

UDC 81'367.625:62+811.111'367.625:521.396  
DOI 10.32837/2312-3192/12/7

## STATISTICS OF LOW-FREQUENCY KERNEL (SUBORDINATING) MODELS OF THE VERBAL WORD-GROUPS IN THE TEXT CORPUS "RADIO ELECTRONICS"

Ludmila Shapa<sup>1</sup>, Maria Nevreva<sup>2</sup>, Marina Tsinovaya<sup>3</sup>

### Abstract

The paper deals with one of the most widely used types of verbal word combinations, namely kernel (subordinating) models of the verbal word-groups functioning in the "Radio electronics" text corpus, which refers to the technical and scientific discourse. The text corpus was compiled on the basis of scientific and technical journals "Radio electronics" published in the USA by the method of a continuous sampling. In order to select this kind of models a theoretical approach describing the problems of combinatorics and valance is considered. This permits to take into account both contact and distance elements, which are necessarily connected with the verb by grammatical (subordinating) bond and form a word-group with it. For compilation of the inventory of kernel models of the subordinating verbal word-groups 10698 sentences are subjected to the distributive analysis. The models occurring with the low frequency (below 180) are exclusively chosen. The total amount of low frequency model is 41 units. Except the distributive methods the statistical ones are applied in the research. The quantitative analysis of kernel models used in the text corpus with a low frequency has shown that the number of different models is much bigger as compared to the same units with a high frequency of usage – 77, 4% and 22, 6%, respectively. But the total frequencies comparison goes in the favour of high frequency models – 0, 4% to 99, 6%. Besides quantitative characteristics the simplicity/complexity parameter of the model structures are analysed. In studying the high frequency kernel models the direct dependence of frequency of usage on the model structure was observed – the simpler the structure is the higher the frequency is. But the low frequency kernel models demonstrate more complex and contradictory functioning as regard to this parameter. In most cases the tendency to indirect dependence is seen quite distinctly, i.e. the simpler the structure is the lower the frequency is, and the chaotic distribution of frequencies occurs in the frequently highest among low-frequency models.

### Keywords

Close and distant components, correlation, low frequency models, model structure, quantitative characteristics, simplicity/complexity parameter, subordinating bonds, statistical methods.

### 1. Introduction

In statistical linguistics the analysis of low-frequency units is considered to be not less important and interesting task than the study of the units included in the high-frequency list<sup>4</sup>. This point can be explained by the fact that, firstly, the units with a low frequency exceed in quantity the high-frequency ones in the terms of the same statistical parameter; secondly, dialectics of changes occurring in systems and objects (including linguistic ones) operates in such way that a phenomenon which reveals a low frequency of usage yesterday or at the moment can become extremely widely used, and on the contrary

the characteristics and concepts that were considered to be preferable for use go to the periphery of public consciousness and life. Therefore, the results of the analysis of low-frequency units can provide a basis for their further study and use by theorists and practitioners specializing in the following humanitarian fields: linguistics of texts of any type of discourse; method of foreign language teaching, which provides presentation of not only the most frequent text units, but less frequent ones as well; statistical linguistics and general statistics.

This paper aims to describe the results of the analysis of low-frequency units operating in specialized technical texts, namely, in texts on Radio electronics. The sublanguage of this field of knowledge was chosen based on the fact that Radio electronics is one of the most promising and rapidly developing industries and sciences, huge volume of information about its achievements presented in the English language.

If to try to make an excursion into the history of studies that are devoted to the analysis of this type of units (low-frequency one), it can be noted that their development is based on material of a different nature. Thus, statistical data on different parts of speech<sup>5</sup>, units, functioning in several areas of legal discourse<sup>6</sup>, word-formation components of nouns<sup>7</sup> and others are considered in detail.

However, data on the analysis of low-frequency word-combinations used in text corpora were

<sup>1</sup> L. N. Shapa, Assoc.Prof. PhD, Odessa National Polytechnic University, Tenistaya, 7, 16, 65009-Odessa, Ukraine, E-mail: shapa.od@gmail.com ORCID ID: 0000-0002-3502-4577

<sup>2</sup> M. N. Nevreva, Assoc.Prof. PhD, Odessa National Polytechnic University, Shevchenko av., 10/7, 21, Odessa-58, Ukraine, E-mail: m.n.nevreva@gmail.com ORCID ID: 0000-0001-9923-1041

<sup>3</sup> M. V. Tsinovaya, Assoc. Prof. PhD, Odessa National Polytechnic University, Profsoyuznaya, 10b, 65076-Odessa, Ukraine, E-mail: marinatsinovaya@gmail.com ORCID ID: 0000-0003-2633-8416

<sup>4</sup> Alekseev 1975; Shubik 1980; Tuldava 2007; Bogdanova 2012; Комарова 2012; Piotrovsky 2005; Summers 2005; Захаров 2005; Krishnamurthy 2006

<sup>5</sup> Shapa 1991; Borisenko 1989; Dyachenko 1984

<sup>6</sup> Trofimova 1988

<sup>7</sup> Nevreva 1986

not found in the available literature. Therefore, the undoubted novelty of this article is that its subject is verbal phrases, and specifically kernel (subordinating) models of verbal word-groups whose elements are attached to the verb with the help of the subordinate link. The goal of this article is to study the statistical characteristics of the subordinating models of verbal word-groups functioning with low frequency in text corpora of scientific discourse.

