Diachronic Dialectal Classification of Mansi: Evidence from Basic Vocabulary

Idalia Fedotova

Diachronic dialectal classification of Mansi: evidence from basic vocabulary

Idalia Fedotova[1,2]

[1]HSE University

[2]Ivannikov Institute of System Programming of the RAS

This paper deals with diachronic classification of Mansi dialect groups based on new resources. The previous research on this data was focused on highlighting the phonetic isoglosses between the discovered dialects, so now it is possible to compare the results based on phonetics with the results gained from lexicostatistics. This research aims to define the dialect groups of Mansi from the 18th to 21st centuries, based on the lists of basic vocabulary, using new expeditional and archival data published on the linguistic platform LingvoDoc and field materials gathered by the author. As a result, for the first time, lists were compiled of the basic vocabulary of seven archival and three contemporary dictionaries that became known to researchers in recent years, an appropriate questionnaire was also completed by a native speaker of Sosva Mansi. Judging by a comparison with the 19th century data, the most distinctive dialect group is North Mansi, as it has retained many features from the 18th century. Other dialect groups have not been so stable. In the late 18th century there were dialects of the Perm province, although not homogenous enough as a group, and out of them the Kungur and Solikamsk dialects were closer to the Tavda (southern) dialect, which is known from recordings from the 19th-20th century. The eastern dialects disappeared at the beginning of the 21st century. Judging by the data obtained from the last native speakers, the Yukonda (eastern) dialect shared 93% of basic vocabulary with the northern dialects.

1 Introduction

The Mansi language belongs to the Ugric group of the Uralic language family. Mansi and Khanty are traditionally thought to form the Ob-Ugric subgroup within Ugric. The present-day Mansi people live mainly in Siberia, in the Khanty-Mansiysk Autonomous district, and there are also a few groups of Mansi in the north of Sverdlovskaya Oblast and Perm Krai. Currently, the Mansi language is on the verge of extinction: in the 2010s, sound recordings data from the last native speaker of the eastern (Yukonda) dialect were made, whereas southern and almost all western dialects disappeared even earlier.

Classificationary division of dialects is usually carried out using three main criteria: phonetic isoglosses, grammatical isoglosses and analysis of common lexicon. The Mansi dialect groups were identified on the basis of materials collected in the second half of the 19th – early 20th century by

Kannisto and Munkácsi. Judging by these materials, all dialects are divided into four large groups: 1) southern, 2) western, 3) eastern, 4) northern, which is shown in Table 1.

	Southern	Eastern	Western	Northern
1	$\bar{a}$	$\bar{o}$	$\bar{o}$	$\bar{o}$
2	$\bar{\underset{.}{\varepsilon}}$	$\bar{a}/\bar{\underset{.}{e}}$	$\bar{\underset{.}{e}}$	$\bar{a}$
3	∂w	$\bar{\imath}$	$\partial \gamma$	$\partial \gamma$
4	$\acute{c}$	$\acute{s}$	$\acute{s}$	$\acute{s}$
5	$\check{s}$	$s(\check{s})$	$\check{s}$	$s(\check{s})$
6	k	x/k	k	x
7	k	k	k	k

Table 1. Main differentiating Mansi isoglosses according to Honti (1988, p.149)

However, having conducted phonetic analysis of the recently discovered materials of the Mansi dialects of the 18th century, Normanskaya and Koshelyuk arrived at the conclusion that dialectal division could have been historically different (Normanskaya 2020, Normanskaya&Koshelyuk 2020). The data of the extinct dialects of the Perm province showed that the third feature (Table 1) is not differentiating for dialect grouping (Normanskaya 2020). The materials from the 18th century, for example, Pallas's unpublished dictionary, reveal such language varieties that combine phonetic and morphological features of different groups (Normanskaya, Koshelyuk 2020). The lexicon of these sources has not yet been studied and/or documented.

