UDC 81’367.625:62+811.111’367.625:521.396
DOI https://doi.org/10.32782/tps2663-4880/2026.46.2.22

CORRELATION OF VERB LEXEMES AND VERB COMPATABILITY MODELS (ON THE BASIS OF TEXT CORPUS ON RADIO ELECTRONICS)

СПІВВІДНОШЕННЯ ДІЄСЛІВНИХ ЛЕКСЕМ ТА ДІЄСЛІВНИХ СУМІСНИХ МОДЕЛЕЙ (НА ОСНОВІ ТЕКСТОВОГО КОРПУСУ З РАДІОЕЛЕКТРОНИКИ)

Shapa L.M.,
orcid.org/0000-0002-3502-4577
Candidate of Philological Sciences, PhD,
Associated Professor at the Foreign Languages Department
National University “Odesa Polytechnic”

Mikeshova G.P.,
orcid.org/0000-0003-1963-0013
Senior Lecturer at the Foreign Languages Department
National University “Odesa Polytechnic”

The paper deals with the continuation of scientific research carried out on the basis of one of the texts referring to scientific and technical discourse – radio electronics. The object of the work is verbal compatibility models functioning in the texts and their matching with the corresponding verbs used in these models quantitively and qualitatively. The previous article was devoted to the description of verbal lexemes and quantity of models from the viewpoint of verbs, which are used in. The given paper describes the models, their statistical and grammar characteristics, the number of models the verbal lexemes are connected with. That is why all the verbs are grouped according to the number of models they (verbs) are matching with. Since the same typical model is found in several verbs, all analyzed verbs can be grouped according to the commonality of the typical models they share: verbs with one, two, three, etc. common models are distinguished. But first of all, the compatibility models are arranged in frequency decreasing order. For example, the first group of verbs ('find', 'set') with one common model is identified by VN model (F = 7730), which has the highest total frequency. The next group of verbs with two matching models is identified by the VN and VNprpN models, since the VNprpN model is the second according to total frequency after the VN model. This procedure continues until all typical models for each verb are verified to match. As a result, the two rows of sets of models and the groups of verbs models are obtained: the first is established by increasing ranks of the models with a continuous sequence – 1, 2, 3, 4, etc. The second row is also established by increasing ranks of the models, but with a discrete sequence of the latter (models), for example, 1, 3, 4; 1, 3, 5, etc. The classification of verbs taking into account the frequency of typical model for each verb was performed. It allows to construct several scheme-trees in which all verbs that have the similar most frequent models are distinguished into separate groups. Based on this characteristic, four groups were identified in the study material. The authors select two groups of verbs: a group of verbs represented by one common most frequent model, and a group of verbs identified by two common most frequent models.

Key words: frequency, commonality, classification, subgroup, distribution.

Стаття присвячена продовженню наукового дослідження, проведеного на основі одного з текстів, що стосуються науково-технічного дискурсу – радіоелектроніки. Об'єктом роботи є моделі вербальної сумісності, що функціонують у текстах, та їх зіставлення з відповідними дієсловами, що використовуються в цих моделях, кількісно та якісно. Попередня стаття була присвячена опису дієслівних лексем та кількості моделей з точки зору дієслів, що використовуються в них. У цій статті описано самі моделі, їх статистичні та граматичні характеристики, кількість моделей, з якими пов'язані дієслівні лексеми. Саме тому всі дієслова згруповані за кількістю моделей, з якими вони (дієслова) збігаються. Оскільки одна й та сама типова модель зустрічається у кількох дієслів, усі аналізовані дієслова можна згрупувати за спільністю типових моделей, які вони поділяють: виділяються дієслова з однією, двома, трьома тощо спільними моделями. Але перш за все, моделі сумісності розташовані в порядку зменшення частоти. Наприклад, перша група дієслів (‘find’, ‘set’) з однією спільною моделлю ідентифікується за моделлю VN (F = 7730), яка має найвищу сумарну частоту. Наступна група дієслів з двома відповідними моделями ідентифікується за моделями VN та VNprpN, оскільки модель VNprpN є другою за сумарною частотою після моделі VN. Ця процедура продовжується доти, доки не буде перевірено, що всі типові моделі для кожного дієслова збігаються. В результаті отримано два ряди наборів моделей та групи дієслів: перший встановлюється шляхом зростання рангів моделей з безперервною послідовністю – 1, 2, 3, 4 тощо. Другий ряд також встановлюється шляхом зростання рангів моделей, але з дискретною послідовністю останніх (моделей), наприклад, 1, 3, 4; 1, 3, 5 тощо. Було проведено класифікацію дієслів з урахуванням частоти типової моделі для кожного дієслова. Це дозволяє побудувати кілька схем-дерев, в яких усі дієслова, що мають подібні найчастіші моделі, виділяються в окремі групи. На основі цієї характеристики в досліджуваному матеріалі було виділено чотири групи. Автори як приклад виділяють дві групи дієслів: групу дієслів, представлену однією спільною найчастішою моделлю, та групу дієслів, визначену двома спільними найчастішими моделями.

