UDC 81'32

THE INFLUENCE OF QUANTITATIVE AND QUALITATIVE FEATURES ON STATISTICAL INDICATOR OF MATERIALITY/IMMATERIALITY OF DISCREPANCIES IN THE NUMBER OF VERBS IN FREQUENCY DICTIONARIES OF DIFFERENT TYPES

ВПЛИВ КІЛЬКІСНИХ ТА ЯКІСНИХ ОЗНАК НА СТАТИСТИЧНИЙ ПОКАЗНИК МАТЕРІАЛЬНОСТІ/НЕМАТЕРІАЛЬНОСТІ РОЗБІЖНОСТЕЙ У КІЛЬКОСТІ ДІЄСЛІВ У ЧАСТОТНИХ СЛОВНИКАХ РІЗНИХ ТИПІВ

Shapa L.N.,
PhD, Associate Professor,
Associate Professor of the Foreign Language Department
Odessa National Polytechnic University

Tomenko M.G.,
Senior Lecturer of the Foreign Language Department
Odessa National Polytechnic University

Lebedeva E.V.,
Senior Lecturer of the Foreign Language Department
Odessa National Polytechnic University

Gvozd O.V.,
Senior Lecturer of the Foreign Language Department
Odessa National Polytechnic University

The article deals with one of the phenomena of statistical lexicography – the materiality/immateriality indication of discrepancy in the numbers of frequency dictionary units in the course of their (dictionaries) comparing. In the presented research a verb is considered. To describe it seven frequency dictionaries dissimilar both in quantitative (different sizes of text corpora they are based on, different threshold frequencies, etc.) and qualitative (various fields of technical and humanitarian discourses, different contents – only terms or all units occurred in a text corpus, etc.) characteristics have been used. The extracted verb frequency lists are compared at several threshold frequencies (from 1 to 22). Matching of various parameters both statistical and linguistic allows to obtain the data, which enable us to come to the following conclusions: in statistical analysis if it concerns one of the products of social activity of a man (in the given case it is the language) the statistical results are completely influenced by social conditions, i.e. the qualitative (humanitarian) characteristics.

Key words: frequency dictionary, matrix, threshold frequency, text corpus, style.

У статті розглядається одне з явищ статистичної лексикографії – показник істотності/неістотності розбіжностей у кількості деяких одиниць частотних словників у процесі їх (словників) порівняння. У представленому дослідженні розглядається дієслово. Щоб описати його, були використані сім частотних словників, які відрізняються один від одного як кількісними (різні обсяги текстових корпусів, на яких вони базуються, різні порогові частоти тощо), так і якісними (різні сфери технічного та гуманітарного дискурсів, різний зміст (лише терміни чи всі одиниці, які функціонують у текстовому корпусі) тощо) характеристиками. Витягнуті частотні списки дієслів порівнюються на декількох порогових частотах (від 1 до 22). Зіставлення різних параметрів, як статистичних, так і лінгвістичних, сприяє отриманню даних, які дають нам змогу зробити такі висновки: у статистичному аналізі, якщо це стосується одного з продуктів соціальної діяльності людини (у цьому випадку мови), на статистичні результати повністю впливають соціальні умови, тобто якісні (гуманітарні) характеристики.

Ключові слова: матриця, порогова частота, текстовий корпус, стиль, частотний словник.

В статье рассматривается одно из явлений статистической лексикографии – показатель существенности/несущественности расхождений в количестве некоторых единиц частотных словарей в процессе их (словарей) сравнения. В представленном исследовании рассматривается глагол. Чтобы описать его, были использованы семь частотных словарей, отличающихся друг от друга как количественными (разные объемы текстовых корпусов, на которых они основаны, разные пороговые частоты и так далее), так и качественными (различные области технического и гуманитарного дискурсов, разное содержание (только термины или все единицы, встречающиеся в текстовом корпусе) и так далее) характеристиками. Извлеченные частотные списки глаголов сравниваются на нескольких пороговых частотах (от 1 до 22). Сопоставление различных параметров (как статистических, так и лингвистических) дает возможность получить данные, которые позволяют нам сделать следующие выводы: в статистическом анализе, если это касается одного из продуктов социальной деятельности человека (в данном случае языка), на статистические результаты полностью влияют социальные условия, то есть качественные (гуманитарные) характеристики.

