The Language of Altruism: Corpus-Based Conceptualization of Social Category for Management Sociology

Mariia Rubtcova¹, Oleg Pavenkov², Vladimir Pavenkov² & Vasileva Elena³
¹ Faculty of Sociology, Saint Petersburg State University, Saint Petersburg, Russian Federation
² Saint Petersburg State University of Cinema and Television, Saint Petersburg, Russian Federation
³ Academy of science of Republic of Sakha (Yakutia), Yakutsk, Sakha Republic (Yakutia), Russian Federation

Correspondence: Mariia Rubtcova, Department of Social Management and Planning, Faculty of Sociology, Saint Petersburg State University, 7-9, Universitetskaya nab., Saint Petersburg, 199034, Russian Federation. Tel: 79-21-381-1889. E-mail: m.rubtsova@spbu.ru

Received: March 6, 2015   Accepted: April 8, 2015   Online Published: May 16, 2015

Abstract
Management sociology poses the problem of the quantitative interpretation of qualitative research. The article deals with the corpus-based method, which can be considered as one of the solution tools. Based on ‘grounded theory’ methodology (Strauss & Corbin, n. d.) and partly debating with conceptual analysis (Sartory & Goertz, n. d.), we propose to elaborate a definition of the concept using quantitative research. The authors identified useful areas of corpus linguistics in the analysis of social and management phenomena and distinguished between corpus linguistics and sociological content analysis methods:
- Direct appeal to the everyday use of the language increases the objectivity of the research;
- A corpus provides a large quantity of representative data;
- The possibility of diachronic and synchronic comparative studies;
- The method itself is not time-consuming and expensive.

We chose the category ‘altruism’ as an example to demonstrate the possibilities of the method. The analysis shows features in the representation of altruism in Russian that the field of management sociology needs to address for the preparation of questionnaires, interview guides and transcript analysis.

Keywords: conceptualization, social categories, management sociology, corpus linguistics, altruism, Russian

1. Introduction
This article presents a study of the conceptualization of social and management categories from the sociological and corpus linguistics perspective and includes a small-scale study that analyses how the concept of altruism and its Russian equivalent, ‘vzaimopomosh’ - mutual help’, is used in Russian. The conceptualization of social categories remains one of the most difficult stages of sociological research. Despite the fact that it has been long regarded as a necessary element (Lazarsfeld, 1962), discussions about the clear conceptualisation and operationalisation of concepts have continued (Leydesdorff, 2013). Simultaneous use of traditional (mono-cultural) and international terms, combined with large-scale cross-national research (for example, the World Values Survey and European Values Study) exacerbates this problem. In this work, we propose to complement existing approaches to the conceptualisation and operationalisation of social and management categories with the achievements of modern linguistic science, primarily corpus linguistics.

Traditionally, conceptualisation begins with a literature review, accompanied by a clarification of the concept’s meaning (Benoit, 2012, p. 219), thus creating concept preceded quantification (Sartori, 1970). However, we are more inclined to adopt ‘grounded theory’ opportunities (Strauss & Corbin, 1997) and would suggest that we start to build a concept using corpus-based quantitative research.

Modern corpus linguistics is the study of language based on examples of real language use (McEnery & Wilson, 2001). Corpus linguistics involves the creation of linguistic corpora, which may have different goals and means of ensuring representativeness. The main corpus represents the whole area of language usage, including fiction,
poetry, the newspaper corpus focusing on newspaper vocabulary, and the parallel corpus, which covers several languages.

An oral (spoken) corpus as one of the most interesting forms for sociologists allows them to approach the everyday life of respondents and their daily vocabulary, perceptions and actual language-use activities. For example, the British oral corpus contains recorded transcriptions of informal oral conversations of volunteers from different cultures, demographic groups and social classes in modern English society. The second part of the British oral corpus is spoken language collected in a variety of contexts, ranging from formal business meetings to entertainment radio (British National Corpus, 2015). Thus, the oral corpus allows us to qualitatively and quantitatively analyse information about the use of language in different social groups, which is very important for the sociological study of social management systems.

