EasyChair Preprint

№ 8989

Shared Semantic Patterns in the Basic Vocabulary of the Uralic Languages in Siberia

Idalia Fedotova

October 5, 2022

Shared semantic patterns in the basic vocabulary of the Uralic languages in Siberia*

Idalia Fedotova[1,2]

[1]HSE University

[2]Ivannikov Institute of System Programming of the RAS

Nowadays there is a trend for compiling cross-linguistic databases on semantic shifts (DatSemShift 2.0) and colexifications (CLICS), as it serves objectives of typological studies and provides vast data for comparative studies. However, there are no such cross-linguistic databases for diachronic semantic shifts that would describe shifts from a proto-language meaning to the meaning in descendant languages. Semantic reconstruction is usually done within one language family. But now with the appearance of a cross-linguistic platform LingvoDoc it is possible to simultaneously compare data from the Uralic, Turkic and Tungusic languages and build maps that show areal patterns. There has not been any catalogue that would show areal distribution of diachronic semantic shifts yet. In this study, preliminary results are shown in the form of 11 maps.

This research was first presented as a talk at the final colloquium of REMODUS Uralic Spring School, Tartu, 9 April 2022.

1 Introduction: The context of the study

With the rise of a number of large cross-linguistic databases there has been an increasing interest in typological studies of lexical semantics (Koptjevskaja-Tamm et al. 2008). The phenomenon that most researchers are focusing on is combining two meanings in one word. The most well-known label for this is polysemy (Apresjan 1995). Meanwhile the new field of lexical typology widely uses the concept of colexification (François 2008). Colexification is a wider notion, comprising both polysemy and homonymy. For typological reasons, especially if the history of a language is unknown and it is hard to distinguish between polysemy and homonymy, the concept colexification is a convenient way of approaching such difficult instances.

Several studies have been carried out on the areal colexification patterns in the world. Gast and Koptjevskaya-Tamm (2018) explored macro-areas of certain semantic patterns in the basic vocabulary of the non-related languages, comprising North and South Americas, Africa and

*The study was supported by grant of the Russian Science Foundation (project No. № 20-18-00403 “Digital description of the dialects of the Uralic languages based on the analysis of big data”)

Austronesia. Georgakopoulos et al. (2021) studied universal and macro-areal patterns in the perception verbs using the same databases (CLICS, ASJP) as the abovementioned authors, which results in virtually the same geography (Africa, Australia, Eurasia, North America, Papunesia, and South America). Within existing databases, there are languages and even parts of the globe that are not included, so there are many gaps in our knowledge.

So in these materials Eurasia is not represented fully. Siberia has been so far almost neglected, whereas this region is a home to 8 non-related language families. The sociolinguistic status of Siberia is an ample example of small-scale multilingualism, with more than 3 languages spoken in a single family even today (Amelina, Khanina 2020). The obvious problem for studying this valuable material is the lack of documented speech. Moreover, doing fieldwork is very difficult, as many settlements where native speakers live are remote and hard to access.

This study offers an overview of the semantic patterns in the basic vocabulary of the Uralic languages in Siberia, that is Ob-Ugric and Samoyed branches. In order to identify the areal patterns of polysemy within the scope of basic vocabulary, two other non-related language families of the same region are compared: Turkic and Tungusic. Speakers of the Uralic, Turkic and Tungusic language families have been living on the same territory and have close mutual contacts with each other for hundreds of years.

