# Web-Academic Impact on Terminology:
# A Corpus-Based Study

Larisa Beliaeva[0000-0002-8622-4595] and Olga Kamshilova[0000-0002-1488-2206]

Herzen State Pedagogical University of Russia, Saint-Petersburg,
Moika river emb, 48, 181186, RF
lauranbel@gmail.com, onkamshilova@gmail.com

**Abstract.** The article presents a study of some key issues of borrowed terminology in Russian scientific texts. The study presumes that global web academic intercourse and professional bilingualism of actively publishing Russian authors facilitate the process of borrowing new terminology of English origin. The problem addressed is the manner and methods Russian authors accept and use new terms in their research papers. The research methodology applied is that of corpus technologies. The study is based on corpus findings in two original research corpora. It aims at developing a procedure of detecting and describing new English terminology and its presentation in recent Russian scientific texts of a restricted knowledge domain, namely web and linguistic technologies. Section 1 presents an overview of factors influencing the quality of Russian academic writing. Section 2 describes the research corpora and a corpus-based procedure of detecting and extracting loan words. Section 3 focuses on analysis of loan terms interpretation by Russian authors. Section 4 summarizes the preliminary results.

**Keywords:** Russian Scientific Text, Borrowed Terminology, Corpus-based Analysis.

## Introduction

Scientific communication today actively involves international sources of information and data, including abstract and citation databases such as Scopus and Web of Science, and web search engines like Google Scholar, that help to widespread research results in global web academic society ("web academy"). Since "web academics" read and publish their research preferably in English, the core lexical component of scientific texts – terminology – is picked up and accepted easily (with much aid of global technologies in use). Unlike publications in international English scientific journals and conference proceedings scientific papers in national languages that are designed for publication in national scientific journals and conference proceedings demonstrate a terminological battle between national terminology and loan terms, the outcome is seldom in favor of the former. Nationally published materials of various reliability have a substantial representation in the web, too, which adds to the resulting diversity of term presentation and interpretation.

As text remains (and it always will) the main source of data and knowledge mining, no matter what its language (natural or artificial) and its addressee (a person or a system) are [1; 2; 3] data requirements of modern information society ("Information 4.0") concern every aspect of how a text is created, structured and used, how texts or their structural parts are shared in various technological and humanitarian systems.

The process of Russian scientific text creation today is complicated by "professional" bilingualism of active Russian scientists, which is natural and almost mandatory under the circumstances, and which influences their performance in writing for Russian journals and conferences [4; 5].

In fact, modern scientific communication reflects the internationalization process of scientific domains, however the insufficient academic literacy in native language is exactly that stumbling-block, which interferes with training professional communication in English [6; 7] and Russian languages. Our observations show that even basic philological education does not guarantee textual competences in academic and scientific writing [8]. The problem redoubles as novice researchers are not trained to present information of their research projects in different textual forms in Russian and actually fall out of Russian academic style tradition.

Terminology is one of the key points on the way to master academic style for research publications. In spite of fast developing terminology management systems an active bilingual researcher proves faster: they pick up and transfer new English terminology, and coin new Russian terms in a voluntary manner. Partly they do so, because translating an English term (or text as a whole) has become a redundant and lost practice, though it was professional translation of foreign texts that not only most adequately introduced new terms, but established parity of accepted Russian terminology and the new borrowings (loans) [9; 10]. Such were the famous issues of "Новое в зарубежной лингвистике" with professionally translated linguistic texts and subject indexes that fixed the new terminology.

Considering time necessary for a proper equivalent choice when translating a scientific or technical text (75% of the time needed for text translation as a whole [11]), labor-intensive procedure of a translation equivalent formation and description is dropped out of the academic writing process. Thus, in national Russian scientific spheres, the result of this incomplete terminology knowledge is generation of pseudo-Russian texts in a bizarre mixture of languages, a sort of new Volapük. We doubt whether the text fragments (1-2) below may be considered as a proper scientific style in Russian, where (1) is a definition of "социально-сетевой дискурс" [social network discourse], abounding in complex terms including prefixes of different origin: *макро-, пара-, гео-* [*macro-, para-, geo-* ], *много-, разно-, одно-, меж-* [mnogo-, razno-, odno-, mezh-] and hyphenated constructions, which makes the definition incomprehensible; and (2) is a typical example of borrowing new terms in every manner possible – translating (*энвайронментализм, экологическое искусство, арт-практики*) , transliterating (*ленд-арт, био-арт, арте повера*) and taking "as it is" (*art&science*), which reduces the text addressee to those in the know:

