the genre- and discipline-specific recurrent expressions typical of their native academic discourse.

The researches mentioned deal with academic English as *lingua franca* as it is used by speakers of Arabic, Chinese, German, Finnish, other European and Asian languages. It should be noted that English academic writing by native speakers of Russian is a relatively new field of research, and papers that compare and contrast the academic style writing have been far and few between. One of the most comprehensive comparative research based on the linguistic corpus of paper titles authored by native Russian and English speakers, style-forming properties of academic English [Ryabtseva, 2018a; 2018b; Stebletsova, Torubarova, 2019; Viktorova, 2015] revealed cross-cultural incongruence. Thus, the

titles in the academic discourse of Russian language speakers were noted not mapping the standard stylistic patterns characteristic to English academic speech; word-for-word translation techniques and the genre conventions of the Russian academic discourse are identified as the two major factors that determine the frequency and distribution of linguistic features in the academic paper titles offered for publication by Russian researchers. As compared to the English academic style that prefers a two-part title pattern, which, in its turn, uses the *and*-conjunction, a column, non-finite forms of the verb, prepositions, interrogative constructions, etc., the Russian academic style favors rather extended titles that include multiple abstract nouns, incorporated into a single noun group with repeated genitive constructions [Ryabtseva, 2018b]. Comprehensive contrastive research into two broad classes of discourse markers revealed significant differences in their usage in Russian and English language academic discourse [Viktorova, 2015]. However, the limitations of the studies mentioned are that it does not employ corpus methodology, and conclusions, though they seem to have been plausible, cannot be validated by statistical data.

In our previous studies, we used the corpus approach to contrast and compare the discourse functions of the first person plural pronoun *we* in *the BE2006 (learned)* corpus and the corpus of non-native academic writing that comprised papers written in English by Russian authors (EFLAW) which is also used in the present study [Kochetova, Kononova, 2018]. The pronoun *we* as a generic convention admits exclusive and inclusive references and represents culturally determined discursive strategy of construing communicative categories of solidarity, credibility, politeness. Although the statistical tests did not indicate significant differences in the frequency of the pronoun in the contrasted corpora, qualitative manual analysis of the concordance lines revealed that in the EFLAW corpus the inclusive *we* are employed to persuade the addressee to share and accept the author's arguments whereas in the BE2006 corpus it is used to construe solidarity based on common background, shared beliefs and opinions [Kochetova, Kononova, 2018].

The present study focuses on the use of recurrent lexical units as markers of code-

switching in academic papers written in English by Russian scientists. It uses corpus methodology, and its advances rely on statistical measures thus providing reliability in revealing correlations in the data and giving a precise and accurate picture of how language is used across various genres, text types, and registers.

The linguistic features analysis, which is at the center of the present study, uses the Matrix approach; after distinguishing the lexical layers with some core markers of academic discourse organization it allows to state the frequency and distribution of lexical bundles in the academic writing of the two groups, a group of English language native speakers and a group of non-native English writers, whose mother tongue is Russian. The present study is concerned with the comparative usage of lexical bundles in native and non-native academic discourse performers. The investigation is carried out on a corpus of professional academic writing of the Russian language scholars to find out quantitative data on functional types of lexical bundles typical of expert academic discourse in British and Russian discursive tradition. Lexical bundles as conventionalized expressions are also called *formulaic sequences* [Wray, 2002], *formulaic expressions* [Simpson, 2004], *fixed expressions* [Moon, 1992], *lexical phrases* [Nattinger, 1988], *multiword lexical units* [Cowie, 1999], *n-grams* [Cheng et al., 2009], or *lexical bundles* [Biber et al., 1999; Cortes, 2004; Hyland, 2008]. While the studies into conventionalized expressions apply different explanations and provide various criteria for the identification of multi-word sequences and thus approach their analysis from altered perspectives, they all conclusively show that conventionalized expressions constitute a large proportion of discourse and therefore indicate competent use of language in a particular context. This is in good agreement with Sinclair's 'idiom principle' [Sinclair, 1987] which postulates that speakers and writers do not select the words that

they use one at a time, but rather choose units of meaning expressed by pre-constructed phrases.