The lists of Mansi basic vocabulary have already been compiled by Zhivlov (2011), based on the materials of Munkácsi's dictionary. However, later, new archival and field dictionaries were published on the LingvoDoc platform, which led to a correction of Proto-Mansi reconstruction (Normanskaya 2015). To clarify the classification of Mansi dialects in diachrony, it is necessary to consider these new materials in the light of lexical isoglosses and lexicostatistics, using a unified methodology for compiling and analysing basic vocabulary. The lexicostatistical approach is applied to the "language vs dialect" problem, which was proven effective on the material of a large number of languages of the world by Koryakov (2017). Among the Uralic languages, dialectal classification has already been revised using new data from LingvoDoc for the Samoyed (Koryakov 2018) and Permic (Normanskaya, Bezenova in print) languages.

2 Materials and Methods

This study uses the latest expeditionary and recently published archival dialectal dictionaries. The following expedition dictionaries were analysed: Yukonda dialect (eastern, collected by Amelina (2013), Middle-Ob dialect (northern, collected by Stenin (2013) and Sosva dialect (northern, collected by Idrisov (2016) in the village Hulimsunt). Additionally, in order to collect the most

complete and actual list of basic vocabulary, in 2021, a native speaker of the Lyapin (Sygva) dialect of the northern group completed a special questionnaire with diagnostic contexts (Kassian et al. 2010), and the results are included in this study.

The archival dictionaries under examination were collected by Pallas in the late 18th century (Sjögren's archive, stored at St. Petersburg department of the Archive of the Russian Academy of Sciences): Perm dialects - Verkhoturye, Cherdyn', Kungur, Solikamsk, as well as one unlocated northern dialect. In addition, we analysed the dictionary of the Berezov (northern) dialect (using the printed edition of Pallas's dictionary) and the dictionary of an undetermined dialect of the Mansi language from the archive of Sjogren of the 19th century, which also belongs to the northern group.

The differentiation of dialects of the Mansi language was carried out by the method of lexicostatistics, which relies on a list of basic vocabulary for each language/dialect. To collect basic vocabulary from all the sources listed above, a list of 110 concepts was used, which is a one hundred-word list of Swadesh with additions by Yakhontov. This list is preferrable, because, firstly, a detailed clarification of semantics has been developed (Kassian et al. 2010) specifically for this list, and secondly, it has been used in the etymological dictionary of the basic vocabulary of the Turkic languages (Dybo 2013) and for compiling hundreds of vocabulary lists in the Global Lexicostatistical Database (GLD), including the Ob-Ugric (Zhivlov 2011). Within the Uralic family, the compilation and analysis of material using this technique, with further verification by native speakers, was previously carried out in five idioms of the Baltic-Finnish group (Rozhanskiy & Zhivlov 2019).

We compiled the lists of basic vocabulary for all the mentioned dictionaries, adding reconstructions at different levels of the proto-language: Proto-Uralic, Proto-Finno-Ugric, Proto-Ugric (UEW), and Proto-Ob-Ugric (Honti 1982). Furthermore, we calculated the percentage of lexical coincidences between pairs of dialects: all borrowings were excluded, the total number of words attested in both dialects was recorded, and within this scope the number of cognates was identified. The percentage of coincidence indicates the number of cognates in the ancestral vocabulary. In order to exclude borrowings from the calculation, these data were checked against works on borrowings in the Mansi language from Tatar (Kannisto 1925) and Komi (Redei 1970), as well as by notes on borrowings in the Mansi dialectal dictionary (Munkácsi & Kálmán 1986).

3 Results

The results of the calculations are presented in Table 2. The lower left half of the table shows the total number of original lexemes recorded in both dialects, and the number of lexemes having a common root. The grey-shaded cells indicate the number of lexemes out of 110 that could be found in each individual list.