*(1)    Представленное исследование направлено на выявление вербальной , и*
***пара****вербальной специфики формирования и функционирования в мировой*
*электронной медийной среде социально-сетевого дискурса (ССД), опреде-*

*ляемого нами как дистантный опосредованный,* **многовекторно-разно***направленный,* **одновременно-разно***временный в реальном времени (on-line) и отложенный (off-line),* **много***тематический электронный* **макро***полилог, который отражает* **меж***личностные,* **меж***этнические,* **меж***конфессиональные, социально-экономические,* **гео***политические и т. д. типы отношений, что находит непосредственное выражение в специфике вербальных и* **пара***вербальных коррелятов письменных и устных дискурсивных высказываний*

*(2)  Это справедливо, например, для* **энвайронментализма** *который, в свою очередь, был тесно связан с другими* **арт-практиками***, в частности.* **ленд-артом***, экологическим искусством,* **арте повера***,* **био-артом***,* **art&science***.*

The fact of the matter is that in bilingual web academic environment there is no need of translating scientific texts into Russian, which results in ignoring and, hence, failing to correctly use any dictionary, including electronic ones, that, in its turn, adds to introduction of new terms into a Russian text, even if their official equivalents are already set and dictionary-fixed. Thus, for example, the word *perceptual* in the universal translation dictionary of ABBY LINGVO system (English-Russian) has the following description:

> **perceptual** *[pə'sepʧəl]*
>     *относящийся к восприятию; перцепционный*
> *perceptual [pə(r)septʃəl]*
> *ADJ; ADJ n*
> *Perceptual means relating to the way people interpret and understand what they see or notice.*
> *[FORMAL]*
>     *Some children have more finely trained perceptual skills than others.*
> **perceptual** *[pə'sepʧəl]*
> *перцептивный*
> *perceptual per|cep¦tual*
> *adjective relating to the ability to interpret or become aware of something through the senses*
>     *A patient with perceptual problems who cannot judge distances*

The dictionary information assumes that a Russian writer shall use translations *перцептивный* or *перцепционный* that are registered as terms in the Russian language, and introducing the adjective *перцептуальный* as offered in (3) is absolutely redundant:

*(3)  Личное (или личностное) присутствие (англ. - personal presence), и близкие к нему понятия «физического присутствия», «пространственного присутствия», «**перцептуального** погружения» можно было бы назвать присутствием в чистом виде…*

Reading and processing information from a vast number of scientific texts in global English to support personal research and international communication brings a Russian author to active use of borrowed lexicon. This is not only Russian specific, one can spot similar use in other scientific cultures, for example, in the materials of

tekom-Jahrestagung and Tcworld conferences, where texts written in German include terms in English, see, for example (4):

(4)   *Am Standort Rorschacherberg (Schweiz) setzen die hauseigenen Technischen Redakteure für die Dokumentenerstellung ein **Content-Management-System** ein – anschließend werden die fertigen Dokumente an den Übersetzungsdienstleister geschickt, der die Daten in einem-**Translation Memory-System** weiter verarbeitet.*

This way of including loan terminology into a native text is permitted due to common alphabet systems, which in case of a Russian text seems hardly possible.

The use and borrowing of English terminology in web-academic society is a natural process, the spread of terminological loans in published Russian texts is uncontrolled and the author's choice of presenting a new borrowed item is absolutely voluntary. Thus, we consider that discussion of the linguistic peculiarities of new term introduction is relevant and pertinent. **The purpose** of this study is to suggest and develop a procedure of detecting and describing actual English terminology and its presentation in recent Russian scientific texts of a restricted knowledge domain, namely web and linguistic technologies [12].

## 1      Methodology and Material under Study

In principle the methods of determining words with no translation could be divided into two groups.