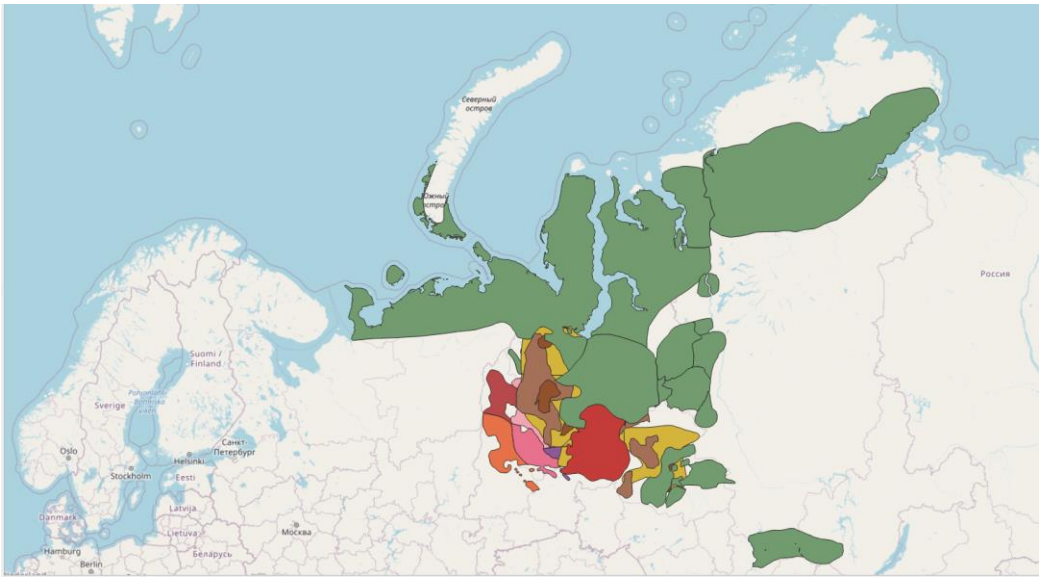

Figure 1. The spread of the Uralic languages of Siberia (map created by using URHIA)

2 Materials and Methods

Basic vocabulary is viewed as the most stable part of the lexicon in the framework of comparative and historical linguistics. In this research, the lists of basic vocabularies of Ob-Ugric and Samoyed languages were compiled using the new dictionaries on LingvoDoc platform and fullest (where possible) printed dictionaries: Khanty (DEWOS, Tereshkin 1961, Solovar 2006), Mansi (Munkácsi, Kálmán 1986), Nenets (Tereschenko 1965, Barmich,Vello 2002), Enets (Sorokina. Bolina 2009), Selkup (Bykonya et al. 2005, Helimski 2007). In addition to the published sources, a survey of native speakers was conducted in order to specify some meanings and usages.

Etymological judgements are based on the widely established etymological dictionaries for Samoyed (Janhunen 1977), Ob-Ugric (Honti 1982, DEWOS) and Uralic languages generally (UEW). Comparisons with Turkic and Tungusic languages are made using the respective etymological dictionaries (Dybo 2013, TMS).

Methodology of this study comprises several approaches, which needs clarification. It is situated at the intersection of synchrony and diachrony, exploring areal semantic patterns and semantic evolution respectively. The first attempt in this framework was made in discovering Samoyed, Turkic and Tungusic parallels (Fedotova 2020).

Turning to defining the key terms, we will explain the main approaches used in this study.

Semantic shifts (defined by Anna Zalizniak): this notion excludes homonymy and can be successfully applied only to the languages whose history is studied well. Database of Semantic Shifts in languages of the world (DatSemShift) is the project and resource which is focused on semantic shifts.

On the contrary, **colexification** (coined by A.François) does not distinguish between polysemy and homonymy and often is applied to the languages whose history is not well-known. Database of Cross-Linguistic Colexifications (CLICS) is the resource for studying colexifications.

In this study, the central term and notion is semantic shift. Even though studying colexifications is a growing trend in general linguistics, this framework is a tangential way of studying lexis on a large scale. Onn the contrary, Uralic languages have a decent body of etymological research and studying semantic shifts is preferrable, as this way we discard instances of homonymy and arrive at more precise conclusions. For a historical linguistics approach this is crutial, as evolution of languages and origins of words need a thorough examination and separation between polysemy and homonymy.

Zalizniak identifies several **types of semantic shifts** (examples are mine):

- **Polysemy:** two distinctive meaning in one word, e.g. ‘hair, fur’ widely spread in Siberia
- **Cognates** (also can be interpreted as protolanguage polysemy): e.g. ‘new’, ‘fresh’ deriving from one root in different Tungusic languages
- **Semantic evolution:** change of lexical meaning which is recorded in language documention, e.g. ‘blood vessel’ > ‘root’ in Ob-Ugric
- **Morphological derivation:** the semantic shifts occurs in wordbuilding via affixes *sēgV- ‘blood’ > *sēgǯe- ‘red’ in Tungusic

The reasons of semantic shifts, as in etymology in general, can be classified as follows:

- Genetic reasons, that is the modern languages inherited polysemy from their ancestor language
- Typological reasons: some semantic shifts are common in many languages (for instance ‘hand’ > ‘handle’)
- Reasons connected to linguistic areas: 1) semantic borrowing (calques?), 2) multilingualism, 3) areal patterns Maps can show the areas of the same semantic patterns.

The theoretical background still needs to be reinforced and more research is going to be done in this framework, this will be followed through in the subsequent articles.

Using Lingvodoc platform for creating maps. Lingvodoc is a linguistic platform which allows to carry out cross-linguistic search across its databases. So far, the array of the languages presented includes Uralic, Turkic, Tungusic and Ket. There are also etymological notes wherever etymological dictionaries are available. Using Lingvodoc platform, it is possible to create maps of various linguistic features. In this study, the results rely on this mapmaking function.

3 Results

The results of the study in the framework of semantic shifts fall into three categories:

1) polysemy in all language families; 2) semantic shifts in all language families; 3) different types of semantic shifts.

We are going to present them accordingly.