Only two lists — Yukonda (eastern) and Lyapin (northern) — turned out to be complete enough to be interpreted from the point of view of the "language vs dialect" problem, so the columns dedicated to them show the percentage. In the course of this work, we arrived at the conclusion that the fragmentary nature of the remaining lists does not allow us to interpret the percentages in a straightforward way (the 91-92% threshold between the language and the dialect), but we can use these calculations in a relative sense to compare the proximity of dialects to each other. In order not to mislead our readers, in the table, we have indicated only the numerical expression of coincidences for all other dialects, and in the analysis of the table below we will operate with percentages, stipulating their relativity.

	Eastern	Northern						Perm			
	Yukonda	Lyapin	Sosva	Middle-Ob	Berezov	19th century (Sjögren)	18th century (Pallas-Sjögren)	Solikamsk	Verkhoturye	Cherdyn'	Kungur
Yukonda	88	93 %									
Lyapin	74 - 69	110									
Sosva	47 - 32	59 - 54	62								
Middle-Ob	45 - 36	57 - 47	38 - 30	60							
Berezov	38 - 29	60 - 54	39 - 36	35 - 32	63						
19th century (Sjögren)	33 - 26	52 - 45	32 - 28	33 - 30	49 - 44	52					
18th century (Pallas-Sjögren)	46 - 34	69 - 59	47 - 41	44 - 37	50 - 49	46 - 39	71				
Solikamsk	36 - 30	55 - 43	35 - 27	33 - 27	40 - 34	36 - 29	48 – 40	60			
Verkhoturye	29 - 22	47 - 32	28-18	31 - 20	42 - 31	39 - 28	39 – 26	36 - 34	49		
Cherdyn'	34 - 25	51 - 35	29 - 21	32 - 25	49 - 37	43 - 30	43 – 33	39 - 36	46 - 38	51	
Kungur	33 - 23	57 - 40	31 - 21	30 - 22	49 - 35	43 - 30	42 – 30	39 - 35	46 - 39	50 - 39	51

Table 2. Percentage of cognates coincidences and overall number of attested concepts from 110-item list in Mansi dialects

4 Discussion

Turning to the coincidences within dialect groups, in the northern 18th century Mansi the archival Berezov dialect almost completely overlapped with the lexicon of Pallas's unpublished dictionary. However, the dialect attribution of the latter is difficult to determine unambiguously: earlier it was established that this manuscript had phonetic isoglosses with northern and eastern dialects, and grammatical ones with Western dialects (Normanskaya, Koshelyuk 2020). So, the Berezov dialect stands out because it is 90% similar to all other dialects of the northern group, whereas Sjogren's 19th century archive has 84-89% of coincidences within them. Contemporary dialects Lyapin and Sosva are close to each other (91% coincidences), while the Middle Ob dialect is much more different (78%-80% of coincidences). A discrepancy of this scale can also be explained by the incompleteness of the data. From the point of view of phonology, contemporary Northern dialects also have differences: Proto-Mansi * $ę$ > Ob. e, Sosva $a/ā$, Proto-Mansi *$ä$ / *$ī$ > Ob. e, Sosva a, also, in some cognates, Proto-Mansi *u > Ob o vs Sosva u (Normanskaya 2015). The percentage of coincidences with archival dictionaries is quite high in all three contemporary dialects (84-92%). Thus, in modern northern Mansi dialects, quite a lot of innovative processes have occurred both at the phonetic and lexical levels, and this group is not homogeneous.

Out of the four extinct 18th century Perm dialects, the Solikamsk list of basic vocabulary is attested most fully (60 words). The remaining lists comprise only 49-51 words. More than 91% of the common lexemes have Solikamsk, Verkhoturye and Cherdyn', and 89% of coincidences are between the Solikamsk and Kungur lists. The biggest distance is found between the Kungur and Cherdyn' dialects (78%). Thus, the Kungur dialect is further away from the rest of the Perm dialects, so the extinct 18th century Perm dialects were not as homogeneous as the northern ones at that time.

