EasyChair Preprint

№ 9308

Semantic Shifts in the Samoyedic Basic Vocabulary and Their Parallels. 1. Sun, Day

Idaliya Fedotova

November 11, 2022

Semantic shifts in the Samoyedic basic vocabulary and their parallels. 1. Sun, day

Idalia Fedotova[1,2]

[1]HSE University

[2]Ivannikov Institute of System Programming of the RAS

This paper is a part of the series which investigate cases of semantic shifts and proto-language polysemy in the basic vocabulary of the Samoyedic languages. This research focuses on the shifts which have analogies in Turkic and Tungusic languages, identified with the help of semantic reconstruction. Special maps were created on LingvoDoc linguistic platform in order to demonstrate areas of similar polysemy and semantic shifts, possibly contact-induced. Using materials from archival and expeditionary dictionaries, the paper proposes a detailed account of the available lexicon of the Samoyed languages within the scope of core lexicon. Our results show 15 semantic shifts in the form of polysemy, semantic evolution and cognates. The present paper investigates polysemy ‘sun, day’ in Samoyed, Turkic and Tungusic languages.

1 Introduction

Semantics have been in the focus of linguistic typology and computational linguistics for the past twenty years. Typology of semantic shifts is one of the branches of linguistic typology of semantics alongside semantic typology and lexical typology (Vanhove 2008). The concept of semantic shift is defined by Zaliznyak (2013: 33) as “a certain conceptual contiguity between two linguistic meanings A and B, manifested in the fact that these two meanings are combined within one word in a broad sense”.

There are five types of realizations of semantic shifts: 1) synchronous polysemy; 2) diachronic semantic evolution of a word from an ancestor language to a descendant language or within the same language; 3) morphological derivation: the meaning of B is represented by a morphological derivative of a word with the meaning of A, or vice versa; 4) cognates: the meanings of A and B belong to two (closely) related languages and they go back to one word of their common language-the ancestor; 5) borrowing: the meaning of B belongs to the borrowed word, and is the meaning of the same word in the source language (Zaliznyak 2013: 33-35). The fifth type is not considered in the preset article, and instead of the term "cognates" as a type of realisation, we use "reflex polysemy".

Studies of linguistic areas in synchrony and diachrony, language contacts and extralinguistic factors which are the probable causes of certain common features of all languages included in the area have a long history and are mainly devoted to lexical borrowings and changes in grammar

(since Weinreich (1953])). Here we are going to mention only some papers which investigate the contact nature of semantic changes or common features of the semantics of words from genetically different languages in the same area. Witkowski and Brown (1985) show the relationship between the absence of polysemy in the words for the upper limbs in languages existing above a certain geographical latitude and wearing long-sleeved clothing (the nomination of parts of the arm with different words is more likely). Dybo (1995) shows how motivation models are distributed over areas of index finger names in Eurasia. According to Dubrovskaya (2014), in the Selkup language, the semantic structure of nouns, the presence or absence of polysemy may be associated with the influence of language contacts. Myznikov (2004) in his monograph points to "interaction and mutual influence in the field of semantics" between the Baltic-Finnish languages and Russian dialects of adjacent areas. However, Myznikov argues that "the development of models of representation of semantically similar units, belonging to different linguistic environments, can be justified only in relation to a carefully studied area" (Myznikov 2004: 36). The typology of the areas based on the maps of the Common Slavic Linguistic Atlas was developed by Vendina (2014). This book shows the evolution of the areas from the proto-linguistic state to the modern time, draws important conclusions about the differentiation of Proto-Slavic dialects and the predominance of convergent processes in the vocabulary.

