Contrastive Corpus Research in Bilingual Phraseography
Guzel Kazbekovna Gizatova¹, Alfiya Rafailovna Baranova²,
Natalya Albertovna Sigacheva³, Gulnara Ildarovna Salyakhova⁴
1 Kazan Federal University, Institute Of International Relations, History And Oriental Studies
2 Kazan Federal University, Institute Of International Relations, History And Oriental Studies, E-mail: baranova.alfiyarafailovna@mail.ru
Tel.: +79033444565
3 Kazan Federal University, Institute Of International Relations, History And Oriental Studies
4 Kazan Federal University, Institute Of International Relations, History And Oriental Studies

ABSTRACT

The paper discusses linguistic issues presenting difficulties for bilingual lexicography. The purpose of the study is to use data of parallel English and Russian corpora for description of non-trivial features of English phraseological constructions *on the brink of* / *on the threshold of* in comparison with their Russian equivalents *на грани* / *на пороге* in the process of compiling of an English-Russian phraseological dictionary. Some ideas of construction grammar are applied as theoretical basis of this research work. The analysis of usage frequency of phraseological constructions applying statistical methods is put forward in this work. The results of the study show that the use of parallel corpora helps to reveal specific character of their functional correlations and non-trivial semantic preferences of English phraseological constructions which do not have standard Russian equivalences. The need for a new dictionary is motivated by the fact that at present there are no English-Russian phraseological dictionaries based on corpora and authentic data in international lexicography. The use of corpora helps to provide new vision of a contextual behavior of phraseological constructions and restrictions of their usage, which is very important for the purposes of lexicographic description.

Keywords: linguistics, parallel corpora, semantic asymmetry, phraseological dictionary, language

.

1. INTRODUCTION

Bilingual lexicography encounters certain problems in connection with treatment of phraseological equivalents in dictionaries. In many cases the generally accepted equivalent of a phraseological unit cannot always be used to translate authentic texts. The paper analyses the semantic asymmetry between English and Russian phraseological constructions with the purpose of compiling a comprehensive English–Russian Phraseological Dictionary.

The primary goal is to analyze cross-linguistic correlations between English and Russian phraseological constructions (PC) which have strong semantic resemblance: *on the brink of (на грани), on the threshold of (на пороге).* They are lexically open and have constant lexical elements *ON THE … OF*, their basic structural feature is the presence of an open slot. This type of phraseological constructions is an important element of communication but it has not been sufficiently studied in phraseography yet. The majority of monolingual and bilingual dictionaries (Kveselevich, 2002; Kuz'min, 2001; Longman Exams Dictionary, 2006; Lubenskaja, 2004; Oxford Advanced Learner's Dictionary, 2004) regard PCs *on the brink of* and *on the threshold of* as complete synonyms. But analysis of their behavior in authentic contexts in the corpora show that it is not the case of full equivalence. In fact, traditional dictionaries are often based on a limited quantity of examples which are not exemplified in many cases. The use of corpora gives additional possibilities compared to the dictionary data, providing more complete and accurate examples of the phraseological constructions being described, organizing the idiom list and arranging dictionary entries. The corpus allows us to determine the degree of frequency of a PC and its variants (Dobrovol'skij, 2014, p. 873).

Some ideas of construction grammar (CxG) are used as theoretical basis of this research work. On the early stages of CxG development the main attention was focused on bordering, compositional fields of the syntax and lexicon. Cf. definition of a construction given by A. Goldberg: "Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition,

patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency" (Goldberg, 1995, p. 5). Similar ideas were expressed in Fillmore et.al (1988).

In recent times the concept of a construction is expanding. This point is reasoned by the fact that many regularly used word combinations and models formed from them are so frequent that they are stored in the memory of language speakers as single blocks, and they are not generated according to the syntactic, semantic or pragmatic rules (Goldberg, 2006). This point of view becomes more and more spread because of the frequency of study of language units using statistical methods, being an important constituent of CxG lines of research (Bybee, 2010; Stefanovich, 2007; Stefanovich, Gries, 2003); Dobrovol'skij, Pöppel, 2016).

In this paper the analysis of usage frequency of constructions applying statistical methods is used.

2. METHODS

The method includes the following stages:

- use of parallel corpora;

- search for all translation equivalents of constructions under study;

- data processing using statistical methods;

- analysis of the results obtained.

