

**LEXICAL, GRAMMATICAL AND STATISTICAL FEATURES OF CONSTITUENTS
IN MODAL CONSTRUCTIONS WITH THE VERB ‘CAN’
(ON THE BASIS OF TEXT CORPORA OF SCIENTIFIC ENGINEERING DISCOURSE)**

**ЛЕКСИЧНІ, ГРАМАТИЧНІ ТА СТАТИСТИЧНІ ОСОБЛИВОСТІ СКЛАДНИКІВ
У МОДАЛЬНИХ КОНСТРУКЦІЯХ ІЗ ДІЄСЛОВОМ ‘CAN’
(НА ОСНОВІ ТЕКСТОВИХ КОРПУСІВ НАУКОВО-ТЕХНІЧНОГО ДИСКУРСУ)**

Borisenko T.I.,

orcid.org/0000-0002-9839-1761

*Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Departments of Foreign Languages
National University “Odessa Polytechnic”*

Tsinovaya M.V.,

orcid.org/0000-0003-2633-8416

*Candidate of Historical Sciences, Associate Professor,
Associate Professor at the Departments of Foreign Languages
National University “Odessa Maritime Academy”*

Ershova Yu.A.,

orcid.org/0009-0006-4251-6891

*Senior Lecturer at the Departments of Foreign Languages
Odessa Military Academy*

The article considers the problem the discourse study deals with – the functioning of text units. Modal constructions with the verb ‘can’ have been chosen as an object of research. The goal of the article is to present lexical, grammatical and statistical features of the elements included in modal constructions with the modal verb ‘can’, which are met in text corpora referred to scientific and engineering discourse. The following methods were applied: contextual method; the method of survey of experts who possess background knowledge in these areas of science and technology; statistical method of rank correlation; quantitative analysis. The material for the experiment is the text corpora of the technical fields “Power Engineering”, “Electrical Engineering” and “Automotive”, the text corpora size is 300 thousand tokens. The usage of the texts that are different as to their research topics in fundamental branches of science and technology will help to make general conclusions, forming style-identifying marks of scientific an engineering discourse. In the analyzed text corpora “Power Engineering”, “Electrical Engineering” and “Automotive” 28 models with a total frequency of 1100 speech units were discovered. The study shows that the highest priority is possessed by the modal verb constructions which have the forms of the passive infinitive, the highest total absolute frequency is possessed by ‘can be Ven’ ($F^* = 481$) construction. Modal constructions with the infinitive in the active voice appear on the second place as to their frequency of usage. Besides the modal constructions functioning with aspectual-temporal forms of the infinitive there four versions of the model “can be + name”, but the amount of such kind of models is so small that these cases could be negligible. The classification of the infinitives entering the modal constructions according to their lexical meanings demonstrates that the majority of the infinitive constituents can be referred to common and general scientific lexical layers, and only a small percentage – to terminological one. As to the modal verb ‘can’ itself, it constantly remains its modal meaning “physical ability to do something” in practically all the modal constructions. And only in the models “can be + name” a new semantic version “probability” appears.

Key words: absolute frequency, aspectual-temporal forms, comparative linguistics, quantitative values, lexical stratification.

У статті розглядається проблема, з якою має справу дискурсологія – функціонування текстових одиниць. Об’єктом дослідження обрано модальні конструкції з дієсловом ‘can’. Метою статті є представлення лексичних, граматичних та статистичних особливостей елементів, що входять до модальних конструкцій із модальним дієсловом ‘can’, які зустрічаються в текстових корпусах науково-технічного дискурсу. Застосовувалися такі методи: контекстуальний метод; метод опитування фахівців, які володіють базовими знаннями в цих галузях науки і техніки; статистичний метод рангової кореляції; кількісний аналіз. Матеріалом для експерименту є текстові корпуси технічних напрямів «Енергетика», «Електротехніка» та «Автомобілебудування», обсяг текстових корпусів 300 тис. слововживань. Використання текстів різних за тематикою у фундаментальних галузях науки і техніки допоможе зробити загальні висновки, сформувавши стильові ознаки науково-технічного дискурсу. У аналізованих текстових корпусах «Енергетика», «Електротехніка» та «Автомобілебудування» виявлено 28 моделей із загальною частотою 1100 мовних одиниць. Дослідження показує, що найвищий пріоритет мають модальні конструкції, які мають форми пасивного інфінітива, найвищу сумарну абсолютну частоту має конструкція ‘can be Ven’ ($F^* = 481$). На другому місці за частотою вживання стоять модальні конструкції з інфінітивом у активному відмінку. Крім модальних конструкцій, що функціонують з видово-часовими формами інфінітива, існує чотири варіанти моделі ‘can be + name’, але кількість

