Espacio Abierto
ISSN: 1315-0006
eabierto.revista@gmail.com
Universidad del Zulia
Venezuela

Rubtcova, Mariia; Vasilieva, Elena; Pavenkov, Vladimir; Pavenkov, Oleg

Corpus-based conceptualization in sociology: possibilities and limits

Espacio Abierto, vol. 26, núm. 2, abril-junio, 2017, pp. 187-199
Universidad del Zulia
Maracaibo, Venezuela

Espacio Abierto Cuaderno Venezolano de Sociología
Vol.26 No.2 (abril - junio, 2017): 187-199

Corpus-based conceptualization in sociology: possibilities and limits

Mariia Rubtcova, Elena Vasilieva** y Vladimir Pavenkov, Oleg Pavenkov****

Abstract

The problem of the quantitative interpretation of qualitative data is one of the most important in sociological research. Textual analysis has placed emphasis on deep and careful study of texts how personal strategies embodied in the concepts. However, quantitative interpretation has always been problematic. Our paper deals with the corpus-based conceptualization method, which can be considered as a method of collecting and organizing data material from linguistic corpora. The corpus-based conceptualization allows us to establish a closer link with the meaning and identify the whole spectrum of meanings. It shows that some sociologists lose essential meanings in the research process because of lack of in-deep immersion in the daily life and speech of the people. We chose the concepts of "altruism" and "mercy" as examples to demonstrate the corpus-based conceptualization and its place in sociological research methodology. Data comes from the Russian National Corpus. The Russian National Corpus consists of 1802 relevant words, with 775 for altruism and 1047 for mercy. Data processing carried out by SPSS 19.0. As the result, we have discussed what difficulties the researcher can meet using this method and have offered Systemic Functional Grammar (SFL) and Role

Aceptado: 14-11-2017 / Recibido: 21-02-2017

* St. Petersburg State University. Russia
** Academy of Science of Republic of Sakha. Yakutia, Russia.
*** University of Film and Television. Russia.

and Reference grammar as a way to accurately determining the context. Our suggestions can be used in the preparation of questionnaires, guides, in an analysis of interview transcripts.

Keywords: Corpus-based conceptualization; sociological operationalization; corpus linguistics

Conceptualización basada en el corpus en sociología: posibilidades y límites

Resumen

El problema de la interpretación cuantitativa de los datos cualitativos es uno de los más importantes en la investigación sociológica. El análisis textual ha puesto énfasis en el estudio profundo y cuidadoso de los textos como las estrategias personales incorporadas en los conceptos. Sin embargo, la interpretación cuantitativa siempre ha sido problemática. El presente artículo trata del método de conceptualización basado en el corpus, que puede ser considerado como un método de recolección y organización de datos de los corpus lingüísticos. La conceptualización basada en el corpus nos permite establecer un vínculo más estrecho con el significado e identificar todo el espectro de significados. Muestra que algunos sociólogos pierden significados esenciales en el proceso de investigación debido a la falta de inmersión profunda en la vida cotidiana y el habla de las personas. Elegimos los conceptos de "altruismo" y "misericordia" como ejemplos para demostrar la conceptualización basada en el corpus y su lugar en la metodología de la investigación sociológica. Los datos provienen del Corpus Nacional Ruso. El Corpus Ruso Nacional consta de 1802 palabras relevantes, con 775 para el altruismo y 1047 para la misericordia. Procesamiento de datos realizado por SPSS 19.0. Como resultado, hemos discutido las dificultades que el investigador puede encontrar utilizando este método y hemos ofrecido la Gramática Funcional Sistémica (SFL) y la gramática de Roles y Referencias como una forma de determinar con precisión el contexto. Nuestras sugerencias pueden ser utilizadas en la preparación de cuestionarios, guías, en un análisis de transcripciones de entrevistas.

Palabras clave: Conceptualización basada en el corpus; Operacionalización sociológica; Lingüística del corpus.

