

RUSSIAN SYNTAX IN THE MIRROR OF CORPUS LINGUISTICS: NEW DATA ON NORM AND USAGE AS A RESOURCE FOR MODERN LINGUODIDACTICS

Mustanov Ubaydillo Zulfanovich

Lecturer, Interfaculty Department of Russian Language, JizPU

Abstract

The article is devoted to the application of corpus linguistics data — primarily the Russian National Corpus (RNC) — in teaching syntax at secondary and higher school levels. The gap between normative grammatical descriptions recorded in textbooks and the actual syntactic usage captured by corpus data is analyzed: the functioning of parcellation, ellipsis, non-standard word orders, and colloquial syntactic constructions across texts of various styles and genres. The necessity of transitioning from a normative-prescriptive to a corpus-descriptive model of syntax teaching is substantiated — one that does not abolish the normative aspect but significantly enriches students' understanding of real language functioning. A system of lessons built on working with corpus data is proposed, ranging from simple RNC searches to independent mini-research of syntactic phenomena. The experience of implementing the corpus approach in classes on contemporary Russian language at Karshi State University is described.

Keywords: Corpus linguistics, Russian National Corpus, syntax, language norm, usage, linguodidactics, parcellation, ellipsis, corpus approach, descriptive grammar.

Introduction

Imagine a typical syntax lesson in high school: students parse sentences, place punctuation marks, identify types of subordinate clauses. The rules are clear, the answer is either right or wrong. The textbook is self-assured. But open any page of a contemporary novel, a magazine article, or even a quality Telegram channel — and a remarkable picture emerges: living Russian prose abounds with constructions that the textbook classifies as erroneous or stylistically substandard. Parcellated sentences. Like this. Word by word. Elliptical utterances without a predicate. Inversions that transform neutral word order beyond recognition.

This gap between the norm recorded in the textbook and usage — that is, actual language practice — is neither accidental nor a sign of linguistic degradation. It is evidence that the living syntactic system is richer and more dynamic than any normative description. And it is corpus linguistics — the branch of linguistics that studies language on the basis of representative text collections — that has for the first time given researchers a tool for the systematic study of this gap.



The Russian National Corpus (RNC), created at the Institute of the Russian Language of the Russian Academy of Sciences and currently comprising over 1.5 billion word occurrences, has been a genuine revolution in the study of the Russian language. However, its didactic potential has been exploited only to a small degree: in school practice, the corpus remains a tool for scholars rather than for students. Bridging this gap is the central task of the present article.

1. Corpus Linguistics and the Problem of Language Norm

The concept of language norm has always been central to the methodology of Russian language teaching. However, norm is not a monolithic and immovable entity: it is a historically variable, stylistically differentiated, and socially conditioned system of preferences. What was an error yesterday becomes an acceptable variant today; what is inappropriate in business correspondence serves as a powerful stylistic device in literary texts.

Corpus linguistics offers an empirical approach to norm: instead of postulating how things "should be," it describes how things "are" — in which contexts, with what frequency, and in which genres and styles a given syntactic phenomenon occurs. This approach does not abolish the normative dimension of language, but significantly complicates and enriches it: norm appears not as a rigid prohibition but as a complex map of probabilities and contextual preferences.

This is of fundamental importance for linguodidactics. A student who is told "parcellation is a syntactic error, it is forbidden" becomes confused when reading Viktor Pelevin or Lyudmila Ulitskaya. A student who is shown corpus data — in which specific texts, for what purpose, and with what frequency parcellation is used — receives a tool for conscious stylistic choice. These represent fundamentally different levels of language proficiency.

2. Syntactic Phenomena in the Corpus Mirror

Let us turn to specific syntactic phenomena that most vividly demonstrate the gap between the textbook and the corpus.

Parcellation — the deliberate fragmentation of a syntactically unified utterance into several intonationally independent segments — is classified by most school textbooks as a stylistically marked device permissible only in literary speech. RNC data, however, show that parcellation occurs across all functional styles, including journalistic and even popular-science texts, with increasing frequency since the 1990s. This is not a degradation of written norms, but a symptom of the convergence of bookish and colloquial syntactic systems in the post-Soviet period.

Ellipsis — the omission of structurally necessary but semantically recoverable sentence elements — is considered in textbooks primarily as a feature of colloquial and dialogic speech. Corpus data refute this rigid stylistic attribution: elliptical constructions are actively used in newspaper headlines, advertising texts, and social media captions. Analysis of these contexts allows students to see how syntactic incompleteness becomes a tool for creating dynamism and intrigue.

Word order is an area in which the discrepancy between normative description and corpus data is perhaps most striking. Textbooks characterize word order in Russian sentences as "free," pointing only to basic theme-rheme patterns. Corpus analysis reveals a far more subtle system:



the position of each sentence element is statistically associated with its semantic role, text style, genre, and even the era of composition. Working with this data transforms the study of word order from a dull topic into a genuine linguistic investigation.