Research questions and methodology

Based on the corpus methodology to the study of language in use, the research aims to reveal extended fixed collocations in academic discourse produced by the two categories of writers. Specifically, this study will address the following research questions: 1. In what ways does the professional academic writing of native and non-native speakers of English differ in the use of lexical bundles typical of English academic discourse? 2. How do these most notable differences report on code-switching strategies in the discursive behavior of non-native writers in English?

To achieve the goals a comparative research design was applied, in which frequently-used word combinations (i.e., lexical bundles) were identified by an automated frequency-driven approach, that is corpus linguistics methods employed as the *SketchEngine* software tool [Kilgariff et al., 2004], then they were compared to the most preferable by the British authors of academic texts functional types of lexical bundles in expert academic discourse.

By definition, **lexical bundles** are groups of three or more words that are the most frequent recurring fixed sequences in a register [Biber et al., 1999, p. 990]. Examples of lexical bundles in English academic writing are expressions like *on the other hand*, as *a result of* or *the end of the*, to mention just a few. To conduct the study, a corpus of English academic writing (EFLAW) that comprises papers written in English by Russian language authors was compiled. As a reference corpus of native academic writing, FLOB-J corpus was taken from [Chen, Baker, 2010]. Corpora information is presented in Table 1.

The results for the FLOB-J were taken from [Chen, Baker, 2010], they were obtained in the comparative study of lexical bundles in BAWE-EN

Table 1. **Comparison of FLOB-J and English as a Foreign Language Academic Writing Corpus (EFLAW)**

Corpus Information	FLOB-J	EFLAW Corpus
The number of texts	80	85
Average words per sample	2,059	3,726.07
Total number of words	164,742	316,716

and BAWE-CH as compared with native professional academic writing. Lexical bundles retrieved from a corpus of published professional academic texts written in English by native Russians EFLAW and the FLOB-J corpus were investigated both quantitatively and qualitatively.

Corpus structure

The corpus of non-native academic writing comprises professional academic papers written in English by Russian authors and published in the online editions of open access journals. The corpus design followed the one adopted in the BNC1994 and included texts dealing with a wide range of scientific topics, engineering, natural sciences, art and humanities, social sciences, law, education, and medicine. The articles were chosen randomly and the number of texts per journal has been adjusted according to the periodical's impact. The structure of the corpus is presented in Table 2.

As the corpora under study are relatively small, certain requirements were imposed, such as the minimum frequency of occurrence for each lexical bundle is five [Altenberg, 1998; De Cock, 1998] with the requirement to occur in at least 3-5 different texts [Biber, Barbieri, 2007; Cortes, 2004]. This criterion for n-grams selection enables to avoid idiosyncrasies from individual writers / speakers and retrieve fixed expression shared by different writers. It is found that the most researched length for writing studies is a four-word sequence, which, probably because the number of 4-word bundles is often within a manageable size (around 100) for manual categorization and concordance checks [Chen, Baker, 2010]. Using the corpus tool *SketchEngine*, 4-word clusters were retrieved automatically and after that word sequences with content words that were present in the corpus EFLAW, or any other context-dependent bundles, usually incorporating proper nouns (e.g., *the Russian Federation*, *Constitution*, *Gogol)*, were manually selected and put on the stop list. To avoid overlapping word sequences, we applied the option that allows to group sub-grams of another longer n-gram together with the longer n-gram. The range of lengths was set between four and five, which allowed us to group 4-word bundles which are derived from a single 5-word combination. For example, *it should be noted*, and *should be noted that* both occur 36 times, coming from the longer expression *it should be noted that*. Another case is when the occurrences of one of the bundles subsume those of the other overlapping bundle(s). For example, *is one of the* occurs 41 times, while *one of the most* occurs 41 times, both of which occur as a subset of the 5-word bundle *is one of the most*. Each case of the above overlapping word sequences was combined into one longer unit to guard against inflated results. The overlaps were grouped automatically and then concordance lines were checked manually.