1. Introduction

The problem of the quantitative interpretation of qualitative data is one of the most important in sociological research (Rubtcova & Pavenkov, 2016); Pavenkov & Pavenkova, 2016). Textual analysis has placed emphasis on deep and careful study of texts and how personal strategies embodied in the concepts. However, quantitative interpretation has always been problematic (Usiaeva et al., 2016).

Quantitative operationalization of concepts can improve the objectivity of the data and exclude some erroneous or obscure use of social categories, as well as determine the particular application in different contexts to identify the relevance of social problems (Rubtcova, Pavenkov & Varlamova, 2017). As stated by P. Lazersfeld, operationalisation of concepts can help to create a model that characterizes the social problems studied in the course of empirical social research (Lazarsfeld, 1962). This model is created on the stage of conceptualization of the concept and is used as a base for research operations.

We propose to use quantitative analysis as the basis for building empirical models, which allows identifying the meanings and trends using corpus linguistics: firstly for the count of key words to define the unit of account institutional contexts, secondly for a qualitative analysis of the identified contexts (Rubtsova & Vasilieva, 2016). It allows increasing the objectivity of the research (see e.g. McEnery & Hardie, 2012). Our paper deals with the corpus-based conceptualization method, which can be considered as a method of organizing data using linguistic corpora. The corpus-based conceptualization allows us to establish a closer link with the meaning and identify the whole spectrum of meanings (Halliday, 1975).

First of all, corpus linguistics studies a language not only as a remarkable social phenomenon (e.g. altruism, power, government). The most significant aim is to explain the language characteristics of a term (concept) (Corpora and Discourse, 2008). Corpus does not initially have such intention; however, it creates for sociology some interesting content that extends standard sociological frames of social categories' knowledge (Abulof, 2015). The attention of the linguistic corpus in natural language gives a path to answers of how people actually discuss social reality (Blei, 2012). If standard sociological content analysis requires conceptual study through the subjective descriptions of a sociologist, the material is already presented in a corpus, and sociologists can only use what is available. This increases the level of objectivity in a researcher`s operationalisation.

Second, a corpus has a meaningful size that allows us to make a big data statistical analysis. The most part of the national corpora contain more than 500 million words (Corpora and Discourse, 2008). Materials are described by years and months over a long period; for example National Corpus of the Russian Language includes the development of language from the 18th to the beginning of the 21st century (National Corpus of the

Russian Language, 2015). A researcher can build the frequency of the use of categories over various periods of time (diachronic word frequency): e.g., the max. frequency of use of the concept ‘mercy’ is in years 2002-2004. The aggregate of phrases also supports the in-depth study of the dynamic context of the concept and its main topics (see, e.g.: Liu, 2012; Abulof, 2015).

Also the main part of national corpora has one principle of structuring that is crucial for cross-cultural studies (Corpora & Cross-linguistic Research, 1998). The corpus linguistics gives us the possibility to recognize differences in the understanding of the social terms in different cultures that should be done before comparative sociological studies (Rawoens, 2010).

Finally, the corpus research is not expensive and time-consuming. It can be done in a user-friendly search format for uploading and downloading data that is compatible with Word, Excel, SPSS etc. (McEnery & Hardie, 2012). So we can do the interpretation procedure quickly. Researchers are excited by the corpus linguistics possibilities because it allows to analyze a huge amount of text inexpensively (Schonhardt-Bailey, 2005)

In our opinion, sociology can lose essential meanings in the research process because of lack of in-deep immersion in the daily life and speech of the people. Context can be understood as institutional frame, which defines the status and role of the concept in social structure . We chose the concepts of "altruism" and "mercy" as examples to demonstrate the corpus-based conceptualisation and its place in sociological research methodology. “Altruism” is defined as the willingness to act selflessly for the benefit of others, without regard to their interests (Pavenkov, Pavenkov, & Rubtcova, 2015). A “mercy” – as the willingness to help someone or to forgive someone out of compassion, humanity (Pavenkov, Shmelev & Rubtcova, 2016).

