

STATISTICAL AND COMPARATIVE ANALYSIS OF THE PHOTOGRAPHERS ENGLISH LANGUAGE LEXEMES USED BY PHOTOGRAPHERS

^aMARGARITA A. MIHAILOVA, ^bALINA F. NIGMATULLINA

Kazan Federal University, 18 Kremlyovskaya street, Kazan 420008, Russian Federation
e-mail: ^agerry-93@inbox.ru, ^binfo@ores.su

Abstract: The paper presents the results of the complex analysis conducted to describe corpus research in the process of analyzing the semantics of codified and non-codified English lexemes of photographers and to identify the quantity and frequency of photographers' terms and professionalism in Corpus of Contemporary American English (COCA). The illustrative and practical parts of the study introduce the method of removing homonymy of lexemes in corpus and a fragment of the statistical study of codified and non-codified lexemes of photographers in photographer's professional corpus of English language and national corpus of English language. In this study, a method for solving the lexical-semantic ambiguity of a lexeme in Corpus of the COCA is presented. Statistical analysis of the material in the COCA was carried out using the empirical concept called "corpus-driven approach"; quantitative method, including a method of calculating the normalized frequency of occurrence of a lexeme in the corpus; method of identifying the meaning of a word. The relevance and novelty of the conducted research consists of the systematic application of the corpus approach in the field of cognitive semantics and sociolinguistics.

Key words: linguistic corpus, statistical analysis, comparative analysis, professional language of photographers, codified lexemes, non-codified lexemes.

1 Introduction

Photography as a specific area of human activity corresponds to a specific type of language, defined by us as the professional language of photographers. "The discourse of a professional community always functions as a normalizer of professional and social values since it has a great impact on the community and the society as a whole" (Mihailova & Solnyshkina 2017). The language of photographers is "a functional variety of the national language, used by a limited group of its bearers in the conditions of both formal and informal communication" (Solnyshkina & Mikhailova, 2017). The professional language uses two groups of units: 1) terms and nomenclature in formal communication conditions (codified lexemes), 2) professionalisms and jargons in informal communication conditions (non-codified lexemes).

2 Sources of the material and data collection

Sources of terms and professionalism of English photographers' lexemes were text materials from the English photo site "Digital Grin Photography Forum" (DGPF).

A photo site is a special software tool (platform, format) where amateur photographers and professional photographers can post their photos, as well as communicate with each other within the framework of a forum, discussion, voting. From a technical point of view, the photo-site is a kind of thematic portal, which is a large web resource that provides exhaustive information on a certain topic.

The subject of the study was the codified and non-codified lexemes of English language of photographers. Analysis of the text data of photo forums allowed selecting the experimental corpus of 622 lexical units of English photographers: 281 collocations and 341 single-word lexemes. The majority of single-word photographers' lexemes is explained by the desire to save, caused by the need to update and reduce the material form and aimed at creating a compact duplicating replacement. "In practice, the key to solving this problem is to create short versions of cumbersome terms" (Teneneva 2013; Villalobos Antunez, 2001) and professionalisms.

When the units were included in the experimental research corpus, the following principles for the selection of terms and professionalism of photographers were the priority: 1) the thematic identity principle, which in our study is applied to the terms and professionalism of photographers, i.e. select units nominating objects, objects and participants in the photo gallery; 2) the chronological principle is directed to the selection of the

photographers' terms, corresponding to the modern level of the photo gallery development; 3) the principle of the "frequency of occurrence" of a lexeme, consisting in the study of terms that occur in the discourse of photographers 2 or more times. This principle helps to exclude the occasionalism of photographers.

Exclusively to the professionalism of the language of photographers, the following criteria were applied:

1. a new, non-existing lexeme. E.g. a new verb de-fish 'to eliminate the effect of wide panoramic or hemispherical image'.
2. a new meaning of existing word. E.g. a lexeme gun, according to CED means 1) a weapon that bullets or shells are fired from; 2) in sport, a device that makes a very loud sudden noise as a signal to start a race в цюпре.; 3) a device that you hold in your hand and use for sending out a liquid; 4) to make an engine operate at a higher speed (CED). The dictionary does not represent the meaning of gun in photographers' professional discourse 'photographic flash'.

