Tetyana BORYSENKO,
orcid.org/0000-0002-9839-1761
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Department of Foreign Languages
National University "Odessa Polytechnic"
(Odessa, Ukraine) tanya47bor@gmail.com

Marina TSYNOVA,
orcid.org/0000-0003-2633-8416
Candidate of Historical Sciences, Associate Professor,
Associate Professor at the Department of Foreign Languages
National University "Odessa Maritime Academy"
(Odessa, Ukraine) marinatsynovaya@gmail.com

Yulia ERSHOVA,
orcid.org/0009-0006-4251-6891
Senior Lecturer at the Departments of foreign languages
Odessa Military Academy
(Odessa, Ukraine) shapa.od@gmail.com

STATISTICAL, LEXICAL AND GRAMMATICAL DIFFERENCES BETWEEN MODAL CONSTRUCTIONS WITH 'CAN/COULD' AND 'MAY/MIGHT'

The article considers the possibility to introduce in the process of teaching the English conversational speech the results of research in the field of corpus linguistics. The object of the study are modal constructions (MCs) with 'can/could' and 'may/might'. The choice of the grammatical topic is connected with the difficulty to understand the fundamental difference between the modal verbs 'can/could' and 'may/might' for the students of non-linguistic universities learning English as a foreign language. The goal is to provide real data (statistical, lexical and grammatical) that can help in solving the problem. The described research is based on the text corpora of the sublanguages of scientific and technical communication "Heating Engineering", "Electrical Engineering" and "Automotive" compiled by a continuous sampling. The text corpora used are formed on the basis of scientific articles in relevant fields of knowledge, published in journals in the UK and USA. The main topics of the texts that were selected for research are included in the discussion of technical inventions, phenomena, and characteristics of devices presented in the articles. The statistical characteristics of the modal constructions with 'may/might' and 'can/could' are different: the 'can/could' constructions surpass 'may/might' ones in both the number of units (20 and 13) and the total frequency of the use (1086 and 446, respectively). The correlation of values of 'can/could' and 'may/might' constructions show the surpass of 'can/could' over 'may/might' more than 1,5. This can be a statistical mark for English learners. The 'can/could' constructions are grammatically more various. The analysis of the lexis of the infinitives included in the both types of modal constructions has demonstrated that from this viewpoint there are some differences between them (modal constructions). The infinitive constituent in 'can/could' constructions possesses both commonly used and scientific lexical meanings. But in 'may/might' ones in most cases the infinitive is referred predominantly to the commonly used lexical layer.

Key words: *frequency, lexical layer, modality, text corpus, semantic structure.*

Тетяна БОРИСЕНКО,
orcid.org/0000-0002-9839-1761
кандидат філологічних наук, доцент,
доцент кафедри іноземних мов
Національного університету «Одеська політехніка»
(Одеса, Україна) tanya47bor@gmail.com

Марина ЦИНОВА,
orcid.org/0000-0003-2633-8416
кандидат історичних наук, доцент,
доцент кафедри іноземних мов
Національного університету «Одеська морська академія»
(Одеса, Україна) marinatsynovaya@gmail.com

Юлія ЄРШОВА,
orcid.org/0009-0006-4251-6891
старший викладач кафедри іноземних мов
Одеської військової академії
(Одеса, Україна) shapa.od@gmail.com

СТАТИСТИЧНІ, ЛЕКСИЧНІ ТА ГРАМАТИЧНІ ВІДМІННОСТІ МІЖ МОДАЛЬНИМИ КОНСТРУКЦІЯМИ З 'CAN/COULD' ТА 'MAY/MIGHT'

