



DOI: 10.18413/2313-8912-2021-7-4-0-2

Olga N. Polshchikova<sup>1</sup>   
Yuliana Y. Genkin<sup>2</sup> 

## Synonymy in the terminology of computational linguistics

<sup>1</sup> Belgorod State National Research University  
14 Studencheskaya St., Belgorod, 308007, Russia  
*E-mail:* polshchikova@bsu.edu.ru  
*ORCID iD:* 0000-0002-0062-9963

<sup>2</sup> Admiral F.F. Ushakov State Maritime University  
93 Lenin Ave., Novorossisk, 353918, Russia  
*E-mail:* yuliana1970@yahoo.com  
*ORCID iD:* 0000-0003-1762-4277

*Received 27 September 2021; accepted 14 December 2021; published 30 December 2021*

**Abstract.** The article presents a study of synonymous relations in the computational linguistics terminology, the relevance of which is due to the need to streamline the corresponding terminology system. The study is focused on identifying the reasons for the presence of synonymous terms in the computational linguistics vocabulary, grouping them in accordance with classification features, analyzing their etymology, morphological nature, forms of variance and interchangeability. The systematization of the terms in question is based on the descriptive method of research. Etymological, definitive and quantitative analysis methods were also applied. As a result of the study, it was found that the main reasons for the presence of synonymous relations in the computational linguistics terminology are associated with a variety of term structure forming methods, the need to select Russian-language correspondences to terms of foreign language origin and the intensive emergence of new concepts due to the rapid development of the professional sphere of automatic processing of natural language. The authors propose a classification of computational linguistics terms-synonyms according to the type of synonymous relation, structure, morphological nature, the components number of the synonymous series, etymological characteristics. Interchangeable word combinations, their truncated verbal forms, abbreviations and syntactic variants of terms in computational linguistics are revealed.

**Keywords:** Terminology; Computational linguistics; Automatic natural language processing; Synonyms; Classification; Variance; Interchangeability

**How to cite:** Polshchikova O. N., Genkin Yu.Y. (2021). Synonymy in the terminology of computational linguistics. *Research Result. Theoretical and Applied Linguistics*, V.7 (4), 23-30, DOI: 10.18413/2313-8912-2021-7-4-0-2

### INTRODUCTION

Computational linguistics is a relatively new professional interdisciplinary field, theoretical and applied developments of

which are formed at the junction of linguistics, mathematics, computational methods and data processing technologies based on artificial intelligence (Bolshakova et al., 2017; Dowell,

Nixon, Graesser, 2019; Ive, Viani, Kam, 2020; Prokhorova et al., 2021). The field of computational linguistics is focused on solving various applied problems related to automatic natural language processing (Kormalev et al., 2004; Dowell, Nixon, Graesser, 2019; Mejía, 2019; Polshchikov, Igityan, 2019; Polshchikov, Lazarev, Konstantinov, 2020; Polshchikov et al., 2020; Qiu et al., 2020; Savin, Drews, Maestre-Andrés, 2020; Aguzumtsyan et al., 2021; Arts, How, Gomez, 2021; Dehouche, 2021; Moura, Lopes Cardoso, Sousa-Silva, 2021; Velikanova et al., 2021).

In the computational linguistics terminology, as well as in other various industry terminologies, synonymy is widespread, which causes "certain difficulties in the field of professional communication" (Taranova, Bubyreva, Taranov, 2016). The computational linguistics terms synonyms are used in oral and professional speech, in textbooks, scientific articles, dissertations, and other research literature. Special attention is paid to the consideration of synonymy in various linguistic resources. In the lexical ontology of WordNet, the words of each part of speech are grouped into *synsets* (synonymic series), which are the dictionary basic units. The presence of synonymy causes "difficulties in identifying different occurrences of terms in the text" (Bolshakova et al., 2017). This is due to the fact that "... terms are often modified when used – truncated, abbreviated, replaced with synonyms, combined, etc.: коммуникативная многозначность запроса (*communicative ambiguity of the query*) – коммуникативная многозначность (*communicative ambiguity*), синтаксическое представление (*syntactic representation*) – СинП (*SynP*), вложенный файл (*attached file*) – вложение (*attachment*). Such textual variants represent different forms of expression of the same concept and, if possible, should be recognized" (Blei, Lafferty, 2007; Blei, 2012; Bolshakova et al., 2017).