Polysemy in all language families

‘new, fresh’ in Ob-Ugric and Tungusic

Proto Ob-Ugric **jiləp* 'new, fresh‘, polysemy in Khanty and Mansi (brown)

Proto-Tungus **ime-* ‘fresh, new’: polysemy in Evenki and Nanai (yellow)

CLICS shows 13 colexifications including Mansi and Komi, DatSemShift displays 0.

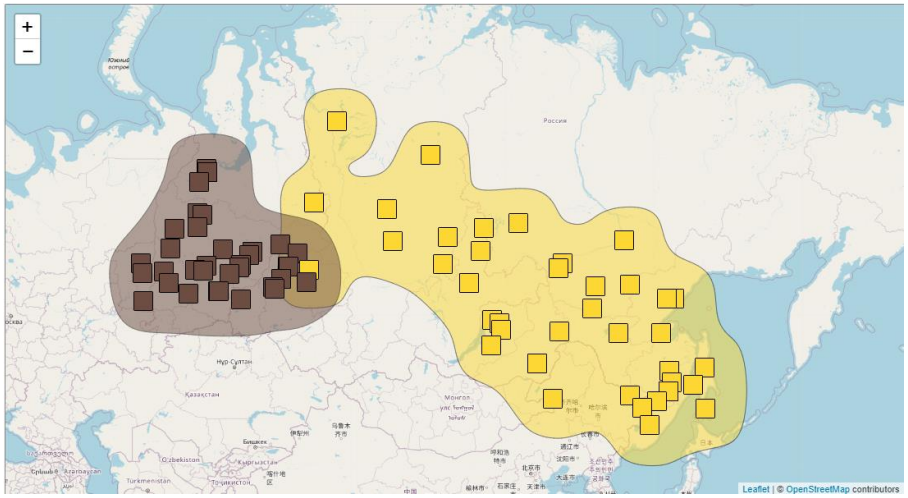

Figure 2. ‘new, fresh’ in Ob-Ugric and Tungusic

These areas are adjacent on the territory of West Siberia, where speakers of the languages in question contact. This polysemy comprises West Siberia and Far East.

‘earth, place’ in Tundra Nenets, Eastern Khanty and Turkic

Khanty (Vakh-Vasjugan) *ta´gai* ‘place, earth’ (contemporary polysemy) (brown)

Protolanguage and contemporary polysemy:

Nenets Tundra *ja, Forest d’a* : Proto-Samoyed **jåå* ‘earth, place’ (red)

Proto-Turkic **jẹr* ‘earth, place’ (blue)

CLICS: 18 including Mansi, Nenets and Selkup, DatSemShift: 0

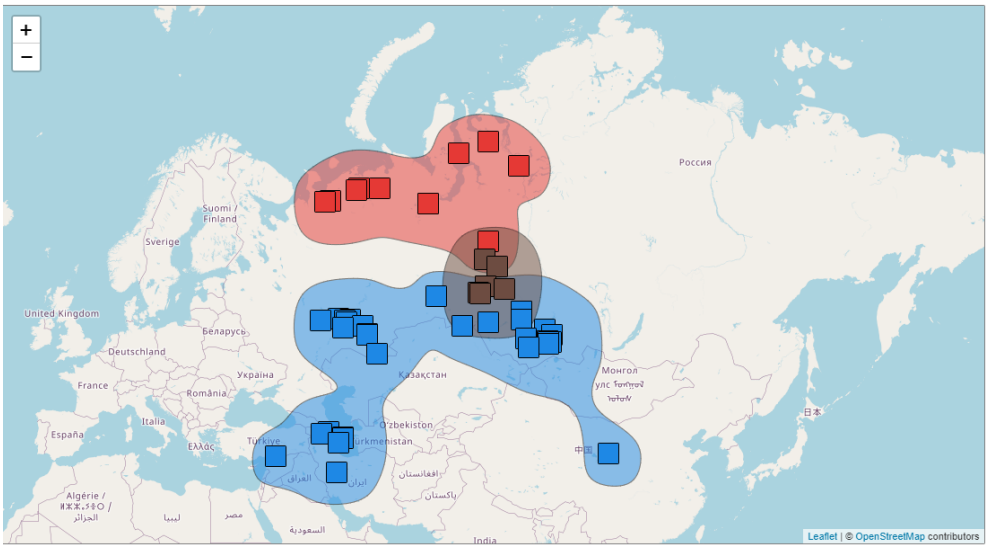

Figure 3. ‘earth, place’ in Tundra Nenets, Eastern Khanty and Turkic

Three areas are adjacent, and the heart of it is the local Khanty area. Taking the size and shape of this area into consideration, one may speculate about the contact-induced change in Khanty.

