

Terminological nomination in modern fields of knowledge

Irina Valujtseva^{1,*}, Olga Ivanova¹, Ilya Khukhuni¹ and Anna Fedosova¹

¹Moscow Region State University, 10A, Radio str, 105005, Moscow, Russia

Abstract. The purpose of the study is to consider the features of the terminological nomination using the example of multiword terms of the English sublanguage of oil and gas processing. The oil and gas glossary Kashagan Development Project Glossary that comprises 1200 units has been used as the research material. Multiword terms have been selected with the application of the continuous sampling method. As a result of the employment of the method of linguistic statistics, it was discovered that, in the sub-language of oil and gas processing, the multiword terms constitute 73% of the entire termbase of the subject area. The structural analysis of terminological phrases demonstrated that the most common type of multiword terms is the two-word terms, comprising 45.2% of the total number of poly-lexemic constructions in the studied sample. This result is consistent with the data obtained by other authors, namely, the information concerning the fact that two-word combinations prevail in various term systems. As the number of components in a multiword term increases, the number of such word combinations in a scientific text decreases. The most common patterns of two-word terminological combinations are N + N and A + N. Three-word terms compose 40% of the total number of the studied terminological phrases. The following patterns of three-word terms are productive: N + N + N and A + N + N. Four-word terminological phrases constitute 12.3%. The most frequently used pattern of four-word combinations is N + N + N + N. Five-word terms comprise 2.4% of the total number of terms. The most common patterns of five-word terms are N + N + A + N + N and A + N + N + N.

1 Introduction

The problem of determining the status and structure of specialized vocabulary has long been one of the most important in terminology studies. The study of the structural and derivational composition of terms, as well as the methods, patterns and features of their formation contributes to the major purpose of terminology studies which is "the study of the features and patterns of formation and development of terminologies in order to develop recommendations for their improvement" [1]. According to Belan E.T., "the identification of the degree of productivity of patterns involved in the formation of specialized units of a certain branch of scientific knowledge, allows ... to predict the paths of its further

* Corresponding author: irinaiv-v@yandex.ru

development" [2]. In addition, the analysis of existing forms of terms allows to elaborate on the particular issues of terminology studies. Thus, it is noted that the pattern of the syntactic structure of a terminological phrase has a significant impact on the motivation of terms – "in a certain semantic model, the terms that are presented according to the most common structural pattern are better motivated" [3].

One of the unresolved issues of the term problem is the question of the terminological status of complex units consisting of terms with constituent relation, and their structural boundaries. The terms most often include one-word, two-word, or three-word units. Moreover, in the actual term systems, the poly-lexemic terminological units consisting of four, five or more words, based on less complex terms, that is compound or derivative terminological units, are observed. Recognition of multiword derived terms in the term system requires a theoretical justification for its termhood, as well as the determination of their grammatical organization, that is the structure, method of the introduction and relations between structural elements.

According to the data provided by Grinev-Grinevich S.V., poly-lexemic units constitute 60% - 95% of the various studied termbases of the European languages [1]. Apparently motivated, multiword terms are mostly peculiar to the new, rapidly developing fields of knowledge. According to Kudinova T.A., "such terms are recognized as the optimal language means in the field of relevant nomination in modern science and, due to the large semantic capacity, are the most important lexical constituent of a scientific text" [4]. Oil and gas processing, which is one of the relatively new subject areas, has significantly expanded the terminological material as a result of significant and fundamentally important changes that have occurred in it over the past years. Thus, the structure of multiword terms of the Russian and English sub-languages of oil and gas processing is regarded as a completely independent field of research and allows to select it as an object of the study.

2 Materials and methods

The material for the research under consideration is the Kashagan Development Project Glossary, the oil and gas glossary comprising 1200 units, which is available on the web site http://www.norma-tm.ru/kdpc-glossary_revised-by-nipi-21.03.05. Through the application of the continuous sampling method, 877 multiword terms have been selected from the glossary for the analysis. The multiword term denotes, according to the definition of Kudinova T.A., "a stable poly-lexemic terminological combination with more than two separately formed independent components" [4]. In foreign terminology studies, such units are called "multiword terms" (MWT) [5]. The remaining 323 units are mono-lexemic terms.

