Nevreva M. M.,
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Foreign Languages Department
Odesa National Polytechnic University

Diachenko G. F.,
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Foreign Languages Department
Odesa National Polytechnic University

Zaitseva O. Yu.,
Candidate of Philological Sciences, Associate Professor,
Professor at the Foreign Languages Department
Odesa State Academy of Building and Architecture

NOMINAL WORD FORMATION MODELS IN THE TEXT CORPORA OF SCIENTIFIC DISCOURSE

Summary. The article discusses the possibility of introducing the results of scientific research carries out on the basis of a survey of real text corpora into the educational process. As an object the English word-formation models functioning in the texts of scientific discourse, which belong to three technical specialties – "Chemical Engineering", "Automotive" and "Electrical Engineering" – were taken. They (specialties) belong to the areas that are different in their scientific and technical topics which allows to draw conclusions about the word formation characteristics that are common to the texts of scientific discourse. Corpora are compiled on the basis of original scientific articles published in American and English scientific journals of the relevant technical fields. In the texts 16 derivational types of nouns were registered (9 mono- and 7 poly-lexeme), however, a comparison of the typology of names allows us to single out only 10 of the most frequent and productive ones which are simultaneously common for the three text corpora: six mono-lexemic R, RS, PR, PRS, RSS, PPRS and four poly-lexemic R + R, RS + R, R + RS, R + R + R. The number of three-word units is so small that it is believed that they are not inherent in the scientific functional style. Mono-lexeme types of nouns with a simple and simple derivative structure (R, RS, PR, PRS) occupy a dominant position in the corpora under study for such important didactical and statistical characteristics as frequency of occurrence and productivity. Their percentage of all word-formation models is more than 90%. The results of studies devoted to the analysis of word-formation models in text corpora of scientific and technical discourse are valuable methodological material for introduction into the training process when teaching English for special purposes, since they represent the results of a survey of real texts and are statistically reliable. This makes it possible not only to enumerate word-formation units and models that generally exist in word-formation system of the English language which was practiced earlier and is being practiced now but to point to real word-formation units used in technical literature.

Key words: text corpus, frequency of usage, productivity, scientific discourse, poly-lexeme units, mono-lexeme units.

Statement of the problem. Most of the educational literature in English used at technical universities of Ukraine for training English the students of technical specialties for special purposes, i.e. the language that they will use during the period of their professional activity, is intended mainly for the so-called passive proficiency of foreign language. For example, to understand the main idea and recognize the information presented in the text in the process of discussing it in the native language; to understand complicated messages used in professional and academic environments, etc. Since the texts are usually taken from the Internet, students of technical universities in Ukraine have a fairly clear idea of the modern directions of scientific research in a particular technical field.

However, future specialists-graduates of technical universities of Ukraine, who seek to start their own scientific developments, present their results in articles, as well as correspond and communicate with colleagues from other countries at conferences and symposia, realize that this is only the first, initial, stage of mastering English, which needs to be further improved and developed. And here they can be helped by research that makes it possible to describe the process of scientists' thinking accurately, i.e. the same frequent semantic, grammatical and statistical standards that are used in writing the scientific articles and with the help of which one can potentially simulate any oral or written utterance. Such results can only be obtained by analyzing authentic texts.

Modeling scientific speech makes it possible to solve one of the most difficult problems faced by students-future specialists – the types of word formation in the English language.

Analysis of recent research and publications. In modern linguistics the issues of English word formation seem to be the most researched. In addition to the fundamental works of such English and American specialists in this field as G.O. Curm [1], J.C. Nesfield [2], H. Marchand [3], E.A. Nida [4] as well as the linguists from other countries – E.S. Kubryakova [4], P.M. Karashchuk [5], O.D. Meshkov [6], etc., in which these issues are covered in details, one can contact modern scientific works describing various parts of the English language in many areas of scientific and artistic discourse [8; 9; 10; 11; 12; 13]. In them, despite the huge variety and multidimensionality of topics, morphemic analysis is present as an obligatory component of the study.

At present, with the emergence of such areas as corpus linguistics, linguistics of functional styles, computational linguistics,

etc., the study of word-formation elements of a text has reached a fundamentally different level, which involves, first, the study of real texts; second, verification of linguistic facts and conclusions by statistical and mathematical methods; third, probability-statistical modeling of various areas of discourse [14; 15; 16].

