

*Svetlana Savchuk (Moscow, Russia)*

## **The Russian National Corpus as a Tool for the Research on Grammatical Variability**

### **Abstract**

The Corpus-based approach seems to be extremely useful and fruitful in research of grammatical variability and correlation between norm and real language usage. The following three aspects of the problem of norm and variation should be distinguished.

If we are interested in its synchronic aspect, we will study the co-existing variants, their distribution among different spheres of functioning, social, professional variants of the language, *etc.*

If we deal with the problem in its diachronic aspect, we will take into consideration some changes in usage, appearance and disappearance of the variants, changing of correlation between several variants, increasing and declining trends, and so on.

The third aspect concerns the evaluation of the different variants with regard to a standard, or so called codified norm fixed in dictionaries and grammar manuals. A variant can be codified in the literary language or can stay beyond the norm. There is quite a natural discrepancy: some variants can fluctuate over a long period of time while normative estimations of these variants change continually. Consequently the evaluation of normative recommendations should also be regarded from synchronical and diachronical points of view, and in relation to real usage.

The paper presents the Russian National Corpus as a tool for the research on grammatical variability. The RNC has all the necessary quantitative and qualitative characteristics to provide an adequate set of examples for various types of linguistic researches. There are two main chronological components in the RNC: 1) the corpus of modern written and spoken texts (of the period from the second half of the XXth century up to nowadays); 2) the diachronic corpus that includes texts dating from the XVIIIth to the first half of the XXth century (the project is still in progress). Being a representative collection of texts, the RNC reflects the Russian language usage in two dimensions: the 'horizontal' one (in functional varieties) and the 'vertical' one (in historic perspective).

Using the corpus data one can consider the problem of grammatical variation in three aspects: 1) to set the correlation of variants in contemporary language usage; 2) to study the evolution that has been occurring during a certain period; 3) to compare these findings with recommendations of dictionaries and grammar manuals in order to evaluate the adequacy of the recommendations to real usage.

The main body of the RNC contains the texts of the last two centuries, therefore it is suitable for studying of short- and medium-term language changes. That is why one can receive rather reliable results in studying such issues as the following:

- morphological variants of nouns, verbs and their evolution;
- variants of government, agreement and other syntactic constructions;
- productivity of word-formative models and means of word-forming, *etc.*

The paper includes also some results of corpus-based research of "weak points" of literary norms, causing the appearance and co-existence of variants. The possibilities of the RNC in linguistic investigations and language learning are demonstrated here by means of qualitative and quantitative analyses of variability of grammar forms and constructions in spoken and written texts in historic perspective.