In the process of characterization of terminological phrases, the structural patterns are usually described, taking into account a part of speech that each component of the phrase belongs to.

When determining the number of components in a terminological phrase, the compound words written with hyphens (composites) were regarded as one single component. Thus, the term, *exposed location single-buoy mooring* is a four-word combination, *oil-producing source rocks* is a three-word combination, *water-tight chamber* is a two-word combination.

In addition to words and composites (that is compound words), abbreviations were considered to be separate components of terminological phrases. The most common abbreviations in the English terminology of the oil and gas industry are initialisms: *ATK* (*aviation turbine kerosene*), *VLCC* (*very large crude carrier*), *MTBE* (*methyl tertiary-butyl ether*) and compounds consisting of initialisms and acronyms: *RI foreexportdelivery*, *Low-BTU Gas*, *O&M manager*, *API gravity*.

In the course of the analysis, symbol-word terms that include a letter or a number and a word have also been observed: *Y-connection winding*, *J-tube for cables and chemical umbilical*.

These types of components of poly-lexemic units have been taken into account in the process of identification of the number of components in multiword terms.

In onomasiological terminology studies, the problem of determining parts of speech of components of phrases (which is relevant for the English language, in which, due to the underdevelopment of inflections, the same word could be both a noun and an adjective, the Ving form could be both a participle I, a gerund, and a verbal noun, and the Ved form could be both a participle II, and an adjective), is solved as follows. In phrases such as *clock tower*, *clay brick*, where the first element could formally be both a noun and an adjective, it is assumed to be a noun and present the structural formula of such phrases in the form of N + N [1]. In phrases such as *bearing wall*, *bowling alley*, *building board*, the first element is given the status of an adjectivized participle, based on the reflection: "since such first elements have the meaning of a quality or a function, there is a reason to consider them to be varieties of adjectives that are originally derived from non-finite verb forms, that is participles" [Ibid]. And the structural formula of such phrases is displayed as Ving + N. The structural pattern Ved + N is attributed to terminological phrases such as *arched bridge* on the grounds that "the first element, which is a past participle, has lost the meaning of the tense and acquired the meaning of a quality, turning, as in the previous pattern, into an adjective" [1].

As for the composite, the part of speech that it belongs to is identified in different ways by various researchers. In the article, the solution proposed by Ivanova O.B. has been applied [6]. On the grounds that parts of speech are categories of a word, discovered on the basis of the common syntactic, morphological and semantic properties [7], it is reasonable to apply a syntactic criterion when identifying the part of speech of a compound word of the English language, which is known to be the analytic language. For instance, an adjective is always syntactically regarded either as an attribute of a noun, forming an attributive structure, or as a predicate, connecting to a noun through a copular verb. In case in the process of identification of syntactic functions of the analyzed composite in texts of a certain database (in this research, the Google Scholar resource was applied), it was discovered that the examined composite could only be either an attribute or a predicate, this composite was considered to be an adjective.

3 Results and Discussion

According to the analysis, in English terminology of oil and gas processing, the ratio of mono-lexemic and poly-lexemic units is as follows: 323 words and 877 word combinations, that is, 27% and 73% of the total number (1200 terms) of the analyzed lexical units.

In terminology studies, there are concepts of the optimal and perfect length of a term. "The optimal length of a term is such a length that reflects the real conditions of its formation in a certain term system" [8].

The increase in the number of components of the multiword term is hindered by the fact that there are certain restrictions on the length of the sentence that includes this word combination. "As part of a sentence, a multiword terminological combination must obey the general communication objectives (transfer and reception of information), since it could not be limitlessly extended" [4].

The object of the analysis is 877 English two-, three-, four- and five-word terms. The most phrases (397) are two-word, the next, in terms of number (351), were three-word terms, four-word terms have been encountered 108 times and five-word units have been encountered 21 times.