Corpus linguistics is similar to content analysis: like content analysis, it involves working with word or concept frequency. However, there are important differences.

Firstly, corpus linguistics has as its immediate goal the study of language but not of a particular social phenomenon (altruism, trust, power, and so on). For the compilers of the corpus the most important aim is to clarify the language features of a word (concept) (Corpora and Discourse, 2008). They do not initially have such intentions; however, they create for the management sociology interesting material that extends typical sociological frames of social categories’ perception (Abulof, 2015). The concentration of the linguistic corpus in natural language provides access to answers of how people actually discuss policy (Blei, 2012).

For example, Russian management sociologists think that the deep Russian folk category of ‘vzaimopomosh’- mutual aid’ is closely connected with the spirit of the traditional Russian community (19th Century). When they construct the ideological concept of consolidating society, they use that word in this context. However, a respondent can associate ‘vzaimopomosh’- mutual aid’ with the Soviet ‘mutual insurance’ (‘cassa vzaimopomoshi’), that confirmed through the study of this context in the Russian National Corpus. Traditional analysis of the literature and definitions does not provide such ‘findings’. However, management sociology needs to know exactly how society can understand the ideological conception. Thus, if content analysis requires serious conceptual design through the subjective interpretations of a sociologist, in the Russian corpus the material is already presented, and the sociologists can only use what is available. To some extent, this adds objectivity to the work of the researcher.

Secondly, the corpus has a significant size that facilitates serious statistical analysis. The majority of the national corpora comprise more than 500 million words (Corpora & Discourse, 2008). Materials are usually represented by years and months over a sufficiently long period; therefore, the 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). Thus, the researcher can establish the frequency of the use of the chosen categories across different periods of time (diachronic word frequency): for example, the maximum frequency of application of the concept ‘manageability’ is in the period spanning the years 2002-2004, which subsequently decreases several times, whilst the concept of ‘trust’, on the other hand, has an entirely different profile. In the 20th Century, peaks occur at times of social upheaval and also during transitions to other directions of development; this can give the researcher context and insight into how people use an appropriate concept. The aggregate of phrases also allows the in-depth study of the changing context of the concept value and main topics (see, for example: Liu, 2012; Abulof, 2015).

Thirdly, most of the national corpora are built on a single principle, which is fundamental for cross-cultural and cross-national studies (Corpora & Cross-linguistic Research, 1998). Before conducting comparative sociological studies, the corpus linguistics allow us to identify differences in the interpretation of the social phenomenon in different cultures (Rawoens, 2010).

Fourthly, the corpus study is not so time-consuming and expensive. Materials have been chosen and provided in mind of a user-friendly search format for downloading and uploading data, and compatibility with Excel, SPSS and other programs (McEnery & Hardie, 2012). This speeds up the interpretation procedure. Some researchers are so inspired by the possibilities of corpus linguistics that they believe it will allow us to analyze a huge amount of text quickly and inexpensively and has the potential to fundamentally change research (Schonhardt-Bailey, 2005)

Like any method of collecting and processing information, corpus-linguistic research has its limitations, which include (but are not limited to) the manual coding of information as regards context. Despite the proliferation of automatic coding systems (Grimmer & Stewart, 2013), coding is often carried out manually, predominantly owing to a lack of processing algorithms in non-English languages. Currently, this problem is solved by
including several independent experts in the coding process; however, it can be assumed that the noted rapid development of software will allow us to solve a certain degree of this problem, and that part of the context coding process will soon be automated.

Thus, we believe that corpus linguistics can provide management sociology with a necessary analytical tool, allowing us to identify potential problems in the use and interpretation of key research concepts. As an example, we propose the creation of such a tool for the concept ‘altruism’.

In previous work, we have noted that the boundaries of social categories in the Russian language are difficult to determine clearly (Volchkova & Pavenkova, 2002; Rubtsova, 2007, 2011; Pavenkov, 2014). In our view, the study of the concept should include the study of synonyms - as sociologists and respondents often use them. It can be said that the theme of the use of synonyms in sociological studies is rarely studied. In Russian the concepts altruism and ‘vzaimopomoshh’ (mutual help) are regarded as synonyms in many Russian conceptions (Kropotkin, 1988).