Specifically, our study answer the following research questions:

- *Are the concepts «altruism» and «mercy» interchangeable as synonyms in sociological research?*

2. Data and Methodology

2.1 Research Design

The proposed methodology of the study of social categories is the combination of quantitative and qualitative analysis of the encoded array of a Corpus. Quantifying the frequencies of word used was based on expert’s context encoding. In accordance with the method, three independent experts performed the encoding. All of them were representatives of St. Petersburg universities, sociologists and were not the article’s authors. They offered the following seven contexts: people’s actions, humans themselves, quality of relations, state, social institutes, organisations, and conception-ideology. Data processing was carried out using SPSS.

2.2 Sampling Procedures

Data comes from the Russian National Corpus. In Fig.1, 2 we can see the frequency of references to the category of "mercy" in the period of 1800-2014 years. The Russian National Corpus consists of 1802 relevant words, with 775 for altruism and 1047 for mercy. Data processing was carried out by SPSS 19.0. Because Russian has two concepts denoting the social phenomena - 'altruism' and 'mercy' (mutual help), which are considered as synonymous, we will explain their most frequent use in the main, newspaper and spoken Russian Corpora (2016).

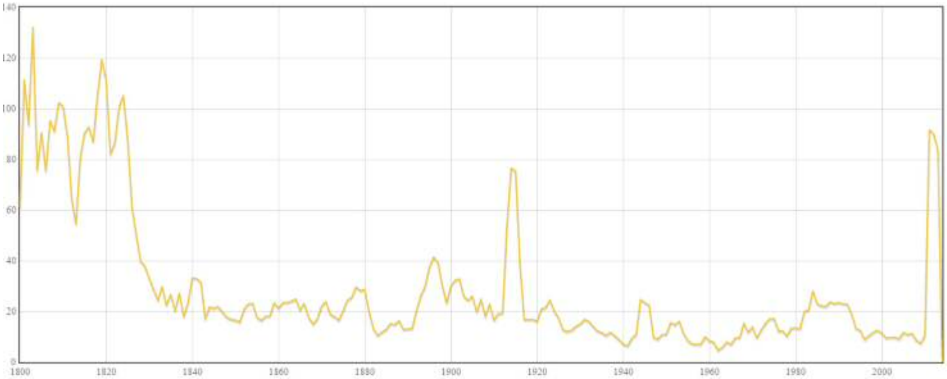

Fig. 1. The frequency of references to the category of "mercy" in the period of 1800-2014 years, the number of references per year (Source: Russian National Corpus URL:. Http: / /www.ruscorpora.ru /) (date Treatment: 02/03/2016)

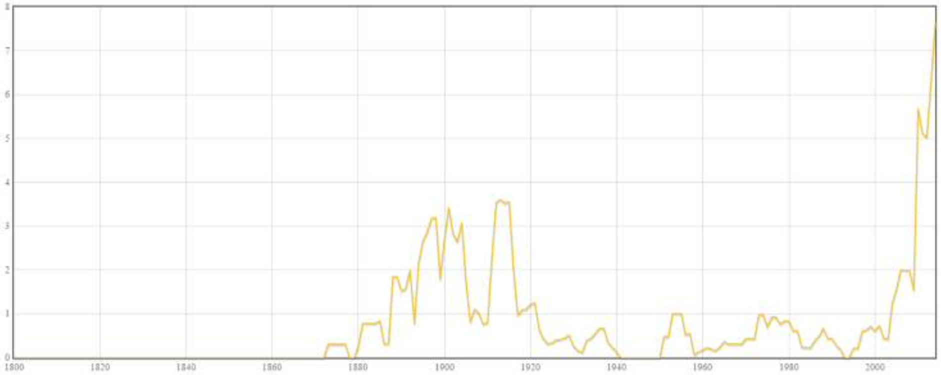

Fig. 2. The frequency of references to the category of "altruism" in the period of 1800-2014 years, the number of references per year (Source: Russian National Corpus URL:. Http: / /www.ruscorpora.ru /) (date Treatment: 02/03/2016)

3. Results

Both categories are mentioned in the following institutional contexts: "People`s actions", "Human himself", "Relations` quality", "State", "Social institutes", "Organisations", "Conceptions and ideology". The studying of the words “altruism” and “mercy”, according to these contexts, led to the results discussed below.