Results and discussion

Three broad categories of bundles, namely Noun-based phrases (NP), Preposition-based phrases (PP), and Verb-based phrases (VP) were retrieved. Noun-based bundles include any noun phrases with post-modifier fragments, such as *the role of the* or *the way in which* (i.e., Category (1) in Table 3). Preposition-based expressions include those starting with a preposition plus a noun-phrase fragment, such as *at the end of* or *in relation to the* (i.e., Category (2) in Table 3). Concerning verb-based bundles, any word combinations with a verb component, such as *in order to make* or *was one of the*, is assigned to this category (i.e., Categories (3) to (8) in Table 3).

Table 2. **The structure of the EFLAW Corpus**

Field	Tokens	Percentage	The number of texts
Art&Humanities	82,305	21.83	19
Education	68,421	18.15	15
Law	41,541	11.01	9
Medicine	20,168	5.35	5
Natural Sciences	32,872	8.75	9
Philosophy	56,217	14.91	10
Social Sciences	50,058	13.27	9
Technical	25,385	6.73	9
Total	376,967	100	85

The proportion of noun-based bundles does not show a great difference in the two corpora. To explore the lexical bundles used in the corpora under study in greater depth, the noun-based combinations were grouped further into two structural subcategories to see if these two corpora could be distinguished from each other. These two subcategories are noun phrase fragments with *of* (e.g., *in the context of*) and any other noun phrase fragments without the preposition *of* (e.g., *the way in which*). Even though the proportion of NP-based bundles when compared with FLOB-J is relatively equal in the corpora under consideration, the EFLAW writing is notably different in the subcategory of NP without the preposition *of* (Table 4). The combinations without the preposition *of* found in this investigation are all part of relative clauses, such as the *extent to which*, *the fact that this*, or *the way(s) in which*. It is evident that EFLAW writing did not use these types of relative clauses as frequently as native speakers did.

Further, we grouped *NP + of* and *PP + of* bundles into two productive frames: [the + Noun + of the/a] and [in the + Noun + of]. The professional writing in FLOB-J showed manifesting a relatively wide range of nouns that collocate with these two frames (Table 5 and

Table 3. **Frequency of lexical bundles in the corpora under study**

Types of *n*-grams	FLOB-J, %	EFLAW, %	Examples
(1) noun phrase with a post-modifier fragment□	32.5	33.9	*the point of view, the way in which, the purpose of the/a, the structure of the/a, the problem of the/a, the extent to which, the fact that this, etc.*
(2) preposition noun-phrase	36	27.5	*As a result of, due to the fact, on the basis of, on the one hand, in the course of, at the same time, as a result of, in the process of etc.*
(3) be + NP/Adjective P	2.6	2.5	*is one of the, is due to the, is a part of the, is a function of*
(4) verbal phrase with active verb	0.9	4.2	*has a number of, take into account the, make it possible to, turns out to be, have a positive effect*
(5) anticipatory *it* + VP/adjective P + (complement clause)	8.8	9.1	*it is possible to, it is necessary to, it is important to, it is impossible to, it is obvious that*
(6) passive verb + PP fragment	7	17.3	*is based on the, is considered to be, is related to the, is regarded as a, is determined by the, is seen as a, is explained by the*
(7) (verbal phrase +) that-clause fragment	2.6	1.2	*should be noted that, can be argued that*
(8) (verb/adjective +) to-clause fragment	7	2.6	*are likely to be, to be able to*
Others	2.6	1.7	*as well as the (a)*
Total	100	100	–

Table 4. **The occurrences of NP phrases without the preposition *of* in the FLOB-J and EFLAW corpora (with raw frequencies)**

FLOB-J	EFLAW
the degree to which (5)	the degree to which (1)
the extent to which (6)	the extent to which (1)
the fact that this (4)	the fact that this (2)
the way in which (14)	
the ways in which (4)	

Table 5. **The occurrences of lexical bundles *the Noun+of the/a* in the FLOB-J, and EFLAW corpora (with raw frequencies)**

