

Datasets for "The causal-noncausal alternation in the Northern Tungusic languages of Russia".

Aralova Natalia

Dynamique du Langage, UMR5596, CNRS & Université de Lyon, Lyon, France

Kazan Federal University, Kazan, Russia

n.aralova@gmail.com

Brigitte Pakendorf

Dynamique du Langage, UMR5596, CNRS & Université de Lyon, Lyon, France

brigitte.pakendorf@cnrs.fr

This deposit contains the datasets from three Northern Tungusic languages: Even, Negidal and Evenki. Our study of causality coding is based on a twenty-verb meaning list proposed by Creissels (to appear) specifically to investigate causal-noncausal alternations. The list comprises the following items:

(1) boil; break; burn; close; run out/use up; dry; fall/drop; get wet/(make) wet; go out/extinguish; increase; melt; move (here: go/bring); open; rise/raise; split; spoil; spread; stop (of humans); turn over; twist

Sources

For Even, we used published dictionaries representing the so-called standard (Cincius & Rišes (1952), verified with Robbek & Robbek (2005)) and a corpus of transcribed oral speech

numbering approximately 90,000 words. This basically represents two dialects, Lamunkhin Even (central Yakutia) and Bystraja Even (central Kamchatka), but also includes a few texts collected from speakers of the Tompo dialect (also central Yakutia). We were unfortunately unable to treat the dialects separately due to lack of data.

The Evenki data came from published dictionaries that represent largely the southern dialects at the basis of the so-called standard language. The main source was Boldyrev (1994), and the data were verified with Boldyrev (2000) and Myreeva (2004).

For Negidal, the study is based on three types of sources: 1) Elicitation of the 20 verb meanings with two speakers (one fluent, one less so). 2) The Negidal-Russian dictionary appended in Cincius (1982) - mainly to find lexemes that the speakers hadn't been able to remember. 3) A corpus of transcribed, translated, and glossed oral recordings of the Upper dialect (Pakendorf & Aralova 2017) numbering approximately 60,000 words at time of data collection.

Principles of coding the data

In the datasets, nC stands for 'noncausal', C stands for 'causal', and the mathematical operator indicates the direction of derivation: $nC > C$ 'causal is derived from noncausal' (causative coding); $nC < C$ 'noncausal is derived from causal' (anticausative coding); $nC \approx C$ 'noncausal and causal are equipollent'; $nC = C$ 'noncausal and causal are expressed by the same item' (i.e. the verb is labile).

- Since we started data collection with elicitation of the Negidal lists, the obtained meaning sometimes deviated from that targeted by Creissels (to appear). Thus, as we were unable to elicit 'move (of an inanimate entity)' we elicited 'move (of an animate being)', i.e. 'go', and the

meaning ‘stop’ was also only obtained for animate beings. These were then the meanings we targeted in the Even and Evenki data collection to ensure comparability.

- In those cases where we found synonymous pairs with different coding patterns, we included them all in the dataset.

- We excluded verbs with very narrow meanings, such as Negidal *boteo(-β)*- ‘dry out’, which refers only to hides that dry out excessively during preparation and then become unworkable.

- We tried to cover the dialectal variation and were inclusive in our choice.

- We attempted to include only ‘basic’ meanings and excluded stems where the derivation seemed to provide additional semantic content.

References:

Boldyrev, Boris V. 1994. *Russko-evenkijskij slovar’* [Russian-Evenki dictionary]. Novosibirsk: Nauka.

Boldyrev, Boris V. 2000. *Evenkijsko-russkij slovar’* [Evenki-Russian dictionary]. Vol. 1&2. Novosibirsk: Izdatel’stvo SO RAN, filial “GEO.”

Cincius, Vera I. 1982. *Negidal’skij jazyk: Issledovanija i materialy*. [Negidal: Analysis and Materials]. Leningrad: Nauka.

Cincius, Vera I. & Ljubov’ D. Rišes. 1952. *Russko-evenskij slovar’* [Russian-Even dictionary]. Moscow: Gosudarstvennoe izdatel’stvo inostrannyx i nacional’nyx slovar’ej.

Creissels, Denis. to appear. The noncausal-causal alternation and the limits of ambitransitivity in the languages of Sub-Saharan Africa. Special issue of *Language Dynamics and Change* on "Lexical valence orientation in cross-linguistic perspective", edited by Grossman, Eitan, Gruenthal, Riho, and Nichols, Johanna.

Myreeva, Anna N. 2004. *Evenkijsko-russkij slovar’* [Evenki-Russian dictionary]. Novosibirsk: Nauka.

Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct Northern Tungusic language of the Lower Amur*. London: SOAS, Endangered Languages Archive. <https://elar.soas.ac.uk/Collection/MPI1041287>. Accessed on 30 Jun. 20.

Robbek, Vasilij A. & Marija E. Robbek. 2005. *Evensko-russkij slovar' [Even-Russian dictionary]*. Novosibirsk: Nauka.