Semantic shifts in all language families

‘top’ > ‘head’ in Samoyed, Mansi and Yakut

Proto-Samoyed **olə̑* ‘top’ > ‘head’ in Selkup and Kamas (red)

Proto-Uralic **päŋe* ‘top, end’ > ‘head’ in Mansi *puŋk* (brown)

Proto-Turkic **tepö* ‘top' > Yakut *töbö* ‘head’ (blue)

CLICS: 37 including Mari and Nenets, DatSemShift: 2

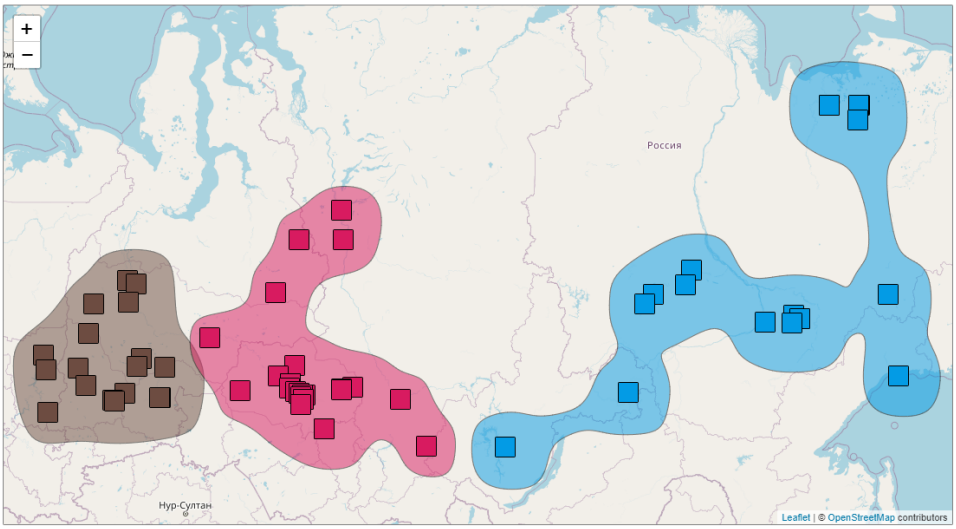

Figure 4. ‘top’ > ‘head’ in Samoyed, Mansi and Yakut

This semantic shift is quite likely typologically, judging by the numbers in the databases. Here at the map no sign of particular areality is shown: areas of Ob-Ugric, Samoyed and Turkic polysemy are evenly distributed across Siberia.

‘blood’ > ‘red’ in Ob-Ugric and Tungusic

‘blood’ Proto-Uralic *kälV and Finno-Ugric *wire > ‘red’ in Ob-Ugric (brown)

Morphological derivation in Tungus-Manchu: *sēgV- ‘blood’, *sēgǯe- ‘red’ (Evenki, Ulcha, Orok, Nanai, Oroch) (yellow)

CLICS: 56 including Mansi, DatSemShift: 9

This is a frequent typological case, according to databases. However, particular realisations in Ob-Ugric and Tungusic are different. Whereas in Mansi and Khanty this shift exists in the form of polysemy, Tungusic languages used a suffix deriving the word ‘blood’ from ‘red’. Like in the case of ‘new, fresh’, these areas are adjacent on the territory of West Siberia, where speakers of the languages in question contact. This polysemy comprises West Siberia and Far East.

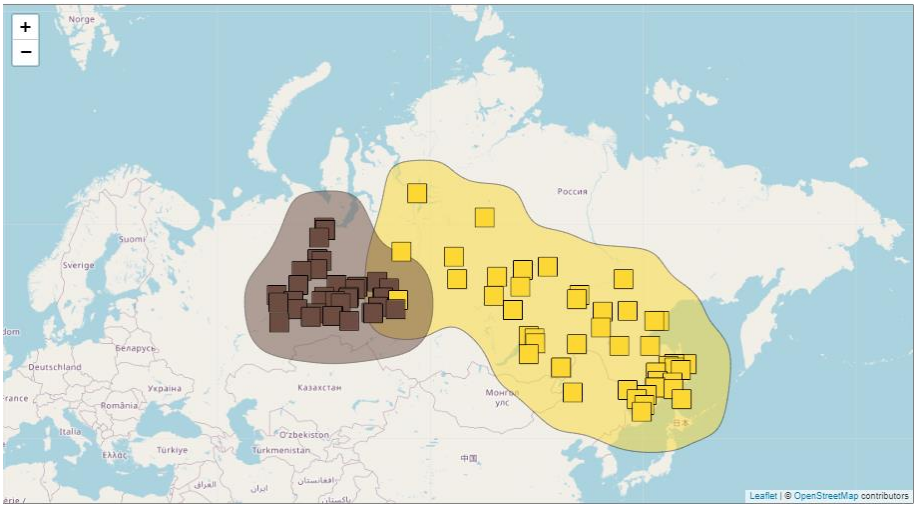

Figure 5. ‘blood’ > ‘red’ in Ob-Ugric and Tungusic

‘fish scales’ > ‘bark’ in Ob-Ugric and Yakut & Dolgan

Mansi , Northern Khanty: semantic evolution **śęme* (**śōme*) [UEW: 476] ‘fish scales’ > ‘bark’ (brown)