At the level of synchrony, Gast and Koptjevtskaja-Tamm (2018) conducted a study of colexification on extensive language material using two databases. They identified areas where cases of colexification are localized (23 in total, the areas of combining the values of ‘feather’ / ‘hair’, ‘fire’ / ‘tree’, ‘fire’ / ‘firewood’, ‘mountain’ / ‘stone’, ‘ear’ / ‘leaf’, ‘bark’ / ‘skin’ are analyzed in detail and shown on the maps). Genetic homogeneity of languages in these areas was also checked. It turns out that contact languages that are not related to each other can form areas of distribution of certain meanings. This paper demonstrates the possibility of studying areal semantics and contact polysemy using databases. Similarly, Zhivlov (2019) observes the areal polysemy ‘earth’ / ‘year’ in North American languages, where the contact origin of this polysemy is established for certain languages.

Contacts of Samoyed, Turkic and Tungusic-Manchu languages are studied by Dybo (2007), Anikin and Helimski (2007) and Terentyev (1999).

2 Materials and Methods

The study aims to identify the cases of polysemy at the level of the Proto-Samoyed language that have parallels in the contact Turkic and Tungusic languages, using the material of the lists of basic vocabulary in the Samoyed languages. The research also shows the cases of parallel semantic developments from the proto-language to the descendant languages.

The search for polysemy and semantic shifts in the Samoyed Swadesh lists was primarily carried out using the Samoyedic etymological dictionary by Janhunen (1977). Upon that the found lexemes were looked up in the dictionaries of separate Samoyed languages: Nganasan (Kosterkina et al. 2001, Helimski Nganasan Dictionary), Tundra Nenets (Tereshchenko 1965), Forest Nenets (Barmich&Vello 2002), Enets (Tundra dialect) (Helimski Enets Dictionary), Enets (Forest dialect) (Sorokina& Bolina 2009), Selkup (Bykonya et al. 2005, Helimski 2007), Kamas (Donner&Joki 1944), Mator (Helimski 1997). The LingvoDoc platform also provided a vast amount of material from the expedition and archival. The reconstruction of the proto-language meaning by Janhunen (1977] was accepted only if the meaning was represented in two distant Samoyed groups: Northern Samoyed (Nenets, Enets and Nganasan) and one of the following

languages: Selkup, Kamas and Mator-Taigi-Karagas (classification according to Helimski (1982, 39)).

The Turkic etymologies that are accepted in the paper follow the "Etymological Dictionary of the basic vocabulary of the Turkic languages" by Dybo (2013). This dictionary represents rigorous study of the material of literary Turkic languages and dialect dictionaries published by that time, so it was chosen for the first stage of our comparative work. In recent years, across the Turkic languages fieldwork has been actively carried out to clarify the compilation procedures of the Swadesh list and the semantics of the words included. Extensive materials and dictionaries are still awaiting publication, therefore, the search for polysemy and semantic shifts in the lists of basic vocabulary of all Turkic dialects is a future task. The outline of the maps with areas of contact changes in meaning are likely to be different. However, at this stage, for preliminary comments, we relied on the above-mentioned dictionary, and in cases where it was necessary to clarify the presence of polysemy, the material was supplemented with data from other sources.

In several cases, in order to confirm the existence of polysemy, we undertook interviewing native speakers of the following languages: Nenets and Selkup (Samoyed languages), Altai, Chelkan, Shor, Tuvan, Bashkir, Tatar, Kazakh, Turkish and Yakut (Turkic languages). We were checking the words with the meanings ‘earth’ (‘soil’, ‘place’, ‘territory’), ‘ashes’ (if it can be labelled with the same word as ‘earth’), ‘clay’ (as a material and as a type of soil), ‘fur’ (as bodily hair and a skin with fur), ‘bark’; additionally, we checked the presence of lexemes with polysemy ‘sand’ and ‘pebble’, ‘feather’ and ‘wing’, ‘hair’ and ‘wool’, ‘skin’ and ‘hide’ (if there is a separate word for ‘tanned hide'), ‘good’ and ‘beautiful', ‘man’ and ‘person, ‘meat’ and ‘body', ‘neck’ and ‘throat', ‘warm’ and ‘soft', ‘sun’ and ‘day'.