To encode and count the frequency of the aforementioned concepts the Russian National Corpus was used. TheRussian National Corpus (2015) is a reference system of Russian electronic texts that is created by linguists for research purposes. National Corpus represents the language’s development in a diversity of genres, registers, and social features, written and oral texts as a balanced and representative composition. The proportion of text types in the Corpus is based on everyday use. It is not a library. Electronic libraries are not suitable for scientific research regarding the nature of language because their purpose is the content of the texts. Unlike an electronic library, the National Corpus is not a collection of ‘interesting’ or ‘useful’ texts. The texts are extracts of words in context and can include new contexts which a social science researcher would not be able to pay attention; however, respondents may perceive them. Therefore, these contexts might significantly affect a meaning that the researcher did not assume.

Using the Russian National Corpus (2015), we can establish similarities and differences in the use of the word (altruism) in the context of (‘vzaimopomoshh’)(mutual help) as the Russian equivalent and try to find answers to our research questions:

- Are the concepts altruism and ‘vzaimopomoshh’ (mutual help) interchangeable as synonyms?
- Has the use of these words in oral and written speech essential differences?

To this aim we can formulate and check the following hypotheses:

Hypothesis 1. The use of these words in contexts has statistically significant differences, so they are not interchangeable.

Hypothesis 2. The use of these words in oral and written speech has statistically significant differences.

2. Method

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 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, conception-ideology.

Data processing was carried out using SPSS.

2.2 Sampling Procedures

Because Russian has two concepts denoting the social phenomena - ‘altruism’ and ‘vzaimopomoshh’ (mutual help), which are considered as synonymous, we will explain their most frequent use in the main, newspaper and spoken Russian Corpora (2015). In Table 1 we can see the different word forms ‘altruism’ and ‘vzaimopomoshh’ in Russian.
Table 1. Wordforms for concepts ‘altruism’ and ‘vzaimopomoshh’-mutual help’ in the Russian National Corpus

<table>
<thead>
<tr>
<th>Russian (transliteration)</th>
<th>English</th>
<th>Russian (transliteration)</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al'truizm</td>
<td>Altruism</td>
<td>Al'truist</td>
<td>Altruist</td>
</tr>
<tr>
<td>Al'truist</td>
<td>Altruist</td>
<td>Al'truisty</td>
<td>Altruists</td>
</tr>
<tr>
<td>Al'truistichnyj</td>
<td>Altruistically</td>
<td>Al'truisticheskaja</td>
<td>Altruistic</td>
</tr>
<tr>
<td>Al'truisticheskaja</td>
<td>Altruistic</td>
<td>Vzaimopomoshh'</td>
<td>mutual help</td>
</tr>
<tr>
<td>Al'truistom</td>
<td>Altruists</td>
<td>Vzaimopomoshhi</td>
<td>mutual help</td>
</tr>
<tr>
<td>Al'truistka</td>
<td>Altruistka</td>
<td>Al'truistichnoj</td>
<td>Altruistically</td>
</tr>
<tr>
<td>Al'truisticheskom</td>
<td>Altruistic</td>
<td>Al'truizma</td>
<td>Altruism</td>
</tr>
<tr>
<td>Al'truizma</td>
<td>Altruist</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Russian National Corpus consists of 1533 relevant words, with 775 for altruism and 778 for ‘vzaimopomoshh’ (mutual help). The numbers are almost equal. At the same time, the concept of ‘vzaimopomoshh’ presented more in main and spoken corpora, the concept ‘altruism’- in the newspapers’ lexis featured at 5% (Table 2.). We have carried out a check test in SPSS and confirmed that the two concepts are presented without statistical differences (p = 0.432) (see Figure 5).

