UDC 81'373.611
DOI https://doi.org/10.24919/2308-4863/75-2-28

Maria NEVREVA,
orcid.org/0000-0001-9923-1041
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Foreign Languages Department
National University "Odesa Polytechnic"
(Odesa, Ukraine) m.n.nevreva@gmail.com

Olena LEBEDEVA,
orcid.org/0009-0000-2419-0804
Senior Lecturer at the Foreign Languages Department
National University "Odesa Polytechnic"
(Odesa, Ukraine) ellebedeva@gmail.com

Olga GVOZD,
orcid.org/0000-0001-9882-8890
Senior Lecturer at the Foreign Languages Department
National University "Odesa Polytechnic"
(Odesa\, Ukraine) gvozdiki24@gmail.com

STATISTICAL CHARACTERISTICS OF ROOT MORPHEMES IN TEXTS OF SCIENTIFIC AND TECHNICAL DISCOURSE

The article examines the units related to the field of word-formation typology – root morphemes that form nouns. The main linguistic space considered in which nouns function, the root morphemes of which are analyzed, is the technical texts of three specialties related to scientific and technical discourse: "Chemical Engineering", "Automotive Engineering", "Electrical Engineering". These three specialties have practically no common scientific topics, so the generalized results obtained can probably be used in the study of word-forming units for any area of scientific and technical discourse. The texts were taken from the scientific articles in journals of relevant fields of technology from the UK and the USA: Chemical Engineering Progress; Process Engineering; Machinery; Chemical and Process Engineering; Automobile Engineering, Auto Industry; IEEE Transactions on Power Apparatus and Systems, Proceedings of the Institution of Electrical Engineers. Based on these texts, three text corpora "Chemical Engineering", "Automotive Engineering", "Electrical Engineering" were compiled. The size of text corpora is 200 thousand tokens for the specialties "Chemical Engineering" and "Electrical Engineering" and 300 thousand tokens for the specialty "Automotive Engineering". Then the frequency dictionaries were created on their basis. The object of the analysis was low-productive and non-productive root morphemes, namely, their statistical characteristics. It was determined that there is some statistical balance between such parameters as productivity, the number of nouns formed by adding affixes to root morphemes and the frequency of their use in texts. For example, low-productive and non-productive root morphemes form nouns that function in texts with a very high frequency of use (no lower than highly productive root morphemes); the number of nouns is one and a half to two times greater than that of highly productive morphemes. The article also gives the reason for the lag in the quantitative indicators of root morphemes from the statistical values of other affixes.

Key words: *frequency, productivity, quantitative values, frequency dictionary, text corpus.*

Марія НЕВРЕВА,
orcid.org/0000-0001-9923-1041
кандидат філологічних наук, доцент,
доцент кафедри іноземних мов
Національного університету «Одеська політехніка»
(Одеса, Україна) m.n.nevreva@gmail.com

Олена ЛЕБЕДЄВА,
orcid.org/0009-0000-2419-0804
старший викладач кафедри іноземних мов
Національного університету «Одеська політехніка»
(Одеса, Україна) ellebedeva@gmail.com

Ольга ГВОЗДЬ,
orcid.org/0000-0001-9882-8890
старший викладач кафедри іноземних мов
Національного університету «Одеська політехніка»
(Одеса, Україна) gvozdiki24@gmail.com

СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ КОРЕНЕВИХ МОРФЕМ У ТЕКСТАХ НАУКОВО-ТЕХНІЧНОГО ДИСКУРСУ

Стаття розглядає одиниці, які стосуються області словотворчої типології – кореневі морфеми, що утворюють іменники. Основний досліджуваний лінгвістичний простір, в якому функціонують іменники, кореневі морфеми яких аналізується, є технічні тексти трьох спеціальностей, що належать до науково-технічного дискурсу: «Хімічне машинобудування», «Автомобілебудування», «Електротехніка». Дані три спеціальності практично не мають загальних наукових тем, тому отримані узагальнені результати можуть бути використані в дослідженні словотвірних одиниць для будь-якої галузі науково-технічного дискурсу. Тексти були взяті з наукових статей у журналах відповідних областей техніки Великобританії зі США: Chemical Engineering Progress; Process Engineering; Machinery; Chemical and Process Engineering; Automobile Engineering, Auto Industry; IEEE Transactions on Power Apparatus and Systems, Proceedings of the Institution of Electrical Engineers. На базі цих текстів було скомпільовано три текстові корпуси «Хімічне машинобудування», «Автомобілебудування», «Електротехніка». Обсяг текстових корпусів був по 200 тис. слововживань у спеціальностей «Хімічне машинобудування» та «Електротехніка» та 300 тис. слововживання у спеціальності «Автомобілебудування». Далі на їхньої основі було створено частотні словники. Об'єктом аналізу стали малопродуктивні і непродуктивні кореневі морфеми, саме – їх статистичні характеристики. Було визначено, що існує певний статистичний баланс між такими параметрами як продуктивність, кількість іменників, утворених за допомогою приєднання афіксів до кореневих морфем, та частотністю їх вживання в текстах. Так, наприклад, малопродуктивні та непродуктивні кореневі морфеми утворюють іменники, які функціонують у текстах з дуже високою частотою вживання (не нижче, ніж високопродуктивні кореневі морфеми), але кількість іменників у півтора-два рази більша, ніж у продуктивних морфем. У статті також наведено причину відставання кількісних показників кореневих морфем від статистичних величин інших афіксів.

Ключові слова: *частота, продуктивність, кількісні величини, частотний словник, текстовий корпус.*

Statement of problem. Literature review. The analysis of units that have a low level of statistical capabilities, for example, low frequency of use or low productivity, is considered in statistical linguistics to be no less important and interesting than the study of units included in the list of high productivity. This can be explained by the fact that, firstly, units operating with low frequency or low productivity, in terms of the quantitative parameter of uses in texts, far exceed this statistical characteristic of high-frequency units; secondly, the dialectics of changes occurring in systems and objects (including linguistic ones) is such that those which yesterday or currently shows very little frequency of use, after some time becomes extremely in demand, and the previous characteristics and concepts that were considered preferable for use, go to the periphery of public consciousness and life (Barber, 1964), (Bybee, 1985). Therefore, the results of analysis of units with insignificant statistical characteristics can provide the basis for their further study and use by theoreticians and practitioners specializing in the following humanities: linguistics of texts of any type of discourse (Wales, 2014); a methodology for teaching foreign languages, which provides for the presentation of not only the most frequent text units, but also less frequent ones (Перебійніс, 2008); statistical linguistics and general statistics (Dornyei, 2007).

However, no data was found in the available literature on the analysis of morphemes (and root morphemes in particular) with low productivity found in text corpora of technical specialties. Therefore, such a linguistic topic represents an undoubted novelty in the field of word-formation typology.

Goal of the article. The purpose of this article is to describe the statistical characteristics of root morphemes that are found in texts as part of lexemes derived from the root stem, i.e. the article considered only units that were subjected to the derivation process. By statistical characteristics we mean the following: 1) frequency of use of root morphemes in texts; 2) the value of the total number of root morphemes; 3) the number of lexemes formed by attaching various affixes or also root elements to root morphemes; 4) productivity, i.e. the number of new words (derivatives) that a root morpheme was capable of creating by adding affixal units and which are fixed in the compiled texts.

Since root morphemes with high productivity have already been considered in other articles by the authors (Nevreva, Grodskaya, Sirotenko, 2019) this article is devoted to presenting the results of the anal-

ysis of low-productive and unproductive root morphemes. Productivity was calculated as follows: productive root morphemes are considered to be capable of forming from 5 to 10 lexemes; unproductive root morphemes form 3–5 lexemes; non-productive roots form 1 lexeme.