The survey was conducted as follows. Lists of diagnostic contexts were compiled in Russian for all of these meanings (lists from the article [Kassian et al. 2010] were used; if the meanings we were interested in were not mentioned in this article, we compiled our own lists for them), and the informants were offered to translate the words or sentences as a whole. In most cases, the data we obtained coincided with the data from the dictionaries.

The primary source for the Tungus-Manchu languages was the "Comparative Dictionary of the Tungus-Manchu Languages" (TMS). The lexical meaning was reconstructed at the proto-linguistic level, if it was preserved in at least two of the three branches: North Tungusic, South Tungusic or Manchu. The protoforms for Tungusic are given according to (EDAL).

Maps showing the distribution of areas of polysemy and semantic shifts were built on the LingvoDoc platform using the functions "select languages" and " areas mode". It should be noted that the function of building areas is based on a mathematical model and does not always perfectly correspond with the tasks of linguistic geography. The squares on the maps correspond to the established places where a particular language is recorded: these are either specific geographical coordinates to which expedition or archival dictionaries are linked on the LingvoDoc platform, or a certain geographic spot (state capitals, regional centers or specific villages and towns) indicated in the published dictionaries for the languages in question. Thus, the area of polysemy or semantic shifts in the strict sense should be understood as the space between the squares representing the settlements. At a certain scale, the area around the squares is drawn too large by the automatic program.

3 Overall Results

The table below shows the results of our study identifying cases of polysemy and semantic shifts are grouped around certain sematic fields:

Celestial bodies	sun, day
Landscape	earth, soil > clay earth, soil > sand sand > pebble earth, soil > ashes
Nature	skin (of an animal), skin (of a human) skin > bark hair, feather feather, wing
Human	man> person
Body parts	flesh, body neck, throat
Qualities	good, beautiful soft > warm

Figure 1. Semantic shifts classified by topics

Further, each semantic shift can be categorized as polysemy or semantic evolution in different language families, as can be seen in the following list:

Proto-Samoyed and Proto-Turkic polysemy: sun, day; earth, soil;

Proto-Turkic polysemy — semantic evolution in Samoyed: sun — day; earth — clay;

Semantic evolution in Turkic — Proto-Samoyed polysemy: soft — warm, earth — soil;

Semantic evolution in the Turkic and Samoyed languages: earth — clay, earth — sand; litter, mud — ashes;

Proto-Tungus-Manchu and Proto-Samoyed polysemy: feather, wing; neck, throat; soft, warm; skin (of an animal), skin (of a human);

Proto-Tungus-Manchu polysemy — semantic evolution in Samoyed: man — person;

Semantic evolution in Tungus — Proto-Samoyed polysemy: skin — bark, hair — feather, good — beautiful;

Semantic evolution in the Tungusic and Samoyed languages: sand — pebbles.

On all maps, the Samoyed area is marked in red, the Turkic area is blue, and the Tungus area is yellow.

3 Polysemy ‘sun, day’

In the Samoyed languages, two roots with synchronous polysemy ‘sun, day' can be identified. Historically, one root — * *jälä* — represents a reflex of proto-language polysemy or syncretism ‘sun, light, day’. The second root — *kåjå, — represents semantic evolution of the meaning of the proto-meaning *‘sun’ to ‘day’.

1) Proto-Samoyed **jälä* ‘light, day, sun' Nenets and Selkup (Janhunen 1977: 40).

In the dictionaries of the Nenets and Selkup languages, reflexes of this proto-form with polysemy are recorded at the synchronous level: Tundra Nenets *яля* ‘day, sun, light’ (Tereshchenko 1965: 838), Forest Nenets *дяӆя* ‘day; light, illumination; sun’ (Barmich&Vello 2002: 35), Taz Selkup. *cēli* ‘day; light; sun’ (Helimski 2007: 24), Selkup. *мел, мēл, мелат, мēлат, мелд, мелде, мэлде, мелыт* ‘day; sun’, ob. Sh, ket. telt, tlt ‘day, sun, weather’ (Bykonya et al. 2005: 235), *чел, чели, чēлы* ‘day; sun’, *челд, чēлдэ̄, челт, челыды, челыт* ‘day; light; sun’ [Bykonya et al. 2005: 279], *челы* ‘day; sun; sky’ [Bykonya et al. 2005: 280], *ʧʲɛ'ɛld* ‘day, sun, light', ʧʲelʲi 'sun; day’ (LingvoDoc).