У статті розглядається можливість впровадження в процес навчання англійської розмовної мови результатів досліджень у галузі корпусної лінгвістики. Об'єктом дослідження є модальні конструкції (МК) із 'can/could' та 'may/might'. Вибір граматичної теми пов'язаний із складністю розуміння принципової різниці між модальними дієсловами «can/could» і «may/might» для студентів нелінгвістичних університетів, які вивчають англійську мову як іноземну. Мета статті – надати реальні дані (статистичні, лексичні та граматичні), які можуть допомогти у вирішенні проблеми. Описане дослідження базується на текстових корпусах підмов науково-технічної комунікації «Теплотехніка», «Електротехніка» та «Автомобілебудування» складено шляхом суцільної вибірки. Використані текстові корпуси сформовані на основі наукових статей у відповідних галузях знань, опублікованих у журналах Великобританії та США. Основні теми текстів, які були відібрані для дослідження, включені до обговорення технічних винаходів, явищ і характеристик пристроїв, представлених у статтях. Характеристики модальних конструкцій із can/could та may/might відрізняються за статистичними, граматичними та лексичними характеристиками. Конструкції з can/could перевершують конструкції may/might як за кількістю одиниць (20 та 13), так і за загальною частота використання (1086 і 446 відповідно). Кореляція величин конструкцій can/could і may/might демонструє перевищення can/could над may/might більш ніж у 1,5 рази. Це може бути статистичним показником для тих, хто вивчає англійську мову. Граматично більш різноманітні є конструкції can/could. Аналіз лексики інфінітивів, що входять до складу обох типів модальних конструкцій, показав, що з цієї точки зору між ними (модальними конструкціями) є певні відмінності. Інфінітивний компонент у конструкціях can/could має як загальновживане, так і наукове лексичне значення. Але в конструкціях з may/might в більшості випадків інфінітив відноситься переважно до загальновживаного лексичного шару.

Ключові слова: частота, лексичний шар, модальність, текстовий корпус, семантична структура.

Statement of the problem and literature review. In the context of constantly developing globalization and the presence of the Internet, when physical borders are no longer an obstacle to communication between citizens of different states, the ability to use the spoken language becomes relevant.

This concerns to a large extent the non-linguistic universities students-future engineers and scientists in the field of science and technology, because conversational skills will allow them to receive the necessary information in a timely manner, as well as be able to communicate with foreign colleagues at conferences and symposia, when discussing private engineering problems.

Teaching English as a foreign language to the non-linguistic university students, and in particular English conversation, is associated with quite great difficulties. Of course, specialized discourse cannot be compared with fiction discourse in its complexity, since it (specialized discourse) contains some conditions common to all technical fields of knowledge that significantly distinguish these two types. These are: narrowly professional aspects of speech behavior; strictly limited use of data from certain levels of language; greater emphasis on analysis rather than synthesis; higher degree of formalization; binding to a limited sublanguage (or complex of sublanguages).

But the base problem is as follows. The process of teaching the English conversational speech is already sufficiently provided with methodologically reliable literature, the amount of which continues to increase. However, with regard to such aspects of society as scientific, technical and industrial human activity and, accordingly, training future engineers in English (colloquial speech) for specialized purposes, it is covered in existing publications to a much lesser extent. We can give the most, from our point of view, useful and progressive examples of such methodological literature, because they include lexicographic information within phrases, which make it possible to record whole phrases for memorization, rather than individual words (Benson, E. Benson & Ilson, 1997; Wouden, 1992; Yorkey, 1969; Kjellmer, Altmann, 2005) which contributes to more rapid development of oral speech. Moreover, these sources present their data in conjunction with quantitative information, i.e. are based on real statistical calculations.

Another main task of modern methods of teaching the speech is the use of accurate and verified linguistic facts in educational process. What specific scientific research are meant? First of all, this concerns the results of the analysis of text corpora. The problem of combining theoretical issues of grammar and scientific research data, or rather, using the results of scientific research to explain issues of theoretical grammar, has long been overdue. Linguistic scientists call corpus linguistics a promising direction that can be widely used in explaining grammatical topics. They have a fairly reasoned approach to the use of this type of data, believing that when introducing theoretical material in schools and universities, artificially created examples or examples that rely solely on the intuition of native speakers cannot be given, while there exist real examples taken from texts for teachers and students (Barnbrook, 1998). Thus, T. Jones (Jones, 2002) directly introduced corpus research data into grammar and vocabulary lessons at the University of Birmingham. There are also attempts by teachers and other universities to apply research data from text corpora into practice. For example, J. Flowerdue (Flowerdue, 1993) believes that data obtained from the study of text corpora allows teachers to introduce exactly those words and situations in which they are used, and which students will later need to work in their subject area.