FLOB-J	EFLAW
rest (11), end (10), **history** (7), size (7), **basis** (5), function (5), **role** (5), rules (5), strength (5), **value** (5), **creation** (4), **existence** (4), example (4), **impact** (4), **level** (4), magnitude (4), **point** (4), **results** (4), status (4), **structure** (4)	**basis** (28), **point** (28), **results** (27), **structure** (22), **end** (22), study (21), beginning (19), problem (18), **level** (15), importance (14), concept (13), **role** (13), definition (12), choice (11), formation (11), purpose (15), **history** (9), **impact** (8), **creation** (14), **existence** (11), **value** (9), use (7)

Table 6). The EFLAW writers also use a variety of lexical items, however, the fact that the non-native authors recognize the importance of these structures leads to the situation when many of them are heavily overused. At the same time, some expressions that are characteristic to the Russian language academic prose, namely phrases formed by the nouns *choice*, *problem*, *role*, *study*, *purpose* are hardly ever found in the native professional academic writing (Table 5) and are considered as markers of code-mixing of Russian performers. The nouns that are shared by the two corpora are in bold.

As for prepositional phrases, the corpora under investigation share a small proportion of nouns used in the pattern [in the + Noun + of]; as is seen in Table 6, the non-native professional academic writing makes use of framing expressions such as *in the context of*, *in the framework of*, authors very often employ place / time / text-deictic expressions, such as *at the same time*, *are shown in fig,*, whereas quantifying expressions (e.g., *the extent to which, percent of the, the size of the*) related to anything potentially measurable, such as size, number, amount or extent are conspicuously absent.

Although the proportion of VP-based bundles in FLOB-J and EFLAW corpora are approximately equal (Table 3), they vary in the subcategories of lexical bundles employed. As the data show, verbal phrases with active verbs and passive constructions are used much more frequently in EFLAW corpus as compared to native professional writing.

In the subcategory Verb / Adjective followed by *to*-clause fragment the EFLAW writers showed a strong preference for the construction *in order to* followed by a variety of verbs: *in order to increase, in order to allow, in order to examine, in order to achieve, in order to show, in order to draw*, etc. which occurs in 38 out of 80 texts whereas the structure *is likely to* followed by *to*-clause fragment is much less evenly distributed as 8 out of 12 occurrences are found in one text so this can be attributed to an individual writer's style. In comparison, in FLOB-J the construction Verb / Adjective followed by *to*-clause fragment occurs 47 times and is found in 25 texts.

Non-native speakers use the "Passive verb + prepositional phrases" form much more frequently than native speakers did. As can be seen in Table 7, there are seven passive-verb

Table 6. **The occurrences of lexical bundles *in the Noun+of the/a* in the FLOB-J, and EFLAW corpora (with raw frequencies)**

FLOB-J	EFLAW
case (19), **context** (19), **form** (8), **presence** (8), absence (7), light (6), **number** (6), **course** (5), hands (5), face (4)	process (83), **course** (50), **context** (19), **form** (36), field (28), framework (26), **case** (26), **number** (17), system (15), structure (14), way (8), **presence** (21), terms (7), variety (7)

Table 7. **Bundles in the Subcategory 'Passive Verb + Prepositional Phrases'**

FLOB-J	EFLAW
are shown in fig. (6)	is based on the/a (45)
be found in the (5)	explained by the fact (19)
be seen in the (5)	is associated with the (18)
be taken into account (5)	is characterized by the/a (18)
can be found in (6)	is connected with the (18)
is concerned with (4)	is determined by the (12)
was found by a (4)	are based on the (12)
	can be regarded as (9)
	can be seen from (8)
	can be found in (8)
	is understood as a (7)
	is shown on fig. (7)
	is seen as a (7)
	is related to the (7)
	is regarded as a (7)
	be taken into account (7)
	is expressed in the (5)
	is confirmed by the (5)

bundles in FLOB-J as compared to as many as eighteen in the EFLAW writing. In comparison to the FLOB-J, where they make up around twenty percent of the VP-based bundle types, the eighteen passive-verb bundles in EFLAW constitute 35 percent of the total VP-based bundle types. Additionally, only the three passive bundles, which are marked in bold in Table 7, were shared by the EFLAW group of writers. The number of those which are not shared exceeds the native writing considerably thus serving as markers of code-mixing. It has been suggested that the more frequent use of VP-based phrases with prepositional passive constructions should be attributed to the conventions of Russian language academic prose that require to present information in a non-personal and detached way thus emphasizing the objectivity of research. Russian language writers refer to a loan translation, supposing such calque to be a mark of English academic writing.