In general, the percentage of multiword terms by the number of components in the studied corpus of the English terms of the oil and gas processing sublanguage could be demonstrated as follows:

Table 1. Percentage of multiword terms by the number of components in the English sublanguage of oil and gas processing.

Number of components							
2		3		4		5	
total	%	total	%	total	%	total	%
397	45.2	351	40.0	108	12.3	21	2.4

The data presented in Table 1 confirm the results obtained by other researchers, which is the fact that two-word combinations prevail in various term systems; as the number of components in a multiword term increases, the number of such word combinations in a scientific text decreases. Long terms, despite the explicit nature of the onomasiological structure, are inconvenient to apply and are less common.

The analysis of English multiword terms allowed to identify the most productive structural pattern. To represent structural formulas, the following abbreviations are applied: N – noun; A – adjective; Adv – adverb; Ving – participle I; Ved – participle II; Abbr – abbreviation.

Among the two-word terms of the oil and gas processing sublanguage, which constitute 45.2% of the total number of multiword terms, the productive structural patterns are:

Table 2. Percentage of two-component terms by structural patterns in the English sublanguage of oil and gas processing.

Structural pattern	Total number	% of 2-word terms	% of all phrases
N + N	270	68.0	30.7
A + N	74	18.6	8.4
Ving + N	23	5.8	2.6
Ved + N	18	4.5	2.0
Abbr + N	6	1.5	0.7
N + prep + N	4	1.0	0.4
N + and + N	2	0.5	0.2

This result differs from the results of other researchers. In computer terminology, according to Akulinina T.V., the most common pattern is A + N and the next most productive pattern is N + N. According to Simonova K.Yu., in the ecological terminology, the most common pattern is also A + N, and the next most frequently used pattern is N + N [10].

The examples for each of the pattern are:

N + N: *concrete aggregate, condensate tank, expansion joint;*

A + N: *deep foundation, static switch, essential generator;*

Ving + N: *drilling island, feeding pump, gathering pond;*

Ved + N: *imported soil, non-woven geotextile, produced water;*

Abbr + N: *RGI train (Railway Gazette International train), BFW pump (Boiler feedwater pump), IC transformer, KO drum (Knockout Drum);*

N + prep + N: *section on gridline, overflow to grade, authority for expenditure;*

N + and + N: *steam and condensate, safety and operability.*

Among the three-word terms that constitute 40% of the total number of poly-lexemic units, the following structural patterns are observed:

Table 3. Percentage of three-word terms by structural patterns in the English sublanguage of oil and gas processing.

Structural pattern	Total number	% of 3-word terms	% of all phrases
N + N + N	138	39.3	15.7
A + N + N	106	30.1	12.0
A + Ving + N	15	4.3	1.7
A + A + N	14	4.0	1.6
N + A + N	13	3.7	1.5
N + N + and + N	10	2.8	1.1
Ved + N + N	10	2.8	1.1
N + Ving + N	10	2.8	1.1
N + and + N + N	7	1.9	0.8
N + Ved + N	5	1.4	0.57
A + N + prep + N	5	1.4	0.57
Ved + A + N	4	1.1	0.45
Abbr + N + N	4	1.1	0.45

In total, 23 structural patterns are applied for the formation of three-word terms in the English sublanguage of oil and gas processing, and 10 of these patterns have only been observed once: *A + N + and + N* (basic sediment and water), *N + prep + Ved + N* (circuit-breaker in moulded case), *Ved + N + and + N* (combined heat and power), *N + and + Ving + N* (communication and signaling system), *A + N + Ving* (computational pipeline monitoring), *Abbr + Ving + N* (LPG loading package), *N + N + Ving* (machinery condition monitoring), *N + N + and + Ving* (plant layout and piping), *A + and + A + N* (sanitary and epidemiological station), *Ving + N + N* (incoming supply assembly).