Didenko and Dubrovskaya (2012) show that, judging by dictionaries, archival and field data, the semantic structure of the mentioned Selkup word is an instance of evolution "from the broadest and most general idea — "when the sun is in the sky = light = day" — to disjoining and opposing the selected specific meanings" (Didenko&Dubrovskaya 2012: 90). In other words, one can see syncretism of meanings or a broad meaning, reflected in dictionaries as polysemy. Based on Tundra Nenets *яля*, Forest Nenets *дяӆя* ‘day, sun, light’, it can be argued that syncretism is also present in Nenets. In the Nganasan language, the reflex of **jälä* is characteristic of another version of polysemy, where the meaning ‘sun’ is lost (a relic of this meaning may be a ‘round pendant’), but the meaning ‘weather’ is present: Nganasan *дялы* ‘day; weather; round pendant, decoration’ (Kosterkina et al. 2001: 54). In the Enets language, the reflexes of this proto-form have the meanings ‘day’, ‘light’, ‘dawn’, but not ‘sun’: Tundra Enets *dɛrɛ* ‘day’ (Helimski Enets Dictionary), Forest Enets *дɛри* ‘day; light, lighting; dawn’ (Sorokina&Bolina 2009: 97). Thus, the data of modern dictionaries confirm the Proto-Samoyed reconstruction of the meaning ‘sun’, ‘day’ and 'light' proposed by Janhunen, as they all occur in both branches of the Samoyed languages, albeit in different combinations.

The Proto-Turkic polysemy is established on the basis of the synchronous polysemy ‘sun, day, sun heat’ in the following languages: Proto-Turkic **gün(-elʲ)* ‘sun, day, sun heat’ > Yakut., Dolgan *kün*, Tuvan, Tophalar *xün*, Khakas, Shor *kün*, Uygur. *kün*, Khalaj. *kin*, *kün*, Turkish., Gagauz *gün*, *güneš*, Azeri *gün*, *günäš*, Turkmen *gün*, Salar *gü:n,* Kumyk *gün*, *güneš*, karachaevo-Balkar. *kün*, Tatar. *kön*, Bashkir *kön*, *könäs*, Nogay, Kazakh, Karakalpak, Kyrgiz., Altay. *kün* (Dybo 2013: 488-489; ESTYA 1980: 100-104).