Ключевые слова: матрица, пороговая частота, текстовый корпус, стиль, частотный словарь.

Linguistic statistics and its section – statistical lexicography has developed in the process of corpus linguistics formation. Wherein statistical lexicography and its resources can be considered as a product of quantitative analysis of a text corpus [1−6]. Thus, a frequency dictionary reflects the statistical structure of texts and can be called a probabilistic-statistical model of a definite field [1].

In order to accurately determine the quantitative regularities of unit, functioning in a probabilistic-statistical model or frequency dictionary, it is necessary to characterize its statistical features as detailed as possible. One of such features is materiality/immateriality of the discrepancies in the number of a unit, occurring in text corpora and, consequently, in probabilistic- statistical models (frequency dictionaries).

This feature is presented in the article as a subject of description. As a unit, whose statistical indicator of materiality/immateriality of the discrepancies in the number is researched, a verb has been chosen.

The point is this feature can be described only in the course of comparing a few frequency dictionaries. According to the assertion of the famous statistical linguists the most reliable statistical data can be obtained in the case if: 1) not one, but several fragments, i.e. corpora that relate to a particular or another genre, discourse or area of discourse should be processed, whereas one sample is too insignificant part of the general population; 2) comparison with the text corpora that are opposite in content or style [1−5].

But the authors have come to the conclusions that observing these conditions is not enough for achieving the goal of the described research. The most productive way of determining the influence of quantitative and qualitative characteristics on the statistical indicator of this type is not only to select the frequency dictionaries belonging to different genres but also the ones possessing the dissimilar quantitative and qualitative characteristics.

The material of the research is as follows. The whole number of dictionaries is seven. These are their quantitative and qualitative characteristics.

First, three dictionaries on “Electronics”. Although they are referred to the common scientific and technical discourse and belong to the same field their characteristics are different:

1) frequency dictionary of the field “Electronics” formed by the authors. The size of the text corpus is 200 000 word tokens, the threshold frequency (minimum frequency) is 1; in the article it is marked as FDau;

2) frequency dictionary of the field “Electronics”. The size of the text corpus is 300 000 word tokens, the threshold frequency is 1; it is marked as FDub;

3) frequency dictionary on “Electronics”. It includes only terminological units. The size of the text corpus is 200 000 word tokens, the threshold frequency of the dictionary is 2; it is marked as FDal.

Secondly, there are two dictionaries that belong to the fields of scientific discourse as well, but which in their subject matter are not associated with “Electronics” absolutely:

4) frequency dictionary on metallurgy. Its size is 200 000 words, the threshold frequency of the dictionary is 22; in the article it is met as FDmet;

5) frequency dictionary of mathematics. The size is 200 000 words, threshold is 2; it is marked as FDmath;

Thirdly, two dictionaries that were formed on the basis of texts describing humanitarian fields: the field of scientific discourse – psychology, and newspaper and journalistic discourse.

6) frequency dictionary-minimum of terms on psychology, its size is 240 000 word tokens, minimum frequency is 2, the mark in the article is FDps;

7) frequency dictionary of newspaper lexis, it was compiled on the basis of the newspaper articles survey in the Great Britain and the USA: Daily Express, Daily Herald, Daily Mail, Daily Mirror, The Daily Telegraph and Morning Post, News of the World, Sunday Mirror, The Times, New York Herald Tribune, The Sunday Times, The New York Times, US News and World Report.