Nevertheless, despite the detailed theoretical development of this issue in textbooks and a solid didactic foundation, which makes it possible to present in the educational process any part of word-formation elements for students to master as well as the successful research in the field of speech modeling, the combination of scientific results in the field of corpus linguistics with didactics has not happened.

Purpose of the article – to describe the nominal types of word formation units functioning in the text corpora as the results of research which can be introduced into training process.

The main research material. Despite the sufficient amount of modern research in the field of corpus linguistics devoted to the description of word-formation typology, the authors can offer some improvement in the methodology determining the implementation of nominal word-formation models in comparison with those already mentioned above [1; 2; 4; 5; 6], which will be more in demand among future specialists for the formation of oral speech and which will give them (specialists) a real picture of the word-formation system of the English language. First, to use as an object of research not one sublanguage as in the already mentioned works but to carry out a comparative analysis of word-formation units in several sublanguages of engineering in order to make further generalizing conclusions about word formation units in scientific discourse. Second, to more widely apply statistical research methods and present not only the nomenclature of derivational models, but also their quantitative and other statistical characteristics. Third, to consider only the most frequent models found in the text corpora as the most statistically reliable.

For the statistical analysis of word-formation models the text corpora of the following areas of scientific knowledge were compiled – "Chemical engineering", "Automobile Engineering", "Electrical engineering". The corpora were formed on the basis of the articles published in scientific journals in Great Britain and the USA of the corresponding engineering fields: "Chemical Engineering" – Chemical Engineering Progress, Chemical and Process Engineering; "Automobile Engineering" – Automobile Engineering, Auto Industry; "Electrical Engineering" – IEEE Transactions on Power Apparatus and Systems, Proceedings of the Institution of Electrical Engineers. The names of the investigated text corpora show that they belong to different areas as to their scientific and technical topics, which will make it possible to draw conclusions about the peculiarities of word formation units common for texts of scientific discourse in general.

As the main method for the word formation analysis of nouns, the interlinear glossing method was used, and specifically, the morpheme-by-morpheme correspondence method, which is considered in Rule No. 2 of the Leipzig Glossing Rules [17]. The vast majority of international linguistic publications use this very method in which each word form is divided into morphemes, and each morpheme, both lexical and grammatical, is provided with a translation.

So, using the method of interlinear glossing, all nouns of the indicated text corpora were segmented into the minimum significant constituents: P – (prefix morpheme), R – (root morpheme), S – (suffixes morpheme), and then the patterns of their organization in nouns were described. Combinations of these morphemes give four types of lexemes: R, PR, RS, PRS with a simple or simple derivative structure, and five with a complex derivative structure: PPR, RSS, RSSS, PRSS, PPRS. In addition to mono-lexemic forms of nouns (9 units), the word-formative types of nouns with a poly-lexemic nature (7 units) are also registered in the studied text corpora.

Among mono-lexemic types of nouns according to such a characteristic as frequency, type R is distinguished. The percentage of all mono-lexemic nouns in the texts of "Chemical Engineering" is 65% (e.g. *tube, form, press*), 67.5% in the texts of "Automobile Engineering" *(e.g. load, pipe, control)* and 62% in the "Electrical Engineering" sublanguage (e.g. *line, type, volt*). The presented data made on the basis of real texts show that the predominance of root words in scientific and technical texts is a kind of linguistic universal for the text corpora of scientific discourse.

The high productivity of root words in the English language was indicated in previous works [3], but the study was carried out in the language as a whole. The results of the analysis on the basis of real texts of the mentioned specialties show that there are very few productive (i.e., used with different nouns) R-morphemes in the text corpora. So, in the "Automobile Engineering" texts only three productive roots are registered *(act, form, press)*. In the texts on "Electrical Engineering" – only one productive R-morpheme is found which forms eight lexemes, for example, *form – performance, transformer, information, deformation*, etc. Not a single productive R-morpheme is found in the "Chemical engineering" texts.

Comparison of the root morphemes inventories of the three investigated text corpora made it possible to identify 153 identical roots, which, of course, is important when presenting the training material on the word-formation typology of the English language.

In addition to the root type of word formation, one of the main sources of replenishment of new words in the English language is suffixation. Therefore, the type of RS (e.g. *scrubber, production, cracking*) ranks second in frequency in all three text corpora and accounts for 24.7% of all mono-lexemic nouns in the "Chemical Engineering" corpus, 23.7% in the texts "Automobile Engineering" and 27.8% in the corpus "Electrical Engineering". However, in terms of productivity the models in which suffix morphemes are attached to the root occupy the first place.

