SPECIFIC AND PREVALENT PECULIARITIES OF THE AUTHORSHIP CORPUS

Khamrayeva Shahlo Mirjanovna

Researcher, Bukhara State University, Bukhara, Uzbekistan

ABSTRACT

The article deals with the relationship of the authorship`s lexicography with the corpus of authorship, the concept of concordance and the frequency of dictionary, the general and unique character of the authorship corpus.

KEYWORDS: *The Corpus of Authorship, Authorship Lexicography, Concordance, Electronic Dictionary*

INTRODUCTION

The rationale of copyright lexicography and the corpus of authorship. The world lexicography having known separate line as author`s lexicography largely includes the theoretical basis of the language dictionary, the experience of investigating author`s idiotype, the ideas for creating new types of dictionaries, in this sense, the peculiarity of the author`s electronic dictionary. Moreover, it deals with issues of the comparative-analytical study of contemporary literary texts on the basis of copyright lexicography (1). The field of authorship lexicography is a rapidly developing area in the world, especially, in Russian lexicography, resulting in establishments of numerous electronic dictionary and corpses. At the Moscow State University named after M. Lomonosov such dictionaries as the Glossary of Krylov slang, the dictionary of Griboedov comedies, the dictionary of Gogol`s "The Government Inspector" comedy, Fonziz`s comedies dictionaries have been created and served as an example of "A. Pushkin`s Glossary". This experience has become popular in Russian lexicography, and similar research has been started in other higher education institutions. As examples among the works of such investigations we can widely give the language dictionary of Lomonosov`s works, the contradictory poems of M. Kuzmin, the ideological dictionary of A. Mitskevich`s poem "Pan Tadeusz". From this field of research, A.P Chekhov`s artworks created by O.V. Kukushkina, A.A. Polikarpov, E.V. Surovtseva are considered to be noteworthy and striking.

METHODS

Principles and derivatives of authorship lexicography. The author`s lexicography is usually made up of two chapters: the basis and the product. In other words, base - case or concordance, product - dictionary. Later on, in lexicography (not in corpus linguistics) the concept of `concordance` is accepted as widespread. Concordance is a list of all the context of the use of this or that phrase in the context of the source and the word concordance is used in this sense in corpus linguistics. Additionally, the term conditional is also used to refer to an alphabetical list of links to a book or work keyword (2). The first example of this controversy was created in the XIII century with the participation of 500 monks and friars. Because of the enormous size of the work (very important concordances: Veda, Bible, Koran or William Shakespeare masterpieces), there were large teams working on. In addition, the "concordance" linguistically is extensively applied to translation, lexicography, language teaching, and learning, as well as textual analysis (3). It is used exclusively

to make linguistic actions, such as compiling a list of words from the cork, comparing the various uses of a word, searching for words and phrases, analyzing keywords, translating various words or combinations, or determining the frequency of use (4). The concordance-dictionary is a separate type of dictionary that is a short context for every word (not semantic interpretation!). The ability to search, explore, analyze, index, compare and mapping of concordance can be used to compile text. The search function allows you to quickly find the desired scene in the text using the given word or combination (5). According to A. Stolyarov, this function of the concordance can also be used to select a citation or to examine the originality of the text. Usually, concordance is often used by reader implementers, in other words, it is important to look up words and idioms in the text and to translate the terms.

An interactive function of concordance allows a linguist or a simple reader with experience new interpretation and commentary, analytical function to analyze different units of the language: lexeme, keywords, their frequency of application in text, indexing function to create a list and index of the word in preparing the text for publication and the comparison function compares all occurrences of speech and word usage in the topic. The dictionary article, in contrast to the ordinary dictionary, provides an example of a concordance word, that is, the context of the application. Consequently, concordance serves as the basis for creating a variety of differential, as well as generalized dictionaries. The contemporaneous works of F. Dostoyevsky, F.T. Tyutchev, AP Chekhov, Griboedov, S. Esen, S. Aksakov, Lomonosov are among these scientific works. It can be said that contrasting is the ability to store large volumes of data over traditional lexicography's. The Constellation Unit is able to track paradigmatic relationships in the text, quantitative analysis of the use of word usage in different parts of her creative heritage, and conclusions on her specific syntactic devices. A glossary is a modern and effective tool for researching linguistic and literary studies, making it an important part of corpus linguistics (4).