The size of the text corpus is 200 000 word tokens, the threshold frequency is 4, the mark in the article is FDnews;

As one can see they differ significantly both in their statistical parameters and contents. Thus, the sizes of the five frequency dictionaries is 200 000 word tokens, while two dictionaries – one of the dictionaries on “Electronics” and on psychology – exceed 200 000 words that, however, is not an obstacle for joint consideration of all frequency dictionaries as in the process of statistical analysis, relative values are used and the relative error is calculated. In addition, according to the threshold frequency parameter the dictionaries on “Electronics” – formed by the authors (FDau) and having 300 thousand units (FDub) – are complete, as all units of text corpora are given in them. All other frequency dictionaries are incomplete, as the minimum frequency is above 1. Thus, the frequency lists have differences in the sizes of text corpora and the threshold frequencies. In content they also have significant differences: five dictionaries contain all the lexical units found in text corpus; two dictionaries include only the terms.

The selection of such dissimilar frequency dictionaries allows to achieve the main goal of the arti-

cle – to analyze how the quantitative and qualitative characteristics are able to influence on the statistical parameter of materiality/immateriality of discrepancies in the number of any single unit functioning in these dictionaries. As we have already mentioned a verb was chosen as a unit to be researched.

Of all mentioned frequency dictionaries, the verbal units have been singled out and the corresponding frequency lists formed. Some general statistical parameters of verbs are presented in table 1.

Table 1

Frequency dictionary	Size of text corpus	Threshold frequency	Total verb frequency	Percentage of total verb frequency to the size of text corpus, %
FDau	200 000	1	25788	12.89
FDub	300 00	1	19485	6.49
FDal	200 000	2	12320	6.16
FDmel	200 000	22	17508	9.4
FDmath	200 000	2	8829	4.41
FDps	240 000	2	14678	6.11
FDnews	200 000	4	17549	8.77

The distribution of verbs around the threshold frequencies in each frequency dictionary is demonstrated in table 2.

Table 2

Threshold frequency	The number of verbs in frequency dictionaries						
	FDau	FDub	FDal	FDmet	FDmath	FDps	FDnews
1	1558	941	-	-	-	-	-
2	1056	880	564	-	139	427	-
4	726	634	425	-	130	409	772
22	268	212	146	234	65	202	188

The consideration of the analyzed frequency lists shows that they differ among themselves by the total number of verbs presented in them. Here we can observe the following paradoxical fact. There are discrepancies in the number of verbs not only in frequency dictionaries compiled on the basis of different types of discourse and different areas of the same discourse, which is natural. But, more importantly, dictionaries, which belong to one area of knowledge, differ significantly in the parameter of frequency of the verbs functioning in them. In our case – dictionaries of the technical field "Electronics" FDau and FDub. For example, the number of verbs in FDub is much less (almost twice) than in FDau, although the size is one and a half times larger.

Such a discrepancy in the quantity between the dictionary units possessing the same threshold frequency and functioning in the same discourse area is very significant and requires additional analysis, since this fact can not be explained by the heterogeneity of the text corpora of the compared frequency dictionaries, i.e. the influence of qualitative characteristics. For such minimal frequencies there are certain regularities concerning the distribution of words with frequency 1, which exist in all frequency dictionaries. In accordance with these regularities, the fraction of vocabulary units with frequency 1 is about 50% of the total number of words in the text corpus [1−3; 7]. These data are confirmed not only in the dictionary formed by the authors, but also some other frequency dictionaries. Thus, in this case it can be assumed that such a difference in the quantity of verbs with F=1 is connected with violation of the mentioned regularities in FDub.

If we continue analyzing the statistics of verbs in these two dictionaries, then we can not fail to note the greater uniformity of the distribution of verbs over the other threshold frequencies in FDub as compared to FDau. In FDub there are no such abrupt jumps in frequencies as in FDau. Perhaps, this is the result of fairly low values in the number of verbs with frequency 1 in FDub.

To calculate the materiality/immateriality indicator of discrepancies in the number of verbs in the selected dictionaries, all frequency dictionaries are matched in pairs. But one can only compare pairs between dictionaries with the same frequency threshold, which ends the dictionary. So we are pairing all the dictionaries in which the minimum frequency is the same.

