

Building a Constructicon for Russian: How to identify families of constructions?

Anna Endresen¹, Anna Klezovich², Olga Lyashevskaya^{2,5}, Daria Mordashova^{3,6},
Maria Nordrum¹, Ekaterina Rakhilina^{2,5}, Francis Tyers⁴ and Valentina Zhukova²

¹UiT The Arctic University of Norway, ²National Research University Higher School of Economics,

³Lomonosov Moscow State University, ⁴Indiana University Bloomington,

⁵Vinogradov Institute of Russian Language (Russian Academy of Sciences),

⁶The Institute of Linguistics (Russian Academy of Sciences)

anna.endresen@uit.no; belkannkl@gmail.com; olesar@yandex.ru; mordashova.d@yandex.ru;
maria.nordrum@uit.no; rakhilina@gmail.com; ftyers@prompsit.com; valentina.zh96@gmail.com

Keywords: Construction Grammar, Russian, families of constructions, linguistic pedagogy, typology

Construction Grammar (Croft 2001; Goldberg 2006) has recently developed a new sub-discipline termed constructicography (Lyngfelt et al. 2018), which unites linguists who are building electronic searchable databases of constructions called constructicons. The Swedish Constructicon (<https://spraakbanken.gu.se/eng/sweccn>) launched this movement, and today comparable projects are underway for other languages including German, Spanish, Brazilian Portuguese, Korean, Japanese, and Russian.

In this talk, we report on a large-scale group project of building a Constructicon for Russian (<https://spraakbanken.gu.se/karp/#?mode=konstruktikon-rus>), focusing on methodological challenges and our strategies and solutions. More specifically, we will focus on how to identify families of constructions that are related in terms of semantics, morphology, and syntax. We will further show how such families of constructions can be represented in the Constructicon. The Russian Constructicon is a free open-access electronic database of Russian constructions accompanied with thorough descriptions of their properties and illustrative examples from the Russian National Corpus (www.ruscorpora.ru). Descriptions are provided by linguists and translated into English and Norwegian. The architecture of the Russian Constructicon is built in parallel with the Swedish Constructicon. However, we are working on additional properties of the interface that will make the Constructicon more user-friendly.

The Russian Constructicon is designed to be a useful research-based linguistic and pedagogical tool for both linguists and L2 learners of Russian. Its current version contains 670 constructions, and more items are constantly being added. One of the challenges that we face is the choice of linguistic material. We collect multi-word constructions of a particular type, namely those constructions that have been termed “partially schematic” (Ehrlemark et al. 2018). Such constructions lie between the two extremes of fully idiomatic expressions on the one hand and fully compositional syntactic sequences on the other hand (Janda et al. 2018). These constructions are difficult to account for in terms of either lexicon or grammar alone. They have both open slots and constraints on the fillers. For example, the *NP čto nado* (lit. ‘NP what needs’, as in *prazdnik čto nado* ‘the holiday is super-duper’) is a construction that has an open slot for a noun phrase (*NP*) restricted to nouns signifying an event, an object, or a person that can be positively evaluated, and a fixed part *čto nado*. Another example is the construction *net-net da i VP-Perf* (lit. ‘no-no and VP-Perf’) used for referring to irregular events, as in *eto slovo net-net da i proskol’znet v razgovore* ‘this word can sometimes appear in a conversation’. Such constructions are more frequent than idioms and present L2 learners with a greater challenge: they are ubiquitous in spoken and written language, but underrepresented in descriptive grammars of Russian.

Having chosen “partially schematic” constructions, we still have to face their diversity and multidimensionality. They are different in terms of their internal structure, degree of idiomaticity, syntactic roles and semantics. Moreover, the initial format of the database represented the collected constructions as a list of unrelated units. In this talk, we will discuss our strategies for turning this list into a structured network of constructions, following the example of Goldberg (2006) and her analysis of the English Subject Auxiliary Inversion family of constructions. In particular, we will present the methodology of how we elaborate a multi-level syntactic and semantic classification of constructions that subdivides our items into meaningful classes and smaller groups and eventually facilitates the identification of constructional families.