2021

Review: *Quantitative Approaches to the Russian Language*

Olesya Kisselev

*University of Texas at San Antonio*

Follow this and additional works at: https://scholarsarchive.byu.edu/rlj

Part of the Slavic Languages and Societies Commons

**Recommended Citation**


DOI: https://doi.org/10.70163/0036-0252.1320

Available at: https://scholarsarchive.byu.edu/rlj/vol71/iss1/10

This Book Review is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Russian Language Journal by an authorized editor of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

The volume edited by Mikhail Kopotev, Olga Lyashevskaya, and Arto Mustajoki is a testament to the reinvigorated interest in quantitative approaches to the study of the Russian language that has marked the past decade and a half. The trend, largely prompted and sustained by the widespread availability of large and well-annotated corpora, that is, digital collections of linguistic data (Kopotev and Mustajoki, 2008), is on full display in the edited volume. The methods and instruments featured in the collection are overwhelmingly corpus-based; however, many of the studies described in the papers showcase how various approaches to the analysis of language data can be successfully combined.

The ten papers that comprise the volume are divided across four parts. The first two introductory papers are followed by Part II: Topics in Semantics, Part III: Topics in Lexicon-Grammar Interface, and Part IV: Topics in Language Acquisition.

The first paper in the collection (written by the editors) provides a detailed overview of quantitative approaches to the study of Russian, a survey of early and contemporary Russian language corpora and other corpus resources, and a short introduction to the basic statistical methods widely applied in the field. The style and organization of the paper make the topic accessible to readers of various degrees of proficiency in linguistic methods and ensure that the readers can fully engage with the rest of the papers in the volume.

Following the lead of the first paper, the paper by Maria Khokhlova provides a user-friendly yet thorough exploration of one of the foundational issues of corpus linguistics research, that of how linguistic structures (broadly conceived) are distributed across corpora of different sizes and composition. The results of the study, highlighting strong correlations between the order of items on frequency lists and the length, composition, and design of a corpus, provide an important insight into interpreting results of studies based on the analysis of frequencies.

The notion of frequency is taken further in Part II, specifically, in the paper by Anastasiya Lopukhina, Konstantin Lopukhin, and Grigory Nozyrev; here, the authors explore the possibility of analyzing the
frequency of word senses rather than word forms that are—more often than not—polysemous. The paper presents and evaluates a model designed to determine noun sense frequency in Russian. The model is based on semantic vectors, that is, an approach to understanding the sense of a word based on the analysis of the contexts in which it is embedded; it is trained on a raw corpus and is subsequently supplied with the data from the Active Dictionary of Russian (Apresjan 2014). The results of the study suggest a wide applicability of sense disambiguation procedures to the study of semantics, to lexicography, and to language teaching, especially in creating learner dictionaries and other teaching and learning resources.

The other two chapters in Part II approach the exploration of word sense from a slightly different methodological perspective. Thus, the paper by Olga Lyashevskaya, Maria Ovsjannikova, Nina Szymor, and Dagmar Divjak explores the possibility of understanding the meaning of Russian modal words through examining formal and semantic context of the modals in the written subcorpus of the Russian National Corpus. The analyses of the formal and semantic features of the extracted contexts for the modal words involved additional annotation carried out by human raters followed by statistical analyses of the results. The value of the paper lies both in its promise to add empirical credence to the many classifications of modal words proposed in the theoretical literature and in its service as an effective model for conducting linguistic research that combines corpus-based and statistical methods with formal and/or logical models for various linguistic categories and concepts.

Another interesting case study is presented in the third paper in Part II (by Andrey Kutuzov and Elizaveta Kuzmenko). Utilizing the methodology of distributed neural embedding models, this paper investigates how changes in the meaning of lexical items occur over time. Analyzing texts created before 1917, during the Soviet era, and after the collapse of the USSR in 1991, the authors tracked semantically similar nouns and adjectives and assessed how the meanings of these words have shifted historically. The paper may be of great interest to anyone attempting a study of the period from either the historic, literary, social, or cultural studies perspectives and is a potent illustration of the value of digital approaches to humanities.