As we can see in fig.1 and fig.2 the concept “altruism” is less popular in Russian language, there are 1047 mentions of the mercy and only 309 mentions of altruism in corpora. The statistical analysis showed that the differences in use of these concepts have statistical significance: Pearson Chi-Square p < 0,000 (see Table 1).

Table 1. Chi-Square Tests

	Value	**df**	**Asymp. Sig. (2-sided)**
Pearson Chi-Square	424,507a	6	,000
Likelihood Ratio	409,543	6	,000
Linear-by-Linear Association	252,619	1	,000
N of Valid Cases	1356		

a. 1 cells (7,1%) have expected count less than 5. The minimum expected count is 4,33.

The analysis of mentioning of the words “altruism” and “mercy” in distinct contexts shows, that these concepts has different semantic load. The altruism is more often used in context of “relations` quality” and “people`s actions”, i.e. as personality characteristic; in opposition, the mercy is used in context of “organisations” and “conceptions and ideology”, i.e. as social feature (see Table 2 and Figure 3).

Table 2. The frequency of references to the category of "altruism" and "mercy" in contexts

			concept		**Total**
			altruism	**mercy**	
context	people`s actions	Count	74	128	202
		% within concept	23,9%	12,2%	14,9%
	human himself	Count	61	31	92
		% within concept	19,7%	3,0%	6,8%
	relations` quality	Count	101	47	148
		% within concept	32,7%	4,5%	10,9%
	state	Count	7	12	19
		% within concept	2,3%	1,1%	1,4%
	social institutes	Count	5	54	59
		% within concept	1,6%	5,2%	4,4%
	organisations	Count	15	405	420
		% within concept	4,9%	38,7%	31,0%
	conceptions and ideology	Count	46	370	416
		% within concept	14,9%	35,3%	30,7%
Total		% within concept	309	1047	1356
		Count	100,0%	100,0%	100,0%

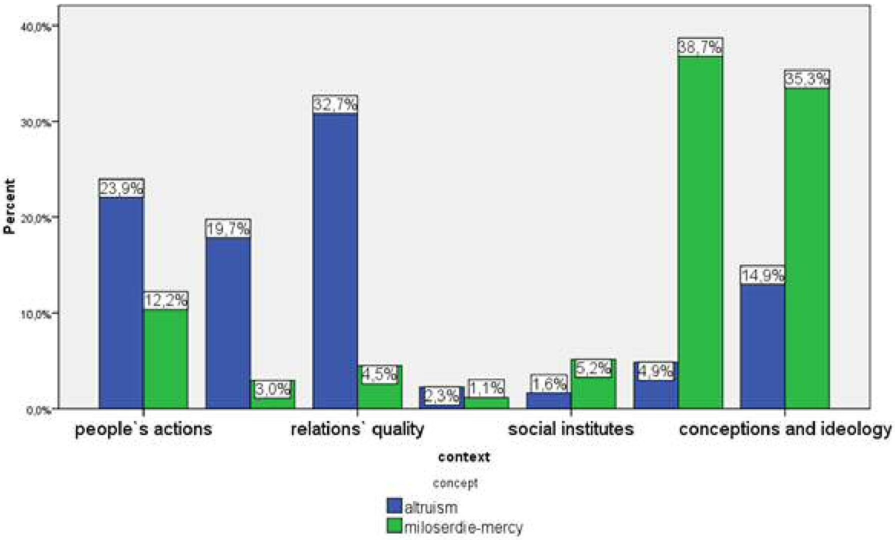

Fig. 3. The frequency of references to the category of "altruism" and “mercy” in different contexts (%)

Also there are different forms of using of analyzed concepts in spoken corpora and in newspaper: the altruism more typical for speeches, while mercy more frequently used in written sources (see Figure 4).