For each group of frequency dictionaries compared in pairs, we build a matrix of the relative difference between the number of verbs in the first and second dictionaries of each pair (in percentage terms). The relative difference is calculated by the following formula

$$E = \frac{(n_1 - n_2) \cdot 100}{n_1}, \qquad (1)$$

where E – is the relative difference; n_1 – the number of verbs in the first frequency dictionary of the pair; n_2 – the number of verbs in the second frequency dictionary of the compared pair (table 3, 4, 5).

We accept the assumption that E = 30% and above is an indicator of a discrepancy materiality in data.

So we are going to analyze each matrix from the viewpoint of materiality/immateriality of discrepancies in the number of verbs in the compared frequency dictionaries at some given frequency threshold.

A comparative analysis of verbs in dictionaries with threshold frequency 1 and higher is limited to two dictionaries – FDau and FDub, so there is no need to build any matrix. The relative difference in the number of verbs between FDau and FDub based on the data in Table 2 shows the materiality of the discrepancy between these dictionaries for the studied parameter.

Such value E is expected here, since as already mentioned, there is a significant difference in the quantities of verbs functioning in dictionaries FDau and FDub with threshold frequency 1 because of some violation of the regularities relating to the number of words with frequency 1.

The next the verbs with frequency 2 and higher are considered. Pairs of dictionaries in the matrix are arranged horizontally and vertically, but the description is made from the viewpoint of horizontally arranged dictionaries. The calculated values of materiality/immateriality of the discrepancies shows the number of pairs. For clarity of presentation, all the values of the number of verbs, which do not exceed the specified E = 30%, will be highlighted in the tables in bold.

Table 3

	FDub	FDal	FDmath	FDps
FDau	**16.7**	46.6	86.8	59.6
FDub	- -	35.9	81.2	51.5
FDal		- -	75.4	**24.3**
FDmath			- -	67.4
FDps				-

Table 3 shows the increase in the threshold frequency of verbs to 2, which leads to the increase in quantity of the compared dictionaries themselves. Now there are 5 of them including FDal, FDmath and FDps.

So we see the dictionaries, which are distinguished by the materiality of the discrepancies in the number of verbs, and also those for which the immateriality of differences is fixed. We are tracing first how the dictionaries are grouped according to the immateriality of discrepancies.

The pair FDub and FDau shows the immateriality of discrepancies in the number of verbs with frequency 2 since the difference in the number of verbs in FDub and FDau becomes much smaller (880 and 1056, respectively) as compared to the values at frequency 1. This could be expected, since the texts of both dictionaries are similar in the field of knowledge and this qualitative feature influences the statistical index.

In another group in accordance with the discrepancy materiality/immateriality indicators FDps and FDal are included, in spite of the fact that these dictionaries refer to different areas of knowledge, and here one could expect the essential differences in the number of verbs. However these two dictionaries are terminological in nature of the units, and this probably allows them to be united in one group.

As can be seen from table 3 significant discrepancies are observed between the frequency dictionaries of sublanguages both different and similar in their subject matter. First, this group includes one of the dictionaries on electronics and an explanation of this can be found in the fact that FDal is a terminological dictionary unlike FDau and FDub. This fact determines the number of verbs in FDal (table 2) since it is known that the terms have a nominal, but not a verbal character.

Among the dictionaries noticeably stands out FDmath, for which the materiality of discrepancies with all other dictionaries is fixed. If compare to dictionaries FDau and FDub, specializing in the field of electronics, which show a sharp jump in the magnitude of the discrepancies at threshold frequency 2, making it immaterial, then with the number of verbs in the dictionary FDmath this does not happen. It is supposed that in the low-frequency zone of dictionaries, mainly terminological units, which commonly belong to the class of nouns, are concentrated. In accordance with P. M. Alekseev`s assertion [2], the more terminological a word is (words referred to terms describing the basic concepts of mathematics), the lower frequency it has.