The younger generation of scientists also contributes to the description and definition of the very subject of corpus linguistics (Finegan, 2014; McEnery; Krishnamurthy, 2006). They convincingly demonstrate that the advantage of corpus linguistics also lies in the fact that it studies almost any speech phenomena that function in a specialized text of almost all types of discourse (scientific, technical or humanities): syntactic phrases of any type (phraseological combinations, multi-component constructions, any types of sentences, structural components); structural syntax; word-formation typology of any parts of speech; forms and functions of parts of speech. It also studies the semantic aspect of speech units, i.e. implementation of semantic definitions of words included or not included in standard dictionaries; traces changes in the semantics of words in the process of their implementation in specialized texts, etc. (Alhasov, Verbytska, & Kolenichenko, 2020). The listed scientific topics clearly show that virtually any theoretical topic can be confirmed by statistical data. This confirms the assertion that modern corpus linguistics is not limited to the compilation of corpora, it also involves large-scale research of languages based on a text corpus research of grammar and vocabulary.

However, not only in European and American universities much attention is paid to the results of the analysis of real texts describing various linguistic phenomena implemented in text corpora. Thus, in Ukraine, at the Odessa Polytechnic National University, the teachers of the Department of Foreign Languages have not only many years of experience working with scientific communication texts, but also the results of corpus research, experience in the formation of probabilistic statistical models and the use of frequency dictionaries. They also gradually introduce, along with theoretical grammar, into the process of teaching English the elements of structural linguistics, and specifically, structural syntax (Shapa, Tomasevich, Dantsetsvich, 2015; Borisenko, Tsynova, 2020).

The article goal. This work presents the results of a study of modal constructions (MCs) functioning in text corpora of several areas of scientific and technical discourse. The choice of the grammatical topic is due to the difficulty of understanding the fundamental difference between the modal verbs 'can/could' and 'may/might' for the students of non-linguistic universities learning English as a foreign language.

Therefore, the goal of the proposed article is the following: to provide real data (statistical, lexical and grammatical) that can help in solving the problem of creating the understandable and stable difference between the modal constructions with 'can/could' and 'may/might'.

Base material. The described research is based on the text corpora of the sublanguages of scientific and technical communication "Heating Engineering", "Electrical Engineering" and "Automotive" compiled by a continuous sampling. The text corpora used are

formed on the basis of scientific articles in relevant fields of knowledge, published in journals in the UK and USA: IEEE Transactions on Power Apparatus and Systems; Power Engineering; Power; Automotive News; Combustion; Control and Optimization; Machine Design; Machinery and Production Engineering; Automotive Engineer. The text corpus of each specialty contains 100 thousand tokens, and the total volume thus amounted to 300 thousand tokens.

The main topics of the texts that were selected for research are included in the discussion of technical inventions, phenomena, and characteristics of devices presented in the articles. It seems that in the future precisely such kind of topics will be necessary in the professional activities of future engineers and scientists.

As already mentioned, the object of the article is constructions with the modal verbs 'can/could' and 'may/might'.

First of all, we present the meanings that normative explanatory dictionaries offer for these modal verbs.

The content of the modal verb 'may/might' and its internal form are precisely analyzed in Oxford Advanced Learner's Dictionary by A. Hornby, where the semantic structure of the modal verb 'may/might' is presented in the following meanings: 2) to indicate permission or a request for permission; 3) to express desires and hopes (may), to express a request (might).

In the same normative explanatory dictionary the verb 'can/could' is fixed as a unit having the following set of modal "meanings": ability or opportunity; "permission" in everyday conversational style; probability and possibility of what is happening; in

the interrogative sentences it gives the shadow of the meaning directed on revealing of surprise, absence of attention; indicates what someone or something is considered possible for the existence or implementation; indicates what is considered to be typical.