The study of the classification of synonymous terms and reasons for their appearance is focused on solving the most important tasks in the practical sphere related to the unification of term systems, their ordering (Taranova, Bubyreva, Taranov, 2016). In this regard, the study of synonymic relations in the terminology of computational linguistics seems to be an urgent task.

### MAIN PART

**The purpose** of the article is to identify the reasons for the synonymous relations existence in the computational linguistics terminology, to group synonymous terms according to various classification criteria, to investigate them from the point of view of etymology and morphological nature, to analyze the forms of variation and interchangeability of terms in the professional sphere under consideration.

### Materials and Methods

The sources of the factual material included texts on the computational linguistics subject, published in reference books, scientific periodicals and collections, in translation and explanatory dictionaries, in particular in the Russian-English Thesaurus on Computational Linguistics (RETCL). The systematization of the considered terms is carried out on the basis of the descriptive research method use. The origin of some terms is revealed using the etymological analysis method. A definitive analysis of the factual material served to trace the special names semantic features. The quantitative analysis method was used as an auxiliary one.

### Results and Discussion

The reasons for the appearance and use of synonyms in professional vocabulary have been investigated in many works (Taranova, Bubyreva, Taranov, 2016; Babalova, Shirobokov, 2018; Dasovkhadzhieva, 2020; Pllana et al., 2020; Vakulik, Sichkar, 2020) In the terminology of computational linguistics, synonymy, in our opinion, is due to the following factors:

1) a variety of ways to form the terms structure;

2) the expediency of selecting Russian-language correspondences to foreign origin terms;

3) the presence of various features of the nomination object that can become the basis for the name;

4) the emergence of new concepts or properties of objects in the process of sciences and technologies constant development, which makes us look for the appropriate exact nominations and the most successful names;

5) available common language synonyms from which terms can be selected;

6) the parallel use of the term and its definition possibility (definitive synonymy);

7) the desire to avoid the same word or phrase repetitions;

8) naming the same concept by different research schools, scientists;

9) the need for language economy, contributing to the use of one-word compound names and abbreviations.

According to the type of synonymic relation, absolute, relative and complex synonyms are distinguished (Fig.). The first type consists of semantically identical terms, which are called doublets: *segmentation (tokenization)* – "splitting text into segments with narrower informational content" (Hobson, Hannes, Cole, 2020); "the process of splitting text into linguistically significant units, i.e. words (word forms), punctuation marks, numbers and alphanumeric expressions" (Mitkov, 2003); *stemmatization (normalization, lemmatization)* – "bringing each word in a document to its normal form" (Bolshakova et al., 2017); "grouping of various forms of a word into clusters" (Hobson, Hannes, Cole, 2020); "the process of grouping various inflectional forms of one word in such a way that when analyzed they

are processed as one word" (RETCL), *sirconstant (adjunct)* – "a unit that fills an active syntactic valence that does not correspond to any semantic valence" (Testeleets, 2001); *word (token)* is "the main structural and semantic language unit that serves to name objects and their properties, phenomena, relations of reality, having a set of semantic, phonetic and grammatical features" (RETCL); "a language unit that serves to name a separate concept" (Kuznetsov, 2000); "a substring in the text located between punctuation marks" (Bolshakova et al., 2017); *linguistic (language) corpus* – "a certain philologically competent array of linguistic data, a set of texts selected in accordance with a certain research task and specially prepared, marked, structured, presented in a unified form" (RETCL); "a representative array of texts collected according to a certain principle (by genre, authorship, etc.) and having linguistic markup (morphological, accentual, syntactic, discursive)" (Bolshakova et al., 2017); "an approximate set of statements selected for analysis and presented in the form of a written text, tape recording, etc." (Akhmanova, 2004); *template (sample)* – "description of a linguistic construct that is used to reflect the desired fact or object in the text and describes various attributes of the text: morphological features of words, their syntactic meaning and relationship, belonging to a separate fragment of a sentence; word order, distance between words, the presence of keywords characterizing the situation" (Kormalev et al., 2004); "a formal description (sample) of a language construct that needs to be found in the text in order to extract the necessary information" (Bolshakova et al., 2017).