3. The Corpus Approach in the Classroom: From Search to Research

The practical integration of corpus data into syntax teaching follows the principle of increasing student autonomy: from observation of ready-made data through guided search to independent mini-research.

At the first level, the instructor independently conducts a corpus search and presents students with a selection of examples for analysis. The task is to identify a pattern in the use of a given syntactic phenomenon — in which styles it occurs more frequently, what the typical context is, whether there is historical dynamics. This level does not require students to have technical skills in corpus work, but cultivates the habit of reasoning about language on the basis of data rather than speculative judgments.

At the second level, students independently formulate a search query in the RNC and analyze the results. The corpus interface is simple enough to master in a single session; at the same time, search capabilities are extensive: one can specify grammatical parameters and restrict the sample by style, genre, or chronological period. Work in small groups followed by discussion creates a genuinely research-oriented atmosphere.

At the third level — in the format of a course paper or project — students conduct a full mini-research: formulate a hypothesis about a syntactic phenomenon, collect corpus data, analyze them, and draw conclusions about norm and usage. At Karshi State University, this format was introduced into practice in the course "Contemporary Russian Language" in 2023. Student works on topics such as "Functions of Parcellation in Contemporary Journalism," "Ellipsis in the Language of Advertising," and "Word Order in Twenty-First-Century Poetry" demonstrated a high level of linguistic reflection and genuine interest in language as an object of scholarly study.

4. Norm and Living Language: Toward a New Didactic Philosophy

The corpus approach to teaching syntax inevitably raises a broader question for the educator: what is the purpose of studying grammar in school and university? If the goal is to develop immunity against errors, then the prescriptive model of "rule — exercise — assessment" is quite functional. But if the goal is to cultivate a linguistic personality capable of conscious, stylistically precise, and communicatively effective language use — then a fundamentally different didactic model is required.

The corpus-descriptive approach does not contradict the normative one: it does not advocate "writing however one pleases" and does not abolish the distinction between literate and illiterate speech. It proposes a more complex and more honest relationship with norm: norm exists and matters, but it is multilayered, historically variable, and stylistically differentiated. A student who understands this is no less literate than one who has simply memorized the rules; they are literate in a different way — more deeply and more flexibly.



Conclusion

Corpus linguistics has opened a new dimension for Russian language teaching: the possibility of speaking about language not on the basis of intuitions and traditions, but on the basis of real data about real usage. The Russian National Corpus is not merely a search engine; it is a mirror in which language sees itself as living, changing, and infinitely rich.

Introducing the corpus approach into syntax teaching requires the educator to be prepared for unexpected data and to engage in an honest conversation with students about the nature of language norm. In return, it offers something invaluable: the transformation of a grammar lesson into a genuine linguistic investigation, in which every student can become a discoverer — of a small but authentic linguistic fact.

Further research prospects are associated with the creation of a methodological manual on corpus linguistics for Russian language teachers, the development of a system of corpus tasks for school and university syntax courses, and the study of the possibilities of AI tools for automating corpus analysis for educational purposes — a task that combines two key challenges of contemporary linguistics.

References

1. Apresyan Yu. D. *Studies in Semantics and Lexicography*. Vol. 1. — Moscow: Languages of Slavic Cultures, 2009. — 568 p.
2. Gasparov B. M. *Language, Memory, Image. Linguistics of Linguistic Existence*. — Moscow: New Literary Review, 1996. — 352 p.
3. Zakharov V. P., Bogdanova S. Yu. *Corpus Linguistics*. — Irkutsk: IGLU, 2011. — 161 p.
4. Lyashevskaya O. N., Sharov S. A. *Frequency Dictionary of Contemporary Russian (based on materials from the Russian National Corpus)*. — Moscow: Azbukovnik, 2009. — 1100 p.
5. Plungyan V. A. *Why Do We Need the Russian National Corpus? // Russian National Corpus: 2003–2005*. — Moscow: Indrik, 2005. — Pp. 6–20.
6. Rozental D. E. *Reference Guide to the Russian Language. Punctuation*. — Moscow: Oniks, 2010. — 256 p.
7. Sirotinina O. B. *Contemporary Colloquial Speech and Its Features*. — Moscow: Prosvetshchenie, 1974. — 144 p.
8. Shvedova N. Yu. *Russian Grammar*. Vol. 2. *Syntax*. — Moscow: Nauka, 1980. — 710 p.
9. Biber D. *Corpus Linguistics: Investigating Language Structure and Use*. — Cambridge: Cambridge University Press, 1998. — 300 p.
10. McEnery T., Wilson A. *Corpus Linguistics: An Introduction*. — Edinburgh: Edinburgh University Press, 2001. — 235 p.