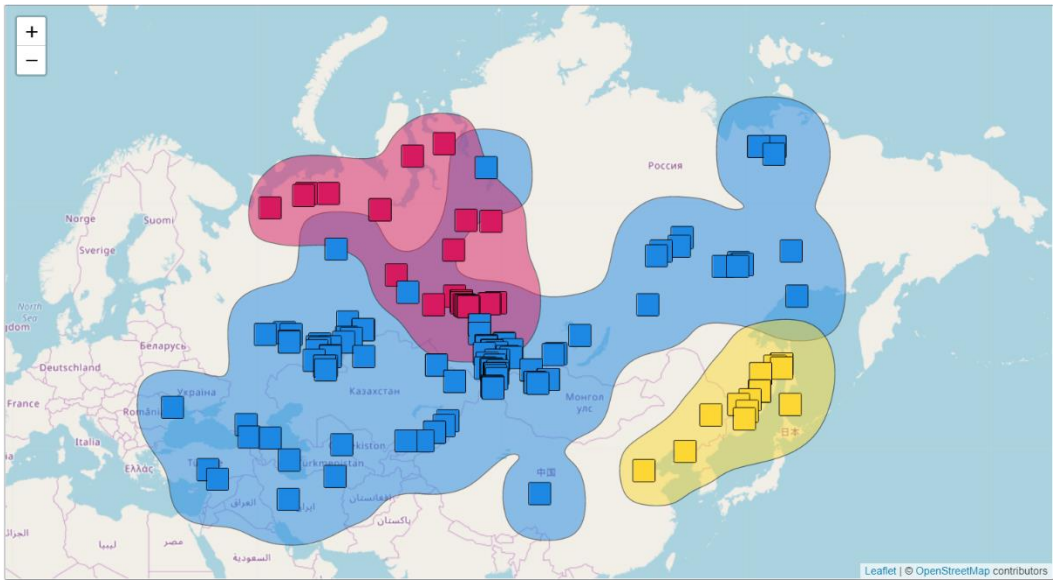

Figure 2. Polysemy ‘sun, day’ in Samoyed, Turkic and Tungusic languages: **jälä*

The yellow colour on the map indicates synchronous polysemy, which was the result of the semantic shift ‘day’ — ‘sun’ in closely related Tungus-Manchu languages: *yīh-nêng-kih* ‘sun, day‘ [TMS 1:319] <Proto-Tungus-Manchu **ine-ŋī* ‘day’ (EDAL: 586), Manchu *ицун* ‘sun; day' [TMS 2:78] < Proto-Tungus-Manchu **sigūn* ‘sun’ [EDAL: 1274]. In the Ulcha and Nanay languages, the reflexes of this proto-form collocate with the elements ‘noon’: Ulcha *сиу*(*н*-) *токоно̀*, Nanay *сиу̃токоно̀* [TMS 2:78] (*токо-но̀*- in both languages is ‘middle’ with an indicator of belonging 3rd P Sg.). These cases can also be drawn to this comparison, since the semantics of ‘noon’ must have emerged as a result of the formation of the phrase ‘day’ + ‘middle’, which indicates the existence of the meaning of ‘day’ in the reflexes **sigūn* in the Ulchi and Nanai languages, at least at the stage of the emergence of these phrases.

The Manchu polysemy ‘sun, day’ (yellow) is located Northeastern China and the Far East, not overlapping with other areas. The area of the Turkic polysemy ‘sun, day’ (blue) extends from Romania in the west to Yakutia in the Northeast and covers most of the territory of the distribution of Turkic languages. The Samoyed area (red) shows the synchronous polysemy ‘sun, day’, so the Nganasan and Enets languages are not displayed on it. This area includes Tundra and Forest Nenets and Selkup languages and covers the Nenets Autonomous District in the Arkhangelsk Region, Yamalo-Nenets Autonomous District, Tomsk Region and the west of the Krasnoyarsk Krai along Yenisei river.

The areas of Samoyed and Turkic polysemy overlap in the contact zone of speakers of the languages under consideration — in the north of Western Siberia in the Yamalo-Nenets Autonomous District (Tundra Nenets and Dolgan). The areas are adjacent, but do not in fact intersect within the Tomsk Region and the Republic of Khakassia, separated from each other by the Kemerovo region, which can be seen with an enlarged map scale.

Now let us turn to the second Samoyed root with the polysemy ‘sun, day’, which was the result of semantic evolution of the protolanguage meaning ‘sun'.

2) Proto-Samoyed **kåj'å* ‘sun’ in Enets, Nenets, Kamas, Mator and Taigi (Janhunen 1977: 58) > Taigi *хая*, mator. *chaja* ‘sun, day’ (Helimski 1997: 260-261). Mator and Taigi are the Samoyed languages of the Sayan Highlands, extinct or ousted by the Turkic languages in the 19th century and known from archival records. The marked area of their distribution is the village of Motorskoye in the Karatuzsky district of the Krasnoyarsk Krai for the Mator language [Helimski 1997: 16] and the territory in the Shushensky district of the Krasnoyarsk Krai for the Taiga (LingvoDoc).