## 2. Methodology

To present a list of kernel (subordinating) models of the verbal word-groups where the verb is connected to the elements around it with the help of subordinating bond, it is necessary to describe the principle of selection of this kind of models. In this case, the problem of combinatorics and valence is touched, i.e. the connection of language elements in speech should be considered from theoretical positions.

The verb, like any other language unit, has the property of valency, i.e. selectivity in relation to its dependent elements, and in accordance with this property it is not combined with any member of the sentence. Thus, the most important problem in constructing the models of verbal word-groups of such kind is the question of the volume of the verbal environment in terms of the nature of the verbal valence, because it is necessary to know which words should and can be included in the model, and which ones should be excluded.

The description of the valence of any word is carried out by modelling its potential combinatorial properties. Consequently, a unit representing a certain valence property of a word is a pattern which reflects a certain type of bonds between classes of words and the structure of a real-life combination of words.

In this article we consider the notion of combinatorics, when grammatical valence of the verb, which denotes its ability to make these or those bonds and form combinations with other classes of words, is described with the help of model sets.

The next important issue that requires consideration when describing kernel models of verbal word-groups is the nature of the bonds within the models themselves. The linguistic literature describes in detail the approach that determines the nature of the bonds in a sentence between related words: interdependence, one-sided dependence, and constellation, on the basis of which subordinating, coordinative, and predicative relationships are established respectively.

Our description of kernel model of verbal word-groups will be limited to the analysis of exclusively subordinating links (which is implied in the article). In this case, it is necessary to give a more detailed presentation about them. Subordinating links are characterized by unevenness of their units, i.e. dependence of one of the components on the other and are divided into kernel and adjunct depending

on the status of the analyzed word – active or dependent. This thesis can be illustrated by the following sentence as an example: *Every computer contains the output/input unit*. Here you can see the kernel connection for the verb *contains*, controlling the noun *unit*, which governs the words *output/input*. And adjunct connection is for the word *every*, which depends on the noun *computer*; *unit*, which is governed by the word *contains*; as well as for the words *output/input*, which depend on the word *unit*. Thus, within the framework of the subordinating verbal word-groups, each verb is in a subordinating relations with the components of a word-group dependent on it (verb) and has a certain number of grammatical connections of the subordinating nature.

And finally, a remark that concerns the different nature of relationships in the models. By the words “different nature” we mean the separation of connections into grammatical and lexical-grammatical. The first one is concentrated on the presence of a grammatical connection between the elements of a combination, without taking into account the possibility/impossibility of being a communicative or nominative unit of a language. The second one involves the formation of a model of a complete phrase, which can be considered as a unit of communication. In this paper we have identified such verb connections that have formed exclusively complete word-groups.

The text corpus “Radio electronics”, which is a part of scientific discourse, served as the material for the analysis of kernel models of verbal word-groups. It was formed by the method of a continuous sampling. The basis for the text corpora are scientific and technical journals “Radio electronics” published in the USA.

In the process of forming the “Radio electronics” text corpus the texts of the articles, their titles and annotations to them, written and compiled by different authors, were plunged into quantitative processing. In addition, the captions under the drawings, charts and graphs were examined. Advertising messages were not subjected to the survey, the equations and formulas were not taken into account when calculating. All main verbs in any form and syntactic functions were recorded (except *to be* and *to have* and verbs with postpositions like “put up”).

For the formation of the frequency list of verbs, which in the future study should form the basis of the kernel verbal models, 25,788 verb usages were singled out of the “Radio electronics” text corpus. There were 1,558 different verbs, whose percentage in the text corpus made 12.9%. Arranging different verbs in the order of decreasing frequency we have compiled a frequency list (dictionary) of the verbs of “Radio electronics” sublanguage. From this frequency list of verbal lexemes the ones with frequency 100 or higher were taken for studying the combinatorial properties of verbs. This list includes 52 units: *use*

(F= 1291), *make* (F=518), *show* (F=500), *see* (F=339), *check* (F=324), *connect* (F=303), *get* (F=303), *operate* (F=283), *work* (F=269), *find* (F=262), *go* (F=254), *do* (F= 231), *provide* (F= 205), *apply* (F=205), *need* (F=205), *give* (F=198), *take* (F=195), *require* (F=189), *set* (F=183), *record* (F=182), *read* (F=176), *cause* (F=165), *produce* (F=160), *know* (F=158), *look* (F=157), *mount* (F=156), *increase* (F=145), *reduce* (F=143), *determine* (F=138), *start* (F=137), *tune* (F=133), *add* (F=131), *change* (129), *develop* (F=128), *measure* (F=127), *adjust* (F=123), *replace* (F=122), *want* (F=122), *build* (F=121), *control* (F=121), *flow* (F=121), *include* (F=120), *note* (F=120), *design* (F=113), *say* (F=112), *come* (F=109), *test* (F=108), *call* (F=107), *try* (F=196), *short* (F=104), *switch* (F=104).