Thus it is obvious that the functioning of units in frequency dictionaries is influenced not only by quantitative, but also by qualitative characteristics.

The calculation results of relative difference between the number of verbs with threshold frequency 4 and higher that function in the frequency dictionaries in question are represented in the matrix of table 4.

Table 4

	FDub	FDal	FDmath	FDps	FDnews
FDau	**12.7**	41.5	82.1	43.7	**6.3**
FDub	- -	33.0	79.5	35.5	**21.8**
FDal		- -	69.4	**3.8**	44.9
FDmath			- -	68.2	83.3
FDps				- -	47.02
FDnews					-

In the obtained matrix six dictionaries are compared since a new one is added – FDnews.

And again we are distinguishing two subgroups, one of them is characterized by the immateriality of

discrepancies, the other by the materiality of differences in the number of verbs.

First of all, we consider a new element of the matrix – FDnews. In investigated parameter FDnews is combines with FDau and FDub. At first glance it may seem unnatural, since these dictionaries refer to different types of discourse, therefore, we should expect the materiality of discrepancies in the number of verbs. However, it seems that the qualitative characteristics of these two discourses, differing in their subject matter but having an equally high dynamics, play their part here. Dynamic nature of the texts, which, in turn, reflects the dynamic one of this area of technical discourse, describing new results of research, the search for new discoveries and making them full of actions and therefore – verbs, is inherent to the texts on electronics. The same characteristics can be given to the newspaper and journalistic texts.

As in the previous matrix, there is also the immateriality of discrepancies in the number of verbs in the pair FDps and FDal. However, one can not but notice a significant difference in the values of indicators of the materiality/immateriality of discrepancies of verbs with frequency 4 compared to the verbs with frequency 2 in FDps and FDal, the value has increased more than six time – 3.8% against 24.3%.

On the other hand, FDnews shows a materiality of discrepancy with the dictionaries FDal, FDps and FDmath. This happens, obviously, due to the following facts: FDnews and FDal belong to different types of discourse, dictionaries FDnews and FDps, although they refer to the humanitarian discourse, but FDps is compiled on the basis of scientific texts as opposed to FDnews. In addition, FDal and FDps compared to FDnews dictionaries, are terminological and the number of verbs in them should be less than in FDnews for the reason that we mentioned above.

As for FDmath this dictionary has its own peculiarities, which differ it from both terminological and nonterminological dictionaries. For FDmath with this threshold of the frequency (4) of using verbs, as well as at the threshold frequency 2, establishing discrepancies with the number of verbs of all frequency lists without exceptions, the materiality value of discrepancies does not practically decrease even with increasing the threshold frequency value.

Thus, a comparative analysis of the number of verbs with threshold frequency 4 and above, functioning in six frequency dictionaries, shows that for pairs of FDnews with two dictionaries FDau and FDub the discrepancies are immaterial, as well as for the pair FDps and FDal. In all other dictionaries compared in pairs the substantial value of materiality of discrepancy in the number of verbs is revealed.

And finally we are describing the analysis results of the matrix of verbs with threshold frequency 22 and above. It is presented in table 5.

Table 5

	FDub	FDal	FDmet	FDmath	FDps	FDnews
FDau	**20.9**	45.5	**12.7**	75.7	**24.6**	**29.9**
FDub	- -	31.1	**10.4**	69.3	**4.7**	**11.3**
FDal		- -	60.3	55.5	**27.7**	**28.08**
FDmet			- -	72.2	**15.8**	**24.5**
FDmath				- -	67.8	65.4
FDps					- -	**7.4**
FDnews						- -

Table 5 demonstrates the matrix which contains FDmet for the analysis as well. So now we are comparing all seven dictionaries in pairs. Twelve (majority) of the compared pairs are characterized by the immateriality of discrepancy in the number of verbs, and 9 (minority) – by the materiality of discrepancies.