The procedure of analysis began from distinguishing in the text corpora the most frequently used modal constructions with the modal words 'can/could' and may/might' to determine the differences of their statistical characteristics.

The obtained structures were classified according to the typological features of structural models with regard to their total absolute frequency (F*). The formalized representation of the structures was 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.

Total amount: 447 units;

As we see in Table 1, thirteen grammatical formulas present the occurrence of the modal verb 'may/might' in the chosen text corpora, the total amount of which is 447 units. The most frequent are 'may V' (292 units in all text corpora) and 'may be Ven' (90 units). These two constructions contain 382 units. This is about 86% of all the modal constructions with 'may/might'.

The category of modality expressed by these formulas appears in reproducible syntactic unities only in one of the possible meanings – "to indicate permission or a request for permission" in various situations, which are reflected in the meanings of the combined lexical component with the verb 'may', for example, *resonance may influence, overvoltages*

Table 1

Modal constructions with 'may/might' in the text corpora

№№	Construction	Area of engineering		
		Power Engineering, F	Electrical Engineering, F	Automotive F
1.	May V	125	43	34
2.	May be Ven	40	42	8
3.	May be A	11	8	2
4.	May be Ven to V	11	17	5
5.	Might V	9	5	4
6.	May D be Ven	9	1	1
7.	Might be Ven	7	5	2
8.	May be Ven D	7	2	1
9.	May be A to V	6	4	3
10	May be D Ven	5	4	3
11.	May not V	4	1	2
12.	May V to V	4	1	3
13.	May not be A	4	4	-

F* 242 137 68

may exist, excitation may result, voltage may occur, etc. All syntactic structures of the type 'may V' and its passive variant 'May be Ven' exhibit the "affinity" in correlating their meaning with the extra-linguistic situation. Thus, according to this models, these two types of modal constructions are reproduced in the meaning: 1) to indicate permission to an inanimate object to perform an action on another object (70% of all analyzed modal phrases of this type), for example, *may cause; may increase; may mix; may effect; may hit; may achieve,* etc. 2) to permit the subject to perform an action on an object, for example, *may debate, may wonder, may write, may suggest, may use, may explain,* etc.

We can see that only two constructions with the verb 'might' (Might V and Might be Ven) function in the text corpora, they have only 7% of all the use of MCs. They are also presented in active and passive variants, for example, *might ultimately facilitate; might eventually change; might also require.* The base syntactic difference between the constructions with 'may' and 'might' is the presence of an adverb.

The grammatical formulas 'may/might + adjective' and 'may/might be A to V' have 8% of all the units. The reproducible modal meaning in such modal constructions is the one "presumability" or "possibility" (actual or theoretical), for example, *may be available, may be true, may be impossible, may be pos-sible, may be significant. etc.; may be easier to talk; may be possible to relate; may be essential to use, etc.*

We see that the last two models implement the modal meaning "presupposition" reasoning about the static nature of the characteristics manifested by an object or phenomenon, which can relate to the entire statement as a whole.

The next modal constructions we are going to consider the ones with 'can/could'. First of all we demonstrate the grammatical formula of the most frequent constructions functioning in the text corpora compiled.

Total amount: 1086 units

The entire list of the constructions with 'can/could' significantly exceeds the one with the verbs 'may/might', it includes 20 different constructions with total amount 1086 units (in the percentage correlation with 'may/might' list the modal verb 'can/could' use surpasses about 1,5).