### 3. Results and discussion

The theoretical substantiations, as well as the frequency list of verbs presented in the “Methodology” section, made it possible to form a corpus of low-frequency kernel (subordinating) models of verbal word-groups that function in the text corpus “Radio electronics”. For the formation of the inventory of compatibility models a distributive analysis of 10698 sentences was carried out, which took into account both contact and distance elements, but necessarily connected with the verb by grammatical (subordinating) bond and forming a word-group with it.

Although 53 kernel verbal models are included in the general inventory list, only the ones with frequency below 180 are selected for this study. They are assigned to low-frequency units and comprised 41 kernel verbal word-groups (models).

Below is a list of models with the corresponding examples recorded in a text corpus, in which the verb models are arranged according to the increasing of their complexity degree. The grammatical connections of the verbs with its surrounding elements were expressed in terms of classes of words, denoted by common markers. We will clarify only those that may cause a reader’s misunderstanding: D – adverb, adverbial phrases, V = – infinitive, S – verb-dependent subordinate clause, which in our study was also equated to the class of words. Models of the type Vlike / as / cnj / N inform about the possibility of choosing between allied words like / as / cnj, which is demonstrated in the examples.

1. VA – *get worse; operate satisfactory.*
2. VV= – *want to build.*
3. VVing – *work standing; start operating.*
4. VVen – *get lost.*
5. VNVing – *get the set working again; increasing the size ... it will reduce the ripple.*
6. VNVen – *find it listed.*
7. VAD – *probably get hot.*
8. VAS – *get smaller until they disappear.*
9. VDV= – *go together to make this really accurate ... thermometer.*

10. VprpVing – *start by moving; insist on receiving.*
11. Vlike/as/cnj/N – *look like Fig. 9.*
12. Vlike/as/cnj/A – *replace if necessary.*
13. Vlike/as/cnj/Ving – *challenges come when considering system design.*
14. Vlike/as/cnj/Ven – *work when plugged in.*
15. AsVenD – *as shown here.*
16. VNND – *usually give you the others.*
17. VNNV= – *give us anything to go on.*
18. VNNVing – *one of ... devices once showed the engineer one of the parameters demonstrating the necessity ...*
19. VNNS – *... and how can an independent technician stop a consumer from giving his company a bad name when the firm is not in the wrong?*
20. VNAD – *often find levels as high as 1000 mV per channel.*
21. VNAV= ... *where you’ll find the more usual switch easier to use.*
22. VNAVing – *make tests easier, saving a lot of time.*
23. VNAS – *control is set low when output drops to about 35 mV ...*
24. VNasV – *adjust the unit as follows.*
25. VNlike /as/cnj /N – *use the temperature as Y axis.*
26. VN like/as/cnj/A – *increase the value if necessary.*
27. VN like/as/cnj/D – *mount Q as before.*
28. VN like/as/cnj/Ving – *short them when soldering.*
29. VN like/as/cnj/Ven – *increase periods if required.*
30. VAprpN – *come loose during the tests.*
31. VDprpN – *differs greatly from the communications system in other countries.*
32. VS prp N – *the viewer, for a price, could see what he wanted.*
33. VprpNV= – *get inside the chip to do anything.*
34. VprpNVen – *work with lead dropped.*
35. asVenprpN – *as shown in Fig.1.*
36. VNNprpN – *give us this data in time.*
37. VNDprpN – *frequently take you into automatic areas ...*
38. VNVasprpN – *power bandwidth be determined as per JHF-A-201.*
39. VprpNprpN – *check from case to ground.*
40. VprpNasN – *look at a tone burst as a carrier.*
41. VprpNprpNVing – *read from 5 mV value to 15 mV value depending on the head.*

Furthermore we present a table with low-frequency kernel (subordinating) models of verbal word-groups, which are arranged in descending order of frequency of occurrence in the text corpus “Radio electronics”. As noted earlier, this article analyzes models starting with frequencies below 180.