One group is usually applied to a specific set of new data, using such language resources as word lists and linguistic models. It is noteworthy that modern lexicological studies rely on analysis of research and national text corpora and, respectively, on automated word lists, which units (tokens) are automatically defined as sequences of symbols between two spaces. Naturally, such tokens are not always words. When detecting non-translated words, the word lists are usually generated from the existing lexicographic paper and web resources, dictionaries or texts corpora, and are then edited with special filters eliminating surplus tokens, such as sets of symbols but not natural language words (dates, URL, formulaic elements), misprints and errors (cf *awaraness, abyssin, ajudicated*), named entities (personal names, organization names, geolocations, time intervals). To assist detecting new words the procedure may apply linguistic models, such as markers of lexical novelty: punctuation marks (different quotation marks, diacritical marks and subscript), application of italics and/or boldface font, etc., they can signal, that the word is a new one (neologism) or is used in a new meaning (see [13]).

The other group of methods, usually applied to multiple data sets, is oriented to statistical evaluation or machine training which are necessary to calculate and evaluate the growth of usage or change of meaning, arisen in due course of time or in various registers [14]. Within this approach one can use national text corpora or specially created corpora (see, for example, Chambers Harrap International Corpus (CHIC), which was specially created for new word extraction and includes more than 500 million words from global English texts, it represents a static, balanced resource [13],

see as well a specially created reference texts corpus in [15]. Naturally, web sources as a whole can be used as a corpus with application of various computer programs of scanning and search, the so-called crawlers. One of such facilities is the extensible program Heritrix with open original texts, the program performs scanning and search in the web archives [14].

The majority of modern studies of new words relies on such automatic scanning of text archives for new words and automatic tracking of their persistent occurrence, which requires to create a list of their formal features in each special domain language, to introduce special timeline for such analysis (weekly, monthly and so forth). In Russia, unfortunately, such methods are not practiced widely, if at all. The analysis of new words that have no translation equivalents, is traditionally performed on a bound material of published translation dictionaries and the absence of a fixed translation is taken as a ground for detecting a new word.

Terminology of web and linguistic technologies domain can't be based on the units already fixed in dictionaries in principle. In this case a most reliable language source is a specialized research corpus.

The present research is based on two original research text corpora including texts recently published on the domain issues. The first one includes representative conference materials published in English: tekom-Jahrestagung und tcworld conference (2013 and 2014), 19th Conference on Computational Language Learning (2015), Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (2015), EURALEX International Congress (2018), Workshop on Natural Legal Language Processing (NAACL HLT, 2015, 2019). The corpus includes 372 papers, written in global English by researchers from Europe, North and South America, Southeast Asia, Australia and New Zealand. The corpus volume is 3 468 000 tokens, dictionary volume is 71 000 word-forms.

For detecting and extracting loan words in Russian scientific texts we built a corpus of papers published in proceedings of two international conferences "Internet and Modern society", namely IMS-2017 and IMS-2018. The research corpus volume is 226301 tokens (34833 word forms), it includes papers, relating to the following conference topics: computer linguistics, applied linguistics, electronic training and online educational technologies, information systems for science and education, cyberpsychology, state and society interaction in digital age, communicative culture of digital age, culture and technologies, information technologies and systems.

Before the papers were added to the corpus they underwent the necessary normalization procedures: formulae and meta information (about authors) were manually removed. References to explicit bibliographic sources were left for the sake of terminology in the texts, though, therefore in the alphabetic-frequency dictionaries (word lists) received with the help of AntConc corpus manager a lot of units are proper names, which then were manually removed from these word lists.

To provide comparability of the English and Russian corpora the English corpus was reduced to a sample of 275056 tokens (13699 word forms), this procedure makes it possible to compare frequencies of new words in both languages that shows prevailing quantity of new or occasional abbreviations in the English corpora (see Table 1)

and comparable quantity of abstract nouns in both corpora. In the studies that will follow this one, we are going to compare new lexical item frequencies in both corpora in order to set the really dominant translation modes for Russian.

The next step in identifying candidates for new terms was translating every item of the automatically produced and manually edited word lists (English and Russian) with the aid of WORD[+] machine translation system. Translated words were removed from the word lists, the words, which received no translation, were checked in ABBY LINGVO lexical system and consequently removed in case their translation was already fixed there. The words remaining without translation after this check were considered candidates for new lexis, possibly, new terms.

## 2      Results and Discussion

The procedures described in section 2 resulted in obtaining two dictionaries of new words based on alphabetic-frequency principle: English and Russian. The new words in the dictionaries are supplied with frequency index and illustration of their use in the context provided by corpus search. For new English words some additional information casting light on their actual meaning was added from their web definitions and description. The dictionary fragments are presented in Tables 1 and 2.

