

A new linguistic database of Uralic nominal structures

Barbara Egedi

HUN-REN Hungarian Research Centre for Linguistics

The aim of the talk is to present a new linguistic database, including the concept behind its creation, the layout as it appears now, and the ways how potential users can utilize it for their own purposes. The database has been developed to support specific comparative linguistic investigations in Uralic studies and beyond. The fundamental concept was to offer an annotated empirical basis for the study of nominal syntax, be it typological or theoretical in nature. Accordingly, rather than being a corpus of texts or text samples, the *Nominal Structures of Uralic Languages* present constructions in a structured and interpreted way. The database is open access and fully available in English including explanations, glosses and translations of the original data.

The research project behind the database closed in 2023, and had the same title, *Nominal Structures of Uralic Languages*. Both have been hosted by the Institute for Historical and Uralic Linguistics, HUN-REN Hungarian Research Centre for Linguistics (former Research Institute for Linguistics, Hungarian Academy of Sciences). The research group working in this project aimed to integrate data analysis, descriptive tasks and theoretical discussion regarding all constructions that have the noun as their core element in a selected group of Uralic languages. The database in its present form collects original language data from Eastern and Northern Khanty, Tundra Nenets, and Udmurt, as well as from two historical stages of Hungarian.

Data have been arranged into a previously established hierarchical structure of linguistic topics, subtopics and parameters. The term *nominal structures* is used to cover two larger modules concerning phrase-level and sentence-level phenomena, respectively. One group of the topics covers all relevant properties of noun phrases, including various adnominal modification patterns, quantification and determination, while the other group of topics is dedicated to copular clauses and nonverbal predication in general. The topics concentrate on syntactic structures, which means that lexical and morphophonological peculiarities are only considered if relevant for syntactic variation. Terminology has been chosen to conform to widely accepted, common descriptive terms and concepts. The talk will explore how the individual items in the database have been designed, and will also explain how to interpret the presentation of its data. Last but not least, the talk will also reflect on this database's relation to already existing corpora and databases with similar or overlapping functions, e.g.

Typological Database of the Ugric Languages (Havas et al. 2015), Typological Database of the Volga Area Finno-Ugric Languages (Havas et al. 2023), Uralic Areal Typology Online, Norvik et al. (2022).

In its present form the *Nominal Structures of Uralic Languages* comprises 12 groups of topics with 446 parameters applied to six different languages. Future extensions are possible either horizontally, by adding further languages, or in depth, through enriching complexity within linguistic topics or adding completely new ones.

Havas et al. 2015. Typological Database of the Ugric Languages. Budapest: ELTE Finnugor Tanszék. (utdb.elte.hu)

Havas et al. 2023. Typological Database of the Volga Area Finno-Ugric Languages (VolgaTyp). Budapest: ELTE Finnugor Tanszék. (volgatyp.elte.hu)

Norvik, Miina et al. 2022. Uralic typology in the light of new comprehensive datasets. *Journal of Uralic Linguistics*. 1:1. 4-42.