The comparison of these data with those that were obtained for dictionaries with threshold frequencies 2 and 4 are represented in the form of a table.

Table 6

Threshold frequency	The entire number of FDs pairs compared	The number of FDs pairs, in which the discrepancy is negligible	Percentage
2	10 pairs	2 pairs	20
4	15 pairs	4 pairs	26.7
22	21 pairs	12 pairs	52.4

From table 6 it follows that the compared pairs of frequency dictionaries, in which there are immateriality of discrepancies in the number of verbs at threshold frequency 2, constitute the smallest part of the total number of comparable pairs, somewhat larger – in dictionaries with threshold frequency 4 of verbs, and more than a half – in dictionaries at the frequency of verbs 22 and above.

If we consider the relationship of these data with the frequency of verbs in the compared dictionaries (based on the hypothesis that dictionaries are similar due to high-frequency words and differ due to words with low frequency), then it is obvious that these two characteristics are interrelated, i.e. with increasing the frequency of verbs in the analyzed dictionaries, the difference among them in terms of the number of verbs decreases.

However, we are to return to the analysis of table 5 which contains a new element – FDmet. First of all we are grouping the pairs of dictionaries on the basis of materiality/immateriality of discrepancies in

the number of verbs. Then we analyze the values of this parameter in those dictionaries, for which a change in the relative difference values in comparison with the previous matrices (table 3, 4) occurs at a given threshold frequency of verbs.

Consider first, as before, a group that includes pairs of dictionaries with the immateriality of discrepancies in the number of verbs.

We begin the analysis from the dictionary forming the greatest number of pairs – FDnews. FDnews has the highest indicator, pairing with five dictionaries – FDau, FDub, FDal, FDmet and FDps. This is more than twice as many pairs as in the matrix with threshold frequency 4. It is obvious that for FDnews as a frequency dictionary that challenges various sides of societies it is quite natural.

In terms of their values there are two highest indicators of the immateriality of discrepancies in FDnews – with FDub (11.3) and FDps (7.4). FDub seems to have overcome its limitations in the number of verbs by which it is characterized at threshold frequencies 2, and already at frequency 4 shows the immateriality of discrepancies but with a small value. And at frequency 22 this value has almost doubled.

The pair FDnews and FDps shows the immateriality of discrepancies that is not observed at threshold frequency 4. Frequency 22 enables to find out the domain of verbs-terms in the dictionary of terms FDps.

The values of this indicator in FDnews in the pairs with the remaining three vocabulary are quite low. Nevertheless, as an example, we will try to explain such an unusual connection as FDnews and FDal. At frequency 4 in FDnews and FDal the indicator of materiality/immateriality of discrepancies is above 30%, i.e. with FDal there is the materiality of discrepancies. But as we mentioned, not only quantitative but also qualitative characteristics should be taken into account, especially when it concerns such dynamically developing social phenomenon (and hence a linguistic phenomenon) as a field of scientific technical discourse "Electronics" and constantly changing newspaper and journalistic discourse.

The second place as to the indicator of immateriality of discrepancies in paring with other dictionaries is occupied by FDps. It shows the immateriality of discrepancies in the pairs with four frequency dictionaries – FDau, FDub, FDal and FDmet. The analysis of the matrix demonstrates that for FDps the immateriality of discrepancies is fixed with all dictionaries, with which at threshold frequency 4 it has the materiality of discrepancies (FDau and FDub), as well as with a new element of the matrix – FDmet. The belonging of FDps to scientific discourse allows it to unite with these frequency dictionaries on the basis of immateriality of discrepancy in the number of verbs at sufficiently high threshold frequencies, although it refers to a field that in its subject is incompatible with electronics and metallurgy: in FDps there are two fairly high values of the indicator of the discrepancy immateriality – in pairs with FDub (4,7) and FDmet (15,8).