3 Methodology

In this study, the empirical and inductive approach of the "corpus-driven approach" to data processing, proposed by Tognini-Bonelli E (Tognini-Bonelli, 2001) was practiced. "The corpus-driven approach <...> aims to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context" (Tognini-Bonelli, 2001). The study used the quantitative method proposed by Biber D., Reppen R. (Biber 2004), which allows to take into account the frequency characteristics of the lexemes under study. The method is aimed at extracting accurate quantitative data on the use of the lexemes, which leads to objective and relevant research results.

Identification of codified and non-codified lexemes of photographers was carried out using a method developed by Kast-Aigner J. (Kast-Aigner 2009), which makes it possible to isolate lexemes nominating photographic equipment, photo processes and participants in photography at the professional corpus of English (DGPH) language.

4 Results and discussion

In this study COCA, founded in 1990, acts as the corpus of national English.

Text material, placed at the photo forums DGPF, and the software of the photo forums are considered by us as professional corpus of the language of photographers. Professional photographers' corpus includes comments of beginner/ professional photographers and amateurs. The professional corpus of the English language of photographers is 74 415 521 words.

However, despite the merit of the DGPF photographers' forums, which is the provision of text material in the professional photographic sphere, we are faced with a problem that was earlier pointed out by Biber D.: «in some cases the interface blocks access to more than a certain number of hits and it will not show the total number, as in the case of words with a frequency of 990–999 in Lexis-Nexis» (Biber & Reppen 2015).

Biber D. argues that 1000 words usages are sufficient for reliable results, because the frequency of lexical categories is manifested in a lesser regression in the corpus of 1000 words: «1,000-word samples reliably represent the distribution of features in a text, as indicated by very high correlations among the feature counts in the samples from each text» (Biber & Reppen, 2015), if the lexeme occurs in the corpus of 1000 words, then it will occur naturally in large texts.

The interface of the English Photographers' (DGPF) language forums provides information on the frequency and use of the word along with the context, with a limit of up to 1,000 uses per unit. E.g., point and shoot lexeme in COCA occurs 40 times, in DGPF the lexeme occurs ≤ 1000 times, i.e. the software of the photo forum does not provide an opportunity to learn the frequency of professionalism in absolute (exact) numbers, if it occurs in the corpus more than 1000 times.

Thus, DGPF software does not provide all the functions and capabilities that the linguistic software of COCA demonstrates, but is sufficient for statistical analysis, which allowed obtaining data on the frequency of English lexemes of photographers.

Comparison of the frequency of the photographers' lexemes in professional (DGPF) and national (COCA) corpuses was conducted on the basis of the normalized frequency of the photographers' lexemes in the corpuses. Following Biber D., Reppen R. (Biber & Reppen, 2015), we introduce the concept of "normalized frequency".

In this study, the normalized frequencies are calculated by converting the number of occurrences of a lexeme in terms of 1 million word-usages. With the help of the formulas developed by us, all absolute numbers were recalculated into a normalized frequency. The formula is the following calculation: $x = (A * 1\,000\,000) / B$, where x is the normalized frequency, A is the absolute frequency, B is the corpus volume (see Table 1).

In the statistical analysis of the English photographers' lexemes in COCA, we are faced with problem of polysemy of words. Search in the corpus allows seeing a list of all the uses of the requested lexeme in context, but not the word's sememe that we are interested in, because at this stage of corpus linguistics, the enclosure software does not provide the ability to automatically filter word values. Accordingly, the corpus demonstrates the statistics of the use of all forms of the word, but not statistics of a certain meaning of the word, which is extremely important for the study of the role of photographers' lexemes in the system of the national language (Kolesnikova & Kamasheva, 2017).

Biber and Reppen note the problem of polysemy as one of the most acute problems for resolution in corpus linguistics (Biber & Reppen, 2015).

Following Biber, we note that differences in significance are manifested in collocations. "In language comprehension and production, collocation plays an important role in disambiguating a polysemous word" (Biber, 2004).

When determining the meaning of a polysemous word, we rely on the method allows us to remove the lexical-semantic ambiguity of the word in the corpus. The method is based on the study of collocates (surroundings) of units of photographers in texts, placed in professional photographers' corpus (DGPF).