The empirical data have been collected from the corpus query system Sketch Engine, subcorpus [enTenTen 2013] (19,7 billion tokens), subcorpus [ruTenTen 2011] (14,5 billion tokens), English and Russian parallel subcorpora OPUS-2. This made it possible to find all instances of the phrases *on the brink of, on the threshold of* and their Russian equivalents *на грани* and *на пороге* and obtain statistically representative data. Parallel subcorpus of The Russian National Corpus was used for comparison.

3. RESULTS

Using English and Russian parallel subcorpora OPUS-2 we compared constructions *on the brink of* and *on the threshold of* in relation to their synonymous closeness. The analysis of their 25 most frequently used co-occurrence partners demonstrate that out of 18,473 cases of usage of the PC *on the brink of* it is most commonly used with the following nouns: extinction – 3,253, collapse – 2, 815,war – 2, 524, bankruptcy– 1, 923, disaster, destruction, ruin. All these nouns express extreme, dangerous, critical, often

unpredictable situations which tend to create problems. The least frequently used noun with the PC *on the brink of* is 'millennium' – 28 cases.

It's quite interesting that out of 4, 881 contexts of usage of the PC *on the threshold of* in the parallel corpus OPUS-2 it is most frequently used with the noun 'millennium' – 976 cases. First three high frequency co-occurrence nouns of the PC *on the threshold of* are: *millennium, century* and *era* which constitutes 70% of all 25 nouns being examined. Other frequently used nouns are: *reforms, changes, success, etc.*

4. DISCUSSION

So the analysis of the representative empirical data allows us to conclude that phraseological constructions *on the brink of / на грани* and *on the threshold of / на пороге* constitute two groups, in which:

1. Interchange of the PCs in the context is impossible or restricted;

2. interchange of the PCs in the context is possible.

Let us analyze both groups.

1. Interchange of the PCs in the context is impossible or restricted

In majority of cases PCs *on the brink of / на грани* and *on the threshold of / на пороге* are not full synonyms and therefore they are not interchangeable. The frequency data from Sketch Engine confirms our observations that *on the brink of/ на грани* is used with abstract nouns which express negative connotations and *on the threshold/ на пороге* is also used with abstract nouns, but has a really "friendly" surrounding in the context. If concrete nouns are used, which happens rather seldom, they are used metaphorically or metonymically in an abstract sense.

Empirical data from corpora shows that most idioms can be translated correctly only if we take the context into account, something that many dictionaries fail to do in a systematic way. As it has been mentioned already, majority of dictionaries postulate a relationship of "full equivalence" between constructions *on the brink of / на грани* and *on the threshold of / на пороге.* But "traditional description … ignores the absence of functional interchangeability between the idioms" (Dobrovol'skij, 2014, p. 872).

Following examples of their usage in the parallel subcorpus of the National Corpus of the Russian Language (NCRL), in the (1) and (2) demonstrate that they are not synonymous, meaning that they are not interchangeable in the context, cf.:

(1) Новая школа строилась *на* самом *пороге* века *(NCRL).*

(2) The new school was built *on the threshold of* this century (NCRL).

(3) You are *on the threshold of* life, you have only known this girl two months and however deeply you think you love her, I appeal to you to break it off at once (NCRL).

(4) Ты стоишь *на пороге* жизни, ты только два месяца знаком с этой девушкой, и какой глубокой ни представляется тебе, твоя любовь к ней, я обращаюсь к тебе с призывом пресечь эту любовь немедленно (NCRL).

In contexts (1, 2) and (3, 4) we cannot translate the Russian PC *на пороге* with the English *on the brink of*, as well as it is impossible to substitute *на пороге* by *на грани* and *on the threshold* of by *on the brink of.*

It is true that most idioms can be translated correctly only if we take the context into account. But in some cases context is not enough and only culture specific information about language can help us give proper interpretation of this or that phrase, cf.:

(5) MEXICAN CULTURAL INSTITUTE "Hina/Jaina: *On the Threshold* of the Mayan Underworld (600-900 A.D.)" [enTenTen2013]