On the third place with respect to the number of pairs with which it has the discrepancy in immateriality indicators is FDmet. It is connected to the frequency dictionaries which both refer to the field of "Electronics" – FDau and FDub. And the values of the indicator are quite large – 12,7 and 10,4, respectively. Some inventions from the field of electronics are apparently used in metallurgical processes, and this gives such values.

Due to the fact that FDmet is not analyzed at frequencies 1, 2, 4, we are unable to follow a certain tendency to increasing or decreasing the discrepancy materiality/immateriality indicator here.

The next we consider a group of dictionary pairs characterized by the materiality of discrepancies in the number of verbs.

On the first place in this group is FDmath. As in the previous matrices with threshold frequencies 2 and 4, in the matrix with frequency 22 FDmath has the materiality of discrepancies with absolutely all frequency dictionaries both for vertical pairs and horizontal ones. We can conclude that the bulk of verbs in the texts of mathematics is used with the higher frequency than 22.

If to analyze the values of the indicators of materiality of discrepancies in the number of verbs for all pairs, in which FDmath is a participant, one can see that the values of the indices in practically all threshold frequencies are in the range 55.5% to 84%, and it is in the matrix with frequency 22 that the minimal value 55.5% is fixed.

The next dictionary we consider is FDal. It demonstrates the discrepancy materiality with FDau and FDub, which is referred to a similar field of knowledge. But here the presence of the materiality of discrepancies is quite understandable, because FDal is a dictionary of terms, while FDau and FDub include all units of texts. The stability of this trend with regard to the dictionaries of similar field of discourse, which we observe at frequencies 2 and 4, could be overcome at threshold frequency 22. However, the terminological nature of the content of dictionary FDal again influences on the number of verbs, even at such a significant threshold frequency. The magnitude of the discrepancies as compared to threshold frequencies 2 and 4 remains practically unchanged in the pairs of FDal with FDau and FDub.

As to the immateriality of discrepancies between the FDmet and FDal it can be explained by the quali-

tative characteristics: belonging to absolutely different fields of technical discourse and the fact that FDal is a dictionary of terms unlike FDmet.

On the basis of the foregoing we can draw the following conclusions.

1. The quantitative values of units in frequency dictionaries of any type (terminological or containing all units of text) are influenced by qualitative characteristics: the content of a frequency dictionary, attribution to various types of discourse or a different fields of the same discourse, dynamics of development of a particular discourse area.

2. The indicator of materiality/immateriality of discrepancies in the number of verbs in frequency dictionaries, as well as the values of this parameter, is directly dependent both on the statistical and qualitative characteristics.

The future research will be devoted to the analysis of the statistical indicator of materiality/immateriality of discrepancies in the number of verbs functioning in text corpora of different time point in order to trace the values of indicator in dynamics, fix the probable changes in them (values) and find out the causes influencing them.

REFERENCES

1. Alekseev P.M. Statistic Lexicography (Typology, compiling and applications of frequency dictionaries). L: Leningrad Pedagogical Institute, 1975. 120 p.

2. Захаров В.П. Корпусная лингвистика: учебное пособие. URL: http://vp-zakharov.narod.ru/VictorZakharov_CorpLingv.doc.

3. Köhler R., Altmann G. Quantitative Linguistics. An International Handbook. Piotrovsky R.G. (Ed.) Berlin. New-York: Walter de Gruyter. 2005. 1027 p..

4. Береснев С.Д. Исследование лексики немецких научно-технических текстов с позиции получателя речи: автореф. дис. … д-ра филол. наук. Л.: АН СССР. Ин-т языкознания. Ленингр. отд-ние, 1974. 35с.

5. Андреев Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л.: Наука, 1967. 404 с.

6. Summers D. Corpus Lexicography – The importance of representativeness in relation to frequency. 2005. URL: http://www.pearsonlongman.com/.../pdfs/corpus-lexicography.pdf.

7. Скляревич А.Н. Возможное обоснование частотных закономерностей лексики текстов / А.Н. Скляревич, Т.А. Якубайтис. Рига, 1984. 62 с.