Texts from three technical specialties that relate to scientific and technical discourse were taken as research material. A few words about the text corpora themselves. They were created on the basis of scientific articles in the following three technical areas of knowledge: "Chemical Engineering", "Automotive Engineering", "Electrical Engineering". The articles used for the text corpora were published in USA and UK scientific journals: Chemical Engineering Progress; Process Engineering; Machinery; Chemical and Process Engineering; Automobile Engineering, Auto Industry; IEEE Transactions on Power Apparatus and Systems, Proceedings of the Institution of Electrical Engineers. The size of the text corpora "Chemical Engineering" and "Electrical Engineering" is 200 thousand tokens, the one of "Automotive Engineering" corpus was slightly larger and amounted to 300 thousand tokens.

It should be noted that these three specialties were not taken arbitrarily, but for a specific purpose – to show the results of statistical analysis obtained on the basis of texts, the specialties of which are completely unrelated to each other in their topics. Of course, one could argue that automotive texts may include some electrical engineering concepts since a car has an electrical part (drive). However, the corpus on electrical engineering itself consists of texts, the scientific direction of which is mainly related to electrical management, the main task of which is the delivery of electricity to industrial or economic facilities, as well as consumption problems.

Thus, we can argue that thematically these three text corpora are independent units that do not have common or related topics. The need for such text objects is associated with the authors' decision to obtain the results of a statistical analysis of a generalized type, which may be suitable for texts of any specialty referred to scientific and technical discourse.

The problem with different sizes of text corpora was solved with the introduction of another additional statistical operation into the process of statistical analysis – converting absolute frequencies into relative ones and presenting data as percentages.

First of all based on all three text corpora the corresponding frequency dictionaries were created, which are a list of word forms used in texts and arranged in descending order of frequency. It was word forms that were represented in the dictionaries, and not words into which all forms of a word are usually combined, as is normally recorded in dictionaries, which allows researchers to analyze the whole variety of forms that function in real texts.

For the convenience of carrying out the future analysis the frequency lists are divided into high-frequency (from the beginning of the list to frequency 7) and low-frequency zones (from 7 to 1). This makes it possible to determine the most reliable, from a statistical point of view, values of tokens found in texts.

Base Material. This part of the article is devoted to the actual description of low-productive and unproductive root morphemes, recorded in both high-frequency and low-frequency zones of frequency dictionaries of three specialties. First of all, it should be noted that many root morphemes that did not show high productivity in the high-frequency zone turned out to be productive in the low-frequency zone, for example, in the corpus of the specialty "Chemical Engineering" the root morpheme 'flow' is registered in two lexemes: flow, overflow. In the low-frequency zone there are five lexemes: 'underflow, counterflow, inflow, microflow, outflow'. In addition, the lexeme 'flow' is included in 14 composites. Representing highly specialized terms, for example, 'air-flow, plug-flow, flow-path, radial-flow, flow-splitter, contraflow, crossflow, flowmeter, flow-properties, flow-ratio, free-flow, parallel-flow, back-flow, steady-flow'.

The statistical characteristics of quantity, productivity and frequency of occurrence of low-productive root morphemes recorded in all zones (high-frequency and low-frequency) of the dictionaries are presented in the table.

The data showing in the table can interpreted in the following way. The lowest occurrence of root

Table 1

Low-productive root morphemes in the texts corpora and their (morphemes) percentage

Text corpora	Quantity		
	Root morphemes (%)	The lexemes with root morphemes (%)	Frequency of use of lexemes with root morphemes (%)
Chemical engineering	121 (0,06%)	254 (0,127%)	16785 (8,39%)
Automotive	125 (0,042%)	335 (0,112 %)	26965 (8,99%)
Electrical engineering	109 (0,055 %)	231 (0,116%)	20972 (10,49%)

morphemes is observed in the texts on automotive industry – only 0.042%. In the text corpora on chemical engineering and electrical engineering the frequency of occurrence is almost similar– 0.055% – 0.06%. The values that show the number of lexemes formed on the basis of root morphemes, i.e. exactly the productivity of root morphemes, practically repeat the order in which the previous values are located, demonstrating the number of root morphemes, and are presented in the following order: the largest number is found in the texts on chemical engineering, in the second place there are morphemes functioning in the texts on electrical engineering, the smallest number of lexemes with root morphemes are observed in the texts on automotive industry.