In this example the symbolic meaning of the concept THRESHOLD is taken into account for a better understanding of the PC. According to the Dictionary of symbols: "It symbolizes the potential of friendship, marriage or reconciliation. This potential can be actualized if the individual who comes is greeted at the threshold and invited in" (Chevalier, 1996, p. 997). So THRESHOLD is the symbol of friendship and welcome. In (5) we are friendly welcomed and invited into Mayan civilization. Here we cannot substitute the construction *on the threshold of* by the construction *on the brink of* because the noun *brink* has different image component and symbolism. Its image component is based on two ancient oppositions: "us – them" and "life – death". It reflects ancient mythological ideas about boundaries in terms of spatial limits, which are divided into two worlds: "us" – "the world of the living" and "them"– "the world of the dead", that is "their" space, which is threatening "us" and that's why it is dangerous. In (6) the spatial metaphor creates an image of an extreme situation which is recognized as potentially dangerous. In (6) someone had "to stop at the act" due to unpredictable and, most probably negative consequences of doing something, cf.:

(6) We must assume, I think, that the forward projection of what imagination he had, stopped at the act, *on the brink of* all its horrible consequences (NCRL).

Constructions *on the brink of* and *on the threshold of* are not interchangeable, we cannot say: "on the threshold of horrible consequences" in this particular or any other context. The examples analyzed demonstrate that though PCs are presented as synonyms almost in all dictionaries, in fact they are not synonymous because each of them has its individual characteristics.

2. Interchange of the PCs in the context is possible

In spite of the fact that many dictionaries and Sketch Engine graphs treat the phrase *on the brink of* having negative connotations, this is by no means always the case. The analysis of the corpora demonstrates that *on the brink of* in many cases has positive meanings and sometimes it can be a semantic synonym to the PC on *the threshold of*. Cf.:

(7) Remember, you are *on the brink of a New Millennium*. Let yourself be inspired, and talk about your inspiration to those who share your life [enTenTen 2013].

(8) Maldives is posed *on the threshold of the New Millennium*, looking forward to it with vigor and enthusiasm. Emboldened with successes of the past and empowered by hope for the future, the Maldives is optimistic in making further strides towards socio-economic development of the country [enTenTen 2013].

Some more examples of the PC *on the brink of* having positive meaning from the British National Corpus (BNC) (9), (10), cf.:

(9) Mr. Ashdown told supporters: 'The prize is within our grasp. We stand *on the brink of* an outstanding result' (BNC).

(10) Another Heineken weekend of records and dazzling rugby has convinced fans that Wales are *on the brink of* a new golden era of backplay (BNC).

In many contexts (11, 12) *on the brink of* and *on the threshold of* are interchangeable, cf.:

(11) Goldstein concludes that a parallel growth slowdown in China and the United States, along with deterioration in global financial conditions linked to a disorderly correction of global payments imbalances, could put a group of emerging markets *on the threshold of economic crisis* [enTenTen 2013].

(12) The conflict between the two states brought them both *on the brink of economic crisis*, as South Sudan stopped producing oil and send it through Sudanese pipelines, accusing the northern state in stealing [enTenTen 2013].

We referred to parallel English and Russian corpora OPUS-2 with the purpose of revealing the most frequently used variants of PCs *on the brink of* and *on the threshold of* and their translations. The results of statistic analysis are presented in the tables 1 and 2.

Table 1

on the brink of	Q	%	награни	Q	%
"0" equivalent	25	19%	"0" equivalent	431	57%
на грани	80	61%	on the verge of	181	24%
на пороге	20	15.3%	on the brink of	92	12%
на краю	5	4.7%	on the point of	17	2,2%
			on the edge of	16	2,1%
			on the cusp of	9	1,2%
			on the fringe(s) of	6	0,8%
			on the precipice of	3	0,4%
Total	130	100%		755	100%

Table 2

on the threshold of	Q	%	напороге	Q	%
"0" equivalent	20	17,1%	"0" equivalent	354	30,5%
			Literal meaning	342	29,5%
на пороге	79	67,5%	on the threshold of	357	30,8%
в преддверии	9	7,6%	on the verge of	26	22,4%
на рубеже	4	3,4%	on the brink of	24	21%
награнице	3	2,5%	on the doorstep of	21	18,1%
назаре	1	0,95%	on/at the point of	9	0,8%
вот-вот	1	0,95%	on the edge of	9	0,8%
			on the cusp of	7	0,6%
			on the eve of	6	0,5%
			at the gates of	3	0,25%
Total	117	100%		1158	100%

These frequency graphs were processed manually to avoid information noise.

The results of statistic analysis in the Table 1 show 130 Russian correlates of the PC *on the brink of*: zero equivalent <25>; "*на грани*" <80>; "*на пороге*" <20>; "*на краю*"<5>.

The results of a dictionary analysis indicate that neither general bilingual nor phraseological dictionaries translate the construction *on the brink of* with "*на пороге*" in their entry in spite of the fact that it has a high level of frequency in OPUS-2 and constitutes 15, 3% of all its correlates.