Ключові слова: частота, спільність, класифікація, підгрупа, розподіл.

Statement of Problem. Literature review. Modern linguistics is characterized by an appeal to the study of not individual text units (word forms) found in text corpora, on the basis of which the so-called probabilistic-statistical models (frequency dictionaries) are created (our dictionaries), but entire constructions connected on the basis of grammatical compatibility [1; 2].

This is also emphasized in literary sources devoted to the theoretical development of this topic [3; 4], and in modern English grammars [5]. Practical attempts have already been made both based on works of fiction [6; 7; 8] and on text corpora related to scientific and technical discourse. One can also cite the rather successful analysis of Modal Verbal Construction [9], which determines the correspondence between the lexical and grammatical characteristics of the elements of a modal phrase.

In addition to providing material for modern English grammars, collocation dictionaries, as well as the results of research on compatibility and combinatorics in English, contribute to the development of automated text translation systems.

Despite the well-developed applied base, which includes works on compatibility and combinatorics, this article may offer a degree of novelty, as it describes the results of an analysis performed using real texts on radio electronics.

This article is a logical continuation of the research described in the article "Kernel Models of Verbal Subordinating Word-Phrases in the Text Corpus "Radio-Electronics" by Shapa L.M., Kudinova T.I. [1], so some facts presented in this work will be repeated to clarify how the data described were obtained and how the research is being further developed.

The main methods used in working with the texts were as follows: contextual analysis, expert assessment (i.e., a survey of specialists in the field of radio electronics), and statistical methods of data calculation.

Goal of the article. The goal of the article is to describe the correlation of the verb lexemes and verb compatibility models functioning in the text corpus on radio electronics and to justify the distribution of verb lexemes around the groups, taking into account matching models.

Basic material. First of all, as mentioned above, it's necessary to reiterate some of the data already described to clarify the presented analysis results:

1) a list of the most frequent verb lexemes selected for analysis. The list included the 52 most frequent verbs found in the "Radio electronics" text corpus: 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).

2) a list of verb combinability models implemented in the text corpus on radio electronics. To create an inventory of compatibility models, a distributional sentence analysis was conducted, taking into account both contact and distance elements, but always connected to the verb by a grammatical (subordinate) relation and forming a phrase with it. The inventory list includes 53 kernel verbal combinability models. Grammatical relations of the verbs with the surrounding elements were expressed in terms of classes of words which are denoted with the help of conventional markers: V0 – a verb that has no subordinating bonds, N – noun, A – adjective, D – adverb, adverbial modifier, V= – infinitive, Ving – Participle I, Ven – Participle II, S – verb-dependent subordinating clause, which in our case can be considered as a word class, prp – preposition, cnj – conjunction.

The list of models will be presented in descending order of frequency: 1. VN (F = 7730); 2. VNprpN (F = 2844); 3. VprpN (F = 1036); 4. VND (F =944); 5.V0 (F = 914); 6. VNV= (F = 715); 7. VD (F = 500); 8. VS(F=422); 9. VNN (F =252); 10. VNprpVing (F = 217); 11. VNS (F = 200); 12. VNA (F = 180); 13. VV= (F = 160); 14. VNlike/as/cnj/N (F = 135); 15. VA (F = 107); 16. AsVenprpN (F = 58); 17. VNVing (F = 55); 18. VprpNprpN (F = 48); 19. VNlike/as/cnj/Ving (F = 33); 20. Vlike/as/cnj/N (F = 32); 21. VVing (F = 29); 22. VNVen (F = 24); 23. VNlike/as/cnj/Ven (F = 23); 24. VprpVing (F = 17); 25. VDV= (F = 16); 26. VVen (F = 13); 27. VAS (F = 10); 28. VNNV= (F = 9); 29. AsVenD (F = 7); 30. Vlike/as/cnj/Ven (F = 6); 31. VNlike/as/cnj/A (F = 6); 32. VNND (F = 4); 33. VNlike/as/cnj/V (F = 4); 34. VprpNV= (F = 4); 35. VprpNasN (F = 3); 36. VNDprpN (F = 3); 37. V/like/as/cnj/A (F = 3); 38. VNNS (F = 2); 39. VNAV= (F = 2); 40. VSprpN (F = 2); 41. VprpNVen (F = 2); 42. VAD (F = 1); 43. Vlike/as/cnj/Ving (F = 1); 44. VNNVing (F = 1); 45. VNAD (F = 1); 46. VNAVing (F = 1);