Based on the data of the number of lexemes formed by attaching affixes to root morphemes (productivity), we can conclude that the authors describing technical phenomena and new inventions, for example, in the field of chemical engineering, do not introduce new terms into the lexicon, but strive to use the old, understandable to all specialists, derivatives with a basic element having a "core" meaning. Conversely, the authors of scientific articles describing inventions in the field of automotive industry prefer to introduce completely new units that have not been used before into the terminology system of this specialty.

It would seem that the order represents a completely stable picture of the distribution of root morphemes across the text corpora under consideration. However the last value which characterizes the frequency of use of lexemes with root morphemes occurred in the text corpora, presents unexpected data. The text corpus on electrical engineering takes the first place in terms of frequency of repetition of lexemes with a significant gap from the other two corpora (10,49%). In the second place there is the text corpus "Automotive Engineering" (8.99%), and in third and last place is the corpus on chemical engineering (8.39%), which was the leader in previously presented values.

First of all, let's consider the nomenclature of root morphemes recorded in the high-frequency zone of the frequency dictionary. Here the low-productive root morphemes account for a total of about a third of the inventory of this type of morpheme in the high-frequency zone in each of the text corpora, with the number of low-productivity roots varying across corpora. Thus, in the text corpus on chemical engineering, 13 high-frequency root morphemes are registered: 'flow, press, vapor, heat, carbon, pipe, form, dense, product, tube, process, case, figure'. In the texts on the specialty "Automotive Engineering" there are 32 root morphemes: 'line, type, load, control, time, engine, car, product, weight, air, combustion, drive, point, sign, test, gear, part, serve, way, wheel, motor, speed, force, operate, power, condition, cylinder, use, cam, emission, heat, percent'. The electrical engineering corpus contains 15 high-frequency root morphemes: 'volt, line, conduct, load, meter, break, resist, test, fact, model, pulse, station, figure, insulate'.

However the bulk of root morphemes, which are part of low-nproductivity roots, have an average frequency of use; they are also included in the high-frequency zone of dictionaries. In the corpus on chemical engineering, their number is 105 units, in the texts on automotive engineering – 93 morphemes, in the texts on electrical engineering – 85 units.

In the low-frequency zone of the dictionaries there is a very small number of root morphemes: in the dictionary "Chemical Engineering" – only 3 morphemes, in the dictionary "Electrical Engineering" – 9 morphemes, in the frequency dictionary "Automotive Engineering" they are absent.

If we consider the entire composition of low-productivity root morphemes from the point of view of their lexical meanings, we can note that they relate mainly to the stylistically neutral and general scientific layers of vocabulary, for example, '*turn, change, effect, place, fact, clear, differ*', etc. However these roots, when combined with affixes and other root morphemes, are capable of producing terms including highly specialized ones. For example, the root morpheme 'change' forms the following word-terms: '*exchanger, interchanging, change-out, change-over; root morpheme 'break' terms 'breaker, breakage, break-down, break-up'*; the root morpheme 'clear' produces the terms '*clearance, closed-clearance, clearance-porket'*, etc. The number of low-productivity lexeme-terms is not very large; in the high-frequency zones of the three frequency dictionaries their total shares are registered from 17.5% to 23%.

The next question concerns non-productive morphemes found in the texts of the three specialties. As noted above, non-productive root morphemes are units that form no more than one lexeme. This type of root morphemes turned out to be the majority in the texts studied – 65.7–68.3% – of the total number of root morphemes. The statistical parameters of non-productive root morphemes are presented in the table, and the statistical indicators (as already mentioned in the work conditions) here were taken from the total number of word forms functioning in the texts, i.e. 200 thousand or 300 thousand ("Automotive") tokens.