The presence of this table allows us to consider such statistical characteristic as the total frequency of using low-frequency kernel verbal models. It contains

Table 1

**The frequency list of the kernel models of subordinating verbal word-groups functioning with low frequencies**

№№	Types of Models	Frequency	№№	Types of Models	Frequency
1.	VV=	160	22.	VprpNV=	4
2.	VNlike/as/cnj/N	135	23.	VprpNasN	3
3.	VA	107	24.	VNDprpN	3
4.	AsVenprpN	58	25.	V/like/as/cnj/A	3
5.	VNVing	55	26.	VNNS	2
6.	VprpNprpN	48	27.	VNAV=	2
7.	VNlike/as/cnj/Vig	33	28.	VSprpN	2
8.	Vlike/as/cnj/N	32	29.	VprpNVen	2
9.	VVing	29	30.	VAD	1
10.	VNVen	24	31.	Vlike/as/cnj/Ving	1
11.	VNlike/as/cnj/Ven	23	32.	VNNVing	1
12.	VprpVing	17	33.	VNAD	1
13.	VDV=	16	34.	VNAVing	1
14.	VVen	13	35.	VNAS	1
15.	VAS	10	36.	VNlike/as/cnj/D	1
16.	VNNV=	9	37.	VAprpN	1
17.	AsVenD	7	38.	VDprpN	1
18.	Vlike/as/cnj/Ven	6	39.	VNNprpN	1
19.	VNlike/as/cnj/A	6	40.	VNasprpN	1
20.	VNND	4	41.	VprpNprpNVing	1
21.	VNlike/as/cnj/V	4			

$F^* = 829$

829 units. This value is several times smaller than the same parameter in 12 high-frequency verbal models – 267702 units, which was calculated earlier in the process of analyzing the statistical characteristics of high-frequency models<sup>8</sup> (article of V. Dragomanov). Their correlation is, respectively, 0, 4% against 99, 6%.

However, if to refer to the values that show the quantitative usage of different models, we can see that in the low-frequency verbal word-groups (models) they are much higher than in the ones used with high frequency – 77, 4% and 22, 6%, respectively .

The next feature of the low-frequency kernel models that the authors would like to explore is simplicity/complexity of model structure, and then perform the analysis of dependence of this feature on the frequency of its usage. When describing high-frequency models it was revealed that such dependence actually exists, and the higher the frequency of usage of the model is, the simpler its structure is.

For low-frequency models we do not completely observe the direct dependence. Nevertheless, it is possible to trace some statistical regularities in the distribution of models of verbal word-groups that function in a text corpus with a low frequency.

As we see, there are two-, three-, four-, five- and six-component models in the table. Let us form a list of models in accordance with the number of compo-

nents included in them, and indicate the frequency of their usage in the texts.

1) four two-component models – VV = (F = 160), VA (F = 107), VVing (F = 29), VVen (F = 13), (F \* = 309);

2) eleven three-component models – VNVing (F = 55), Vlike / as / cnj / N (F = 32), VNVen (F = 24), VprpVing (F = 17), VDV = (F = 16), VAS (F = 10), AsVenD (F = 7), Vlike / as / cnj / Ven (F = 6), V / like / as / cnj / A (F = 3), VAD (F = 1), Vlike / as / cnj / Ving (F = 1), (F \* = 172);

3) twenty four-component models – VNlike / as / cnj / N (F = 135), asVenprpN (F = 58), VNlike / as / cnj / Ving (F = 33), VNlike / as / cnj / Ven (F = 23), VNNV = (F = 9), VNlike / as / cnj / A (F = 6), VNND (F = 4), VNlike / as / cnj / V (F = 4), VprpNV = (F = 4), VNNS (F = 2), VNAV = (F = 2), VSprpN (F = 2), VprpNVen (F = 2), VNNVing (F = 1), VNAD (F = 1), VNAVing (F = 1) VNAS (F = 1), VNlike / as / cnj / D (F = 1), VAprpN (F = 1), VDprpN (F = 1), (F \* = 291);

4) five five-component models – VprpNprpN (F = 48), VprpNasN (F = 3), VNDprpN (F = 3), VNNprpN (F = 1), VNasprpN (F = 1), (F \* = 56);

5) one six-component model – VprpNprpNVing (F = 1), (F \* = 1).

The data describing groups of models containing different numbers of components show that:

– firstly, if to create the plot of dependence of the number of models on quantity of components in them

<sup>8</sup> Shapa, Kudinova 2017

we get an almost perfect curved line with a peak in “four-component models”;

– secondly, the total frequencies (F\*) of usage in the texts of various groups of models have an almost exact inversely proportional dependence on the number of components, i.e. the more components, the lower the frequency of usage of models. The exception is the group with four-component models. However, it can be assumed that since this is the largest group, quantitative values have played a role here.

The next stage of the statistical study will be the conditional division of the entire list (the table) with the low-frequency models into frequency zones:

- an upper zone, which includes the most frequent models – from 160 to 50;
- a middle frequency zone – from 49 to 10;
- a low frequency zone – from 9 to 1.