47. VNAS (F = 1); 48. VNlike/as/cnj/D (F = 1); 49. VAprpN (F = 1); 50. VDprpN (F = 1); 51. VNNprpN (F = 1); 52. VNasprpN (F = 1); 53. VprpNprpNVing (F = 1).

The entire list of models was conditionally divided into frequency zones: high-frequency zone – from 7730 to 200; medium-frequency zone – from 199 to 10 occurrences in the texts; low-frequency zone – from 9 to 1 use in the texts. The high-frequency zone includes 11 models; the medium-frequency zone includes 16 models; the low-frequency zone includes 26 models. As we can see, the total number of occurrences of the entire list of combinability models is 16775 units. Having the number of models, the frequency of use of each model and their total frequency, it is easy to calculate the share of models in each frequency zone in the overall list of models. Models in the high-frequency zone account for 93% of all presented in the list, i.e., the smallest number of models covers the majority of uses; in the medium frequency zone, the percentage of 16 models occupies 5%; in the low-frequency zone, the share of 26 models is 2%.

3) a list of the most frequent verb lexemes with which the combinability models are realized: find (F=262), set (F=183) (2 words) – 11 models; check (F = 324), make (F = 518), read (F = 176), see (F = 339), show (500) (5 words) – 10 models; add (F = 131), build (F = 121), change (F = 129), connect (F = 303), develop (F = 128), do (F = 231), get (F = 303), give (F = 198), increase (F = 145), mount (F = 156), operate (F = 283), replace (F = 122), start (F = 137) (13 words) – 9 models; apply (F = 205), cause (F = 165), determine (F = 138), know (F = 158), measure (F = 127), switch (F = 104), take (F = 195, tune (F = 133) (8 words) – 8 models; adjust (F = 123), come (F = 109), feed (F = 250), go (F = 254), produce (160), provide (F = 205), record (F = 182), reduce (F = 143), require (F = 189), try (F = 196), use (F = 1291) (11 слов) – 7 моделей; call (F = 107), look (F = 157), need (F = 205), note (F = 120), say (F = 112), short (F = 104), want (F = 122), work F = 269) (8 words) – 6 models; control F = 121), design (F = 113), test (F = 108) (3 words) – 5 models; flow (F = 121) (1 word) – 3 models; include (F = 120) (1 word) – 2 models.

Here, we can observe a certain statistical relationship between the number of words and the number of models they are implemented with. Initially, the fewer words, the more models they are used in (the first and the second word groups). Then, the relationship reverses: the more words, the fewer models (groups three through six). Then, the relationship repeats itself: the fewer words, the more models (groups seven through the end of the list).

With all the necessary data, we can present a study that focuses on the distribution of verbs into groups based on the number of correlated models. That is, the focus here will be not on the lexemes (presented above), but on the models. Since the same typical model is found in several verbs, all the verbs analyzed can be grouped based on the commonality of their matching typical models.

The procedure for distributing verbs into groups based on the number of matching models is as follows. Verbs with one, two, three, and so on common models are placed into separate groups. Since all models vary in frequency, we will organize the verbs into groups based on the number of models they correlate, taking into account the descending of their total frequency (in this case, we are not talking about the frequency with which a particular verb appears in texts, but about the frequency of a model in which the verb lexeme functions). For example, the first group of verbs ('find', 'set') with one common model is identified by VN model (F = 7730), which has the highest total frequency. The next group of verbs with two matching models is identified by the VN and VNprpN models, since the VNprpN model is the second according to total frequency after the VN model. This procedure continues until all typical models for each verb are verified to match.