таких моделей настільки мала, що цими випадками можна знехтувати. Класифікація інфінітивів, що входять до модальних конструкцій, за їх лексичними значеннями свідчить про те, що більшість інфінітивних складників можна віднести до загальнонавчаних і загальнонаукових лексичних пластів і лише незначний відсоток – до термінологічного. Що стосується самого модального дієслова 'can', то воно постійно зберігає своє модальне значення «фізична здатність щось робити» практично в усіх модальних конструкціях. І тільки в моделях 'can be + name' з'являється новий семантичний варіант «ймовірність».

Ключові слова: абсолютна частотність, видово-часові форми, компаративна лінгвістика, кількісні значення, лексична стратифікація.

At the present stage of development of theoretical linguistics there is a gradual decrease in the attention of scientists to one of the most interesting and promising areas – discourse studies. Its place was firmly occupied by such unquestionably important and deep topics as cognitive linguistics, based on the luminaries of cognitive science [1; 2], corpus linguistics [3], which focuses on the creation of national corpora [4; 5] and practically creating the same general totality, which was mentioned more than once in works on statistical lexicography [6]. Comparative linguistics is also gaining momentum especially quickly and confidently.

However, according to the authors of the paper the range of problems in theoretical linguistics and in particular discourse studies is still far from being exhausted and requires further research. And this primarily concerns such an area of discourse studies as the research of units of text corpora that relate to scientific and technical discourse.

In this regard the authors offer the results of their research of such text units as modal verbs functioning in scientific and technical discourse, in particular the verb 'can' and the modal constructions in which it is included as a modal element.

The attention to modal verbs as grammatical units functioning in a language with different systems can first of all be noted in works on comparative linguistics [7; 8; 9; 10; 11]. Along with them the studies have appeared in the linguistic literature that describe changes in the semantic and grammatical aspects of modal verbs, possibly occurring under the influence of certain social changes: the development of industry, changing conditions of living standards, closer communication between different nations as a result of migration processes [12; 13].

However, it seems that all of the above works consider already completed processes of development of phenomena occurring in speech (in our case, modal verbs), and present data that are already final results. While working with real text corpora, the description of certain units in discourse makes it possible to consider them as static (i.e. the functioning of already established norms and meanings), as well as dynamics (i.e. the very process of inevitable and irreversible changes in linguistic units, with subsequent future shifts in the language system).

The absence of such studies confirms the novelty and relevance of the presented work.

Aim. The goal of the article is as follows: to present lexical-semantic, grammatical and statistical features of the elements included in modal constructions with the modal verb 'can', which are met in text corpora referred to scientific and engineering discourse.

Statement of tasks. For these aims a linguistic experiment was held, where the following objectives were set: 1) to distinguish all formulas and models of modal constructions functioning in the text corpora considered; 2) to group infinitives of modal constructions in accordance with their lexical meanings, applying the principles of forming the common, general and terminological layers of lexis; 3) to combine all the like modal constructions and calculate their quantitative values.

To execute the first task the methods of distributed and quantitative analysis have been used. The second task requires several methods, they are: contextual method; the method of survey of experts who possess background knowledge in these areas of science and technology as well as statistical method of rank correlation. The third one requires a type of quantitative analysis.

Discussion. The material for the experiment was the text corpora of the technical fields "Power Engineering", "Electrical Engineering" and "Automotive", which are an integral part of scientific and engineering discourse and formed by the method of continuous sampling. The basis of text corpora were scientific and technical journals published in the USA and the UK – IEEE Transactions on Power Apparatus and Systems; Power Engineering; Power; Automotive News; Combustion; Control and Optimization; Machine Design; Industrial and Production Engineering; Automotive Engineer. The text corpora size is 300 thousand tokens. It can be assumed that the usage of the texts that are different as to their research topics in fundamental branches of science and technology will help to make general conclusions, forming style-identifying marks of scientific and engineering discourse.

From the database formed by three text corpora all illustrative examples with the verb 'can' were selected using the method of characteristics correlation. Then the obtained models (constructions) are

classified according to the typological features of structural models with regard to their total absolute frequency (F*). The formalized representation of the models is expressed by the following marking:

- V – infinitive without the particle “to”;
- to V – infinitive with the particle “to”;
- Ven – participle II;
- N – noun;
- A – adjective in the function of predicative;
- prp – preposition.