Five models function in the upper (the most frequent) zone, 10 models – in the middle frequency one, and 26 models are in the low frequency zone. As we can see, there is also a strict dependence of the number of models on the frequency of their usage, i.e. the lower the frequency of the usage is, the higher number of models is included in a particular area of the list.

Now let us see how various groups of models (with different numbers of components in them) are distributed over conditional frequency zones of the list:

upper frequency zone – models VV = (F = 160), VNlike / as / cnj / N (F = 135), VA (F = 107), asVenprpN (F = 58), VNVing (F = 55);

middle frequency zone – models VprpNprpN (F = 48), VNlike / as / cnj / Ving (F = 33), Vlike / as / cnj / N (F = 32), VVing (F = 29), VNVen (F = 24), VNlike / as / cnj / Ven (F = 23), VprpVing (F = 17), VDV = (F = 16), VVen (F = 13), VAS (F = 10);

the low frequency zone – models VNNV = (F = 9), asVenD (F = 7), Vlike / as / cnj / Ven (F = 6), VNlike / as / cnj / A (F = 6), VNND (F = 4), VNlike / as / cnj / V (F = 4), VprpNasN (F = 3), VNDprpN (F = 3), V / like / as / cnj / A (F = 3), VNNS (F = 2), VNAV = (F = 2), VSprpN (F = 2), VprpNVen (F = 2), VAD (F = 1), Vlike / as / cnj / Ving (F = 1), VNNVing (F = 1), NAD (F = 1), VNAVing (F = 1), VNAS (F = 1), VNlike / as / cnj / D (F = 1), VAprpN (F = 1), VDprpN (F = 1), VNNprpN (F = 1), VNasprpN (F = 1), VprpNprpNVing (F = 1).

The description of distribution of the kernel models of verbal word-groups in accordance with conditional frequency zones will be as follows.

In the upper zone, where the models functioning with a high frequency of usage are concentrated:

two models with two components are presented, which are not complicated by prepositions or conjunctions – VV = (F = 160), VA (F = 107);

one three-component model, which is without prepositions or conjunctions either – VNVing (F = 55);

and two models with a complex structure – VNlike / as / cnj / N (F = 135) and as / VenprpN (F = 58).

No definite structural or frequency dependence can be revealed here, because, despite the high frequency of usage, this distribution is rather chaotic. It is obvious here that both simple and complex models can function with a sufficiently high frequency. The same can be said about the models that are located on the border of the selected high-frequency zone – simple VNVing (F = 55) and as/VenprpN model (F = 58) complicated by a preposition.

In the conditional middle frequency zone:

two two-component models are observed – VVing (F = 29), VVen (F = 13);

five three-component models – Vlike / as / cnj / N (F = 32), VNVen (F = 24), VprpVing (F = 17), VDV = (F = 16), VAS (F = 10), two of them – VprpVing and Vlike / as / cnj / N are complicated by prepositions and conjunctions;

two models with four components with prepositions and conjunctions – VNlike / as / cnj / Ving (F = 33), VNlike / as / cnj / Ven (F = 23);

one complex five-component model – VprpNprpN (F = 48).

In the middle frequency zone (from 49 to 10) there is some correlation between such parameters as simplicity/complexity of the model and the frequency of its usage.

We should stress that in the description devoted to the analysis of high-frequency kernel (subordinate) models of verbal word-groups (an article of V. Dragomanov) it was proved that the simpler the kernel model as regard to its structure is, the more often it is used in the text corpus. But when we deal with the low-frequency ones this dependence is violated.

To be precise in this case, it cannot be said that the more complex the structure of the model is, the more often it is used. But the opposite statement would be incorrect either, since it is impossible not to take into account the fact that the majority of simpler two- and three-component models have rather modest frequency values, and the more complex ones (with prepositions and conjunctions) have much higher values. Four-component complex models also have high values of usage in the text corpus, and, finally, the most complex – the five-component model – has the highest frequency of usage.

The low-frequency zone of the list is the most numerous and contains, as it has already been noted, 26 models. According to the degree of simplicity/complexity of the structure they are distributed as follows:

first of all, the list of these low-frequency models of verbal word-groups does not contain simple two-component models;

three-component models are represented by the following six units – AsVenD (F = 7), Vlike / as / cnj / Ven (F = 6), V / like / as / cnj / A (F = 3), VAD (F = 1), Vlike / as / cnj / Ving (F = 1), NAD (F = 1). The major-

ity – 4 units – is structurally complicated, they have a higher frequency of occurrence than models with a simple structure. The only exception is Vlike / as / cnj / Ving model, which has the lowest frequency of usage;

fifteen four-component both structurally simple and complex models – VNNV = (F = 9), VNlike / as / cnj / A (F = 6), VNND (F = 4), VNlike / as / cnj / V (F = 4), VprpNV = (F = 4), VNNS (F = 2), VNAV = (F = 2), VSprpN (F = 2), VprpNVen (F = 2), VNNVing (F = 1), VNAVing (F = 1), VNAS (F = 1), VNlike / as / cnj / D (F = 1), VAprpN (F = 1), VDprpN (F = 1).