The obtained dictionaries allow to fix neologisms in the subject domain in question, describe productive word-building technics, analyze the attempts to adapt the loans to a Russian text, they may be used as a source information for translation dictionaries as well.

Thus, preliminary observation of dictionary items and their morphological structure marks similar derivation technics in both languages:

- the use of Latin and Greek components (*hypernymhyponym, mereological,* cinephotomacrography*, гиперпараметр, дискурсология, интерактив, квазикультурный*, etc);
- the use of English (-ness, -hood, -icity, etc.) and Russian (-ство, -сть, истика etc.) affixes for words which mean properties and characteristics, inherent to new entities (aboutness, termhood, formulaicity, гуманитаристика, дистантность, коммуникационность);
- the use of complex hyphenated constructions in English (-based, -driven, -formed, etc.) and Russian (–коммуникативный, -поисковый, -ориентированный, etc.)
- the use of abbreviation (AAsum - a very complex MDS method which fully exploits the advantages of clustering and the flexibility of matrix factorization; ACT-PTQ - Adding the adjectives ( modifiers ) and settings to the subject, verb and object, which we will called parse tree quintet (ACT-PTQ); MMO - многопользовательские онлайн игры; НБИК – Ипотечное бюро независимого кредитования).

**Table 1.** Fragment of English new word dictionary

| Frequency | Word | Context in the corpus, web description or definition |
|---|---|---|
| 1 | aboutness | determining the aboutness of conversations |
| 1 | ABox | The non-generic sentence (lb) roughly speaking provides ABox content for a machine-readable knowledge base, i.e., knowledge about particular instances. |
| 50 | ABS | The abbreviations c.c., ABS and TLC have various meanings in and out of the field of science. ABS: absolutive case (can be subject or object depending on transitivity). ERG: ergative (subject with transitive verbs). INE: inesive. INS: instrumental. DAT: dative. Absolute evaluation (called ABS in Fig. 1) therefore is required to determine the quality of a given translation. |
| 1 | abstractive-based | The abstractive-based approaches gather information across sentence boundary, and hence have the potential to cover more content in a more concise manner. |
| 7 | Abs-Cl | The first classifier (Comprehensive Classifier / Comp-Cl) is intended to cover dialectal statistics, token statistics, and writing style while the second one (Abstract Classifier / Abs-Cl) covers semantic and syntactic relations between words. |
| 96 | AdaGrad | AdaGrad - An adaptive learning rate method. AdaGrad algorithm (Duchi et al., 201 1) with mini-batch is adopted for optimization. |
| 8 | AdaGradUpdate | Note that AdaGradUpdate (x, g) is a procedure which updates the vector x with the respect to the gradient g. |
| 5 | B-ADDR | mainly common first names, such as John; such names are frequently labeled as B - ADDR across movies. |
| 9 | w.addSVD | We will compare to two other previously studied composition methods: weighted addition (w.addSVD), and lexfunc (Baroni and Zamparelli, 2010). w.addSVD is weighted addition of SVD vectors |
| 1 | add-one-approach | the add-one_approach returned the combination of the three features FORM, PLACE and FREQ_10PSadd-one-approach |

As far as the research focus is borrowed terminology in a Russian scientific text let us consider the ways the loans are treated by Russian authors.

While the recognized methods of new words introduction into a Russian text are tracing (loan translation), transliteration and transcription, the corpus under study demonstrates a sure preference of transliteration to the other two.

In the aspect of terminological system development and new lexical units generation linguistic technologies are very special, since their active progress results naturally in the origin of new terms and permanent updating of the accepted. Terminology of linguistic technologies has been traditionally developing on the English language base ever since 1947 when practical computer systems and algorithmic languages were created in USA and Great Britain.