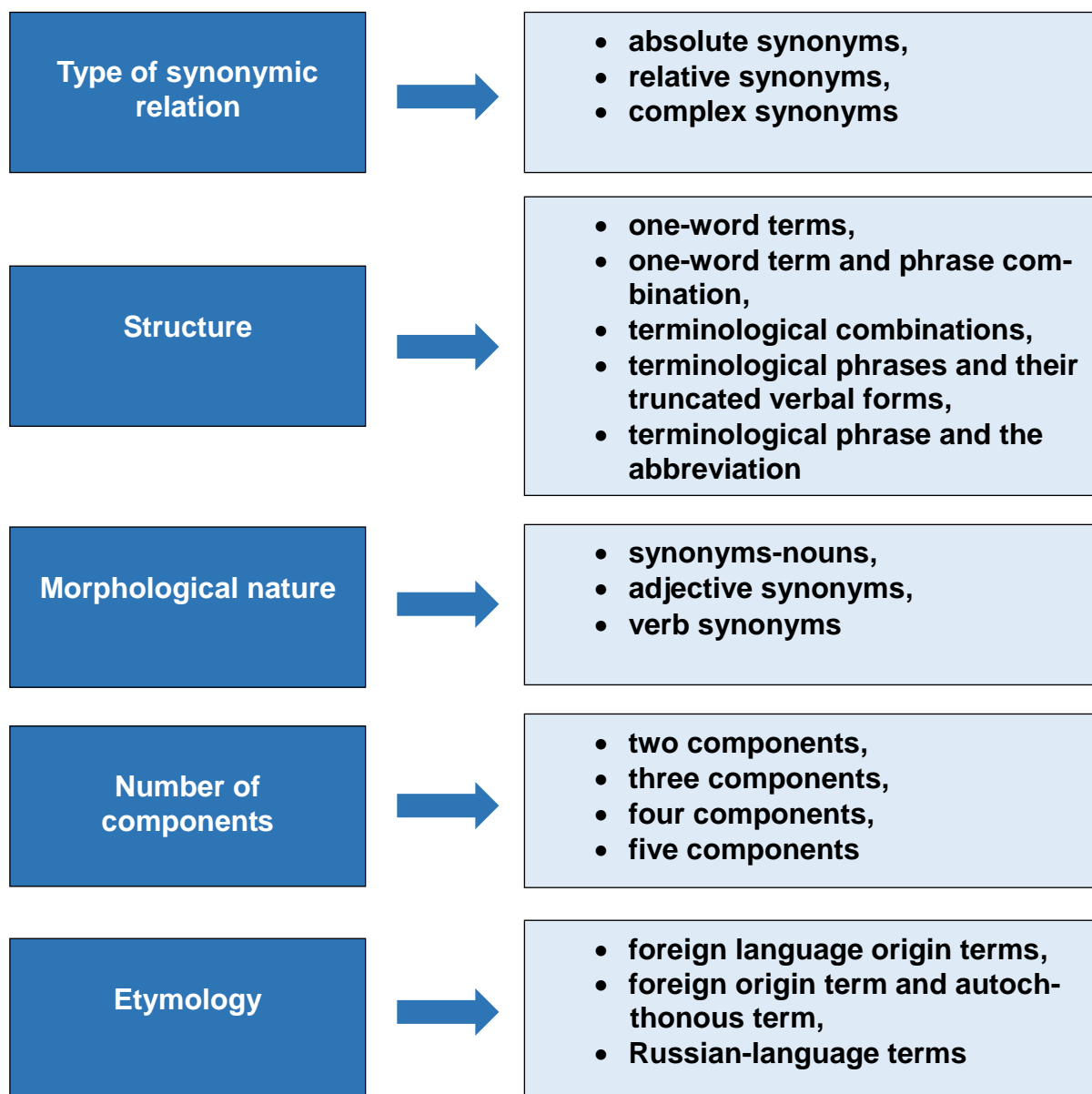


Fig. Classification of computational linguistics terms-synonyms

Semantically similar terms belong to relative synonyms, examples of which are: *computational linguistics* – ("a direction in applied linguistics focused on the computer tools use – programs, computer technologies for organizing and processing data – for modeling the language functioning in various conditions, situations, problem areas, etc., as well as the entire scope of computer models application of language in linguistics and related disciplines" (RETCL) and *automatic natural language processing* (has a narrower meaning: "the research direction dealing

with ... computer processes modeling of the texts in natural language analysis and generation, sounding speech recognition and synthesis, as well as solving applied problems related to the text and sounding speech transformation, for example, the resolution of the words ambiguity in the text, machine translation, automatic abstracting, speaker identification by voice (speaker identification) and many others" (RETCL); *ambiguity resolution* ("competing variants removal from the linguistic object representation (text) while maintaining a consistent interpretation"

(RETCL) and *homonymy removal* (has a refined definition: "the stage of text analysis at which a single variant of morphological analysis for each token is selected" (Bolshakova et al., 2017).