Let us consider the list of four-component models in more details. It should be noted that it is here, in this most distant reserve of future possible high-frequency speech units, there is an almost equal number of simple (without prepositions and conjunctions) and complicated (with prepositions or conjunctions) models of verbal word-groups:

the most frequent pair is VNNV = --- VNlike / as / cnj / A;

in the line of models with F = 4 the balance is slightly disturbed – only one simple model and two complicated VNND, VNlike / as / cnj / V, VprpNV =;

in the line of models with F = 2 there is the same quantity of both simple and complicated units VNNS, VNAV =, VSprpN, VprpNVen;

and finally, the low-frequency four-component models with F = 1 also have an equal number of structurally simple and complicated units – VNNVing, VNAVing, VNAS, VNlike / as / cnj / D (F = 1), VAprpN, VDprpN.

Four five-component models of the lowest sector of the low-frequency zone – VprpNasN (F = 3), VNDprpN (F = 3), VNNprpN (F = 1), VNasprpN (F = 1) – have an extremely complicated structure, and function with F = 3 and F = 1.

The only six-component model of the verbal word-groups – VprpNprpNVing (F = 1) is both complicated and frequently lowest.

#### 4. Conclusions

So, the results of the statistical research of low-frequency kernel (subordinating) models of verbal word-groups described in this article allowed us to come to the following general conclusions.

1. The quantitative analysis of kernel (subordinating) models of verbal word-groups found in the text corpus “Radio electronics” with a low frequency has shown that the value of occurrence of different models is much higher as compared to the same units with a high frequency of usage – 77, 4% and 22, 6%, respectively.

2. Although the correlation of quantitative characteristics goes in favor of low-frequency speech units (models of verbal word-groups) when compared with the same units with a high frequency (the results of a statistical study were obtained earlier), the analysis

showed that the correlation of their total frequencies in the text corpus makes 0, 4% to 99, 6%.

3. A comparative analysis of low-frequency and high-frequency kernel (subordinating) models of verbal word-groups by the simplicity/complexity parameter of the model structure showed the following. As mentioned in the article, the models with a high frequency of usage, that were considered earlier, showed direct dependence of this parameter on the frequency of usage, i.e. the simpler the model is with regard to its structure, the more often it is found in the text corpus.

In verbal word-groups with a low frequency of occurrence there is no such direct dependence, and another statistical phenomenon can be noted. When the entire list of low-frequency models is divided into conditional frequency zones, the units with a frequency from 160 to 50, i.e. which have, as we see, a fairly high frequency of usage in the texts, do not show any statistical dependence of their frequencies on the simplicity/complexity of the model structures. The distribution of simple models and the ones complicated by prepositions or conjunctions in the area from 160 to 50 is simply chaotic.

But going further, as the values of the statistical parameter of frequency decrease, a completely distinct dependence of the frequency of usage on the simplicity/complexity of the model structure begins to appear, with an advantage both in frequency and quantitative indicators in favor of sophisticated multi-component models. This dependence begins to manifest itself quite clearly in the frequency area from 49 to 10, although with some reservations. It is enough to say that here the complex five-component model VprpNprpN (F = 48) turned out to be the most frequent.

In the lowest frequency conditional zone with a frequency range from 9 to 1, in which most models are located (26 out of 41 units), the crowding out of simple models by structurally complicated units, which have not only a quantitative advantage but also a higher frequency of usage in this zone, occurs. Simultaneously one can observe the reduction of frequencies to the minimum, and a gradual balance of the functioning of simple and complicated models, their equal representation in this zone.

Thus, based on the described results of statistical analysis of low-frequency kernel (subordinating) models of verbal word-groups, it can be assumed that the same structural and frequency dependence may be characteristic not only for them, but also for low-frequency word-groups of other types.

This assumption requires further statistical study of text units in order to carry out an appropriate comparative analysis with the results already obtained.