We distinguish the following stages in the implementation of the method for resolving lexical-semantic polysemy:

1. Compiling a vocabulary collocation of lexical units of photographers, which includes a constant and / or frequent encirclement of the professionalism of photographers. The collocation dictionary was compiled on the basis of illustrations recorded in the DGPF. F.g. the codified lexeme AF 'the object detection system, in which the lens itself focuses on the induced object' build following collocations AF with, AF within, use AF, AF system, AF on, Nikon AF, AF focus, Canon AF, AF error, Nikkor AF.
2. Entering collocations into the search system of the national language corpus in order to identify the frequency of lexeme of photographers in the COCA. If the collocation is fixed in the COCA, the illustrations, which include the collocation of the language of photographers, are read by the researcher manually in order to exclude coincidences with phrases meanings of which are not realized in the professional language of photographers. F.g. the frequency of collocations with lexeme AF are following; AF within (1),

AF system (4), Nikon AF (1), AF focus (2), Canon AF (3), Nikkor AF (4). The AF does not demonstrate the meaning 'the object detection system, in which the lens itself focuses on the induced object' in collocations AF with, use AF, AF on, AF error.

3. Addition of the frequency of recorded collocations. The frequency of AF 'the object detection system, in which the lens itself focuses on the induced object' in COCA is 15.

Table 1 represents the absolute frequency of lexeme (AFL) in the national language corpus, the absolute frequency of occurrence of the lexeme in the meaning of professional photographers' discourse (AF) and the normalized frequency (NF) of the lexical units of English photographers' lexemes in national and professional corpuses.

Table 1. Statistics of absolute and normalized frequency of English photographers' lexemes in national (COCA) and professional (DGPF) corpuses

English photographers' lexemes	Corpus of national language COCA			Corpus of professional language DGPF	
	AFL	AF	NF (1 million word-usages)	AF	NF (1 million word-usages)
diaphragm	872	11	0,019643	123	1,652881
infinity	176	41	0,073214	702	9,433516
grip and grin	5	5	0,008929	25	0,335951
bigma	-	-	0	761	10,22636
body	147125	273	0,4875	≤ 1000	13,5

Non-codify lexeme grip and grin 'photo on which a group of people is standing, embracing, and all the photographed are smiling' fixed in COCA 5 times, in DGPF 25 times. The normalized frequency of professionalism grip and grin in COCA is 0.008929 (5), in DGPF - 0.335951 (25).

Comparative analysis of the normalized frequency of occurrence of professional English lexemes in COCA and DGPF shows: allows to make the following conclusions:

1. 418 (67%) photographers' lexemes of English professional language are homonymous to the lexemes of the national language. 166 (39%) of these lexemes are recorded in the professional meaning in COCA, the remaining 252 (61%) photographers' lexemes are not found in COCA.
2. 348 photographers' lexemes of English professional language (57%) recorded in the English national corpus (COCA). E.g., the professionalism of photographers photo op 'a specially equipped area for a photo session and autograph session with famous personalities' can be met 303 times in COCA, in DGPF - 996 times.
3. 270 photographers' lexemes (43%) are absent in COCA. E.g., the professionalisms B&G 'newlyweds (bride and groom)', digivangelist 'the one who always uploads photos, videos, posts in electronic form', pentaxian 'adherent of the Pentax cameras' are absent in the COCA.
4. The normalized frequency of 364 photographers' lexemes that are found in COCA is lower in the corpus of English national language than in the professional one. E.g., the normalized frequency of the term parallax error 'the difference between what the lens sees and what the person sees through the viewfinder of the camera (color difference, volume, etc.)' is higher in the professional corpus of photographers - 0.215009 (12) than in the national one - 0.001786 (1).
5. 4 professionalisms of the language of photographers are met in the national corpus more often than in the professional English language corpus of photographers. The normalized frequency of professionalism cameraman 'photographer who

always takes a camera with him' equals 2,210,714 (1238), in the professional language of photographers the normalized frequency of cameraman is significantly lower - 0.631589 (47).