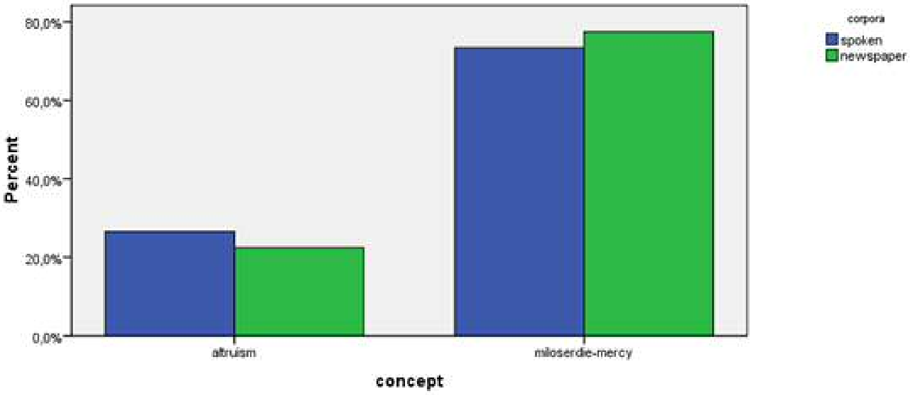

Fig. 4. The frequency of references to the category of "altruism" and “mercy” in spoken corpora and in newspaper (%)

The altruism in spoken corpora is used in context of “conceptions and ideology” (56% of all mentions in spoken corpora) and “human himself” (44%), while in newspapers it increasingly used in context of “relations` quality” (35.6%) and “people`s actions” (26.1%). This could means that people treat this concept as a part of ideals, beliefs and principles of social life, but in official sources it is handled as an impulsive cause.

The mercy in spoken corpora is also used in context of “conceptions and ideology” (58%), but in newspapers it is frequently mentioned in context of “organisations” (40%). It possibly means, that this word in written sources is used as reflection of action of charities. At the same time, a significant level of mentions of mercy in newspapers in context of “conceptions and ideology” (40%), shows that this concept has big value for morality and ethical basis of society (see Table 3).

Table 3. The frequency of references to the category of "altruism" and “mercy” in different corpora

concept				**corpora**		**Total**
				spoken	**newspaper**	
altruism	context	people`s actions	Frequency	0	74	74
			% in corpora	,0%	26,1%	23,9%
		human himself	Frequency	11	50	61
			% in corpora	44,0%	17,6%	19,7%
		relations` quality	Frequency	0	101	101
			% in corpora	,0%	35,6%	32,7%
		state	Frequency	0	7	7
			% in corpora	,0%	2,5%	2,3%
		social institutes	Frequency	0	5	5
			% in corpora	,0%	1,8%	1,6%
		organisations	Frequency	0	15	15
			% in corpora	,0%	5,3%	4,9%
		conceptions and ideology	Frequency	14	32	46
			% in corpora	56,0%	11,3%	14,9%
Total			Frequency	25	284	309
			% in corpora	100,0%	100,0%	100,0%

Table 3. The frequency of references to the category of "altruism" and "mercy" in different corpora (Cont.)