As can be seen on the map, the compact area of the Motor and Taigi languages is absorbed by the vast Turkic area. In addition, it is adjacent to the area of the Proto-Samoyed polysemy described above, i.e. ‘sun, day’ in Selkup dialects. It is reasonable to assume semantic evolution of **kåjå* ‘sun’ under the influence of polysemy ‘sun, day’ in contact languages.

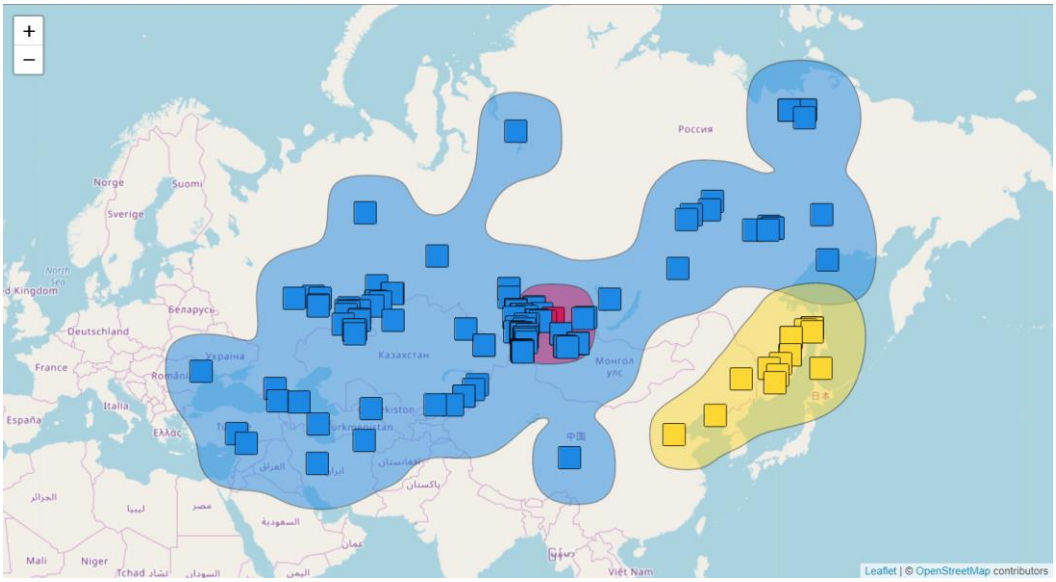

Figure 3. Polysemy ‘sun, day’ in Samoyed, Turkic and Tungusic languages: **kåjå*

4 Discussion and Conclusion

A close detailed survey of the Samoyed polysemy ‘sun, day’ in comparison with contact languages has revealed an instance of a contact-induced semantic change. One root with this polysemy (*jälä*) inherited this semantic pattern from the protolanguage stage, while the other root (*kåjå*) underwent a change, being in close contact with the languages with polysemy ‘sun, day’. This method comparison has proven to be effective and illustrative, so further research in the whole set of the semantic patterns in basic vocabulary needs to be done.

Existence of contact-induced semantic change is something that is implied in many studies in Linguistics, however, theoretical background has not met applied and evidence-based approach in its entirety. Our series of publications aims to make the first step in this direction.

6 References

Anikin, A. and Helimski, E. (2007). Samoyed- Tungus- Manchu lexical links. Moscow

Barmich, M., Vello, I. (2002) Nenets-Russian and Russian-Nenets dictionary (Forest dialect). Textbook for elementary school students. 2nd edition, extended. Sankt-Peterburg.

Bykonya, V., Kuznetsova, N., Maksimova, N. (2005) Selkupsko-russkiy slovar'. {Selkup-Russian dialectal dictionary}. Tomsk.

Didenko, A., Dubrovskaya, N. (2012) Problems and tasks of complex studying and survey of semantic derivation of nouns in Selkup dialects // Bulletin of Tomsk State Pedagogical university. 1 (116). P. 87—91.