In the analyzed text corpora “Power Engineering”, “Electrical Engineering” and “Automotive” 28 models with a total frequency of 1100 speech units were discovered. They are represented in the table below in the order of decreasing the frequencies.

The list of modal constructions functioning in the text corpora

“Power Engineering”, “Electrical Engineering” and “Automotive”

The study showed that the highest priority is possessed by the modal verb constructions which have the forms of the passive infinitive. There appeared to

be only 8 of them, but their total frequency (594 units) covered more than a half (54%) of all number of the models obtained. They show almost all methods of variation of constituents in syntagmatic presented in the text corpora. The highest total absolute frequency is possessed by ‘*can be Ven*’ (F*= 481) construction, it accounts for 88% of all usages of modal constructions with the passive infinitive. Within this group the vast majority of models – 522 units – is used with the modal verb ‘can’ in the present tense and only 72 structures with the modal verb in the past tense.

Modal constructions with the infinitive in the active voice appear on the second place as to their frequency of usage (F*=361, which is 33 % of all modal constructions). Data of the table show that these aspectual-temporal forms of the infinitive are diversified enough in this type of the voice.

Besides the modal constructions functioning with aspectual-temporal forms of the infinitive there four versions of the model “can be + name”, the frequency of their usage is 56 units, i.e. 5% of all modal constructions.

№№	Models (constructions)	F*	Text corpora		
			Power Engineering, F*	Electrical Engineering, F*	Automotive, F*
1	can be Ven	481	228	175	78
2	can V	231	113	22	96
3	could V	83	28	24	31
4	could be Ven	65	18	27	20
5	can V N	51	19	22	10
6	can V prp	26	11	10	5
7	cannot be Ven	24	11	7	6
8	can be A	21	11	8	2
9	cannot V	19	11	2	6
10	can be N	17	1	14	2
11	can be Ven to V	15	1	4	10
12	could be A	10	1	4	5
13	could be N	8	2	2	4
14	can have N	8	3	3	2
15	could have Ven	8	3	1	4
16	could not V	7	-	1	6
17	can V N prp	6	5	-	1
18	cannot V prp	5	-	2	3
19	could not be Ven	4	1	1	2
20	can V to V	3	1	-	2
21	cannot be Ven to V	2	-	-	2
22	can V Ven	2	2	-	-
23	can V N to V	2	1	1	-
24	could have been Ven	2	2	-	-
25	can N be Ven	2	-	1	1
26	cannot be Ving	1	1	-	-
27	cannot V to V	1	1	-	-
28	could not have been Ven	1	-	1	-

The next stage of the research concerns the description of lexical characteristics of the infinitives entering modal verb constructions and differentiation them according to their lexical meanings into common, general scientific and terms.

First of all the principles of grouping the infinitive should be presented. As mentioned above the following statistical and survey approaches of forming the common, general and terminological layers of lexis were applied:

- the verbs that in case of comparing the range of the words in the frequency dictionaries of the three text corpora and Word book by E. Thorndike, J. Lorge [14] gave a small difference of ranks and did not show terminological inclination neither in the direction of general literary dictionary nor in the direction of the remaining two dictionaries were attributed to commonly used ones [15; 16], e.g. *make, use, announce, observe, think, find, take, expect, become, see, do, meet, occur, climb, etc.*;

- the verbs that showed a significant difference between ranges when frequency lists of three analyzed text corpora and Word book by E. Thorndike, J. Lorge were compared, but a slight difference when the three frequency lists were compared, and turned out to be terminological enough compared to common literary lexics, but non-terminological for three technical frequency dictionaries were attributed to the layer of the scientific lexis, e.g. *add, operate, set, mount, control, draw, solve, determine estimate, calculate, construct, verify, approximate, regard, apply, carry out, achieve, vary, verify, predict, compare, judge, accomplish, effect, assume, substitute, employ, state, etc.*;

- and finally, the following lexemes belong to the terms that were determined not only by the method of rank correlation, but a survey of experts in the fields of science and technology, e.g. *turn over, interconnect, simulate, cut, digitize, process, store, gauge, isolate, time, coat, plot, automate, feed back, shield, etc.*

The results of the contextual analysis of the text corpora “Power Engineering”, “Electrical Engineering” and “Automotive” carried out by the authors to confirm this or that lexical meaning of the infinitive element of the modal construction with the verb ‘can’ are given below.