concept				corpora		Total
				spoken	newspaper	
mercy	context	people`s actions	Frequency	2	126	128
			% in corpora	2,9%	12,9%	12,2%
		human himself	Frequency	5	26	31
			% in corpora	7,2%	2,7%	3,0%
		relations` quality	Frequency	4	43	47
			% in corpora	5,8%	4,4%	4,5%
		state	Frequency	1	11	12
			% in corpora	1,4%	1,1%	1,1%
		social institutes	Frequency	3	51	54
			% in corpora	4,3%	5,2%	5,2%
		organisations	Frequency	14	391	405
			% in corpora	20,3%	40,0%	38,7%
		conceptions and ideology	Frequency	40	330	370
			% in corpora	58,0%	33,7%	35,3%
Total			Frequency	69	978	1047
			% in corpora	100,0%	100,0%	100,0%

4. Discussion and conclusion

We explored the everyday using of the two Russian words, 'altruism' and 'mercy', which originally had the same meanings. The concepts can be considered as synonymous in questionnaires and guides of interview. Based on an analysis by the Russian National Corpus, we have described seven contexts of word use for 'altruism' and 'mercy' in the following institutional contexts: "People`s actions", "Human himself", "Relations` quality", "State", "Social institutes", "Organisations", "Conceptions and ideology". The quantitative analysis shows differences in the use of these concepts. The differences between 'mercy' and 'altruism' are statistically significant for all three Russian corpora. Thus, their use as a synonym in sociological research is wrong.

The analysis shows both advantages and disadvantages of the proposed method of corpus linguistics for the operationalization of social categories. The positive side is our possibility to give a description of the context of its usage. We put emphasis on the noun (concept) that avoids the loss of meaning, which the respondents may use in answering the survey questions, and especially in the case of in-depth semi-structured interviews.

However, the proposed method has several disadvantages. Despite the fact that it is aimed at avoiding subjectivity and it is trying to raise objectivity in the operationalization of social categories, the encoding process itself continues to depend on the will of researchers. Existing techniques to reduce subjectivity, such as the coding of several (usually three) independent researchers has the elements of subjectivity because they are usually representatives of one culture and affiliated (related) persons. One attempt to avoid subjectivism is appeal to the Systemic Functional Linguistics (SFL) (Halliday, 1985; Rubtcova & Pavenkov, 2016) and Role and Reference grammar (RRG) (Van Valin, & LaPolla, 1997)

In our example with «altruism» and «mercy», we used conceptualization of a noun, whereas SFL and especially RRG offer us to change the focus on the process (verbal group). Verbal group exists regardless and independent of the investigator and can show the context of the roles of participants. In this case, we can avoid subjectivity in coding and may analyse grammatical structure of sentence. Our suggestions can be used in the preparation of questionnaires, guides, in an analysis of interview transcripts at the first stage. In the following studies, we need to further explore the possibilities of RRG and conduct double coding: with emphasis on nouns (as it is shown in this article) and with an emphasis on verbs (as suggested by RRG).

References

Abulof, U. (2015). Normative concepts analysis: Unpacking the language of legitimation. **International Journal of Social Research Methodology**, 18(1), 73-89. http://dx.doi.org/10.1080/13645579.2013.861656

Aijmer, Karin & Bengt Altenberg (eds.) (1991**) English Corpus Linguistics: Studies in Honour of Jan Svartvik**. London: Longman

Corpora and Discourse: **The Challenges of Different Settings** (2008) T. 31. P. 1–297. http://www.corpus-linguistics.com

Gladkova, A., & Romero-Trillo, J. (2014). Ain't it beautiful? The conceptualization of beauty from an ethnopragmatic perspective. **Journal of Pragmatics**, 60, 140-159. http://dx.doi.org/10.1016/j.pragma. 2013.11.005

Halliday, M.A.K. (1975). Sociological aspects of semantic change. In Luigi Heilman (ed), **Proceedings of the Eleventh International Congress of Linguists.** Bologna: Mulino.

Halliday, M.A.K. (1985) **Introduction to Functional Grammar**, London: Edward Arnold.

Halliday, M.A.K. (1991). "Corpus studies and probabilistic grammar". In **English Corpus Linguistics** ed by K. Aijmer & B. Altenberg, 30-43. London: Longman.