The types of PR (e.g. *dioxide, polyester, research*) and PRS (e.g. *deformation, degradation, replacement*) demonstrate approximately the same activity in all three text corpora – 7.6–8% of all mono-lexemic nouns.

Mono-lexemic nouns with a complex derivative structure – RSS (e.g. *conductivity, effectiveness*), PRSS (e.g. *reactivation, compressibility*), PPRS (e.g. *noncondensables, reinforcement*), RSSS (e.g. *rationalization*), PPR (e.g. *nontransposition*) are essentially complicated variants of types with a simple derived structure. It follows from this that when studying the structure of a noun in order to extract the information contained in a word correctly one should pay attention first of all to the types R, RS, PR, PRS.

Let us further consider the typology of poly-lexemic nouns the proportion of which in the general inventory of nouns is insignificant and amounts to 4.5% in the texts "Chemical Engineering", 6.6% in the texts "Automobile Engineering" and 5.3% in the text corpus "Electrical Engineering". Poly-lexemic words in the three corpora mainly consist of two lexemes, each

of which is related to an independent word (e.g. *heat-transfer, flow-rate*). The results of the analysis show that the components of two-lexeme nouns are usually the two most productive and frequent mono-lexeme types – R and RS. For example, R is included in almost all two-lexeme formations: R + R (e.g. *time-phase*), RS + R (e.g. *falling-film*), R + RS (e.g. *tower-footing*), R + PRS (e.g. *high-performance*), R + PS (e.g. *earth-return*). The most productive and frequent type in all cases is the R + R type which is 83.3–85, 7% of all two-lexical nouns. Three-lexemic nouns are represented by only one type R + R + R (e.g. *short-line-fault*); in the texts under consideration only six nouns of this type were found.

Thus, the results of statistical and contextual analysis make it possible to mark not only the most frequent units in word-formation models but also the least ones used in text corpora of technical disciplines.

The obtained results of the analysis of the text corpora "Chemical Engineering", "Automobile Engineering" and "Electrical Engineering" allow us to draw the following conclusions.

1. The results of scientific research devoted to the analysis of word-formation models that function in the text corpora of scientific and technical discourse are a valuable methodological material for introduction into the process when teaching English for special purposes since they represent the results of a survey of real texts. This makes it possible not only to enumerate derivational elements and models that exist in the English language in general which was practiced earlier and is being practiced now but to point to real word-formation units used in technical literature.

2. Although 16 derivational types of nouns (9 mono- and 7 poly-lexemic) were registered in the texts of the specialties "Chemical Engineering", "Automobile Engineering" and "Electrical Engineering", a comparison of the typology of nouns has given the possibility to single out from them only 10 most frequent and productive ones which are simultaneously common to three text corpora: six mono-lexemic R, RS, PR, PRS, RSS, PPSS and four poly-lexemic R + R, RS + R, R + RS, R + R + R. The number of three-word units is so small that it is believed that they are not inherent in the scientific functional style.

3. Mono-lexemic types of nouns with a simple and simple derivative structure (R, RS, PS, PSS) occupy a dominant position in the studied corpora in terms of such didactically and statistically important characteristics as frequency of occurrence and productivity. Their percentage of all word-formation models is more than 95.

4. The presence of common linguistic and statistical parameters of the word-formation typology of nouns found in the text corpora of scientific discourse suggests that the selected types could be considered as common in this particular style, i.e. serve as certain stylistic markers since it is known that it is statistical characteristics that are signs of belonging to one or another type of discourse.

In the future the authors envisage to perform statistical and lexical analysis of individual morphemes, of which the word-formation types of nouns functioning in text corpora of other technical areas are composed, with the goal of possible introduction of scientific results into the training process for students of technical universities.