The concordance lines showed that verbal phrases followed by *that*-clause occur 61 times in 35 texts in the EFLAW corpus. Most of the expressions have the verb *note* in combination with the modal verb *should,* and five instances use the verb *argue* in combination with the modal verb *can.*

Table 8 shows the difference between the two corpora in the category (8) that can be explained by the fact that although there is no significant difference in the use of expressions that denote ability (*it is difficult to, to be able to*), non-native speakers tend to use fewer lexical bundles that help to minimize categorical propositions as compared to the native speakers writing, namely *would have to be, seems to have been, would be difficult, it has been suggested,* etc.

In the non-native writing, phrases with anticipatory *it* are used nearly as frequently as in the native speakers' writing. Most of the phrases with the anticipatory *it* are used to express the writer's attitude about proposition (obligation/directive) (e.g., *it is important to*) or ability (e.g., *it is difficult to*). However, the native writers avoid using negative phrases that are too categorical and show a preference for hedging expressions.

Conclusion

The comparative corpus-based study of the two corpora of native and non-native academic writing in English has revealed lexical markers that are used as pre-constructed phrases to provide coherent discourse. Due to the Matrix approach, that allowed us to draw structural and functional comparisons of lexical markers in their syntagmatic environment, it has been proved that the use of discourse organizing lexical bundles in native and non-native academic writing in English demonstrates some generic similarities and differences. Both corpus data contain NP-based bundles to organize academic discourse, VP-based bundles composed by the anticipatory *it-clause* (*with a complement clause*). To sum up, non-native professional academic writing by a group of academic researchers was found to exhibit the widest range of reoccurring lexical bundles with abstract nouns that are markers of code-mixing, whereas native writing showed a much lower range. It should be noted that non-native writing lacked noun-based phrases without prepositions, such as *the extent to which, to a large extent,* whereas it overused certain expressions (e.g., *all over the world)*, which native academics rarely used. Some high-frequency expressions in Russian academic writing that are typical of genre conventions in Russian are loan translations from the Russian academic discourse, namely *from our point of view, with the help of, by means of.*

Table 8. **Bundles in the Subcategory "Anticipatory It+ VP/adjective P + (complement clause)"**

FLOB-J	EFLAW
It is necessary to (14)	**It is necessary to** (46)
It is clear that (11)	**It is important to** (30)
It is difficult to (6)	**It is possible to** (19)
It is possible to (6)	It is impossible to (12)
It is important to (6)	It is obvious that (7)
It can be seen that (5)	**It is difficult to** (7)
It has been suggested that (4)	**It is clear that** (6)

As the analysis shows, native academic writers, exhibit a wider range of VP-based bundles composed by the *(verbal phrase) + that clause fragment* or *(verb/adjective) + to-clause fragment*, whereas Russian researchers overuse verbal phrases with active verbs as well as constructions with passive verbs or participle clauses (*is confirmed, is expressed*), which is an evident cognitive mark of code-mixing. The non-native academic writers rarely modify their writing by hedges to mitigate propositions and the statements imply strictness and needless categoricalness. In general, non-native writers relied on formulaic expressions most, while the expert writers used the fewest clusters, which confirms lower academic proficiency.

Thus, the application of the Matrix approaches jointly with the corpus techniques allowed us to retrieve word combinations that are used as generic building blocks in constructing the logical development of academic discourse. In particular, lexical bundles that make a core of native English expert academic discourse were retrieved and used for comparison with the non-native writing. Further studies on the frequency-driven formulaic expressions and syntagmatic combinations of words found in native expert writing can help in identifying extra valid and authentic data for reconstructing the models of English academic discourse and monitor its variability as code-mixing in global academic exchange in the English language.

REFERENCES

Altenberg B., 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. *Phraseology: Theory, Analysis and Applications*. Oxford, Oxford University Press, pp. 101-122.