Yakut, Dolgan contemporary polysemy ‘fish scales’, ‘bark’ < semantic evolution from Proto-Turkic **Kaδiř* ‘bark’ (blue)

CLICS: 17, no Uralic examples, DatSemShift: 0

The areas of the detected polysemy are isolated and do not point to any areal patterns. However, there should be a potential historical and cultural explanation of such semantic shift, which deserves further research.

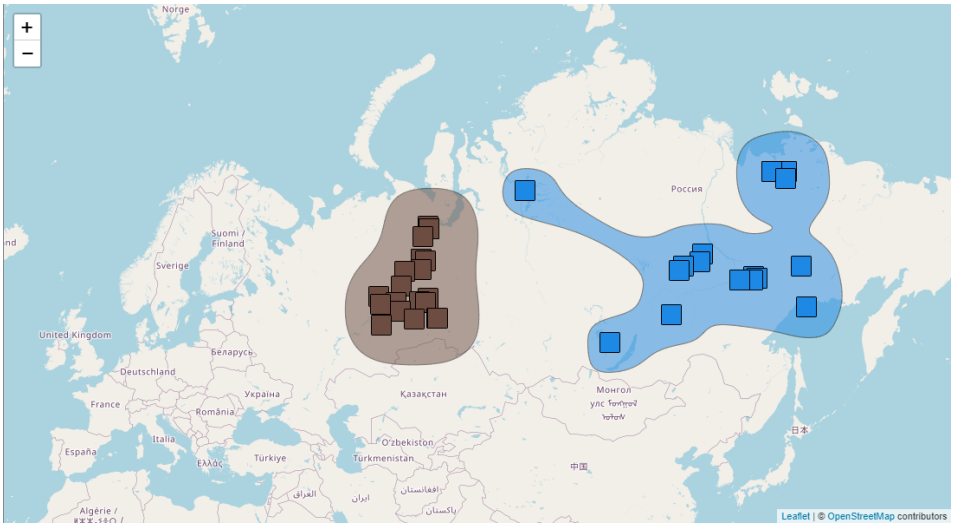

Figure 6. ‘fish scales’ > ‘bark’ in Ob-Ugric and Yakut & Dolgan

Different types of semantic shifts

‘old woman, wife’ in Samoyed and Khanty

Proto-Samoyed polysemy**menəsä* ‘wife, old woman’: polysemy in Enets *méńise* (red)

Evolution: Khanty *imi* ‘old woman, wife, grandmother’ < Proto-Uralic **imV* ‘old woman, grandmother’ (brown)

CLICS: 5 including Northern Saami, Kildin Saami and Nenets, DatSemShift: 9

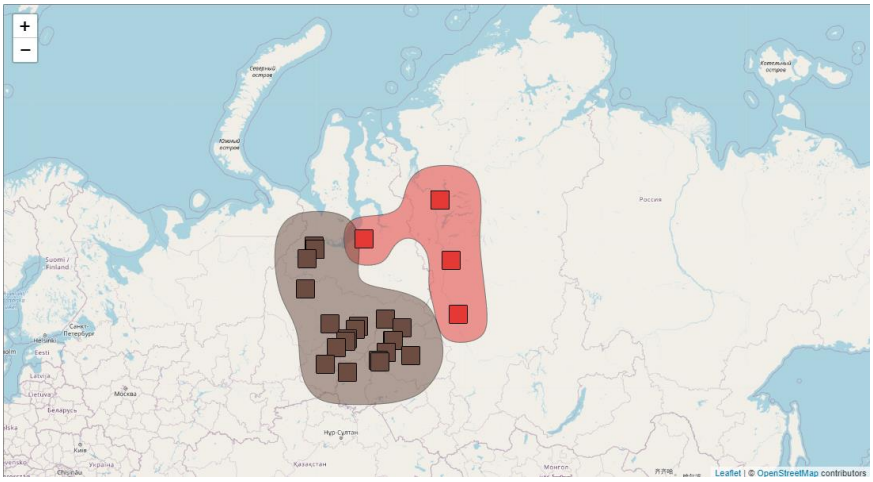

Figure 7. ‘old woman, wife’ in Samoyed and Khanty

‘blood vessel’ > ‘root’ in Ob-Ugric and Turkic

Finno-Ugric *särV ‘blood vessel’ > Ob-Ugric ‘root’ (brown)