References:

1. Curme G.O. English Grammar. New York, 1957. 308 p.
2. Nesfield J.C. English Grammar Past and Present. London, 1910. 470 p.
3. Marchand H. The categories and types of present-day English word-formation. Wiesbaden, 1960. 375 p.
4. Nida E.A. Morphology. The descriptive analysis of words. Michigan, 1946. 221 p.
5. Кубрякова Е.С. Словообразование (Общее языкознание. Внутренняя структура слова). Москва,1972. С. 344–386.
6. Каращук П.М. Словообразование английского языка. Москва, 1977. 176 с.
7. Мешков О.Д. Словосложение в современном английском языке. Москва, 1985. 175 с.
8. Анюшкин Е.С. Субстантивные термины и их экстралингвистическая детерминированность в немецкой терминологии технологии сахаристых веществ : автореф. дис. … канд. филол. наук : 10.02.04. Воронеж, 1984. 20 с.
9. Архангельский Т.А. Принципы построения морфологического парсера для разноструктурных языков : автореф. дис. … канд. филол. наук : 10.02.21. Москва, 2012. 22 с.
10. Гинсбург Е.М. Словообразование и синтаксис. Москва, 1979. 264 с.
11. Мартемьянова М.А. Особенности формирования современных научных технических терминологических систем (на примере терминов нанотехнологий) : автореф. дис. … канд. филол. наук : 10.02.04. Ижевск, 2011. 22 с.
12. Худинша Е.А. Особенности становления и развития английских базовых терминов в подъязыке экономики : автореф. дис. … канд. филол. наук : 10.02.04. Омск, 2011. 22 с.
13. Чукреева Е.И. Ономасиологическое исследование политических терминов и общественно-политической лексики в американском английском языке (на материале выступлений бывшего госсекретаря США Мадлен Олбрайт). *Лингвокультурология*, 2010. № 4. С. 222–224.
14. Beresnev S.D. Practice of word-formation typology. *Philological Sciences*, 1975. No. 2. P. 30–38.
15. Неврева М.Н. Словообразовательная типология имен существительных в подъязыках техники (на материале английского языка) : автореф. дис. … канд. филол. наук : 10.02.04. Одесса, 1986. 21с.
16. Шапа Л.Н. Формы и функции имен прилагательных в научно-технических текстах (на материале английского подъязыка «Электроснабжение») : автореф. дис. … канд. филол. наук : 10.02.04. Одесса, 1989. 20 с.
17. Leipzig Glossing Rules. URL: http://www.eva.mpg.de›lingua› (Last accessed: 04.10.2016).

Неврева М. М., Дяченко Г. Ф., Зайцева О. Ю. Номінальні словотвірні моделі в текстових корпусах наукового дискурсу

Анотація. У статті розглядається можливість введення в навчальний процес результатів наукових досліджень, проведених на основі обстеження реальних текстових корпусів. Як об'єкт було взято словотвірні моделі англійської мови, що функціонують у текстах наукового дискурсу, які належать до трьох технічних спеціальностей – «Хімічне машинобудування», «Автомобілебудування» і «Електротехніка». Вони належать до різних за своєю науково-технічною тематикою галузей, що дозволяє зробити висновки про особливості словотворення, загальні для текстів наукового дискурсу. Текстові корпуси складені з оригінальних наукових статей, опублікованих в американських і англійських наукових журналах відповідних технічних галузей. У текстах були зареєстровані 16 словотворчих типів іменників (9 моно- і 7 полілексемних), проте порівняння типології імен дозволяє виділити з них тільки 10 частотних і продуктивних, які одночасно є загальними для трьох текстових корпусів: шість монолексемних К, КС, ПК, ПКС, КСС, ППКС і чотири полілексемні К + К, КС + К, К + КС, К + К + К.

Кількість трехлексемних слів настільки мала, що вважається, що вони не властиві науковому функціональному стилю. Монолексемні типи іменників з простою і простою похідною структурою (К, КС, ПК, ПКС) займають домінуюче становище в досліджуваних корпусах за такими дидактично і статистично важливими характеристиками, як частотність виявлення і продуктивність. Їх частка від усіх словотворчих моделей становить майже 90%. Результати досліджень, присвячених аналізу словотворчих моделей у текстових корпусах науково-технічного дискурсу, є цінним методичним матеріалом, цілком статистично надійним для введення в навчальний процес під час викладання англійської мови для спеціальних цілей, оскільки являють собою підсумки обстеження реальних текстів. Це дозволяє не просто перераховувати словотвірні елементи та моделі, які трапляються взагалі в англійській мові, що практикувалося раніше і практикується зараз, але вказувати на реальні словотвірні одиниці, які використовуються в технічній літературі.

Ключові слова: текстовий корпус, частотність вживання, продуктивність, науковий дискурс, полілексемні одиниці, монолексемні одиниці.