Table 2. The distribution of the concepts ‘altruism’ and ‘vzaimopomoshh’ (mutual help) in the Russian National Corpus

<table>
<thead>
<tr>
<th>Total documents, sentences and words in Corpus</th>
<th>Corpus Total documents, sentences, words</th>
<th>newspaper documents, sentences, words</th>
<th>spoken documents, sentences, words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total documents, sentences and words in Corpus</td>
<td>85996 19362746 229968788</td>
<td>332720 12920590 173518 798</td>
<td>3525 623625 10 754 403</td>
</tr>
<tr>
<td>Word ‘altruism’ in the Corpus, including wordforms</td>
<td>775 446</td>
<td>284</td>
<td>25</td>
</tr>
<tr>
<td>Word ‘vzaimopomoshh’ (mutual help) in the Corpus, including wordforms</td>
<td>778 481</td>
<td>269</td>
<td>28</td>
</tr>
<tr>
<td>Total words ‘altruism’ and ‘vzaimopomoshh’ (mutual help) in the Corpus, including word forms</td>
<td>1533 927</td>
<td>553</td>
<td>53</td>
</tr>
</tbody>
</table>

3. Results

Table 3. The spheres of application (context) of ‘altruism’ and ‘vzaimopomoshh’-mutual help

<table>
<thead>
<tr>
<th>The spheres of application - contexts</th>
<th>People’s actions</th>
<th>Humans themselves</th>
<th>Relations’ quality</th>
<th>State</th>
<th>Social institutes</th>
<th>Organizations</th>
<th>Conceptions and ideology</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>vzaimopomoshh ‘mutual help’</td>
<td>Count Expected</td>
<td>124,3 8,4%</td>
<td>49,7 13,6%</td>
<td>149,2 10,2%</td>
<td>47,2 7,3%</td>
<td>34,0 40,2%</td>
<td>166,5 20,3%</td>
<td>207,1 100,0%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>8,4%</td>
<td>13,6%</td>
<td>10,2%</td>
<td>7,3%</td>
<td>40,2%</td>
<td>20,3%</td>
<td>100,0%</td>
</tr>
<tr>
<td>altruism</td>
<td>Count Expected</td>
<td>120,7 23,8%</td>
<td>48,3 13,0%</td>
<td>144,8 13,6%</td>
<td>45,8 19,9%</td>
<td>33,0 1,3%</td>
<td>161,5 2,0%</td>
<td>200,9 1,3%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>23,8%</td>
<td>13,0%</td>
<td>13,6%</td>
<td>19,9%</td>
<td>1,3%</td>
<td>2,0%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count Expected</td>
<td>245,0 16,0%</td>
<td>98,0 6,4%</td>
<td>294,0 19,2%</td>
<td>93,0 6,1%</td>
<td>67,0 4,4%</td>
<td>328,0 21,4%</td>
<td>408,0 26,6%</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>16,0%</td>
<td>6,4%</td>
<td>19,2%</td>
<td>6,1%</td>
<td>4,4%</td>
<td>21,4%</td>
<td>26,6%</td>
</tr>
</tbody>
</table>
With a fleet of almost balanced corpora, we want to see how people utilise these words. Table 3 shows the differences in the context of ‘altruism’ and ‘vzaimopomoshh’-mutual help’. There are no contexts in which these concepts can be used as synonyms (Figure 1). A Chi-Square test demonstrated the significance differences in the spheres of application (context) of the words ‘altruism’ and ‘vzaimopomoshh’-mutual help’ (p < 0.001). The more rigorous test of Cramer’s V confirmed these results (p < 0.001). We can make the conclusion that the use of these words in contexts has statistically significant differences, so they are not interchangeable (see Hypothesis 1).

According research question 2, we need to see how people utilize these words in speech and writing. Russian Corpus (2015) does not have a separate written corpus. However, we can use the newspaper and the main corpora, because they consist of written sources.

To check oral or written influence on the use of ‘altruism’ and ‘vzaimopomoshh’-mutual help’ we have used the ANOVA-test. It allows us to identify the factors influencing the use of concepts. The results are in Table 4.