The composition of A.P. Chekhov`s works includes all the completed works of A. Chekhov and is based on the author`s collection of excellent works. There are 600 texts (works) in the body: 17 plays and 583 prose works, 1 271 664 words of usage (176 308 words of play and 1 095 356 words of prose), 101 282 different words (23 093 drams; 141 prose works), 36,419 lexeme (11,802 dramatic pieces, 34,763 prose works). The dramatic and prose works of a single title are considered separately.

The works of A.P. Chekhov are encoded on the basis of the linguistic body marks: word (1); the initial form of the word (2); the word combination (3); special applications (4) and extra linguistics: work title (1); title and dedication (2); year of publication (3); switching (4); Text type: prose or dramatic (5); signature (6); the genre of work (7); the type of work of the genre (8); speaking genre (9); place of writing (10); type of story (11); scale of topics (12); the main theme of the work (13); semantic group (Geo; Person; Name; Color; Sound) (14); volume and page number in the collection of excellent works (15).

Keywords(forms) include the number and highlighted phrases (abnormal, foreign language). *The initial form of the word* could be derived from the mirror of "word". The words in the dictionary are somewhat different from the dictionary (5): 1) the word of corps is also included in the drama "Ivanov" (1987); 2) does not distinguish between functional ambiguity (predictive) index, and the dictionary is indexed by this index; you will need to look up information about a category to know which category to enter; 3) option information is as follows: the main option and the second option are presented side by side. Mas, b, by, b. When the main option is selected, the second option is not output. If the unit you are looking for cannot be found in the mirror of "*word*", it is best to apply to the "*Keywords(forms)*" window.

The name of the work includes all the information related to a particular work: the year and place of the work, the idea of the work, and the description of A. Chekhov`s collection of works. "*The title of the heading (supplementary title) and dedication*" can be found in the second chapter (if any) and it is possible to find out about the dedication of the work.

A. Pushkin`s corpus. While speaking about the creation of the Electronic Encyclopedia of Pushkin (6), A. A. Polikarpov mentioned that not only was Pushkin`s corpse, but also that there was no such authoritative work of a writer, and regrets that this publication was a first-rate experience and that the author`s lexicography is now slowly growing. In his opinion, the corpus of authorship is an area that has been neglected in the world of corpus linguistics. The authorship corpus created till this day (F. D. Dostoevsky`s publicity corpus, made by scientists from the Petrozavodsk State University) shows that they are narrower (mainly informative) as an electronic corpus, representing one or two aspects of the linguistic feature. This issue has not been solved in the world of lexicography. The main (even uniqueness) unit of search in the world of copyright was unspecified, the uncertainty of the events are to be keywords(forms). A.S. Pushkin`s creations had been remaining a unique experience in defining the complex features of the author`s works while creating the copyright corps, and the Pushkin Electronic Encyclopedia was a serious step forward. The language of Pushkin`s works became the basis for learning the modern Russian literary language and grammatical principles. Pushkin`s works reflect the linguistic features of the electronic version of the electronic collection, which is only a word-based search engine holding some of Pushkin`s works and illustrations. The need to create a linguistic database is based on the fact that Pushkin`s works have not been comprehensively studied in whole or in part. Not only Pushkin`s lexicon and phraseology are semantic unspecified, but Pushkin`s grammar-morphological, morphemic, word-formation, and syntactic features are also not systematized. The peculiarity of the use of the word and phrase, in general, the vocabulary of Pushkin in different genres, has not been studied as a single system. These problems can be solved by integrating Pushkin`s corpse into a whole (through a comprehensive linguistic analysis of the word, morpheme, and combination of grammatical and semantic classifications). Combining these features, Pushkin`s works must have a perfect database that can be used by all experts, conducting independent research and producing the necessary scientific conclusions. Having broad linguistic information of such a body unit (even the meaning of the word, even the meaning of the core) extends the scope of research in the field of the linguistic, linguistic, language-forming and evolutionary process. It is also a foundation for the study of the new phase of Pushkin language traits, Pushkin texts, syntactic units, and structure as a corpus database.

