

What happens if you try to build a constructicon for a whole language?

Valentina Zhukova¹, Anna Endresen¹, Laura A. Janda¹, Daria Mordashova^{2,3}, Ekaterina Rakhilina^{4,5} & Olga Lyashevskaya^{4,5}

¹UiT The Arctic University of Norway, ²Institute of Linguistics of the Russian Academy of Sciences,

³Lomonosov Moscow State University, ⁴National Research University Higher School of Economics,

⁵Vinogradov Russian Language Institute RAS

valentina.a.zhukova@uit.no, anna.endresen@uit.no, laura.janda@uit.no, mordashova.d@yandex.ru, rakhilina@gmail.com, olesar@yandex.ru

Keywords: Construction Grammar, Constructicography, Constructicon

Our points of departure are Fillmore's (1988: 37) claim that "[t]he grammar of a language can be seen as a repertory of constructions" and Goldberg's (2005: Chapter 8) observation that constructions are related to each other in radial category networks. To answer Fillmore's challenge to describe an entire language in terms of constructions, we have built a large-scale constructicon resource that represents a major portion of the grammar of Russian. Throughout this process we have endeavored to represent the semantic and formal relationships among constructions, following Goldberg's model. The result is the Russian Constructicon (<https://constructicon.github.io/russian/>), a structured inventory of over 2200 constructions and their organization into families, clusters, and networks. Constructions in this resource are supplied with extensive linguistic annotation, definitions, and corpus-based examples of use.

We present the challenges faced and decisions made in the concept and design of a (relatively) comprehensive constructicon resource. We focus on three issues: 1) choice of constructions, 2) target users, and 3) extensions to other languages and language pedagogy.

Choice of constructions

If a constructicon is to model the entirety of the grammar of a language, then it must represent constructions at varying levels, from schematic "macro-constructions", through mid-level "meso-constructions", to individual "micro-constructions" (terms introduced by Traugott 2008). Many of the most schematic constructions, such as the transitive verb construction, are already represented in reference grammars, and many micro-constructions, such as those found in lexemes containing derivational morphology, are already represented in dictionaries. While the Russian Constructicon focuses primarily on otherwise underrepresented constructions at the meso-level, it contains constructions at varying levels of schematicity. We will discuss in detail some of the difficult decisions concerning the granularity of representation and whether to join constructions in a single higher-level entry, or split them across multiple entries, or combine both strategies.

Target users

A resource that represents an entire language has many potential users: linguists, NLP professionals, language teachers, and language learners. These users have different and sometimes conflicting needs and expectations. This is particularly relevant for conventions and abbreviations. Linguists are most likely to be familiar with Leipzig Glossing Rules. NLP specialists need annotation of constructions in terms of Universal Dependencies. Language teachers and learners require more user-friendly accommodations.

Extensions to other languages and language pedagogy

The programming design is open-source and deliberately lean and simplistic to facilitate portability to other languages with minimal programming support. For language learning, we have created add-on resources, among them Construxercise! (<https://constructicon.github.io/construxercise-rus/>), a bank of over 150 exercises that target strategic sets of Russian constructions, and instructional videos (<https://www.youtube.com/playlist?list=PLUzLnIT3QLjKhelfFsryUT1nXvxdb9b75>).

References

Fillmore, Charles J. 1988. The Mechanisms of "Construction Grammar". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.

Traugott, Elizabeth C. 2008. The grammaticalization of NP of NP patterns. In: A. Bergs and G. Diewald (Eds.), *Constructions and Language Change*. Berlin: Mouton de Gruyter, 23-45.