Proto-Turkic **ḍamor* ‘blood vessel, root’ (blue)

CLICS: 25, no Uralic, DatSemShift: 5

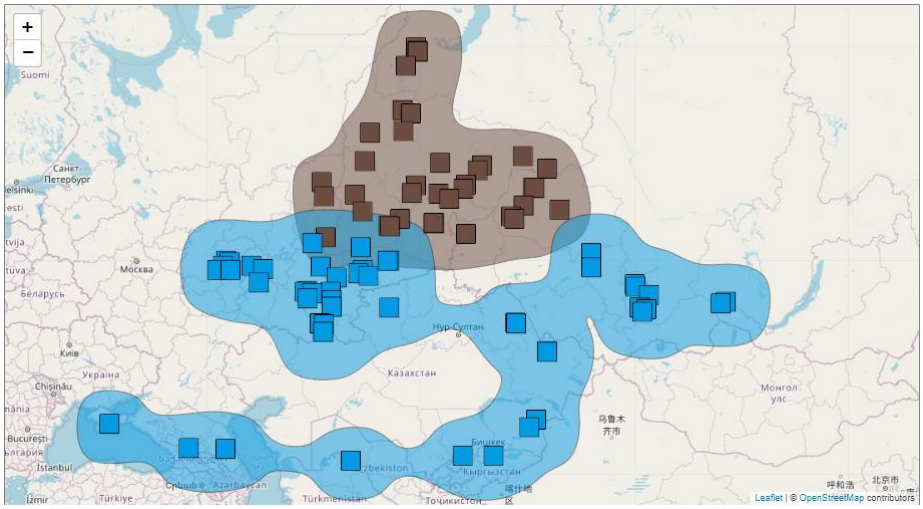

Figure 8. ‘blood vessel’ > ‘root’ in Ob-Ugric and Turkic

‘head’, ‘end’ in Ob-Ugric and Turkic

Proto-Uralic **päŋe* ‘head, end’ > Mansi *p'aŋk ‘head’* & Proto-Uralic **ukV (*okV)* ‘end*’* > Khanty *ŏχ* ‘head’ (brown)

Proto-Turkic **baλč* ‘head’ > Northwestern (Kipchak) Turkic ‘end’ (blue)

CLICS: 15 including Finnish, North Karelian and Hill Mari, DatSemShift: 0

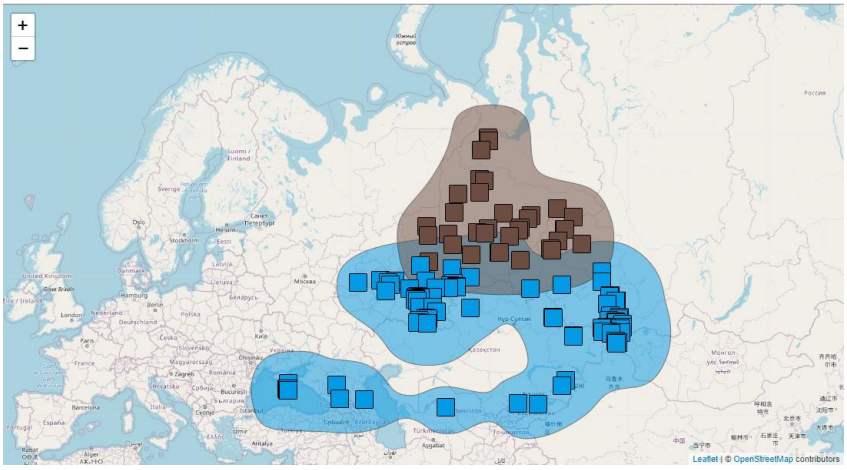

Figure 9. ‘head’, ‘end’ in Ob-Ugric and Turkic

‘earth’, ‘clay’ in Mansi, Samoyed and Turkic

‘earth’ > ‘clay’ evolution and polysemy in Nenets, Enets, Selkup, Kamas, Taigi

Proto-Samoyed **jåə̑* , **tʲə̑pə̑* (red)

evolution ‘clay’ > ‘earth’ in Southern Mansi < FU **śawe* ‘clay’ (brown) This evolution may be contact-induced.

evolution and polysemy ‘earth’, ‘clay’ in Turkic < Proto-Turkic **bąl-čïk, -gač* ‘dirt, clay’ (blue)