Complex synonyms combine the terms of the first two types, for example, the phrases *opinion extraction* and *tonality analysis*, meaning "identifying emotionally colored vocabulary and emotional evaluation of the author in relation to the objects in question in the text" (RETCL), are absolute synonyms, and the term *automatic text processing* ("text conversion in artificial or natural language using a computer" (RETCL) is a relative synonym of the first two.

Synonymous terms differ in structure. Parallel use of the following terms has been found in the computational linguistics terminology:

1) one-word terms: *фонация (phonation)* – *голосообразование (vocalization)*; *тегирование (tagging)* – *разметка (markup)*; *репрезентативность (representativeness)* – *сбалансированность (balance)*;

2) one-word term and phrase combination: *ресинтез (resynthesis)* – *вокодерный синтез (vocoder synthesis)*; *тема (theme)* – *логический субъект (logical subject)*; *тон (tone)* – *мелодика речи (speech melody)*, *форманта (formant)* – *полюс спектра (pole of the spectrum)*; *цель (goal)* – *конечная точка (endpoint)*; *чтение (reading)* – *репродуцированная речь (reproduced speech)*; *тезаурус (thesaurus)* – *семантический словарь (semantic dictionary)*; *оцифровка (digitization)* – *цифровое кодирование (digital coding)*; *коллокация (collocation)* – *устойчивое словосочетание (stable phrase)*; *кластеризация (clustering)* – *кластерный анализ (cluster analysis)*;

3) terminological combinations: *извлечение информации (information extraction)* – *выделение концептов (concepts identification)*, *группа числительного (a numerals group)* – *количественная конструкция (quantitative construction)*; *референциальный анализ*

(*referential analysis*) – *разрешение анафоры (anaphora resolution)*; *система управления терминологией (terminology management system)* – *терминологический менеджер (terminology manager)*.

In the computational linguistics context, in addition to synonyms, various forms of same terminological phrases expression are used interchangeably:

- terminological phrases and their truncated verbal forms: *автоматическое реферирование (automatic referencing)* – *автореферирование (autoreference)*;

- terminological phrase and the corresponding abbreviation: *компьютерная лингвистика (computational linguistics)* – *КЛ (CL)*, *машинный перевод (machine translation)* – *МП (MP)*, *лингвистический процессор (linguistic processor)* – *ЛП (LP)*, *автоматическое распознавание речи (automatic speech recognition)* – *АПП (ASR)*, *информационно-поисковая система (information retrieval system)* – *ИРС (ИПИС)*, *искусственная нейронная сеть (artificial neural network)* – *ИНС (INN)*; *терминологический банк данных (terminological data bank)* – *ТБД (TDB)*.

By morphological nature, synonyms-nouns are the most common in the studied vocabulary (75.2%), for example, *артикуляция (articulation)* – *произнесение (pronunciation)*, *определение (definition)* – *толкование (interpretation)*. Adjective synonyms are less common (18.9%), for example, *сконструированный (constructed)* – *искусственный (artificial)*, *автоматический (automatic)* – *машинный (machine)*, *речевой (speech)* – *голосовой (voice)*. A small number of verb synonyms (5.9%) were revealed, for example, *тестировать (to test)* – *испытывать (to check)*; *корректировать (to correct)* – *исправлять (to repair)*.

A different number of components may belong to a synonymous series in the vocabulary computational linguistics:

1) two components: *именной (nominal)* – *субстантивный (substantive)*; *просодическая (prosodic)* – *суперсегментная (supersegmental)*;

2) three components: *онтологический* (*ontological*) – *энциклопедический* (*encyclopedic*) – *внелингвистический* (*extralinguistic*), *тональный* (*tonal*) – *интонационный* (*intonation*) – *мелодический* (*melodic*);

3) four components: *рамка* (*frame*) – *структура* (*structure*) – *схема* (*diagram*) – *формат* (*format*);

4) five components: *накопитель* (*storage*) – *память* (*memory*) – *архив* (*archive*) – *банк* (*bank*) – *база* (*database*).