Altenberg B., Granger S., 2001. The Grammatical and Lexical Patterning of MAKE in Native and Non-Native Student Writing. *Applied Linguistics*, no. 22, pp. 173-194.

Biber D., Johansson S., Leech G., Conrad S., Finegan E., 1999. *The Longman Grammar of Spoken and Written English*. London, Longman. 1203 p.

Biber D., 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam, Philadelphia, John Benjamins. 261 p.

Biber D., Barbieri F., 2007. Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*, no. 26, pp. 263-286.

Canagarajah S., 2002. Multilingual Writers and the Academic Community: Towards a Critical Relationship. *Journal of English for Academic Purposes*, no. 1, pp. 29-44.

Canagarajah S., 2013. *Translingual Practice: Global Englishes and Cosmopolitan Relations*. London, New York, Routledge, Taylor and Francis Group. 240 p.

Canagarajah S., 2016. Mul'tilingval'naja pedagogika neoliberal'nogo vremeni: raschistka territorii i planirovanie dvizhenija vpered [Multilingual Pedagogies in Neoliberal Times: Clearing the Terrain and Charting the Way Forward]. *Social'nye i gumanitarnye nauki na Dal'nem Vostoke* [The Humanities and Social Studies in the Far East], no. 2 (50). URL: http://www.eastjournal.ru/journalE.htm (accessed 23 January 2020).

Chauncey K., Grainger J., Holcomb P.J., 2008. Code-Switching Effects in Bilingual Word Recognition: A Masked Priming Study with Event-Related Potentials. *Brain and Language*, vol. 105, iss. 3, pp. 161-174. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2684792/ (accessed 23 January 2020).

Chen Yu., Baker P., 2010. Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, vol. 14, no. 2, pp. 30-49. URL: https://www.researchgate.net/publication45681690_Lexical_Bundles_in_L1_and_L2_Academic_Writing (accessed 23 January 2020).

Cheng W., Greaves Ch., Sinclair J., Warren M., 2009. Uncovering the Extent of the Phraseological Tendency: Towards a Systematic Analysis of Concgrams. *Applied Linguistics*, no. 30, pp. 236-252.

Cortes V., 2004. Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, no. 23, pp. 397-423.

Cowie F., 1999. *What's Within? Nativism Reconsidered*. Oxford, Oxford University Press. 334 p.

Crystal D., 2003. *English as a Global Language*. Cambridge, Cambridge University Press. 212 p.

De Cock S., 1998. A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus Linguistics*, no. 3 (1), pp. 59-80.

Deuchar M., 2006. Welsh-English Code Switching and the Matrix Language Frame Model. *Lingua*, vol. 116, iss. 11, pp. 1986-2011.

Fricke M., Kootstra G.J., 2016. Primed Codeswitching in Spontaneous Bilingual Dialogue. *Journal of Memory and Language*, vol. 91, pp. 181-201.

Gabdreeva N.V., Ageeva A.V., Timirgaleeva A.R., 2013. *Inojazychnaja leksika v russkom jazyke novejshego perioda* [Foreign Lexis in the Russian Language of a Modern Period]. Moscow, Flinta Publ., Nauka Publ. 328 p.

Gao X., 2016. A Cross-Disciplinary Corpus-Based Study on English and Chinese Native Speaker' Use of Linking Adverbials in Academic Writing. *Journal of English for Academic Purposes*, vol. 24, pp. 14-28.

Gardner-Chloros P., 2009. *Code-Switching.* Cambridge, New York, Cambridge University Press. 242 p. URL: https://www.academia.edu/12864556/Code-Switching_by_Gardner-Chloros (accessed 23 January 2020).

Gevers J., 2018. Translingualism Revisited: Language Difference and Hybridity in L2 Writing. *Journal of Second Language Writing*, vol. 40, pp. 73-83. URL: https://www.sciencedirect.com/science/article/abs/pii/S1060374317302965 (accessed 23 January 2020).

Gullberg M., Indefrrey P., Muysken P., 2009. Code-Switching Between Structural and Sociolinguistic Perspectives. *The Cambridge Handbook on Linguistic Code-Switching. Vol. 43.* Cambridge, Cambridge University Press, pp. 21-39.