The results of statistic analysis in the Table 2 show 117 Russian correlates of the PC *on the threshold of*: zero equivalent <20>; "*на пороге*"<79>; "*в преддверии*" <9>; "*на рубеже*" <4>; "*на границе*" <3>. Two equivalents: "*на заре*" and "*вот-вот*" are used only once each. The data in the table 2 indicate that the Russian PC *в преддверии* is used in 7, 6% cases of all its correlates usage. But dictionaries don't provide such translation for the PC *on the threshold of*.

The evidence suggests that the Russian PCs *на пороге* (table 1) and *в преддверии* (table 2) should be included in the dictionary entries of the English PCs *on the brink of* and *on the threshold of* accordingly, because they prove to be more frequent than other PCs which are included in dictionaries, cf.: *на краю, на границе, на рубеже, на заре* (tables 1, 2).

The frequency graphs demonstrate vividly that there is a great range of variants actually represented in texts which are missed out in the dictionaries, name but a few of them: *on the cusp of, on the fringe of, on the precipice of*; *on the doorstep of, on the eve of, at the gates of, at the dawn, at the start of, at the turn of, at the beginning of, at the onset of, at the outset of* and some others. These variants should definitely be included in the dictionary entries of phraseological constructions *on the brink of* and *on the threshold of*.

5. SUMMARY

The empirical data presented in the article proves that in English PCs considered here, synonymy is not as complete as it seems at first glance. In many cases *on the brink of* and *on the threshold of* are semantically asymmetrical. Very often the generally

accepted equivalent of a PC cannot always be used to translate authentic texts and there are restrictions or impossibility of its substitution by its near-synonyms. But in some cases context is not enough and only culture specific information about language can help us give proper interpretation of this or that phrase.

6. CONCLUSIONS

Parallel corpus analysis allows to reveal full diversity of variants actually represented in texts which is practically impossible using only monolingual and bilingual dictionaries. Using corpora broadens phraseographer's resources while arranging the illustrative part of the dictionary entry and searching translation correlates of each unit of the source language being under consideration.
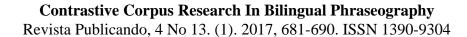
7. ACKNOWLEDGEMENTS

8. REFERENCES

Bullon S., 2009.Longman Exams dictionary. England: Pearson Education Limited.

Bybee J., 2010.Language, usage and cognition. Cambridge: Cambridge University Press Publ.

Chevalier J. & Gheerbrant J.,1996. Dictionary of symbols. London: Penguin Books.

Dobrovol'skij, D. 2014."The use of corpora in bilingual lexicography", In A. Abel, C. Vettori & N. Ralli. Proceedings of the XVI EURALEX International Congress: the User in Focus (pp. 867-884). Bolzana/Bozen: EURAC Research.

Dobrovol'skijD ,2016. Grammatika konstruktsij i frazeologija. Voprosy jazykoznanija, 3, p. 12..

Dobrovol'skij D. & Pöppel L., 2016.Diskursivnaja konstruktsija N v tom, što i eyo paralleli v drugih jazykah: kontrastivnoje korpusnoje issledovanije. Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta, 6(34), pp.164-175,.

Fillmore C.J., Kay P., O'Connor M.C., 2012.Regularity and idiomaticity in grammatical constructions. The case of let alone. Language, , 22 (1), p.41-78.

Goldberg, A. 1995.Constructions: A Construction Grammar Approach to argument structure. Chicago and London: The University of Chicago Press Publ.

Goldberg,A. 2006. Constructions at Work. Oxford: Oxford University Press.

Kveselevich, D. 2002.Sovremennyj russko-anglijskij frazeologicheskij slovar'. M.: Astrel'.

Kuz'min S., 2006.Russko-anglijskij frazeologicheskij slovar'. M.: Flinta.

Lubenskaja S.,2004. Bol'šoj russko-anglijskij fraseologicheskij slovar'. M.: Ast-PressKniga.

Stefanovich A., 2007.Linguistics beyond grammaticality. Corpus Linguistics and Linguistic Theory, 3(1), p.57-71.

Stefanovich A. & Gries S., 2003.Constructions: Investigating the interaction between words and constructions. International Journal of Corpus Linguistics, 8(2), p.209-243.

Wehmeier S., 2000.Oxford Advanced Learner's Dictionary. Oxford: Oxford University Press.