It is advisable to distinguish the computational linguistics terms synonyms from the point of view of etymology. The parallel use of foreign language origin terms is revealed: *актант* (*actant*) – *аргумент* (*argument*), *синтаксический анализ* (*syntactic analysis*) – *парсинг* (*parsing*). Quite often you can observe a synonymous pair, which is made up of a foreign origin term and an autochthonous term: *вокабула* (*vocabula*) – *заголовок* (*title*), *пассивная* (*passive*) – *внешняя* (*external*). In addition, in the studied professional vocabulary, Russian-language terms are used in parallel: *словарный вход* (*dictionary entry*) – *заглавное слово* (*capital word*), *прямой* (*direct*) – *пословный* (*word-by-word*).

In addition to synonymy, many industry terminologies are characterized by variation, that is, formal modification of the same phrase without violating the meaning identity. In the computational linguistics terminology, the use of syntactic variants – word combinations with various possible types of grammatical connection is revealed: *корпус речи* (*the corpus of speech*) – *the речевой корпус* (*speech corpus*), *корпус текстов* (*the corpus of texts*) – *the текстовый корпус* (*text corpus*), *мера ассоциации* (*the measure of association*) – *ассоциативная мера* (*the associative measure*), *ресурсы лингвистики* (*the resources of linguistics*) – *лингвистические ресурсы* (*linguistic resources*), *фрейм целей* (*the frame of goals*) – *целевой фрейм* (*the target frame*).

## CONCLUSIONS

The presented research allowed us to identify reasons for the presence of synonymous relations in the computational linguistics terminology, which, first of all, are related to the variety of term forming structure ways, the intensive emergence of new concepts due to the automatic natural language processing professional sphere rapid development. In addition, terminological synonymy is due to the formation of special names from existing language synonyms, the desire to avoid repetition and save language resources, as well as the parallel various scientific schools terminological databases functioning.

Computer linguistics synonyms are classified by the synonymic relation type (absolute, relative and complex), by structure (one-word terms, one-word terms and phrases, terminological combinations), by morphological nature (nouns, adjectives, verbal synonyms), by the synonymic series components number (two-, three-, four-, five-component), by etymological feature (foreign origin synonyms, a term of foreign origin and synonymous autochthonous term, Russian-language synonymous terms).

In the computational linguistics terminology, interchangeable phrases, their truncated word forms and abbreviations, as well as syntactic variants in the form of phrases with various grammatical connection types are revealed.

So, despite the synonymy undesirability in professional speech, there are synonymous words and phrases in the computational linguistics terminology. In our opinion, synonymy of terms should not be interpreted only as a negative phenomenon. Sometimes it is appropriate, as it helps to clarify the wording of the thought, avoiding unnecessary repetitions. The presence of synonymy and variation is an integral feature of the terminology under study and its development consequence.