If to compare the data in this table to the data in Table 1, we can note undoubtedly more significant

Table 2

Non-productive root morphemes in the texts corpora and their (morphemes) percentage

Text corpora	Quantity		
	Root morphemes (%)	The lexemes with root morphemes (%)	Frequency of use of lexemes with root morphemes (%)
Chemical engineering	232 (0,12%)	232 (0,12%)	14195(7,1%)
Automotive	273 (0,09%)	273 (0,09%)	18855 (6,3%)
Electrical engineering	238 (0,12%)	238 (0,12%)	15758 (7,9%)

quantitative values, which demonstrate, firstly, the presence of a larger number both in absolute frequencies and in percentage terms (more than twice) of the inventory of root morphemes, recorded in the list of non-productive morphemes of this type. Moreover, there is a coincidence of numerical values in the texts on chemical engineering and electrical engineering, and a slightly smaller value for root morphemes found in the texts on automotive engineering. Secondly, some clarification is needed so that there are no misunderstandings with the understanding of the same numerical values in the indicators of non-productive root morphemes and nouns derived from them – they are the same, since non-productive root morphemes are capable of creating no more than a single noun. And, finally, the frequency of use of non-productive morphemes is quite high, practically no lower than that of low-productivity morphemes, because all non-productive morphemes are realized in texts with a very high frequency of use and are located in the high-frequency zone of the frequency dictionaries.

Thus, there is a certain balance between the very insignificant amount of root morphemes realizing their function of forming derivatives and the very large amount of frequency of use in the texts, which shows a certain compensatory ability in non-productive morphemes. Thus, among non-productive morphemes in the texts on chemical engineering, 10 high-frequency ones are identified: 'type, liquid, size, corrosion, rate, acid, time, condition, equip, unit'; in the texts of the specialty "Automotive Engineering" there are 17 units of such morphemes: 'valve, road, level, brake, develop, model, system, fuel, industry, vehicle, flow, automobile, fluid, rear, steel, economy, standard'; in the electrical engineering texts – 6 morphemes: 'current, time, circuit, system, case, condition'.

Medium-frequency non-productive roots are registered: 127 units in the frequency dictionary for chemical engineering; in the automotive industry 125 units; in electrical engineering 191 units. The remaining non-productive roots have a low frequency of occurrence.

Comparison of the inventories of root morphemes of the three frequency dictionaries under study allowed us to identify 153 identical roots. Eighteen of them are low-productivity, for example, '*charge, control, face, heat, limit, line, lead, meter'*, etc. They form 28 lexemes, which are the same for frequency dictionaries. In addition, the inventory of identical root morphemes includes 54 common non-productive roots, for example, *'character, class, complex, function, metal, motor, strength, plant'*, etc., which form 51 lexemes common to all three text corpora, for example, '*characteristic, complexity, functionality, metallurgy*', etc.

The productivity of other common root morphemes manifests itself in the texts of different specialties in different ways. So, for example, the root 'mix' in the texts on chemical engineering is a low-productivity one, and in the texts on automotive and electrical engineering it is non-productive in general; the root 'ground' is of low productivity in the texts on automotive and electrical engineering, but has the rank of non-productivity in the texts on chemical engineering.

In general 81% of the same root morphemes for the three text corpora form lexemes of stylistically neutral and general scientific vocabulary, for example, '*process, instrument, search, space'*, etc.; the terms are created by 19% lexemes, for example, '*curve, engine, pulse, metal, load, tube'*, etc.

It would be of interest to conduct a comparative analysis of the inventory lists of nouns implemented in the three text corpora under study, because knowledge of the words of one frequency dictionary of any specialty can significantly facilitate the task of a recipient to master the basic vocabulary units of texts of other specialties. Thus, as a result of comparison of inventory lists of nouns, 355 non-derivative and derived units were identified common to the three frequency dictionaries, as well as 124 words common to the texts on chemical engineering and automotive engineering, as well as 110 words common to the texts on chemical engineering and electrical engineering and 99 nouns for specialties "Automotive Engineering" and "Electrical Engineering".