English lexical units with transparent morphological structure are well motivated not only for native speakers of English, they are understandable for speakers of other languages, bilingual "web academics", by virtue of knowledge of their components. Providing that translation practice is lost in "web academy" there arises a set of loan terms either transliterated or transcribed lexical units, that are not motivated in the Russian morphological context. If morphological components are either transparent, or correlated with a Russian component (irrespective of their Greek or Latin origin), the loan term is motivated and its use is justified, though such terms permit parallel introduction of specialized translation equivalent, i.e. tracing, see, for example (5):

*(5)    "Наука о данных" или **даталогия** ("Datalogy"), начиная с 70-х годов прошлого века, рассматривается как академическая дисциплина, а с начала 2010-х годов, во многом благодаря популяризации концепции "больших данных", — и как практическая межотраслевая сфера деятельности.*

At the same time, a transliterated compound borrowing does not always permit to trace over the neologism meaning, as in (6):

*(6)    Данный **датасет** [dataset] уже был получен готовый с сайта британского проекта Mendeley, направленного на хранение и распространение научных трудов и баз данных по всему миру*

**Table 2.** Fragment of Russian new word dictionary

| Frequency | Word | Context in the corpus |
|---|---|---|
| 1 | агрессивно-девиантный | Наиболее часто на экспертное исследование поступают содержащие лингвистические признаки экстремизма и ксенофобии материалы, опубликованные пользователями социальных сетей «ВКонтакте», «Одноклассники», «Facebook», транслирующие модели агрессивно-девиантного речевого поведения |
| 1 | аддикт | Относительно небольшой процент всех геймеров являются аддиктами |
| 1 | адъективность | адъективность: отношение числа прилагательных к числу словоформ в тексте; |
| 1 | Азиопа | «Азиопа» для обозначения союза Азии и Европы |
| 1 | айпад | Такими орудиями являются, к примеру, компьютеры, смартфоны, айпады, гаджеты и виджеты, |
| 1 | Аквафон | мобильная компания «Аквафон» |
| 7 | аккаунт | регистрация соответствующего *«альтер-эго»*, в виде различных аккаунтов, чтобы исключить возможные недоразумения и возможные комментарии. |
| 1 | аксиологичный | Традиционные этические представления аксиологичны и антропоцентричны, то есть они, в основном, описывают взаимоотношения между людьми |
| 1 | акторно-сетевой | делает чрезвычайно эвристичным обращение к акторно-сетевой теории Б. Латура, в рамках которой технические устройства получают статус субъектов в совместной деятельности. |
| 1 | акустико-тактильный | В настоящее время исследователи выделяют несколько видов синестетической метафоры: слухо-зрительная; <…>, акустико-тактильная; … |

The component *set / сет* is actively borrowed by Russian authors, however, it is homonymous to the Russian root morpheme *сет- (сеть, сетевой)* for *network,* which can cause misunderstanding of *сетература* [netliterature]*, сетикет* [netiquette]*, сетинг* [setting] on the one hand and *синсет* [synset] on the other.

In some cases even the transparent structure of an original English word does not guarantee understanding of a transcribed or transliterated loan, in such cases the loan word is used with a definition to avoid incorrect interpretation (7):

*(7)     В то же время, пользователи часто указывают, что их сообщение имеет личный, а не официальный характер, это так называемый* **«дисклеймер»** *- письменный отказ от ответственности за возможные последствия в результате действий человека (или организации), заявившего данный отказ.*

The English word *disclaimer* is polysemantic, its different meanings are fixed in the ABBY LINGVO system as follows:

*1) отказ, отклонение, отречение*

*2) письменный отказ от ответственности*

*The company asserts in a disclaimer that it won't be held responsible for the accuracy of information. — В разъяснительном замечании компания предупреждает, что она отказывается нести ответственность за точность информации. Syn: denial, disavowal, rejection, renunciation*

*3) оговорка о случайном характере совпадений (имён персонажей в книге или фильме с именами реально существующих людей)*

*4) отказ (от права на что-л.), отречение*

The loan word *дисклаймер* [disclaimer-2] is borrowed in only one of these meanings.

A productive word-building model is used in case, when the borrowed word belongs to a different part of speech than the original word, for example, the English word *interactive* is an adjective, while it is borrowed as a noun *интерактив,* obviously due to analogy with *актив, позитив, диминутив, etc.*:

*(8)     Однако при использовании новых технологий необходимо помнить, что, несмотря на все новации,* **интерактив** *в цифровом формате не отменяет правила хорошего тона, тем более, что, как уже отмечалось, тема самопрезентации в сети на персональном и корпоративном уровнях - одна из наиболее актуальных тем современной деловой коммуникации.*

Sometimes the adjective to noun conversion is prior to borrowing and the new term is a loan transliteration, for instance *(аффективный) диспозитив* [affective disposition] as a loan term from social science domain:

(9)     **Аффективный идеологический диспозитив** *не восприимчив к идущим от Просвещения моделям борьбы с идеологией через рациональное (научное) объяснение фактов. <...>* **В аффективном диспозитиве** *популистской идеологии герой-одиночка борется с силами, намного превосходящими его, к тому же эти силы никогда не играют честно.*

The list of the examples can be continued, but they only confirm the fact, that Russian scientific texts are overloaded with borrowed terms and their meaning is not always clear even to the author, makes the Russian text inaccurate and precarious.