Halliday, M.A.K. (1992). "Language as system and language as instance: the corpus as a theoretical construct". Mouton de Gruyter, **Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82**, Stockholm, 4-8 August 1991, ed. Jan Svartvik (Trends in Linguistics Studies and Monographs 65).

Lazarsfeld, P. F. (1962). **American Sociological Review**, 27(6), 757-767. http://dx.doi.org/10.2307/2090403

McEnery T., Hardie A. (2012) **Corpus linguistics: Method, theory and practice**. Cambridge: Cambridge University Press.

McEnery T., Wilson A. (2001) **Corpus Linguistics**. Edinburgh: Edinburgh University Press.

Pavenkov, O., Pavenkov, V. & Rubtcova, M. (2015). The altruistic behavior: characteristic of future teachers of inclusive education in Russia. **Procedia - Social and Behavioral Sciences. International Conference Psiworld 2014** - 5th Edition. Volume: 187 Pages: 10-15.

Pavenkov O., Rubtcova M. Love as a concept in the religious philosophy of Pavel Florensky. **Anales del Seminario de Historia de la Filosofia.** 2016. V. 33. Issue 1. P. 163-180.

Pavenkov Oleg, Shmelev Ilya & Rubtcova Mariia (2016). Coping Behavior Of Orthodox Religious Students In Russia. **Journal for the study of religions and ideologies**. 15(44), 205-224.

Pavenkov, O. & Pavenkova, M. (2016). Discourse analysis based on Martin and Rose's taxonomy: a case of promoting student discourse on the CLIL PhD programme in religion philosophy. **Revista Electrónica Espacio Teológico.** 10: 17 Pages: 129-139

Rubtcova, M. P.& Pavenkov, O. V. (2016). Interaction of Russian and English academic genres in CLIL doctoral programmes of Management Sociology: a gap in the process of implementation. **Dilemas Contemporáneos: Educación, Política y Valores**. 3: 2.

Rubtcova, M. & Vasilieva, E. (2015). Managing Human Capital: How Public Servants Support the Governance's Performance Conception in Russia. **Proceedings of 2015 International Conference on Public Administration**, 237-247.

Rubtcova M., Pavenkov O. & Varlamova J. (2017) Multisemiotic Analysis of Orthodox Patriarchs' Photographs: Cross-Cultural (Indian and Russian) Differences in Interpretation of Interactive Meanings. In: **Linguistics: Past, Present and Future Perspectives**. Ed. Harry Barnes. Nova Science Publishers Inc., New York

Rubtcova, M. & Pavenkov, O. (2016). Systemic Functional Linguistics as a Macro-sociolinguistics Framework: The Stages of Development. **Studies in Linguistics**. 38: 471-492.

Rubtsova, M. V. & Vasilieva, E. A. (2016). Conceptualization and operationalization of notion «trust» for applied sociological research. **Sotsiologicheskie Issledovaniya**. Issue: 1 Pages: 58-65.

Russian National Corpus. (2016). Retrieved from http://www.ruscorpora.ru/

Rubtsova, M. V. 2007. Manageability: sociological theoretical analysis of notions. **Sotsiologicheskie Issledovaniya** (12), 32-38.

Rubtsova, M. V. 2011. Governmentability In interactions of subjects. Traditional and new practices. **Sotsiologicheskie Issledovaniya** (2), 46-53.

Volchkova, L.T. and Pavenkova, M.V., 2002 Sociology of management. Theoretical principles. **Sotsiologicheskie issledovaniya**. Issue: 3:141-144

Van Valin, R.D. Jr. & R. J. LaPolla, 1997. **Syntax, Structure, Meaning and Function.** Cambridge: Cambridge University Press

Usiaeva, A., Rubtcova, M., Pavlenkova, I.; et al. (2016). Sociological Diagnostics in Staff Competency Assessments: Evidence from Russian Museums. **International Journal of Production Management and Engineering**. 4: 1 Pages: 29-33.