can be Ven – center can be located, the devices can be interconnected;

can V – motor can run; turbine can generate; could V – council could take into account, computer could control;

could be Ven – electrical circuit could be broken; short circuit could be stopped;

can V N – computer can solve the problem; it can generate the power 200 kW;

can V prp – device can switch on; it can deal with; **cannot be Ven** – console cannot be mounted; error cannot be found;

cannot V – circuit cannot become; wattmeter cannot determine;

can be Ven to V – tire can be supposed to be engaged; calculation can be used to obtain;

can have N – system can have a device; transformer can have a coil;

could have Ven – generator could have energize; motor could have achieved;

could not V – carburetor could not include; brake could not work;

can V N prp – cylinder can remain the valve in; piston can push the exhaust gases out;

cannot V prp – circuit cannot connect to; we cannot draw in;

could not be Ven – capacitor could not be included; temperature could not be measured;

can V to V – engineer can try to control; ammeter can operate to obtain;

cannot be Ven to V – transformer cannot be used to reduce; calculation cannot be assumed to be the same;

can V Ven – it can become abstracted; sound wave can appear transmitted;

can V N to V – four-cycle diesel engine can push the piston to obtain;

could have been Ven – engine could have been named; system could have been modeled;

can V N be Ven – part can possess a detail be used;

cannot be Ving – current cannot be flowing;

cannot V to V – wattmeter cannot measure to achieve;

could not have been Ven – control apparatus could not have been determined.

The examples also show that the verb ‘can’ in its turn implements its only modal meaning “physical ability to do something”. We can also conclude that the variation of the morphological characteristics of constituents in these structures does not influence the implementation of the modal meaning of the entire phrase, and the main modal meaning of “physical ability” is just clarified in time (compare: *can be designed – could be designed; can be measured – could be measured, etc.*), focuses on the possibility or impossibility of the action taken by the subject (*can be checked – cannot be checked; can be estimated – cannot be estimated*), and not any additional semantic (connotative) features are added to the modal meaning of the mentioned above structural types.

The research performed on the basis of three text corpora, which do not have the same scientific subject, enables us to assume that the modal constructions formed due to the models represented in the table reveals, basically, only one dictionary meaning “can be done” in other languages in the fields of science and technology, and the structural diversity of the infinitive does not expand the range of modal meanings or make it more complicated.

The number of lexical units in each stratification layer calculated for all three text corpora is as follows. The largest number of lexemes is in the commonly used layer, which confirms the results of other studies in the field of vocabulary stratification, lexical layers, its share is 52% of all lexemes. In second place there is the layer of general scientific lexemes – 35% of all lexemes, the terminological layer includes only 13% of the units calculated.

The analysis of the compatibility of the verb ‘can’ with the verbs belonging to the above mentioned lexical layers in the forms of passive and active voices within each of the three text corpora shows that when the allocation is performed with the lexemes of common and scientific layers, then the limitations in the tokens of these verbs do not exist. If the verb ‘can’ combines with the infinitives of verbs-terms, then there is a semantic correlation, which is characteristic for each area of knowledge. This is naturally explained by the fact that the lexemes of the first two layers are common to virtually all areas of science and technology, while in combination with verbs-terms the modal construction indicates the possibility of performing actions over phenomena, characteristic only to one of the three subject areas – “Power Engineering”, “Electrical Engineering” and “Automotive”.

The construction models “can be + name” (noun, adjective, prepositional attributive structures) were considered next. Regarding its frequency of usage as well as the variety of the formulas used in the text corpora it occupies the last place in the list of modal constructions. The model “can be + name” is represented in the fields of “Power Engineering”, “Electrical Engineering” and “Automotive” in only four versions:

can be A – *circuit can be open; system can be elementary;*

can be N – *they can be blades of the turbine;*

could be A – *plot can be sinusoidal; temperature could be high;*

could be N – *fuel could be the form of spray; fuel could be basin.*

The observations show that reconsidering and transformation of the main meaning of the verb ‘can’, which is “possibility”, take place and as a result of its contact with different parts of speech a new semantic version “probability” appears, for example, *However, it has suggested that it can be open to diluting its stake to support growth at the carmaker.* The shift of semantics in this construction in comparison with the structures with the infinitive is quite natural, since it includes a nominal element.

As for the lexical meaning of nominal lexemes in the modal constructions of this type, it can be seen from the examples that two lexemes belong to the commonly used vocabulary layer (*open, high*); three lexemes (*elementary, sinusoidal, form*) – to the general scientific layer; two ones – to the terminological layer (*blades, basin*).