CLICS: 56, no Uralic, DatSemShift: 4, including Udmurt

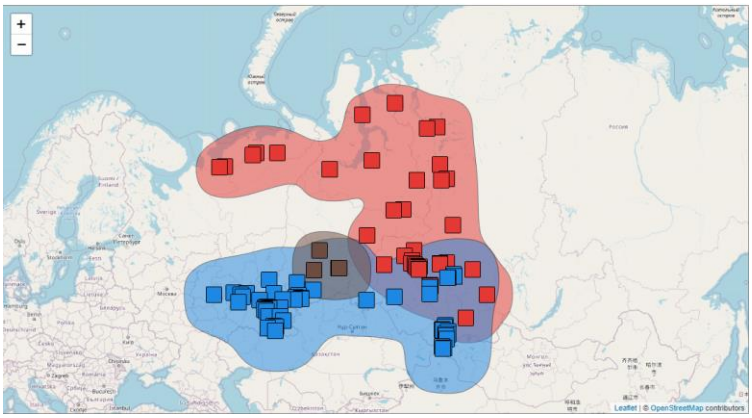

Figure 10. ‘earth’, ‘clay’ in Mansi, Samoyed and Turkic

‘inside’ > ‘belly’ in Mansi, Turkic and Tungusic

Mansi *kaχr-* ‘belly’ < Finno-Ugric **kirkV* ‘inside’ (brown). This shift may be contact-induced, as the Mansi area is surrounded by such type of polysemy in neighbouring languages.

Proto-Turkic **ič* ‘inside' > ‘belly’ Yakut *is*, Dolgan *uc*, K-Balkar *ич*, Tatar *эч*, Bashkir. *es*, Nogai *iš*, Kazakh *iš*, Karakalpak *iš*, Kyrgyz *ič*, Altai *ič* (blue)

Proto-Tingus-Manchu **xemu-gde* cognate polysemy ‘inside’ (Evenki, Even, Orok), ‘belly’ (Negidal, Udighe, Ulcha, Nanai) (yellow)

CLICS: 10, no Uralic, DatSemShift: 4

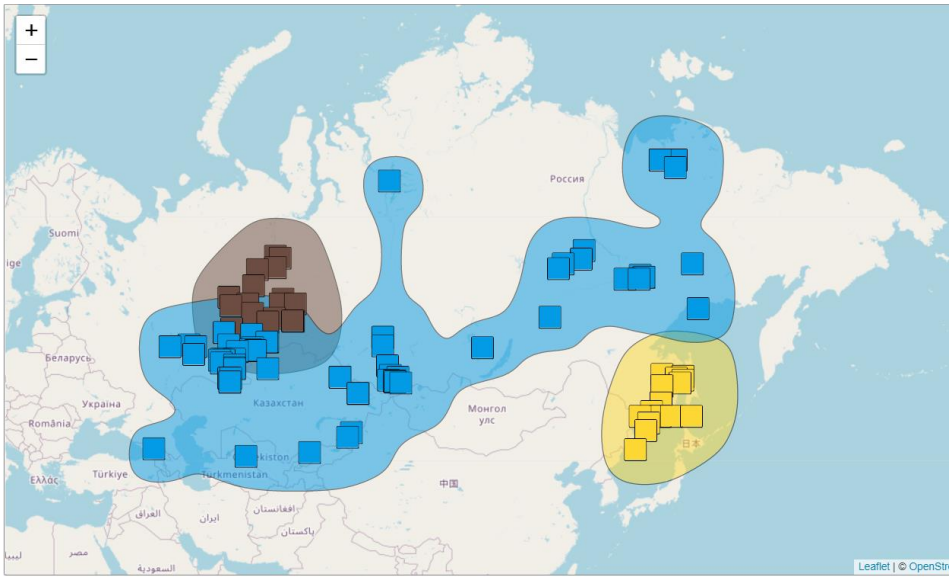

Figure 11. ‘inside’ > ‘belly’ in Mansi, Turkic and Tungusic

‘hair, fur’ > ‘feather’ in Ob-Ugric, Samoyed and Turkic

Proto-Samoyed polysemy **tə̑r* ‘(bodily) hair, feather’ (red)

Finno-Ugric **puna* ‘hair, fur’ > Ob-Ugric ‘feather ’ (brown) Proto-Turkic **tük* ‘(bodily) hair’ > ‘feather’ Proto-Tungus-Manchu **xin-ŋa-* ‘hair, fur, feather’ (yellow)

