

Neuer samojedischer Wortschatz – Compiling an updated etymological dictionary of the Samoyed languages

Kaisla Kaheinen (presenter)

University of Helsinki

Tapani Salminen (co-presenter)

University of Helsinki

Published in 1977, Juha Janhunen's *Samojedischer Wortschatz* (SW = Janhunen 1977) is one of the cornerstones of diachronic research into the Samoyed languages, and into Uralic etymology in general. SW has retained its status as a general reference work for scholars looking for information on the lexicon and reconstruction of Proto-Samoyed, as well as the spread of individual cognates of Proto-Samoyed lexemes in the modern Samoyed languages. The publication of SW can be characterized as a revolutionary event for the reconstruction of Proto-Samoyed, since, for the first time in history, a complete account of the PS sound system, based on regular sound correspondences, became widely available to researchers. At the same time, Eugene Helimski was independently working on his own reconstruction (see Хелимский 1979), the striking resemblance of which to the one used by Janhunen in SW is certainly an impressive showcase of the effectiveness of the comparative method.

Nearly half a century after the publication of SW, several changes and additions have been made to the reconstruction of Proto-Samoyed, especially concerning the vowel system (Helimski 2005; Salminen 2024). New lexical materials, such as dictionaries, corpora, and specialized studies have appeared, (cf., for instance, Helimski 1997 for Mator, Alatalo 2004 for Selkup, and Kaheinen 2023 for Nganasan), containing previously unknown cognates of Proto-Samoyed lexemes, and new etymological suggestions have been made, most of which have not been thoroughly evaluated yet. These have not been incorporated into a single publication but remain scattered across publications, some of which are difficult to access, or are passed on in the form of manuscripts or even as word-of-mouth between specialists. This makes it very difficult for more generalist researchers to familiarize themselves with diachronic Samoyed studies, and even from a specialist perspective, the disorderliness of data is very impractical.

In this talk, we present our project of compiling an updated and enhanced etymological dictionary of the Samoyed languages. The dictionary, which is to be published both as an online database as well as in print, will contain all lexical stems that can be reconstructed for Proto-

Samoyed, their cognates from the most up-to-date sources, and further etymological references. The reconstructions will be updated to match the recent advances in the field. The main goal of our work is to create a widely available, comprehensive corpus of Proto-Samoyed lexemes that can be used as a reference work by specialists and non-specialists alike and contribute to the diachronic study of the Samoyed languages by highlighting previously neglected findings and by making new ones.

Our presentation addresses several acute questions that have arisen over the course of our work so far. What needs to be determined before the work can proceed further, is, for example, to what extent should etymological discussion be included in the work in the case where the authors disagree with an existing etymological proposal for a given lexeme, how much explicit argumentation to include in defence of the new reconstructions, as well as a few specific details of the reconstructions themselves, especially concerning non-first syllable vowels.

With a project of this scale, there are also questions related to matters of funding, publication, and workflow management. Gathering an etymological dictionary is a long-term project, the likes of which are difficult to complete in the current era of project-based short-term employment. The pressure to gain 'merit' as quick as possible may favour writing shorter papers over long-term work, or the premature publication of unpolished results. Digital databases make data more widely accessible in an updatable format, but on the other hand, they require maintenance and are vulnerable to digital obsolescence. For the work to reach completion, compromises need to be made regarding the amount of detail and number of sources vs. simplicity and time constraints. With this presentation, we wish to open the discussion concerning these matters in Uralic etymological research and dictionary-compilation.

Alatalo, Jarmo. 2004. *Sölkupisches Wörterbuch*. Helsinki: Société Finno-Ougrienne.

Helinski, Eugen. 1997. *Die matorische Sprache: Wörterverzeichnis – Grundzüge der Grammatik – Sprachgeschichte*. Unter Mitarbeit von Beáta Nagy. Szeged.

2005. The 13th Proto-Samoyedic vowel. In Beáta Wagner-Nagy (ed.), *Mikola-konferencia 2004*, 27–39. Szeged: SzTE Department of Finnougristics.

Janhunen, Juha. 1977. *Samojedischer Wortschatz. Gemeinsamojedische Etymologien*. Helsinki: Helsingin Yliopisto.

Kaheinen, Kaisla. 2023. *Etymologia ex silentio: Nganasanin äännehistoria ja kielikontaktit*. Helsinki: Helsingin yliopisto.

Salminen, Tapani. 2024. The Samoyed languages. In Edward Vajda (ed.), *The languages and linguistics of Northern Asia*. Vol. 1. 167–251. Berlin: De Gruyter Mouton.

Хелимский, Е. А. 1979. *Древнейшие угорско-самодийские языковые связи. Анализ некоторых аспектов генетических и ареальных взаимоотношений между уральскими языками*. АКД. Тарту.