All analyzed verbs differ in the sets of inherent models in which they (the verbs) are realized in speech. If a verb or a group of verbs lacks a given model, the model with the next highest total frequency is taken into account when sorting the verbs into groups. This allows us to identify subgroups with identical quantitative but different qualitative models. For example, the verbs 'come' and 'include', although they possess the same number of models (2), are classified into different subgroups because the models VN and VprpN are recorded for the verb 'come', while VN and VNV= are recorded for the verb 'include'. As a result of this procedure of dividing verbs into groups, we obtain two rows of sets of models and the groups of verbs they unite: the first is established by increasing ranks of the models with a continuous sequence – 1, 2, 3, 4, etc. The second row is also established by increasing ranks of the models, but with a discrete sequence of the latter (models), for example, 1, 3, 4; 1, 3, 5, etc. Thus, with the help of the described procedure, all the verbs under study are distributed among models, and the models, in turn, among the verbs that are used in them, with each model in the group taking a place depending on its rank.

The data obtained show that, through various combinations of models, it is possible to identify 26 sets that correspond to the same number of verb groups. All of these verb groups can be characterized both by the sets of models registered for them and by the number of verb lexemes.

All model groups are characterized by high diversity in the number of verbs recorded within them.

It's noteworthy that the subgroups identified with a continuous sequence of models represent larger groupings of verbs than those identified with a discrete sequence of models. Thus, among the subgroups of the first classification, the largest grouping is the one consisting of 51 lexemes. Among the subgroups identified with a discrete sequence of models, however, there are 11 verbs. Furthermore, among the subgroups identified with a continuous sequence of models, not a single subgroup is identified that contained a single verb. Among those established with a discrete sequence, there are seven subgroups represented by only one verb. These include subgroups (with a continuous sequence) that cover the verbs: flow (model VprpN), come (models VN, VprpN), include (models VN, VNV=), say (models VN, VNprpN, VprpN, VNV=), look (models VN, VNprpN, VprpN, VD), test (models VN, VNprpN, VprpN, VND, VNprpVing), short (models VN, VNprpN, VprpN, VND, VNV=, VNS), show (models VN, VNprpN, VprpN, VND, VNV=, VD, VS, VNN, VNS).

It's interesting to note a certain regularity of subgroups identified in a continuous sequence of models. It consists of the fact that as the number of matching models increases, the number of verbs in the subgroups decreases. Thus, if a subgroup represented by one model contains 51 verbs, then a subgroup represented by two models contains 38 lexemes, and so on, up to the subgroup containing two verbs–'find' and 'set'–united by eleven models.

In subgroups organized with a discrete sequence of models, the number of verbs does not depend on the number of models identified in them. Thus, there are subgroups with the same number of verbs but different numbers of models. For example, separate subgroups include the verb 'flow' with one model, the verb 'short', for which six models are identified, and the verb 'show', which implements nine models.

Among the resulting verb groups, four are highly enriched (from 51 to 34 lexemes), two are moderately enriched (from 33 to 16 lexemes), and twelve are lightly enriched (from 15 to two verbs). Eight verbs form groups containing one lexeme each.

From all of the above, it can be concluded that the classification is characterized primarily by its ability to generalize linguistic units, and secondarily by their individualization. Consequently, this classification focuses primarily on the combinability of the aforementioned verb groups, and less on the combinability of individual lexemes. The sets of models, correlated with each of the established verb groups characterize them in terms of combinability as parts of speech as a whole (within the verbs under study).

Let's present a classification of verbs based on the frequency of typical models for each verb. Each of the typical models used to classify verbs has a different frequency for different verbs. By arranging the models for each verb in descending order of frequency, we can derive a classification of verbs based on the commonality of models and their ranking.

For clarity, the classification of verbs based on this feature can be represented as a classification tree.

The procedure for constructing a tree consists of several stages and can be summarized as follows:

1. All verbs that have the similar most frequent model are distinguished into separate groups. Based on this characteristic, four groups were identified in the study material.

The first group identified, based on the most frequent VN model, includes 42 verbs.

The second group of verbs, also united by one of the most frequent VprpN model, is represented by the following seven verbs: 'come, flow, go, look, operate, start, and work'.

The third group includes two verbs, 'note, say', united by the VS model.

The verb 'call', for which the VNN model is the most frequent, forms the fourth group.

2. Within each of the identified groups, all verbs are selected based on the second most frequent model, then the third, and so on, until all typical models for each verb in the identified groups have been ranked. As an example, we will select two groups of verbs: a group of verbs represented by one common most frequent model, and a group of verbs identified by two common most frequent models. We will then observe how the grouping of verbs on the basis of models and the models organized verbs changes.