A suitable way out could be introduction of new terms followed by their working definitions, which permits to avoid possible misunderstanding or incomprehension.

However, interpretations of brand-new terms (even from the standpoint of the text authors) are rarely included into the text, they are rather an exception, as in (10) with the author's orthography and grammar:

*(10) Д. Берман и Д. Уэицнер ещё в 1997 году писали о детецентрилизованной структуре Интернета, которая предоставляет выбор из многочисленных* **вебсайт-хостингов (платформ, поддерживающих непрерывную работу сайта)***[site hostings (platforms that support the nonstop operation of the site)], что тем самым исключает необходимости получения «разрешения со стороны власти»*

At the same time in some cases we can see within the limits of one sentence new term introduction, which is not followed with explanations or interpretation, and introduction of terms interpretation with use of neologisms with no definitions at all, see,

But very often authors fail to introduce new terms correctly, avoiding definitions of loans even if their interpretation is part of research issue, for example, when classifying new objects of study as in (11). In this paragraph one can notice an ineffectual attempt to explain one loan by another one, which proves the assumption made above of authors' incomplete awareness of the loan term meaning: *дигитальный перформанс* [digital performance] could be hardly considered as a definition or equivalent for *медиаперформанс* [media performance]:

*(11) При этом культурные трансформации, новые требования к искусству, новая эстетика, социально-культурные потребности стали основаниями появления множества видов искусств. Это «электронное искусство», «цифровое искусство», видео-арт (в том числе, виджеинг, саунд-арт, медиаинсталляция и медиаскульптура),* **медиаперформанс (дигитальный перформанс)***, медиаландшафт (или медиасреда), сетевое искусство (интернет-арт или нет-арт, иногда также* **веб-арт***) и др.*

The paragraph is a happy illustration of the research focus issue since it demonstrates almost every possible way of a loan term presentation in a Russian scientific text: verbal translation (*«электронное искусство»* [electronic art]*, «цифровое искусство»* [digital art], *сетевое искусство* [web art]), transliteration (*саунд-арт* [sound art]*, интернет-арт* [internet art]*, нет-арт* [net art]*, веб-арт* [web art]), transcription (*виджеинг, cf.:* VJing) and tracing (*медиаперформанс* [media performance]*, дигитальный перформанс* [digital performance]) with a morphological adaptation to Russian word building rules (cf.: *медиаинсталляция, медиаскульптура, медиаландшафт, медиасреда*).

In this paragraph there is an attempt of pointing to new terms synonymy: "*медиаландшафт (или медиасреда), интернет-арт или нет-арт, иногда также веб-арт* [media landscape (or media environment, internet-art or net-art, sometimes even web-art]" that can't be interpreted as sufficient or reliable as the author does not clearly state that *медиаперформанс* and *дигитальный перформанс* are synonymous too. What is more, obviously synonymous *электронное искусство* [electronic art] and *цифровое искусство* [digital art] are introduced as different kinds of art (*множество видов искусств* [many kinds of art]).

# Conclusion

Contemporary English-oriented web academic intercourse and mandatory, under the circumstances, professional bilingualism of actively publishing Russian authors facilitate the process of borrowing new terminology of English origin. The lost practices of professional (within a particular knowledge domain) term translation and proper academic writing instruction result in a "scientific Volapük" mainly affecting the terminological component of Russian scientific texts published in national journals and conference proceedings.

The research focus being the manner and methods Russian authors accept and use new terms in their research papers, we suggest a corpus-based procedure for compiling dictionaries of new words (terms). The two original research corpora of contemporary English and Russian scientific texts restricted to web and linguistic technologies domain present a reliable material for detecting and describing new English terminology and its presentation in recent Russian scientific texts.

The corpus findings demonstrate that of the three recognized methods of borrowing (tracing or loan translation, transliteration and transcription), Russian authors prefer transliteration to the other two.