Table 4. The factors influencing the use of ‘altruism’ and ‘vzaimopomoshh’-mutual help’

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Hypothesis</td>
<td>364,731</td>
<td>1</td>
<td>364,731</td>
<td>1,470E3</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1,770</td>
<td>7,132</td>
<td>,248a</td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>Hypothesis</td>
<td>44,377</td>
<td>6</td>
<td>7,396</td>
<td>20,216</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8,432</td>
<td>23,047</td>
<td>,366b</td>
<td></td>
</tr>
<tr>
<td>Corpora</td>
<td>Hypothesis</td>
<td>1,024</td>
<td>2</td>
<td>,512</td>
<td>,892</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8,175</td>
<td>14,236</td>
<td>,574c</td>
<td></td>
</tr>
<tr>
<td>context * corpora</td>
<td>Hypothesis</td>
<td>9,031</td>
<td>12</td>
<td>,753</td>
<td>4,791</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>237,520</td>
<td>1512</td>
<td>,157d</td>
<td></td>
</tr>
</tbody>
</table>

a. .256 MS (corpora) + MS (Error); b. .351 MS (context*corpora) + MS (Error); c. MS (context*corpora) + .299 MS (Error); d. MS (Error)

As we can see, the choice of a linguistic corpus (main, newspaper or spoken) has no influence (p = 0.432) on the results of the concepts ‘altruism’ and ‘vzaimopomoshh’-mutual help’. From another side, the context and interaction between context and corpus are strong - statistically significant (Sig., p < 0.001). The Chi-Square Test shows the significant differences between the use of the word ‘altruism’ and ‘vzaimopomoshh’-mutual help’ (p < 0.001) in main, newspaper or spoken linguistic corpora. We can see them on a graph (Figure 2&3).
Thus, we find out that there is a statistically significant difference between the use of concepts in oral and written speech (see Hypothesis 2).

4. Discussion

As we can see, quantitative operationalization of social category can improve the accuracy of the obtained data, eliminating erroneous or ambiguous use of these categories, and determine the frequency of their use in different contexts to identify the relevance of social problems. As P. Lazarsfeld (1962) has argued, operationalization of concepts can contribute to the creation of a model describing a social problem, which can be studied through empirical social research. This model is created during operationalization of the concept and is used as a basis for operations research.

We have explored the use of the two Russian words, ‘altruism’ and ‘vzaimopomoshh’-mutual help’, which originally had the same meanings. The concepts are regarded as synonymous in many Russian social science theories. Based on an analysis by the Russian National Corpus, we have described seven contexts of word use for ‘altruism’ and ‘vzaimopomoshh’: people’s actions, humans themselves, quality of relations, state, social institutes, organisations, and conceptions/ideologies.

The quantitative analysis shows differences in the use of these concepts. The differences between ‘vzaimopomoshh’-mutual help’ and ‘altruism’ are statistically significant for all three Russian corpora: main,
spoken and newspaper. Only in a few contexts can these concepts be used as synonyms: as features of quality of relations, as a conception and ideology. This confirms our hypotheses.

The concept of altruism is often related through the use of pronouns and is more personal, whereas the concept of vzaimopomoshh’ originally stemmed from life in local Russian communities. However, in the 20th century, the state has adopted the term as its template, which explains its frequent mention in the context of organizations, social institutions, and the state itself. In these patterns, the lyrical original meaning of the term, “happy life together,” is almost lost.

We have processed the data corpus that gives us to work with in questionnaires and interviews. We use corpus linguistics to identify the contexts of categories use. This allows us to develop ‘grounded theory’ and to expand our knowledge about the use of concepts in questionnaires. Thus, we operationalized ‘concept- in- context’ not by applying the General theory to a particular case, but as the basis of our particular theoretical framework, which was from the outset included and has developed within our case. This approach is very important when studying different cultures - when we want to conduct a survey in very different countries with one questionnaire. In General, we believe this methodology is extremely important in all cases of intercultural differences. The conceptualisation of social categories prevents the loss of meaning, which is important when we conduct in-depth interviews, focus groups and in data processing of transcripts.

In future studies, we would have paid more attention to the replenishment of the spoken Russian National Corpus with sociological interview transcripts, which improves our ability to seek social differences in the use of these concepts.

References


Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).