So the common nouns make up 23-24% of the lexicon of each of the three corpora. If we take into

account the common words of other parts of speech, as well as common words from the class of words-filler (prepositions, conjunctions, articles, pronouns, etc.), then the proportion of common words will be much higher. More or less common vocabulary for technical specialties is inevitable, since they are based on both a single language system and more or less general technical concepts.

Conclusions. Thus all of the above allows us to draw the following conclusions. Root morphemes do not exhibit high combinative ability. The low productivity of root morphemes is a natural phenomenon and can be explained as follows. The root morpheme, as the main part of the word, is obligatory for every word, while, for example, affix morphemes are not an obligatory part of the word. A root morpheme can exist independently, without being combined with other morphemes, and in the text corpora under consideration, most words are used without derivational morphemes. The sets of prefixes and suffixes are relatively limited and vary slightly between corpora in both composition and productivity. Consequently, the same prefix or suffix morpheme can theoretically combine with an infinite number of root morphemes, while the same root morpheme does not have such capabilities.

The results of the analysis of quantitative characteristics suggest that there is a certain statistical balance between the very insignificant amount of implementation by root morphemes of their function of forming derivatives and the very large value of frequency of their use in texts, which is demonstrated on the basis of unproductive units (in our case, in unproductive and unproductive root morphemes) some compensatory ability.

The results obtained make it possible to continue the research begun in the field of word-formation typology and move on to the analysis of determining the differences in statistical characteristics of nouns formed from both noun and verbal root morphemes.

BIBLIOGRAPHY

1. Barber Ch. L. Changes in Present-Day English. Edinburgh and London, 1964.
2. Bybee J. L. Morphology: A study of the relation between meaning and form. Amsterdam: Benjamins, 1985. 235 p.
3. Wales K. The Stylistic tool-kit: Methods and sub-disciplines. *Cambridge of handbook of stylistics.* Cambridge: Cambridge University Press, P.Stochwell & S.Whiteley. 2014. 13–31.
4. Перебійніс В. І. Морфологія англійського дієслова: система та функціонування. Київ, 2008. 545 с.
5. Dornyei Z. Research methods in applied linguistics. Oxford: Oxford University Press, 2007. Pp. 336.
6. M. N. Nevreva, E. B. Grodskaya, T. V. Sirotenko Genesis of nominal root morphemes in the English text corpora of scientific and technical discourse. *Вісник Маріупольського державного університету. Серія Філологія.* Маріуполь, 2019. Вип. 20. С. 225–233.

REFERENCES

1. Barber Ch. L. (1964) Changes in Present-Day English. Edinburgh and London. Changes in Present-Day English. [Changes in Present-Day English]. [in English].
2. Bybee J. L. (1985) Morphology: A study of the relation between meaning and form. Amsterdam: Benjamins. 235 p.
3. Wales K. (2014) The Stylistic tool-kit: Methods and sub-disciplines. Cambridge of handbook of stylistics. Cambridge: Cambridge University Press, P.Stochwell & S.Whiteley. 13–31.
4. Perebiinis V. I. (2008) Morfolohiia anhliiskoho diieslova: systema ta funktsionuvannia. [English verb vorphology: system and functioning]. Kyiv. 545 s. [in Ukrainian]
5. Dornyei Z. (2007) Research methods in applied linguistics. Oxford: Oxford University Press, 2007. Pp. 336.
6. M. N. Nevreva, E. B. Grodskaya, T. V. Sirotenko (2019) Genesis of nominal root morphemes in the English text corpora of scientific and technical discourse. Visnyk Mariupolskoho derzhavnoho universytetu. Seriia Filolohiia. – Bulletin of Mariupol State University. Series Philology, 20, 225–233.