

Sepideh Haghdast Majidi (SZTE)

Discourse Markers and Non-canonical Patterns in Russian: A Corpus-based Study

Russian word order is highly flexible, exhibiting a wide range of canonical and non-canonical patterns, yet it is simultaneously constrained by information-structural principles and pragmatic cues that govern discourse organization and guide the listener's interpretation of focus and topic. Building on the influential work of Russian scholars such as Yu.D. Apresjan (1988), Y.V. Paducheva (2016), and others, this study applies corpus-based methods to extend their insights into the relationship between discourse markers and word order in contemporary Russian. Despite extensive study, the precise ways in which these interacting factors shape word order remain only partially understood and often debated in the literature. The present study investigates the interaction between Russian discourse markers and syntactic arrangement by examining whether commonly used markers, including *же* ("zhe" *indeed*), *ведь* ("ved'" *after all*), *ну* ("nu" *well*), and *кстати* ("kстати" *by the way*), display systematic preferences for particular clause positions and whether their presence is associated with subtle but measurable variations in word-order patterns, such as SVO, OVS, OSV, SOV, VSO, and VOS. To achieve this, we employed the Russian Universal Dependencies (UD) treebanks, which provide richly annotated syntactic structures for contemporary Russian texts across multiple genres and registers. Sentences containing at least one of the target discourse markers were automatically extracted from the corpus, and the linear positions of subjects, finite verbs, and objects were determined using dependency annotations within RStudio and verified through additional automated checks. These marker-bearing sentences were then compared with a baseline of sentences lacking discourse markers in order to quantify the distribution of canonical and marked word-order patterns and to identify potential correlations with marker placement and genre-specific tendencies. Over 500 sentences were analyzed in this preliminary investigation, providing initial descriptive statistics regarding marker frequency, their positions relative to core syntactic elements, and the associated patterns of word order. The results are presented in terms of both absolute counts and relative distributions, offering an empirical basis for understanding how discourse markers contribute to the organization of information in Russian sentences. These findings indicate that discourse markers exert a subtle but systematic influence on Russian word order, supporting the view that information structure and syntax are tightly intertwined and pointing to fruitful directions for future corpus-driven research.

References:

1. Апресян Д. Типы коммуникативной информации для толкового словаря. *Язык: система и функционирование*. Москва, 1988. С. 10–22. [Apresjan, D. *Types of*

Communicative Information for an Explanatory Dictionary. Language: System and Functioning. Moscow, 1988, pp. 10–22.]

2. **Падучева Е. В.** Коммуникативная структура предложения. *Материалы для проекта корпусного описания русской грамматики (rusgram.ru)*. На правах рукописи. Москва, 2015. Дата последнего изменения: 2024-03-31 17:43:02 MSK. [Paducheva, E. V. *Communicative Structure of the Sentence. Materials for the Project of a Corpus-Based Description of Russian Grammar (rusgram.ru)*. Manuscript. Moscow, 2015. Last modified: 2024-03-31 17:43:02 MSK.]