The study shows that the highest priority is possessed by the verb constructions which have the forms of the passive infinitive. There appeared to be only 8 of them, but their total frequency (594 usages) covered more than a half (54%) of all tokens of modal constructions. They show almost all methods of variation of constituents in syntagmatic text corpora. The highest total absolute frequency is possessed by 'can be Ven' (F*=481) construction, for example, center

Table 2

Modal constructions with 'can/could' in the text corpora

№№	Construction	Power Engineering, F*	Electrical Engineering, F*	Automotive, F*
		Area of engineering		
1.	can be Ven	228	175	78
2.	can V	113	22	96
3.	could V	28	24	31
4.	could be Ven	18	27	20
5.	can V N	19	22	10
6.	can V prp	11	10	5
7.	cannot be Ven	11	7	6
8.	can be A	11	8	2
9.	cannot V	11	2	6
10.	can be N	1	14	2
11.	can be Ven to V	1	4	10
12.	could be A	1	4	5
13.	could be N	2	2	4
14.	can have N	3	3	2
15.	could have Ven	3	1	4
16.	could not V	-	1	6
17.	can V N prp	5	-	1
18.	cannot V prp	-	2	3
19.	could not be Ven	1	1	2
20	can V to V	1	-	2

F* 462 329 295

can be located, the devices can be interconnected. It accounts for 88% of all tokens of modal constructions with the passive infinitive. Within this group the vast majority of structures – 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 appeared on the second place as to their frequency of usage (F*=361, which is 33%) for example, *motor can run; turbine can generate.* Data of the table show that these aspectual-temporal forms of the infinitive are diversified enough in this type of the voice.

The results of the contextual analysis of text bodies of "Power Engineering", "Electrical Engineering" and "Automotive" sublanguages, which fragments are presented in the examples, show that the verb 'can' implements the only modal meaning of "physical ability to do something". 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.), and not any additional semantic (connotative) features are added to the modal meaning of the mentioned above structural types.

The lexical characteristics of the infinitives included in the modal constructions with 'may/might' and 'can/could' as semantic constituents are as follows. The most verbs with the highest frequent values are the commonly used ones, for example, 'use, make, see, take, find, to meet, do, occur, achieve', etc. And they are equally used both in 'may/might' and 'can/could' constructions. The infinitives in the 'can/could' constructions can be quite often used with the units which are referred to the scientific layer of the vocabulary, which are applied for the engineering phenomena description, for example, 'determine, reduce, cause, design, operate, control, produce, calculate', etc. While the infinitives in 'may/might' constructions which are referred to this layer according

their semantics occur very rare, they are for example, 'derive, program, generate, select', etc. Their frequency values are quite high.

So the methods for constructing models of the 'can V' (can be Ven) and 'may V' (may be Ven) types can be considered to be identical. In them the verbs 'can' and 'may' should be supposed the functional substitutes, but, as the study of constructions shows, only if they are used with the same verb or with lexemes that are interchangeable in meaning. For example, in the article about an engine the following synonymous constructions are implemented: 'can run – may work, can work – may run'. a significant number of syntactic constructions with the verb 'can' realize the same meaning – "ability", and this demonstrates the "freedom" of choice made by the author, means of expression from the language system in speech. Consequently, constructions with different lexical content, built according to the same model, can have the same content plan and reproduce the same modal meaning.

Conclusions. The analysis presented above allowed to draw the following conclusions.

1. The statistical characteristics of the modal constructions with 'may/might' and 'can/could' are different: the 'can/could' constructions surpass 'may/might' ones in both the number of units (20 and 13) and the total frequency of the use (1086 and 446, respectively). The correlation of values of 'can/could' and 'may/might' constructions show the surpass of 'can/could' over 'may/might' more than 1,5. This can be a statistical mark for English learners.

2. As it as mentioned the 'can/could' constructions are grammatically more various.

3. The analysis of the lexis of the infinitives included in the both types of modal constructions has demonstrated that from this viewpoint there are some differences between them (modal constructions). The infinitive constituent in 'can/could' constructions possesses both commonly used and scientific lexical meanings. But in 'may/might' ones in most cases the infinitive is referred predominantly to the commonly used lexical layer.

BIBLIOGRAPHY

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

2. Ton van der Wouden Prolegomena to a Multilingual Description of Collocations. EURALEX'92 I-II., Proceedings, Tampere: University of Tampere, Finland, 1992. p.p. 449–456.