The corpus of F. Dostoyevsky`s texts was created as a source for the dictionary of Dostoevsky`s language concept, and the composition of the dictionary was the basis for the corpus-setting parameters. As the dictionary aims to fully encompass Dostoyevsky`s lexicon vocabulary, the body of the corpus includes all the artistic and publicist works of the author. The text storage unit is defined as a separate work (from novel to a mere letter). Nowadays, almost all electronic carriers have possessed his works of art, publicity, epistolary heritage, and stored in ASCII format with minimal scroll to the worksheet and page. There are several dictionaries based on the case of the corpus frequency, created by A. Shaykevich. There is also a remarkable database of idiomatic information that is based on the authors` idioms. Dostoevsky`s body corpus has also a CD-ROM shape ("Dostoevsky: Texts, research, materials").

Principles of compilation F. Dostoyevsky`s "Dictionary-Concordance" contains the ability of using any word and phrase in all original texts of the entire 30-volume collection of masterpieces(fiction, journalism and writer`s own diary) by the author, also holds within itself the frequency capability and the ability to access the main text. Hence, all completed works are taken as text, but works ("A number of articles on Russian literature" - "The question of universities" to name

but a few) that had not been completed or unsupported by the writer have not been incorporated in the 30-volume collection of excerpts from the editorial commentary and links, writers` articles and drafts. In AP Chekhov`s works, the editorial notes in the collection of excellent works have also been considered as a corpus body unit. This dictionary-concordance is taken into account as the main source for F. D. Dostoevsky`s workbook and frequency dictionary. All stages of writing the text were completed. Besides, its main unit by way of wordy and lemma, the frequency of words written by the author in his work as well as in citation was fully defined. The common nouns as word combinations have been conjoined as one unit (The journal "Russian Herald" is given as a name; search also applies to this procedure), which helps to determine the frequency of reference to the work of an author. Another advantage of dictionary-concordance is the ability to identify the word context and its address. The user can easily view the word phrase given to search in the context of dozens of words from the right and left sides, and quickly select the original text (the body of the 30-volume work based on the corps). This allows you to compare the text with electronic text. The peculiarity of the casing can be arranged as follows:

1. The primal word in the main catalogue is given in the capital letter (ex: the word "*Белинский*" in the shape of "*БЕЛИНСКИЙ*"), but the letter shape is insignificant when searching for words.

2. For technical reasons, the italic words in the original source of the text are neutralized.

3. Lexical homonym is provided not in different dictionary articles, but in one simple article (the user can easily identify the homonym depending on the context).

4. The complex binder and the phrase are separated.

5. Abbreviations, abbreviations, numbers are not outlined in word.

6. A frequency of use, abbreviated magazine and name of the newspaper has been interpreted and edited (ex*: Совр. - "Современник", Вест.Европы- "Вестник Европы"*)

7. Words with hyphens are normalized *без-дар-ного – "бездарный".*

8. The spelling of the words has not been changed (the period of the author's epoch is preserved: *"парке", "Толоза"* lemmas (modern*"паркет","Тулуза"*)).

9. Foreign words (unmanaged, elsewhere) have not been embedded (ex: Fr. n'est), and the word markup does not exist.

A. Griboyedov's dictionary of language is the first full-text lexicographical description of the linguistic character of the great writer's works. The dictionary contains all the published works of the lexical composition, the main part of which consists of alphabetical and frequency concordances. The concordance contains more than 12,000 dictionary articles (150000 words), and the dictionary contains a lexeme (main word), its grammatical character (word of mouth, rod, hint of live / lifelessness), the quantitative index in all texts, and a brief explanation. Each word is expressed in context and reflects a link address to the source (chapter, department, event, case, etc.). The text of the work "The Woes of Wit" along with the dictionary concordance are composed of an additional alphabetical index, frequency dictionary, the frequency dictionary of prosaic and poetry, inverse and grammatical dictionary. Additional dictionaries are part of the basic glossary being component of the different layout (basic form, frequency, grammar group). They are only made up of lexemes, and the utilization word is not included. This dictionary is intended for all linguists dealing with the history of the author's

language, Russian literature and the history of the literary language. Due to its excellent size, scale and structure extensive philological research: a specific author or linguistic character of the period, comparative analysis of linguistic features of various works, the use of words in the Russian language could be widely used in the history of the Russian language and the study of the grammar form of certain words.