Hyland K., 2008. As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, no. 27 (1), pp. 4-21.

Kilgarriff A., Rychlý P., Smrž P., Tugwell D., 2004. *The Sketch Engine. Proceedings of the XI EURALEX International Congress.* Lorient, Universite de Bretagne-Sud, pp. 105-116. URL: https://www.sketchengine.co.uk/wp-content/uploads/The_Sketch_Engine_2004.pdf (accessed 11 January 2015).

Kochetova L.A., Kononova I.V., 2018. Corpus-Based Contrastive Study of Discursive Strategy of Construing Interpersonal Relations in English Language Academic Discourse. *Journal of Siberian Federal University. Humanitites and Social Sciences*, pp. 1-11. DOI: 10.17516/1997-1370-0353.

Lefringhausen K., Marchall T.C., 2016. Locals' Bidimensional Acculturation Model: Validation and Associations with Psychological and Sociocultural Adjustment Outcomes. *Cross-Cultural Research*, no. 50 (4), pp. 356-392. DOI:10.1177/1069397116659048.

Leonard R., Nowacek R., 2016. Transfer and Translingualism. *College English*, vol. 78, no. 3, pp. 258-264.

Maksimova Ju.S., 2016. Razvitie kontaktov kul'tur cherez jazykovoj kompromiss [Evolution of Cultural Contacts by Means of Language Cooperation]. *Social'nye i gumanitarnye nauki na Dal'nem Vostoke* [The Humanities and Social

Studies in the Far East], no. 2 (50), pp. 120-124. URL: http://www.eastjournal.ru/journalE.htm (accessed 23 January 2020).

Marinova E.V., 2008. *Inojazychnye slova v russkoj rechi konca XX – nachala XXI v.: problemy osvoenija i funkcionirovanija* [Foreign Words in Russian Speech of the End of the 20th – Beginning 21st: Issues of Adaptation and Functioning]. Moscow, Jelpis Publ. 509 p.

Martin P.M., 2003. A Genre Analysis of English and Spanish Research Paper Abstracts in Experimental Social Sciences. *English for Specific Purposes*, no. 1, pp. 25-43.

Molina C., 2011. Curricular Insights into Translingualism as a Communicative Competence. *Journal of Language Teaching and Research*, vol. 2, no. 6, pp. 1244-1251.

Moon R., 1992. Textual Aspects of Fixed Expressions in Learners' Dictionaries. *Vocabulary and Applied Linguistics.* Palgrave, Macmillan, pp. 13-27.

Moore A.M., Barker G.G., 2012. Confused or Multicultural: Third Culture Individuals' Cultural Identity. *International Journal of Intercultural Relations*, vol. 36, pp. 553-562.

Moradi H., Chen J., 2019. Structural Analysis of Persian-English Reverse Codeswitching and Code-Mixing. *Science Journal of Volgograd State University. Linguistics*, 2019, vol. 18, no. 1, pp. 122-131. DOI: https://doi.org/10.15688/jvolsu2.2019.1.10.

Myers-Scotton C., 2006. *Multiple Voices: An Introduction to Bilingualism.* New York, Blackwell Publisher. 473 p.

Nattinger J., 1988. Some Current Trends in Vocabulary Teaching. *Vocabulary and Language Teaching.* London, New York, Longman, pp. 62-82.

Pennycook A., 2008. English as Language Always in Translation. *Australian Review of Applied Linguistics, International Forum on English as an International Language, Special Forum Issue.* Sydney, vol. 32, no. 3, pp. 30.1-30.9.

Proshina Z.G., 2016. Problemy i perspektivy translingval'nyh i transkul'turnyh kontaktov [Issues and Challenges of Translingual and Cross-Cultural Contacts]. *Social'nye i gumanitarnye nauki na Dal'nem Vostoke* [The Humanities and Social Studies in the Far East], no. 2 (50), pp. 6-9. URL: http://www.eastjournal.ru/journalE.htm (accessed 21 January 2020).