Donner, K., Joki. (1944) A. Kai Donners Kamassisches Wörterbuch: nebst Sprachproben und Hauptzügen der Grammatik. (Lexica Societatis Fenno-Ugricae, 8.) Helsinki.

Dubrovskaya, N. (2010) Nominal semantic structure in relation with typology of derivative meanings: words with broad and general meanings in Selkup // Bulletin of Tomsk State Pedagogical university. 7 (97). P. 84—87.

Dybo, A. (1995) Die Namen des Zeigefingers in den Türk- und den altaischen Sprachen // Laut- und Wortgeschichte der Türksprachen: Beiträge des Internationalen Symposiums Berlin, 7. bis 10. Juli 1992 (Turcologica, Band 26), Harrassowitz Verlag 1995. S. 17—43.

Dybo, A. (2007) Linguistic contacts of Early Turks. Lexicon of Proto-Turkic era. Moscow

Dybo, A. (2013) Etimologicheskiy slovar'bazisnoj leksiki t'urkskikh yazykov. [Dictionary of the basic vocabulary of the Turkic languages]. Astana.

Gast, V. and Koptjevskaja-Tamm, M. (2018) The areal factor in lexical typology: some evidence from lexical databases. *Aspects of linguistic variation.* P. 43—81.

Helimski, E. (1982) The earliest Hungarian-Samoyed linguistic parallels. Linguistic and ethnogenetic interpretation. Moscow.

Helimski, E. (1997) Die Matorische Sprache. Szeged.

Helimski, E. (2007) Severnosel'kupskij slovar'. {Northern Selkup dictionary}. Hamburg.

Helimski, E. Materials for the Nganasan dictionary (unpublished).

Helimski, E. Materials for the Enets dictionary (with additions by A. Urmanchieva on the Forest dialect) // https://www.slm.unihamburg.de/ifuu/download/helimski-enzisch.zip

Janhunen, J. (1977) Samojedischer Wortschatz: Gemeinsamojedische Etymologien. Helsinki.

Kosterkina et al. 2001 — Nganasan-Russian and Russian-Nganasan dictionary / N. T. Kosterkina, A. Ch. Momde, T. Yu. Zhdanova, edited by V. Yu. Gusev. Sankt-Peterburg.

LingvoDoc – Linguistic platform LingvoDoc http://lingvodoc.ispras.ru/ (accesed 3/11/2022)

Myznikov, S. (2004) Lexis of Finni-Ugric origin in Russian dialects of the North-West. Etymological and linguageographical analysis. Sankt-Peterburg.

Sorokina, I. and Bolina, D. Eneckij slovar' s kratkim grammaticheskim ocherkom : okolo 8 000 slov. {Enets dictionary with a short grammatical survey}. Sankt-Peterburg.

Terentyev, V. (1999) The earliest Turkic-Samoyed language contacts // JSFOu Vol. 88. P. 173—200.

Tereshchenko, N. (1965) Nenetsko-russkiy slovar'. {Nenets-Russian dictionary}. Moscow.

TMS — A comparative dictionary Tungus-Manchu languages: Materials for an etymological dictionary / Ed. by V. I. Tsintsius. Vol. 1—2. L., 1975—1977.

Weinreich, U. (1953) Languages in contact: Findings and problems. The Hague.

Witkowski, S. and Brown, C. (1985) Climate, clothing, and body-part nomenclature // Ethnology. Vol. 24. P. 197—214.

Vanhove, M. (ed.) 2008. From polysemy to semantic change. Towards a typology of lexical semantic associations. Amsterdam — Philadelphia.

Zaliznyak, A. (2013) Semantic shift as an object of typology // Issues in linguistics. 2013, 2. P. 32—51.

Zhivlov, M. (2019) Areal polysemy 'earth/year' in North American languages: historical implications // Этнография / Etnografia. Vol. 3. P. 167—180.