Comparative analysis of the normalized frequency of occurrence of professional English lexemes in COCA and DGPF shows:

1. 418 (67%) photographers' lexemes of English professional language are homonymous to the lexemes of the national language. 166 (39%) of these lexemes are recorded in the professional meaning in COCA, the remaining 252 (61%) photographers' lexemes are not found in COCA.
2. 348 photographers' lexemes of English professional language (57%) recorded in the English national corpus (COCA). E.g., the professionalism of photographers photo op 'a specially equipped area for a photo session and autograph session with famous personalities' can be met 303 times in COCA, in DGPF - 996 times.
3. 270 photographers' lexemes (33%) are absent in COCA. E.g., digivangelist 'the one who always uploads photos, videos, posts in electronic form', pentaxian 'adherent of the Pentax cameras' are absent in the COCA.
4. The normalized frequency of 364 photographers' lexemes that are found in COCA is lower in the corpus of English national language than in the professional one. E.g., the normalized frequency of the term parallax error 'the difference between what the lens sees and what the person sees through the viewfinder of the camera (color difference, volume, etc.)' is higher in the professional corpus of photographers - 0.215009 (12) than in the national one - 0.001786 (1).
5. 4 professionalisms of the language of photographers are met in the national corpus more often than in the professional English language corpus of photographers. The normalized frequency of professionalism cameraman 'photographer who always takes a camera with him' equals 2,210,714 (1238), in the professional language of photographers the normalized frequency of cameraman is significantly lower - 0.631589 (47).

5 Conclusion

The proposed method for resolving the lexical-semantic ambiguity allowed in a short time and with minimal errors to search for the contexts of the necessary lexeme in a certain meaning in English national textual corpus (COCA). Comparative analysis of the normalized frequency of occurrence of professional English lexemes in COCA allows to make the following conclusions: 1) the majority of exemplified lexemes of photographers (418 (67%)) are homonymous to the lexemes of the national language; 2) the minority of photographers' lexemes (270 (33%)) photographers' lexemes are absent in national English language corpus; 3) the normalized frequency of 364 (61%) photographers' lexemes demonstrated in COCA is lower in national corpus than in photographers' professional corpus. Thus, the low frequency of use or absence of photographers' lexemes in the national language corpus testifies that the language of photographers is the language, to a certain extent, of an isolated group of people. The concepts and phenomena of the professional language are specific and narrow.

The methods, algorithms and tools developed and applied by the authors may be used to describe structural, semantic and functional parameters of any limited group of vocabulary.

Acknowledgements

This research was supported by the subsidy of the Russian Government to support the Program of Competitive Growth of Kazan Federal University.

Literature:

1. CED – Cambridge Dictionary: English Dictionary, Translations & Thesaurus, URL: <https://dictionary.cambridge.org/>

2. COCA – Corpus of contemporary American English , URL: <https://corpus.byu.edu/COCA>,
3. DGPF – Digital Grin Photography Forum, URL: <https://dgrin.com/categories>
4. Kast-Aigner J. (2009). Terms in context: A corpus-based analysis of the terminology of the European Union's development cooperation policy, *Fachsprache – International Journal of LSP*, No. 3-4, pp. 139-152.
5. Mihailova M. A., Solnyshkina M. I. (2017). Photographers' Nomenclature Units: A Structural and Quantitative Analysis, *Tarih kultur ve sanat arastirmalari dergisi-journal of History culture and Art research*. 6(5), pp.166-172.
6. Solnyshkina M.I., Mikhailova M.A. (2017). Processes of semantic derivation in the language of photographers , *M.I. Solnyshkina, M.A. Mihailova, Philology and Culture*. 2(48), pp.114-120.
7. Teneneva I.V. (2013). Morphological and syntactical compression in phototechnical terminology , *South-Western State University Edition. Series: Linguistics and Pedagogy*. No. 4, pp. 7-12.
8. Tognini-Bonelli E. (2001). *Corpus Linguistics at Work. Studies in Corpus Linguistics*, Benjamins, No. 6, p. 224.
9. Biber D., Reppen R. (2015). *The Cambridge Handbook of English Corpus Linguistics* , Cambridge University Press, pp. 623.
10. Biber D. (2004). Representativeness in corpus design. In G. Sampson and D. McCarthy (Eds.), *Corpus linguistics: Readings in a widening perspective*, pp. 97-174.
11. Villalobos Antúnez J.V. (2001). La ética y el derecho ante la filosofía intercultural y la globalización, *Unica: Revista de Artes y Humanidades*, No. 4, pp. 71-76.
12. Kolesnikova J., Kamasheva A.V. (2017). The alienation of the rights to life and health: the institutional dimension, *Astra Salvensis, Supplement No. 2*, pp. 59-65.