## BIBLIOGRAPHY

- Богданова С. Ю. Возможности корпусной методологии в решении лингвистических задач / С. Ю. Богданова. Иркутск.: Вестник ИГЛУ. № 2 (18). – 2012. – С. 47-50.
- Борисенко Т.И. Английские модальные глагольные конструкции в подъязыках техники: дис. ... канд. филол. наук: спец. 10.02.04 Германские языки / Т. И. Борисенко. – Одесса, 1989. – 180 с.
- Дьяченко Г. Ф. Исследование семантики глагола в английских текстах подъязыков техники: автореф. дисс. ... канд. филолог. наук: спец. 10.02.04 “Германские языки” / Г. Ф. Дьяченко. – Одесса, 1984. – 16 с.
- Захаров В. П. Корпусная лингвистика / В. П. Захаров // Учебное пособие. – 2005. – (Электронный ресурс: available at [http://vp-zakharov.narod.ru/VictorZakharov\\_CorpLingv.doc](http://vp-zakharov.narod.ru/VictorZakharov_CorpLingv.doc))
- Комарова З. И. Методология, метод, методика и технология научных исследований в лингвистике / З. И. Комарова // Учебное пособие. – Екатеринбург: Изд-во УрФУ, 2012. – 818 с. (Электронный ресурс).
- Неврева М. Н. Словообразовательная типология в подъязыках техники (на материале английского языка): дисс. ... канд. филол. наук: 10.02.04 / М. Н. Неврева. – Одесса, 1986. – 257 с.
- Трофимова А. С. Синтаксические единицы современных английских текстов делового стиля: дисс. ... канд. филолог. наук: спец. 10.02.04 “Германские языки” / А. С. Трофимова. – Одесса, 1988. – 200 с.
- Шапа Л. Н. Формы и функции имен прилагательных в научно-техническом тексте (на материале английского подъязыка электроснабжения): дисс. ... канд. филол. наук: 10.02.04 / Людмила Николаевна Шапа. – Одесса, 1991. – 201 с.
- Shapa L. N. Kernel (subordinating) models of verbal word-phrases in the texts “Radio-electronics” // L. N. Shapa, T. I. Kudinova // К: Науковий часопис НПУ ім. М. П. Драгоманова. – Серія 9 “Сучасні тенденції розвитку мов”. – Вип. 15. – 2017. – С. 218–229.
- Шубик С. А. Статистические методы в лингвистике / С. А. Шубик // Сб. науч. работ Статистика речи и автоматический анализ текста. – Ленинград: Наука, 1980. – С. 52–64.
- Alekseev P. M. Statistic Lexicography (typology, compiling and applications of frequency dictionaries) / P. M. Alekseev. – Leningrad, L: Leningrad Pedagogical Institute, 1975. – 120 p.
- Krishnamurthy R. Corpus Lexicography / Ramesh Krishnamurthy // Elsevier Encyclopedia of Language & Linguistics. – Birmingham: Aston University, UK. – 2nd Edition. – 2006. – P.p. 250-254. (Electronic resources: available at [https://www.researchgate.net/publication/291110989\\_Corpus\\_Lexicography](https://www.researchgate.net/publication/291110989_Corpus_Lexicography)).
- Quantitative Linguistics An International Handbook [edited by Köhler R., Altmann G., Piotrovsky R.G.]. – Berlin, New-York, Walter de Gruyter. – 2005. – 1027 p.
- Summers D. Corpus Lexicography – The importance of representativeness in relation to frequency / D. Summers. – 2005. – (Electronic resources: available at <http://www.pearsonlongman.com/.../pdfs/corpus-lexicography.pdf>).
- Tuldava Ju. Исследования по сопоставительному и прикладному языкознанию / Ju. Tuldava. – Tartu: Tartu Riiklik Ülikool, 1983. – 164 p. (Electronic Resources).