Ryabtseva N.K., 2018a. Academic Papers Titles: A Russian – English Perspective. *Nauchnyy dialog* [Academic Dialogue], no. 6, pp. 32-42. DOI: 10.24224/2227-1295-2018-6-32-42.

Ryabtseva N.K., 2018b. Academic Paper Titles and Their Dominating Patterns: A Russian-English Perspective. *Vestnik Volgogradskogo*

gosudarstvennogo universiteta. Seriya 2, Yazykoznanie [Science Journal of Volgograd State University. Linguistics], vol. 17, no. 2, pp. 33-43. (in Russian). DOI: https://doi.org/10.15688/jvolsu2.2018.2.4.

Simpson R., 2004. Stylistic Features of Academic Speech: The Role of Formulaic Expressions. *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam, Philadelphia, John Benjamins, pp. 37-64.

Simpson-Vlach R., Ellis N., 2010. An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, no. 31 (4), pp. 463-512.

Sinclair J., 1987. Collocation: A Progress Report. *Language Topics: Essays in Honour of Michael Halliday. Vol. 2.* Amsterdam, Philadelphia, John Benjamins, pp. 319-332.

Siyanova A., Schmitt N., 2008. The Canadian Modern Language Review. *La Revue canadienne des langues vivantes*, no. 64 (3), pp. 429-458.

Spencer-Oatey H., Franklin P., 2009. *Intercultural Interaction: A Multidisciplinary Approach to Intercultural Communication*. London, Palgrave, MacMillan. 367 p. DOI: 10.1057/9780244511.

Spencer-Oatey H., 2018. Transformative Learning for Social Integration: Overcoming the Challenge of Greetings. *Intercultural Education*, no. 29 (2), pp. 301-315. DOI: 10.1080/14675986.2018.1425828.

Stebletsova A.O., Torubarova I.I., 2019. Formality as a Style-Forming Property of Academic English: Synchronic Analysis of Medical Research Texts. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 18, no. 4, pp. 157-173. (in Russian). DOI: https://doi.org/10.15688/jvolsu2.2019.4.12.

Tlostanova M.V., 2008. *Ot filosofii mul'tikul'turalizma k filosofii transkul'turacii* [From Philosophy of Multiculturalism to Philosophy of Transculturation]. Moscow, RUDN Publ. 251 p.

Viktorova E.Yu., 2015. *Vspomogatel'naya systema diskursa* [Supportive System of Discourse]. Saratov, Izdatel'skiy tsentr Nauka. 404 p.

Wray A., 2000. Formulaic Sequences in Second Language Teaching: Principle and Practice. *Applied Linguistics*, no. 21 (4), pp. 487-489.

Wray A., 2002. *Formulaic Language and the Lexicon*. Cambridge, Cambridge University Press. 348 p.

Yang X., 2018. A Corpus-Based Study of Modal Verbs in Chinese Learners' Academic Writing. *English Language Teaching*, vol. 11, no. 2, pp. 122-130.

Zamel V., 1996. Transcending Boundaries: Complicating the Scene of Teaching Language. *College English*, vol. 6, no. 2, pp. 1-11.

Zamel V., 1997. Towards a Model of Transculturation. *TESO: Quarterly*, vol. 31, no. 2, pp. 341-351.

Information About the Authors

Larisa A. Kochetova, Doctor of Sciences (Philology), Associate Professor, Head of the Department of English Philology, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, kochetova@volsu.ru, english_philology@volsu.ru, https://orcid.org/0000-0002-5278-7373

Elena Yu. Ilyinova, Doctor of Sciences (Philology), Professor, Department of English Philology, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, ilynov@volsu.ru, english_philology@volsu.ru, https://orcid.org/0000-0002-3310-4020

Информация об авторах

Лариса Анатольевна Кочетова, доктор филологических наук, доцент, заведующая кафедрой английской филологии, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, kochetova@volsu.ru, english_philology@volsu.ru, https://orcid.org/0000-0002-5278-7373

Елена Юрьевна Ильинова, доктор филологических наук, профессор кафедры английской филологии, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, ilynov@volsu.ru, english_philology@volsu.ru, https://orcid.org/0000-0002-3310-4020