Analysis of dictionary items and their morphological structure revealed similar productive technics for new terminology in both languages, such as the use of Latin and Greek components, affixes meaning properties and characteristics, the use of complex hyphenated constructions and abbreviation.

Individual authors' methods of adapting English terms to a Russian text vary from a suitable definition of the borrowed term to a voluntary, often incorrect or erroneous (cf.: *гаждет* instead of *гаджет*) transcription / transliteration or mere borrowing a term "as it is" (*теневой DOM (ShadowDOM), HTML-импорты ( HTML-imports), art&science*).

The obtained dictionaries allow to fix neologisms in the subject domain in question. They help to highlight productive word-building technics, which may contribute to lexicographic (terminographic) practices. The dictionaries may be as well used as a reliable source information for specialized translation dictionaries.

# References

1. Beliaeva, L.N., Almazova, N. I., Kamshilova, O.N. Towards Textproductive Competences of a Language Worker and Novice Researcher. In: 18th PCSF 2018 - Professional Culture of the Specialist of the Future: The European Proceedings of Social & Behavioural Sciences EpSBS, vol. LI, pp.103-109 (2018).
2. Belyaeva, L. Scientific Text Corpora as a Lexicographic Source In: SLOVKO 2009. NLP, Corpus Linguistics, Corpus Based Grammar Research, Proc. from the Intern.Conference, November 25 – 27 2009, pp. 19-25. Smolenice, Slovakia (2009).
3. Delgado, M., Martin-Bautista, M.J., Sanchez, D., Vila, M.A. Mining Text Data: Special Features and Patterns. In: Lecture Notes in Computer Science, vol. 2442, pp. 140-151. Springer-Verlag GmbH (2002).
4. Beliaeva, L., Chernyavskaya, V., Nefedov, S. Language Worker in the Framework of Information 4.0. In: 18th PCSF 2018 - Professional Culture of the Specialist of the Future:

The European Proceedings of Social & Behavioural Sciences EpSBS, vol. LI, pp. 608-614 (2018).

5. Beliaeva, L., Kamshilova, O. Problems and Perspectives of Language Worker Professional Training. International Journal of Open Information Technologies 6(12), 35-42 (2018).

6. Tarp, S. Beyond Lexicography: New Visions and Challenges in the Information Age. In: Bergenholtz, H., Nielsen, S. & Tarp, S. (eds.) Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow, pp. 17-32. Peter Lang AG, International Academic Publishers, Bern (2009).

7. Neuman, Y., Nave, O. and Dolev, E. Buzzwords on Their Way to a Tipping Point: a View from the Blogo-sphere'. Complexity 16(4), 58-68 (2010).

8. Beliaeva, L.N, Kamshilova, O.N. Textproductive Competences of a Research Beginner: the Russian Academic Discourse Today. The Humanities and Social Studies in the Far East. XVI (1), 31-35. (2019). [in Russian].

9. Haspelmath, M. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In: T. Stolz, D. Bakker & P.-R. Salas (eds.), Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes, pp. 43-62.Mouton de Gruyter, Berlin (2008).

10. Tadmor, U. Loanwords in the world's languages: Findings and results, In: Haspelmath, M. & Tadmor, U., (eds). Loanwords in the world's languages, pp. 55-75. De Gruyter Mouton, The Hague (2009).

11. Vasiljevs, A., Pinnis, M., Gornostay, T. Service model for semi-automatic generation of multilingual terminology resources. In: Terminology and Knowledge Engineering, 19-21 Jun 2014, pp. 67-76. (2014).

12. Cook, P., Lau, J.H., Rundell, M., McCarthy, D & Baldwin, T. A lexicographic appraisal of an automatic approach for detecting new word-senses. In: Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference: Tallinn, Estonia, pp. 49-65. (2013).

13. O'Donovan, R., O'Neill, M. A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary. In: J. D. E. Bernal (ed.) Proceedings of the 13th EURALEX International Congress. Barcelona, Spain: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra, pp. 571-579. (2008).

14. Abel, A., Stemle, E.W. On the Detection of Neologism Candidates as a Basis for Language Observation and Lexicographic Endeavors: the STyrLogism Project. In: Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, pp. 535-545 (2018).

15. Kerremans, D. A Web of Words. A Corpus-Based Study of the Conventionalization Process of English Neologisms.: Peter Lang GmbH, Frankfurt (2015).