The group of verbs united by the VN model is the most numerous and includes the following verbs: add, adjust, apply, build, cause, change, check, connect, control, design, determine, develop, do, find, feed, get, give, include, increase, know, make, measure, mount, need, produce, provide, read, record, reduce, replace, require, see, set, show, short, switch, take, test, try, tune, use, want. The verbs 'call, come, go, look, note, operate, say, start, and work', for which the VN model cannot be typical based on their valence properties, are excluded from the list.

The verbs that are correlated on the basis of the two most frequent models are presented in the table.

Verb groups defined by sets of two common most frequent models

Common models	Verbs combined with common models	Amount of words
VN, VNprpN	add, adjust, apply, build, change, check, connect, determine, develop, do, feed, find, get, measure, mount, produce, provide, read, record, reduce, replace, require, set, short, show, take, test, try, tune, use	30
VN, VS	design, know, see	3
VN, VNV=	cause, include, need, want	4
VN, VprpN	increase, switch	2
VN, VNA	make	1
VN, VND	control	1
VN, VNN	give	1

Conclusions. All of the above allows us to draw the following conclusions.

1.A description of verb lexemes from the perspective of the verb models in which they (the verbs) function in the texts on radio electronics allows us to distinguish two types of sequences: continuous (1, 2, 3, 4, etc.) and discrete (1, 3, 4; 1, 3, 5), which makes it possible to rank-order verbs and models. The continuous sequence encompasses a larger number of verb lexemes (51 units) than the discrete sequence (11 units). Verb groups identified using the continuous sequence contain more verb lexemes than those identified using the discrete sequence.

2. The mutual combination of models and verb lexemes allows us to identify 26 sets of models and 26 sets of verb lexemes.

3. The ability to arrange models in descending order of frequency allows us to identify groups of verbs that are distributed according to the commonality of models and their ranking.

4. The most frequent models also include the most frequent verb lexemes (only groups of verbs that share one or two most frequent models are presented).

REFERENCES:

1. Shapa L. N., Kudinova T. I. Kernel Models of Verbal Subordinating Word-Phrases in the Text Corpus “Radio-Electronics”. *Науковий часопис Національного педагогічного університету им. М.П.Драгоманова*. Київ, вид-во НПУ ім. М.П.Драгоманова. Сер. 9 «Сучасні тенденції розвитку мов». Вип. 15. 2017. С. 211-221.

2. Shapa L., Nevreva M., Tsinovaya M. Statistics of low-frequency kernel (subordinating) models of the verbal word-groups in the text corpus “Radio electronics”. *Одеський лінгвістичний вісник*. Одеса, 2018. С.47-55.

3. Benson M., Benson E. and Ilson R. The BBI combinatory dictionary of English: a guide to word combinations, Amsterdam-Philadelphia, 1997.

4. Перебійніс В.С. Моделювання в лінгвістиці. Проблеми та методи структурної лінгвістики: лекції, прочитані на респ. *Семінарі викладачів мовознавчих кафедр пед ін-тів УССР*, 30.09 – 2.10. Київ, 1964. С.34-47.

5. English Course. Phrase Grammar. *Hollywood Vocabulary*, available at: hollywoodvocabulary.com›grammar.php

6. Сердюков П.І. Граматична сполучність дієслів в англійської мові: Дис. … канд. філол. наук: Германські мови 10.02.04, Київ. 1978. 257 с.

7. Ткаченко Ю.М. Сполучність дієслова в сучасній англійської мові. Дис. … канд. філол. наук: 10.02.04 Германські мови. Харків, 1972. 305 с.

8. Bobohodjaeva M. Таҳлили муқоисавии ибораҳои номӣ: дар асоси маводи забонҳои тоҷикӣ ва англисӣ: дисс. ... Номзади илмҳои филологӣ.: 10.02.20 «Забоншиносии муқоисавӣ-таърихӣ, типологӣ ва муқоисавӣ». Душанбе, (Comparative analysis of the nominal word combinations: on the material of the Tadjik and English languages: dys. ... Candidate of Philological Science: 10.02.20 “Comparative-historical, typological and comparative linguistics”. Dushanbe, 2000. 126 p.

9. Borisenko T. I., Tsinovaya M. V., Tsapenko L. E., Sirotenko T. V. The influence of component semantics in modal verb constructions with the modal verbs of “obligation” on their grammatical and statistical features (on the basis of the technical discourse texts). *Проблеми семантики слова, речення та тексту*. Київ: Київський національний університет ім.Тараса Шевченка. 2018. С. 82-100.

Дата першого надходження статті до видання: 30.04.2026
Дата прийняття статті до друку після рецензування: 22.05.2026
Дата публікації (оприлюднення) статті: 29.05.2026