CLICS: 111 including Khanty and Komi, DatSemShift: 4

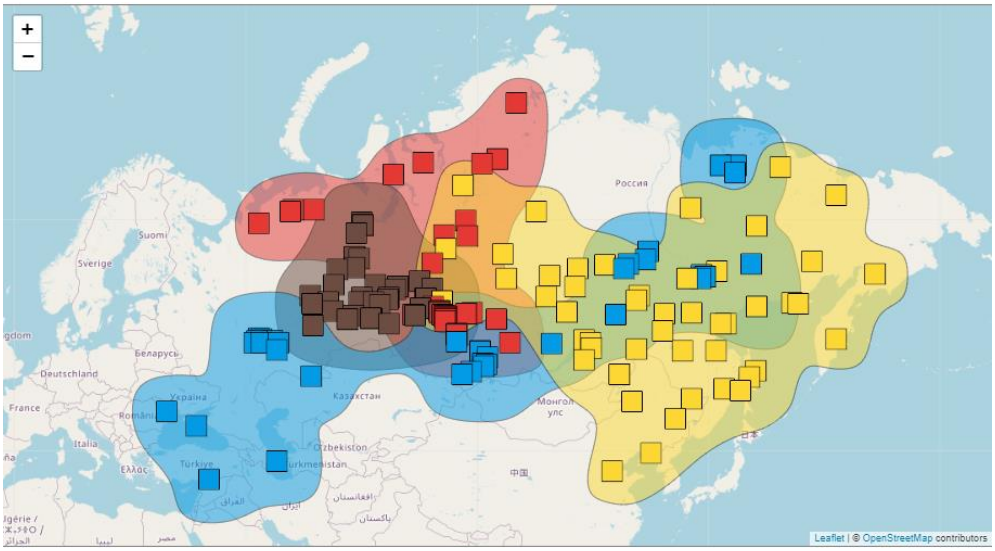

Figure 12. ‘hair, fur’ > ‘feather’ in Ob-Ugric, Samoyed and Turkic

4 Discussion and Conclusion

So far, we have developed 11 maps with preliminary results. Having identified the shared patterns in a specific area, we need to find out if this similarity is a typological or an areal issue. In cases with no evident extralinguistic factors, the reason for shared semantic patterns can be small-scale multilingualism.

Possible contact-induced semantic changes in basic vocabulary (such as ‘earth, clay’ and ‘earth, place’ in Mansi) might reflect very early language contacts or established multilingualism of the speakers. In order to go deeper into the origins of these semnatic processed, diachronic geography of the languages and history of migrations in Siberia should be thoroughly considered.

Thus, the research is based on the intersection of synchrony and diachrony, of typology and semantic reconstruction. It yields the results that would otherwise not be obtained and poses questions for further inquiries.

5 References

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. Selkupsko-russkiy slovar’. {Selkup-Russian dialectal dictionary}. Tomsk.

CLICS - Database of Cross-Linguistic Colexifications https://clics.clld.org/ (accesed 3/10/2022)

DatSemShift — Zalizniak, A. et al (2016-2022). “Database of Semantic Shifts”. Moscow: Institute of Linguistics, Russian Academy of Sciences. // https://datsemshift.ru/

DEWOS — Steinitz, W. (1966–1993) Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache. Berlin.

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

Fedotova, I. (2020). Polysemy in the Samoyed core lexicon lists and language contacts. *Ural-Altaic studies*. 2020; 2 (37): 77-113.

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. Severnosel'kupskij slovar'. {Northern Selkup dictionary}. Hamburg.

Honti, L. (1982) Geschichte des obugrischen Vokalismus der ersten Silbe. Budapest: Akadémiai Kiadó.

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

Kassian, A., Starostin, G., Dybo, A., Chernov, V. (2010). The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship*. no. 4, pp. 46–89.

LingvoDoc – Linguistic platform LingvoDoc http://lingvodoc.ispras.ru/ (accesed 3/10/2022)

Munkácsi, B. and Kálmán B. (1986) *Wogulisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Rédei, K. (1986–1991) *Uralisches Etymologisches Wörterbuch*. Budapest: Akademiai Kiado, 1986–1991. Bd. I–III.

Solovar, V. (2006) Hantyjsko-russkij slovar'. {A Khanty-Russian dictionary} Sankt-Peterburg.

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

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

Tereshkin, N. (1961) Ocherki dialektov hantyjskogo yazyka. Chast' pervaya. Vahovskij dialekt. {A survey of Khanty dialects. Part 1. Vakh dialect} Moskva, Leningrad: Izdatel'stvo Akademii nauk SSSR.

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

UEW — Rédei, K. (1986-1989) Uralisches Etymologisches Wörterbuch. Budapest.

URHIA — Uralic Historical Atlas https://sites.utu.fi/urhia/

Zaliznyak, A. (2018) Katalog semanticheskikh perekhodov: 20 let spustya // Russian Journal of Linguistics. Vol. 22, №4 (2018). P. 770-787. {Zaliznyak Anna A. Catalog of semantic transitions: 20 years later // Russian Journal of Linguistics. Vol. 22, no. 4 (2018). P. 770-787}

Zhivlov, M. (2011). Annotated Swadesh wordlists for the Ob-Ugrian group (Uralic family). Online version at: Starostin, George (ed.) 2011-2020. *The Global Lexicostatistical Database* Moscow: Higher School of Economics, & Santa Fe: Santa Fe Institute. URL: http://starling.rinet.ru/new100/ (accesed 3/10/2022)