The concept and methodology of AS Griboyedov's dictionary. The dictionary of the author's work should provide users with the full and comprehensive information about a particular author language (what word or phrase should be used in which publication work, in what grammatical form, how often the frequency they use). Like other copyrighted dictionaries, the body corpus includes all published works of the author. The process of creating the dictionary is characterized by its excellent perfection, based on a new methodology and technology. Because the dictionary of compilations of the dictionary lies linguistic examination, formed on the basis of perfect structure, philological and grammatical markup body. A markup setting body is a reliable source of subjective objectivity and illustratively. Corps of A.Griboyedov's works are marked up with the layout in the process of working on the project "Fundamental Electronic Library: Russian Literature and Folklore". The main feature of this body is its full and high level, deep philological training. The texts are customized in the text of the text (heading, notes, replication, line, page parameters) that connects all texts to the text address. The grammatical markup of dictionary is based on the A.A. Zakiznyak dictionary (7), which uses every word in the body of the word lexeme (dictionary form), grammatical sign of lexeme (word sequence, genus, species, live / lifelessness), grammatical sign of word- shape, (number, preposition shape, verbal conjugation, time). The first markup scale was carried out by a company of "Yandex", the second one was manually executed: ammonia has been solved, the situations computer "misunderstood" were marked up again and errors were reconfigured, the difference between current Russian literary language and the language features of the Griboyedov period was shiny highlighted. Different forms of markup-outlined corpus dictionary have laid the basis for creating separate dictionaries of contexts like the concordance, frequency, inverse, grammatical, speech, text, and personality types to name but a few, holding each within itself.

Due to the fact that Griboedov's works had not been perfect (with the exception of contemporaries of that time Pushkin or Chekhov's works), the selection of text for the body was deemed to be one of the remarkable processes, and the most authoritative editions of the author were taken as a body corps unit. A similar aspect of this particular corps with its analogs is that as the published dictionary, the amount and content of the information would not be limited. The basis of the dictionary is the alphanumeric-frequency concordance of author's texts. The lexemes were proposed in alphabetical order, and the words were given in the grammatical form. As a rule, it is possible to create a dictionary of concordance based on different types: alphabets, frequencies (placed in order of lowering usage frequency of words), inverse (based on opposite word alphabet), and grammatical dictionary (lexemes were arranged on a grammatical basis). All of these features are available in an electronic version that meets all the requirements of the glossary for convenience.

In contrast to other corpora of authorship, however, it should be pointed out that the text was structured in the description of the text as structural elements such as headline, text remark, and footnote. The name grouping of proper nouns may be separated from each other, including the name, surname, patronymic, individual pseudonyms, geographical name, separate grouping of the origin or derived attributes (even written and capitalized in contrast to the current orthography), the name of the work and the separation of the publications *(Бахчисарайский, Гюлистанский,*

Карабахский). Because of the barrier to scale growth, in the paper version of the dictionary, proper nouns are not included in the lexical content, but they are available in the electronic version of the dictionary.

The purpose of the corpus is research and illustrative, and the copyright corpus can be created for both targets. The corpus of the investigative body creates his own body to solve a particular linguistic problem and, in this regard, draws conclusions for his research. The author of the illusory corpus prefers the body to be widely used by the users to carry out various activities (teaching assignments, experiments on the body, use of linguistic background in translation and textual editing). That is to say, the author's corps is a *researching* and *illustrative* body corpus.