3. Yorkie R. Which Desk Dictionary Is Best for Foreign Students of English? Havaii: TESOL Quarterly, 1969. vol. 1, № 3.

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

5. Barnbrook G. Language and computer. A practical introduction to the computer analysis of language. Edinburgh: Edinburgh University press, 1998. 209 p.

6. Johns T. Data-driven learning: the perpetual challenge. In: Kittemann/Marko, 2002. P. 107–117.

7. Flowerdew J. Concordancing as a tool of course design. *System*. Vol. 21, Issue 2, May 1993, P. 231–244.

8. McEnery T. Corpus Linguistics: Method, Analysis, Interpretation. Lancaster University, URL: www.futurelearn.com/courses/corpus-linguistics

9. Finegan, E. Language: its structure and use (7th ed.). Boston, MA: Cengage Learning, Inc., 2014. 613 p.

10. Alhasov, Y., Verbytska, A., & Kolenichenko, T. (2020). Teaching English to adult learners within extracurricular activities at university: barriers and motivation factors. *Advanced Education, 7*(15), 12–19. https://doi.org/10.20535/2410-8286.195696

11. Borisenko T. I., Tsynova M.V. Negative modal verb constructions in the text corpora of scientific and technical discourse. *Південний архів (філологічні науки)*. Херсон, 2020. № 81/2020. С. 61–65.

12. Krishnamurthy R. Corpus Lexicography. *Encyclopedia of Language & Linguistics.* 2006. 250-254. DOI:10.1016/B0-08-044854-2/00416-8

13. Shapa, L.N., Tomasevich, N.P., & Dantsevich L.G. Terminologization of adjectives in the texts of scientific communication (on the material of the sublanguages of Electrical Engineering). *Вісник харківського національного університету імені В.Н. Каразіна.Серія «Філологія»*. 2015. Вип. 73. С. 172–179.

REFERENCES

1. Benson M., Benson E & Ilson R. (1997) The BBI combinatory dictionary of English: a guide to word combinations, Amsterdam-Philadelphia

2. Ton van der Wouden (1992) Prolegomena to a Multilingual Description of Collocations. EURALEX'92 I–II., Proceedings, Tampere: University of Tampere, Finland, p.p. 449–456.

3. Yorkie R. (1969) Which Desk Dictionary Is Best for Foreign Students of English? Havaii: TESOL Quarterly, vol. 1, № 3.

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

5. Barnbrook G. (1998) Language and computer. A practical introduction to the computer analysis of language. Edinburgh: Edinburgh University press,. 209 p.

6. Johns T. (2002) Data-driven learning: the perpetual challenge. In: Kittemann/Marko, P. 107–117.

7. Flowerdew J. (1993) Concordancing as a tool of course design. *System*. Vol. 21, Issue 2, May P. 231–244.

8. McEnery T. Corpus Linguistics: Method, Analysis, Interpretation. Lancaster University, URL: www.futurelearn.com/courses/corpus-linguistics

9. Finegan, E. (2014) Language: its structure and use (7th ed.). Boston, MA: Cengage Learning, Inc., 2014. 613 p.

10. Alhasov, Y., Verbytska, A., & Kolenichenko, T. (2020). Teaching English to adult learners within extracurricular activities at university: barriers and motivation factors. *Advanced Education, 7*(15), 12–19. https://doi.org/10.20535/2410-8286.195696

11. Borisenko T. I., Tsynova M.V. (2020) Negative modal verb constructions in the text corpora of scientific and technical discourse. Pivdennyi arkhiv (filolohichni nauky). Kherson, № 81/2020. С. 61–65.

12. Krishnamurthy R. (2006) Corpus Lexicography. *Encyclopedia of Language & Linguistics.* 250-254. DOI:10.1016/B0-08-044854-2/00416-8

13. Shapa, L.N., Tomasevich, N.P., & Dantsevich L.G. (2015) Terminologization of adjectives in the texts of scientific communication (on the material of the sublanguages of Electrical Engineering). Visnyk kharkivskoho natsionalnoho universytetu imeni V.N. Karazina.Seriia «Filolohiia». № 73. С. 172–179.