## References

- Aguzumtsyan, R.V., Velikanova, A.S., Polshchikov, K.A., Igityan, E.V. and Likhosherstov, R.V. (2021). Application of intellectual technologies of natural language processing and virtual reality means to support decision-making when selecting project executors. *Economics. Information technologies*, 48 (2), 392–404. (In Russian)
- Akhmanova, O.S. (2004). *Slovar' lingvisticheskikh terminov* [Dictionary of linguistic terms], Editorial URSS, Moscow, Russia. (In Russian)
- Arts, S., Hou, J. and Gomez, J.C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, 50, 104-144. (In English)
- Babalova, G.G. and Shirobokov, S.N. (2018). Synonymy of Computer Terminology. *Science of Man: Humanitarian Research*, 4, 34-39. (In Russian)
- Blei, D. and Lafferty, J. (2007) A correlated topic model of Science. *Annals of Applied Statistics*, 1, 17-35. (In English)
- Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, 55, 77-84. (In English)
- Bolshakova, E.I., Vorontsov, K.V., Efremova, N.E., Klyshinsky, E.S., Lukashevich, N.V. and Sapin, A.S. (2017). *Automatic Natural Language Processing and Data Analysis*. HSE Publishing House, Moscow, Russia. (In Russian)
- Dasovkhadzhieva, A.A. (2020). On the question of synonymy in sports terminology. *Lingua-Universum*, 3, 14-16. (In Russian)
- Dehouche, N. (2021). Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics. Sci. Environ. Polit.*, 21, 17-23. (In English)
- Dowell, N.M.M., Nixon, T.M. and Graesser, A. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, 51, 1007-1041. (In English)
- Hobson, L., Hannes, H. and Cole, H. (2020). *Natural Language Processing in Action*. Piter, Saint Petersburg, Russia. (In Russian)
- Ive, J., Viani, N. and Kam, J. (2020). Generation and evaluation of artificial mental health records for Natural Language Processing. *npj Digital Medicine*, 3, 69. (In English)
- Kormalev, D.A., Kurshev, E.P., Suleimanova, E.A. and Trofimov, I.V. (2004). The architecture of instrumental systems for extracting information from texts. *Software systems: theory and applications*, 2, 49-70. (In Russian)
- Kuznetsov, S.A. (2000). *Great Dictionary of Russian language*. Norint, Saint Petersburg, Russia. (In Russian)
- Mejía, J. M. M. (2019) Toward the analysis of Richard Wagner's work by means of computational linguistics *Forma y function*, 32(1), 125-148. (In English)
- Mitkov, R. (2003). *The Oxford handbook of computational linguistics*. Oxford University Press, New York, US. (In English)
- Moura, R., Lopes Cardoso, H. and Sousa-Silva, R. (2021). Automated Fake News Detection Using Computational Forensic Linguistics. *Lecture Notes in Computer Science*, 12981, 788-800. (In English)
- Pllana, S., Pllana, G., Pllana, E. and Pllana, Z. (2020). Synonymy and terminological doublet in economic terminology. *Russian linguistic bulletin*, 3, 117-120. (In English)
- Polshchikov, K.A., Lazarev, S.A., Konstantinov, I.S., Polshchikova, O.N., Svoynkina L.F., Igityan, E.V. and Balakshin, M.S. (2020). Model for assessing the effectiveness of the robotic system of communicative functions. *STIN*, 6, 4-7. (In Russian)
- Polshchikov, K. and Igityan, E. (2019). Evaluating the effectiveness of text analyzers. *Proceedings of the 2nd International Conference on Mathematical Modelling in Applied Sciences*, Belgorod, 97-98. (In English)
- Polshchikov, K.A., Lazarev, S.A. and Konstantinov I.S. (2020). Assessing the Efficiency of Robot Communication. *Russian Engineering Research*, 40, 936-938. (In English)
- Polshchikov, K., Lazarev, S., Polshchikova, O. and Igityan, E. (2019). The Algorithm for Decision-Making Supporting on the Selection of Processing Means for Big Arrays of Natural Language Data. *Lobachevskii Journal of Mathematics*, 40, 1831-1836. (In English)
- Prokhorova, O.N., Polshchikova, O.N., Polshchikova, A.K. and Deev, A.V. (2021). Systematicity of Computational Linguistics Terminology. *Proceedings of the Southwest State University. Series: Linguistics and Pedagogics*, 11(1), 29-39. (In Russian)

Qiu, X.P., Sun, T.X., Xu, Y.G., Shao, Y.F., Dai, N. and Huang, X.J. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63, 1872-1897. (In English)

Russian-English thesaurus on computational linguistics (RETCL). URL: <https://uniserv.iis.nsk.su/thes/search.php> (Accessed 18 October 2021). (In Russian)

Savin, I., Drews, S. and Maestre-Andrés, S. (2020). Public views on carbon taxation and its fairness: a computational-linguistics analysis. *Climatic Change*, 162, 2107-2138. (In English)

Taranova, E.N., Bubyreva Zh. A. and Taranov, A.O. (2016). The problem of synonymy in special terminology. *Bulletin of TSPU*, 2, 55-60. (In Russian)

Testelets, Y.G. (2001). An introduction to general syntax. Russian State Humanitarian University, Moscow, Russia. (In Russian)

Vakulik, I.I. and Sichkar, I.Yu. (2020). Medical terminology and equivalent synonymy:

features of education. *Humanitarian research*, 11, 1. (In Russian)

Velikanova, A.S., Polshchikov, K.A., Likhosherstov, R.V. and Polshchikova, A.K. (2021). The use of virtual reality and fuzzy neural network tools to identify the focus on achieving project results. *Journal of Physics: Conference Series*, 2060, 012017. (In English)

**Конфликты интересов: у авторов нет конфликта интересов для декларации.**

**Conflicts of Interest: the authors have no conflict of interest to declare.**

**Olga Nikolaevna Polshchikova**, Candidate of Philological Sciences, Associate Professor, Department of the Russian Language and Cross-cultural Communication, Belgorod State National Research University.

**Yuliana Yurievna Genkin**, Candidate of Philological Sciences, Associate Professor, Foreign Languages Department, Admiral F.F. Ushakov State Maritime University.