Depending on the stability of the case, the fixed and dynamic type is differentiated, the corrugated body is a solid body, full and full-text body according to the fragment view were ranked in the list. It does not require a permanent update, as it is in the case of a national building and once created it does not need to be repaired again and again. The reason for the recognition of the full-text body is that the author's body is entirely covered by the creator's creativity and that all his creative heritage (full text of the work) is included. The corpus can vary according to the discretion of the creator: written, parallel to one's speech, combined with the textual specificity of the text (since the author's entire genre is included), and according to the input method: payment, free and closed. The description is characterized by morphological, syntactic and semantic, full text, text size, chronologically; diachronic is regarded as a single and commonly-respected body according to its totality (number of builders).

It is worthy of attention that an important part of the copyright is the interface, regardless of the magnitude and magnitude of the accuracy. A perfectly designed interface makes it easy for the user to understand the body's accuracy and performance.

Besides, another example of vital importance's in the correction of the author's body is the degree of correction and its type. The ability to generate a linguistic markup setting of the author's works can be the basis for the creation of an ideographical dictionary if it is possible to create a frequency and inverse dictionary based on the linguistic mapping and parsing.

In addition, the authorship corpse is closely linked to author's lexicography. Therefore, development level of the author's lexicography is a bilateral relationship with corpus linguistics. The up growth of one cannot be seen without the other. Unfortunately, these two areas of linguistics are a slow-moving branch in the world linguistics. While much has been done in the field of copyright lexicography, there is a multitude of quite considerable issues waiting for its solution. The development of the copyright corpus is in a stagnant position the same as that. The reason is that the body capacity expansion does not affect elaborateness of the author's corporeality. The existing copyright corps (European Languages Corps) are staying in very primitive forms. However, on the basis of a well-established and deeply-arranged copy of the book, it can open up new opportunities for lexicology, lexicography, language history, dialectology, sociology and other fields. With the help of the author's corpse, it is possible to create a concordance database of a particular writer or poet inside a national corps. The achievements of Russian linguistics in this regard are ambitious. Corps of great writers and poets such as A.P. Chekhov, A. Pushkin, F. Dostoevsky, the Shakespeare of the British National Corpus, the Goethe Corps of the University of Mangeymus, and the Honorable representatives of Corps of Persian-Tajik Classical and Fiction Literature serves as an example for further development. Having come to terms with the principles of creation of the above-mentioned copyright corpus, we have come to the conclusion that they should pay attention to the following as their

specific, different and common aspect. Creating various types of alphabetic, alphanumeric, inverse, semantic-grammatical dictionaries; development of a separate methodology and concept of corpus creation; choosing source for the body and dictionary; finding signs of corpus arc, dictionary article structure, descriptive characteristic of the lexeme (general word, brief description, grammatical character, frequency); grammatical phrase of speech, character set usage (context) are the main issues of creating a copyright corpus.

CONCLUSIONS

Last but not least, it is grandiosely significant to select a material for the body (whether the writer's texts are taken or not) as the mark of each author's corpse. The parameters such as the structure of the dictionary article (the dictionary of the trustworthiness and the equality of opportunity in a dictionary article or separately), body interface the manager of the corpus (the manager can have different programs in each case) are unique.

REFERENCES

1. www.ruslang.ru/seminar_aut_lexocorg020413 *сайтидаги "Теория и практика авторской лексикографии" семинар материаллари.*

2. *М.А.Бобунова. Конкорданс XXI В.: новая старая форма // Вопр. лексикографии. 2016. №2 (10) С.41-54.*

3. *К.И.Якубовский, К.А. Якубовская. Обзор современных лингвистических технологий и систем // Вестник МГУП. 2015. №2 С.315-319.*

4. *А.И.Столяров Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. 2017. № 2 [Электронный ресурс]. URL: http://human.snauka.ru/2017/02/21074 (дата обращения: 29.09.2017).*

5. *Кукушкина О.В. и др. Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова с электронным приложением. М., 2012.*

6. *"Вып. 1. Поэзия и драматургия А.С.Пушкина. Путеводитель по Пушкину". М., 2006.*

7. *Поляков А.Е. Словарь языка А.С. Грибоедова. М., "Языки славянской культуры", 2008.*

8. *Зализняк А.А. Грамматический словарь русского языка. М.: Русский язык, 1980. – 880 с.*