## REFERENCES

- Alekseev P. M. (1975) Statistic Lexicography (typology, compiling and applications of frequency dictionaries) / P. M. Alekseev. – Leningrad, L: Leningrad Pedagogical Institute, 120 p.
- Bogdanova S. Yu. Vozmozhnosti korpusnoy metodologii v reshenii lingvisticheskikh zadach [Возможности корпусной методологии в решении лингвистических задач] / S.Yu. Bogdanova. Irkutsk.: Vestnik IGLU. № 2 (18). 2012. S. 47-50.
- Borisenko T.I. Angliyskie modalnye glagolnye konstruksii v pod'yazykah tehniki [Английские модальные глагольные конструкции в подъязыках техники]: dis. ... kand. filol. nauk: spets. 10.02.04 Germanskie yazyiki / T. I. Borisenko. – Odessa, 1989. – 180 s.
- Dyachenko G. F. Issledovanie semantiki glagola v angliyskikh tekstah podyazykov tehniki [Исследование семантики глагола в английских текстах подъязыков техники]: avtoref. diss. ... kand. filolog.nauk: spets. 10.02.04 “Germanskie yazyiki” / G. F. Dyachenko. – Odessa, 1984. – 16 s.
- Komarova Z. I. Metodologiya, metod, metodika i tehnologiya nauchnykh issledovaniy v lingvistike [Методология, метод, методика и технология научных исследований в лингвистике] / Z. I. Komarova // Uchebnoe posobie. – Ekaterinburg: Izd-vo UrFU, 2012. – 818 s. (Elektronnyiy resurs).
- Krishnamurthy R. Corpus Lexicography / Ramesh Krishnamurthy // Elsevier Encyclopedia of Language & Linguistics. – Birmingham: Aston University, UK. – 2nd Edition. – 2006. – P.p. 250–254. (Electronic resources: available at [https://www.researchgate.net/publication/291110989\\_Corpus\\_Lexicography](https://www.researchgate.net/publication/291110989_Corpus_Lexicography)).
- Nevreva M. N. Slovoobrazovatel'naya tipologiya v pod'yazykah tehniki (na materiale angliyskogo yazyika) [Словообразовательная типология в подъязыках техники (на материале английского языка)]: diss. ... kand filol. nauk: 10.02.04 / M. N. Nevreva. – Odessa, 1986. – 257 s.
- Quantitative Linguistics An International Handbook [edited by Köhler R., Altmann G., Piotrovsky R.G.]. – Berlin, New-York, Walter de Gruyter. – 2005. – 1027 p.
- Shapa L. N. Formy i funktsii imen prilagatelnykh v nauchno-technicheskom tekste (na materiale angliyskogo pod'yazyka elektrosnabzheniya) [Формы и функции имен прилагательных в научно-техническом тексте (на материале английского подъязыка электроснабжения)]: diss. ... kand. filol. nauk: 10.02.04 / Lyudmila Nikolaevna Shapa. – Odessa, 1991. – 201 s.
- Shapa L. N. Kernel (subordinating) models of verbal word-phrases in the texts “Radio-electronics” // L. N. Shapa, T. I. Kudinova // К: Naukoviy chasopis NPU im. M. P. Dragomanova. – Seriya 9 “Suchasnl tendentsiyi rozvitku mov”. – Vyp. 15. – 2017. – С. 218–229.
- Shubik S. A. Statisticheskie metody v lingvistike [Статистические методы в лингвистике] / S. A. Shubik // Sb. nauch. robot Statistika rechi i avtomaticheskii analiz teksta. – Leningrad: Nauka, 1980. – S. 52-64.
- Tuldava Ju. Issledovaniya po sopostavitel'nomu i prikladnomu yazykoznaniiyu [Исследования по сопоставительному и прикладному языкознанию] / Ju. Tuldava. – Tartu: Tartu Riiklik Ülikool, 1983. – 164 p. (Electronic Resources).
- Zaharov V. P. Korpusnaya lingvistika [Корпусная лингвистика] / V. P. Zaharov // Uchebnoe posobie. – 2005. – (Elektronnyiy resurs: available at [http://vp-zakharov.narod.ru/VictorZakharov\\_CorpLingv.doc](http://vp-zakharov.narod.ru/VictorZakharov_CorpLingv.doc)).

### Анотація

Робота присвячена одному з найбільш широко використовуваних типів дієслівних словосполучень, а саме ядерних моделей підрядних дієслівних словосполучень, які функціонують в текстовому корпусі «Радіоелектроніка», що відноситься до технічно-наукового дискурсу. Текстовий корпус був скомпільований на основі американських науково-технічних журналів «Радіоелектроніка» методом суцільної вибірки. Для відбору цього типу моделей був розглянутий теоретичний підхід, який описує проблеми комбінаторики та валентності. Це дозволило включити як безпосередньо так й дистанційно розташовані елементи, які обов'язково повинні з'єднуватися з дієсловом граматичним (підрядним) зв'язком та формувати з ним словосполучення. З ціллю скомпільовати інвентар ядерних моделей підрядних дієслівних словосполучень були підвергнуті дистрибутивному аналізу 10698 речень. Відбирались тільки моделі, які з'являються у текстовому корпусі з низькою (нижче ніж 180) частотою. Загальна кількість моделей з низькою частотою досягала 41 одиниць. Крім дистрибутивних методів використовуються також статистичні. Кількісний аналіз ядерних моделей, використаних у текстовому корпусі з низькою частотою, показує, що кількість різних моделей є набагато більше у зрівнянні з такими ж одиницями, які мають високу частоту використання – 77, 4% та 22, 6 %, відповідно. Але ж порівняння сумарних частот говорить на користь моделей з високою частотою – 0, 4 % до 99, 6 %. Крім кількісних характеристик аналізується параметр простоти/складності структур моделей. При вивченні високочастотних ядерних моделей спостерігається пряма залежність частоти використання від структури моделі – чим простіша структура, тим вище частота використання. Але низькочастотні ядерні моделі демонструють більш складне та суперечливе функціонування по відношенню до цього параметру. У більшості випадків досить чітко видна тенденція до зворотної залежності, тобто чим простіша структура моделі, тим нижче її частота; у найчастотніших серед низькочастотних моделей відбувається хаотичний розподіл по частотах.

### Ключові слова

Безпосередньо і дистанційно розташовані елементи, кореляція, низькочастотні моделі, структура моделі, кількісні характеристики, параметр простоти/складності, підрядні зв'язки, статистичні методи.