The notion of frequency, as well as the notion of lexical ambiguity, is picked up again in the paper by Alexander Pipersky, which opens the section on lexicon-grammar interfaces. Focusing on Russian bi-aspectual verbs, that is, an aspectral pair “containing two homophonous verbs having largely intersecting sets of forms” (116), the author tests three quantitative methods used to determine the status of a bi-aspectual verb. The first
one involves assessing relative frequencies of a verb’s perfective and imperfective gerunds; the second method analyzes grammatical profiles of bi-aspectual verbs (as reflected in the distribution of its grammatical forms in the corpus); while the third method involved a perception experiment with native speaker participants. The triangulation of the methods can reliably indicate whether a verb is more imperfective- or perfective-like, helping disambiguate verb senses.

The classic example of lexicon–grammar interfacing, namely collocations, are the topic of the paper by Lidia Pivovarova, Daria Kormacheva, and Mikhail Kopotev. This work addresses methodological concerns of extracting collocations by comparing the efficiency of different statistical methods traditionally used to extract collocations—specifically, t-score, log-likelihood, and Dice. By cross-evaluating the results of the statistical extraction procedures with dictionary entries and human rater evaluations, the authors empirically establish the advantages and limitations of the statistical approaches to the extraction of collocations from a corpus. The results of the study are important for collocational studies of either standard or learner varieties of Russian.

The paper concluding Part III, by Anastasia Bonch-Osmolovskaya, demonstrates how quantitative corpus methods can be used in linguistic research. By studying the distribution of dative subject constructions with predicative and adjective forms, the author illustrates that frequency rates of dative subjects differ depending on the type of predicate. The results also differ depending on the subcorpora use, showing clear tendencies across historic periods.

The last two chapters in the volume break away from the collection’s larger methodological bend toward corpus analytic procedures and illustrate the more traditional (psycho)linguistic approaches to the analysis of elicited data. The study by Alexei Korneev and Ekaterina Protassova focuses on the development of literacy in bilingual Finnish-Russian children. By using a computer-based digitized tablet that records the process of handwriting, the researchers assessed the development of writing skills, taking into consideration factors such as different home literacy practices. The results of rmANOVA tests reveal that home language may support language skills, but literacy training in another language/script supports the quality of writing as well.

The paper concluding the volume, written by Robyn Orfitelli and Maria Polinsky, explores the limits of the use of grammaticality judgment tasks (GJT) in cases where the use of GJT implies higher demands on metalinguistic decision-making, such as in the case of heritage speakers.
subjected to elicited judgments. Discussing the results of the two
studies where heritage speakers of Russian were subjected to GJT or a
comprehension task, the authors draw on statistical measures to point out
the limits of acceptability judgment testing procedures.

Overall, the volume lives up to the stated goal of presenting “current
trends in researching Russian quantitative linguistics, to evaluate the
research methods vis-à-vis Russian data, and to show both the advantages
and the disadvantages of the methods” (3). The papers are written in a
user-friendly and accessible way, making the collection a great resource for
anyone looking to expand their knowledge of current trends in linguistic
research.

Olesya Kisselev
University of Texas at San Antonio

References
In A. Mustajoki, M. Kopotev, L. Birjulin, & Ju. Protasova (Eds.),
Instrumentarij rusistiki: Korpusnye podxody (pp. 7-24). Helsinki:
Helsinki UP.

Intermediate Course in Russian Language for International Relations,
National Security, and Socio-Economics, 7th edition. Dubuque, IA:
Kendall Hunt. 552 pages.

The newest, seventh edition of Political Russian continues to serve the needs
of students specializing in international relations, Russian affairs, politics,
economics, government, national security, and related fields who wish to
improve functional proficiency in reading, listening and speaking at the
intermediate and advanced levels. However, the book’s reach is much
wider. Several features of this course, including those new to this edition,
make it a great resource for Russian-language instructors teaching a variety
of courses and proficiency levels.

The book is organized into two main parts, or “concentric circles,”
that target learners at different starting proficiency levels, beginning
with novice for second-year students in Circle 1. Circle 2 builds on and
expands the skills and competencies acquired in Circle 1. The book aims
to develop students’ functional proficiency in reading, listening, and
speaking; teach them effective strategies for working with authentic texts;