This outcome is partially consistent with the results obtained when analyzing other terminologies. According to the findings of Grinev-Grinevich S.V., the most common structural pattern of terminological combinations in English construction terminology is *N + N + N*, followed by *A + N + N* [1]. In the English "Internet" terminology, the most frequently applied pattern is *N + N + N*, which constitutes 48% of the number of English three-word combinations [9]. However, in the terminology of ecology [10] and in the biotechnology terminology [4] the most common is the pattern *A + N + N*, while the next most productive in both of the above-mentioned terminologies is the pattern *N + N + N*, which comprises is 33% of the three-word combinations in the terminology of ecology and 20.4% of the three-word terms of biotechnology.

The examples for the most common patterns of three-word terms are:

N + N + N: fuel gas line, gas air cooler, gas reinjection header;

A + N + N: electrical power equipment, formal safety assessment, individual equipment shutdown;

A + Ving + N: global positioning satellite, retroreflective warning sign;

A + A + N: main auxiliary room, main central building, maximum permissible concentration;

N + A + N: fiber optic cable, field terminal assembly, process electric heater.

Four-word terms in the English sublanguage of oil and gas processing are formed on the basis of 23 structural patterns, and 12 of these patterns have only been observed once. The most frequently used patterns are demonstrated in Table 4:

Table 4. Percentage of four-word terms by structural patterns in the English sublanguage of oil and gas processing.

Structural pattern	Total number	% of 4-word terms	% of all phrases
N + N + N + N	36	33.3	4.1
A + N + N + N	27	25.0	3.0

N + N + and + N + N	6	5.5	0.68
Ving + N + N + N	4	3.7	0.45
N + and + N + N + N	4	3.7	0.45
Abbr + N + N + N	4	3.7	0.45
Ved + N + N + N	4	3.7	0.45
Ved + N + and + N + N	3	2.7	0.3
N + N + Ving + N	3	2.7	0.3
N + A + A + N	3	2.7	0.3
N + N + A + N	2	1.2	0.2

The structural patters that have appeared once are: A + N + Ving + N (*aqueous film forming foam*), N + Ved + Ving + N (*category – 5 enhanced cabling system*), A + Ving + and + Ving + N (*chemicals receiving and dispensing unit*), N + Ved + A + N (*demand assigned multiple access*), A + A + Ving + N (*differential global positioning system*), A + N + A + N (*double pole double throw*), Ved + N + A + N (*induced current cathodic protection*), Adv + Ving + A + N (*normally occurring radioactive material*), A + A + N + N (*private automatic branch exchange*), A + Ved + N + N (*public switched telephone network*), Ved + Ved + N + N (*sealed crimped input terminal*), A + N + N + and + N (*sanitary discharge treatment and storage*).

The examples of the most common patterns of four-word terms are:

N + N + N + N: *asset information management system, fire water jockey pump, fire water pump house;*

A + N + N + N: *clean air compressor module, electronic document management system, intelligent motor control system;*

N + N + and + N + N: *analyzer maintenance and data acquisition system, fire explosion and gas protection, integral control and safety system;*

Ving + N + N + N: *operating switchgear supply system, rotating data acquisition system;*

N + and + N + N + N: *explosion and fire hazard class, fire and gas detection system;*

Abbr + N + N + N: *SE project approvals group, SE project services group.*

In the formation of five-word terms in the English sublanguage of oil and gas processing (2.4% of the total number of poly-lexemic units), 16 patterns are applied, and 13 of these patterns have only been observed once:

Table 5. Percentage of five-word terms by structural patterns in the English sublanguage of oil and gas processing.

Structural pattern	Total number	% of 5-word terms	% of all phrases
N + N + A + N + N	3	14.2	0.3
A + N + N + N + N	3	14.2	0.3
N + N + N + N + N	2	9.5	0.2

The examples of the most common patterns of five-word terminological combinations are:

N + N + A + N + N: *fire water central foam package, storage contamination oil water drain;*

A + N + N + N + N: *sour test production fluid pipeline, early oil production system pipeline;*

N + N + N + N + N: *flash gas compression inlet manifold, naphtha splitter reboiler condensate drum.*

4 Conclusions

The conducted research allowed to discover the patterns of the structural organization of the English terminology of oil and gas processing and to establish the degree of productivity of the identified structural patterns of terminological phrases.