Conclusions. The most frequent modal constructions with the verb ‘can/could’, functioning in the corpora “Power Engineering”, “Electrical Engineering” and “Automotive” and discussed in this article are the ones with the passive infinitive in various combinations of the constituents. The second place as to occurrence in the text bodies is occupied by the infinitive in the active voice. The most insignificant as to the frequency of usage turned to be constructions “can/could + name”.

The predominance of the passive forms over active ones is considered to be one of the main characteristics of the scientific discourse texts, because the absence of the doer of the action can be observed there. However, it can be seen in our study that the active voice forms are also quite numerous. This point is explained by the fact that the objects themselves (*turbines, electrical circuits, brakes, motors, etc.*) reveal their functioning activity.

The classification of the infinitives entering the modal constructions according to their lexical meanings demonstrates that the majority of the infinitive constituents can be referred to common and general scientific lexical layers, and only a small percentage – to terminological one.

As to the modal verb ‘can’ itself, it constantly remains its modal meaning “physical ability to do something” in practically all the modal constructions. And only in the models “can be + name” a new semantic version “probability” appears. But the amount of such kind of models is so small that these cases could be negligible.

REFERENCES:

1. Fillmore Ch. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*. New York, 1976. Vol. 280. P. 20–32.

2. Miller G. A. Images and Models, Similes and Metaphors. *Metaphor and Thought*. Cambridge-London: Cambridge University Press, 1993. P. 357–400.
3. Newman J. Aiming low in linguistics: Low-level generalizations in corpus based research. *Proceedings of the 11th International Symposium on Chinese Languages and Linguistics*. (Taiwan, Hsinchu May 24, 2008). Taiwan, Hsinchu: National Chiao Tung University, 2008. P. 50–58.
4. National corpus of the American English. URL: <http://corpus.byu.edu/coca> (дата звернення: 20.03.2020)
5. National corpus of the British English. URL: <http://corpus.byu.edu/bnc> (дата звернення: 5.10.2020)
6. Перебийніс В. І., Муравицька М. П., Дарчук Н. П. Частотні словники та їх використання. Київ: Наукова думка, 1985. 202 с.
7. Бобохоҷаева Мухаббат Таҳлили муқоисавии ибораҳои номӣ: Дар асоси забонҳои тоҷикӣ ва англисӣ (Comparative analysis of nominal phrases: On the material of Tajik and English languages): дис. ... кандан. филол. Илмҳо: 10.02.20. Забоншиносии муқоисавӣ-таърихӣ, типологӣ, муқоисавӣ. Душанбе, 2000. 126 с.
8. Umarova M. B. Comparative analysis of modal verbs in the English and Uzbek languages. *Young scientist*. Херсон, 2014. No. 7. P. 641–642.
9. Исмоилзода Эрачи Содик Хусусиятҳои грамматикии конструкцяхо бо феълҳои модалӣ дар забонҳои тоҷикӣ ва англисӣ (Grammatical features of constructions with modal verbs in Tajik and English): Реферат дис. ... Номзади фанҳои илм. Илмҳо: 10.02.20 (Abstract dis. ... Ph.D. Sciences: 10.02.20). Душанбе, 2011. 26 с.
10. Собир А.М. Типологія муқоисавӣ ва муқоисавии синтаксис ва сохтори ҷумлаи содда дар забонҳои системаҳои гуногун (дар асоси забонҳои тоҷикӣ ва англисӣ) (Comparative Typology of Syntax and Structure of a Simple Sentence in Different System Languages (Based on Tajik and English Languages): дис. ... кандан. филол. Илмҳо: 10.02.20. Забоншиносии муқоисавӣ-таърихӣ, типологӣ, муқоисавӣ. Душанбе, 2000. 252 с.
11. Малявін Д. В., Королева Т. М. Способи вираження модальних відносин в англійській та українській мовах. Одеса: ОДУ, 1986. 86 с.
12. Coates J. The semantics of the modal auxiliaries. L. Canberra: Croom Helm, 1993. 246 p.
13. Ney J. W. The modals in English. *Journal of English Linguistics*. London, 1976. vol. 10. P. 8–20.
14. The Teacher's Word Book of 30 000 words / edited by E. Thorndike, J. Lordge. New York: Columbia University Teacher's college press, 1968. 274 p.
15. Oxford Advanced Learner's Dictionary. 7th edition. / edited by A. S. Hornby. Oxford: Oxford University Press, 2005. 1539 p.
16. New Polytechnic dictionary / edited by Ishlinskij A.U. M: Scientific Publishing house GRE, 2000. 656 p.