It was found that, in the sublanguage of oil and gas processing, multiword terms prevail in the terminological vocabulary and constitute 73% of the entire terminological composition of the subject area, which indicates the intensive development of the studied term system.

The analysis of structural patterns of compound terms revealed that the most common pattern of poly-lexemic units are the two-word terms, that are represented by 390 units and constitute 45.2% of the total number of multiword terms in the study sample. The most common patterns of the two-word terminological combinations are N + N and A + N.

The three-word terms are represented by 351 units and comprise 40% of the total number of the studied terminological phrases. The most productive structural patterns of the three-word terms are N + N + N and A + N + N.

Four-word terminological combinations compose 12.3% (108 terms). The most productive pattern of four-word units is N + N + N + N, which is represented by 36 units, which constitute 33.3% of the number of four-word terms and 4.1% of the entire sample. In the formation of the four-word terms, 23 structural patterns are applied, but 12 of these patterns have only appeared once.

The five-word terms in the English sublanguage of oil and gas processing comprise 2.4% of the total number of multiword terms. The five-word terms are formed on the basis of 16 structural patterns, 13 of which have only been observed once.

Since "the purposeful nature of the appearance of terms implies the possibility of conscious regulation and implementation of such processes as normalization, unification and harmonization on a national and international scale" [11], and based on the fact that the terminology of the studied industry clearly reflects a number of typical processes inherent in modern scientific and technical terminology in general, it is hoped that the conducted analysis would facilitate the solution of the problems of unification and improvement of the terminology of the industry, as well as contribute to the general theory of terminological nomination.

References

1. S.V. Grinev-Grinevich, *Terminovedenie: uch. pos. dlia stud. vyssh. ucheb. zavedenii* (M., Akademiia, 2008)
2. E.T. Belan, *Osobennosti formirovaniia novykh terminosistem (na materiale angliiskoi i russkoi terminologii mezhdunarodnogo turizma): dis. ... kand. filol. Nauk* (M., 2009)
3. T.R. Kiiak, *Strukturno-semanticheskie osobennosti otraslevoi terminologii: mezhvuz. sb. nauch. tr.* (Voronezh, Izd-vo Voronezh. un-ta, 1982)
4. T.A. Kudinova, *Strukturno-semanticheskie osobennosti mnogokomponentnykh terminov v pod"iazyke biotekhnologii: na materiale russkogo i angliiskogo iazykov: dis. ... kand. filol. nauk* (Orel, 2006)
5. D. Maynard, S. Ananiadou, *5th International Congress on Terminology and Knowledge Engineering* (Innsbruck, Austria, 1999)
6. O.B. Ivanova, *Dinamika stanovleniia terminologii novoi predmetnoi oblasti: na materiale terminosfery nanotekhnologii v angliiskom i russkom iazykakh: dis. ... kand. filol. Nauk* (M., 2010)

7. *Lingvisticheskii entsiklopedicheskii slovar'* (M., Sovetskaia entsiklopediia, 1990)
8. V.M. Leichik, *Terminovedenie. Predmet, metody, struktura* (M., Editorial URSS, 2006)
9. I.G. Kudriavtseva, *Osobennosti formal'noi struktury i semanticheskie kharakteristiki terminologicheskikh slovosochetanii (na materiale angliiskoi i russkoi spetsial'noi leksiki nauchno-tekhnicheskoi oblasti «Internet»): avtoref. dis. ... kand. filol. nauk* (M., 2010)
10. K.Iu. Simonova, *Stanovlenie i razvitie terminologii angliiskogo pod"iazyka ekologii: dis. ... kand. filol. Nauk* (Omsk, 2004)
11. I.I. Valuitseva, I.G. Khukhuni, *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Serii: Lingvistika* **5**, 6 (2019)