Now let us turn to the differences between different groups of dialects. 18th century: generally, between the northern and Perm dialects, the percentage of coincidence ranges from 66% to 76%, with the exception of the Solikamsk dialect, as mentioned above. Contemporary dialects: the last attested eastern (Yukonda) dialect has 70% and 80% of coincidences with Sosva and Middle-Ob (northern) dialects respectively. However, according to the most complete list collected from a native speaker of the Lyapin (northern) dialect, the percentage of coincidences between contemporary northern and eastern dialects is 93%.

As the Cherdyn', Kungur and Verkhoturye dialects are closest to the Tavda (southern) dialect by phonetic correspondences (Normanskaya 2020), we compared Perm dialects with the list of Tavda (from materials collected at the end of the XIX century) (Zhivlov 2011). A relatively high percentage of correspondences in the basic vocabulary confirms the proximity of these dialects (from 75% to 83%). The Solikamsk dialect coincides with the Tavda dialect by 90% and, apparently, belongs to the same group.

It is valuable to compare the results obtained with Table 3 – the data collected by Zhivlov from Munkácsi's dictionary. It should be noted that this dictionary was compiled from field materials recorded in 1888-1889, and is separated by a century from both Pallas's dictionary and the new field dictionaries of the 2010s.

	Northern	Western			Eastern	Southern
	Northern	Middle Lozva	Lower Lozva	Pelym	Konda	Tavda
Northern		92 %	89%	88%	87%	82%
Middle Lozva			98.%	96 %	96%	90 %
Lower Lozva				97%	97 %	91 %
Pelym					97 %	91 %
Konda						92 %
Tavda						

Table 3. Percentage of coincidences in Mфnsi dialects according to Zhivlov (2011)

The Northern and Tavda (southern) dialects are the most distant (82%), while the eastern Konda and western dialects have virtually no differences (96-97% coincidences). The distance between the northern dialects and the eastern (Konda) is 87%. Taking this analysis into consideration, Koryakov (2017) identified two languages between Mansi varieties, claiming the 91-92% threshold in the "language vs dialect" problem. However, data from the beginning of the 21st century demonstrate 93% similarity between the eastern and northern dialects of Mansi, which indicates the unity of the Mansi language.

Also, a comparison of Tables 2 and 3 shows that in the 18th century a number of Perm dialects existed, which did not form a united group and disappeared back in the 19th century. So by the beginning of the 21st century, of all the dialects except northern, only Yukonda (eastern) was still spoken.

5 Conclusion

Comparing the percentages of coincidence and lexical and phonetic isoglosses of the 18th century, two dialectal groups are clearly distinguished: northern and southwestern (Solikamsk-Kungur). The rest of the dialects of that time, apparently, cannot be grouped into one unit. Nevertheless, in the 21st century, the northern (Lyapin) and eastern (Konda) overlap by 93% in basic vocabulary, which confirms the unity of the contemporary Mansi language. The major obstacle to conducting this research in full is the impossibility of collecting complete lists of basic vocabulary from other archival and field sources. Therefore, the discussed percentages of coincidence between dialects other than Lyapin and Yukonda cannot be understood literally, but they still give an idea of the relation between different idioms. Study of the material of Perm dialects changes the traditional classification of Mansi dialects (Honti 1988) and highlights significant changes that they underwent in the 19th century.

6 References

Amelina, M., Normanskaya, Yu. *Slovar' yukondinskogo dialekta mansiyskogo yazyka* [Dictionary of Yukonda dialect of Mansi]. URL: http://lingvodoc.ispras.ru/dictionary/657/3/perspective/657/4/view (accessed 06.10.2022)

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

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

Honti, L. 1988. Die wogulische Sprache. *The Uralic Languages. Description, History and Foreign Influences. Handbuch der Orientalistik*. [Sinor Denis ed.] E. J. Brill, Leiden; New York; København; Köln: Brill, 1988. pp. 147─171.

Idrisov, R., Normanskaya, Yu. *Slovar' sos'vinskogo dialekta mansiyskogo yazyka* [Dictionary of Sosva dialect of Mansi]. URL: http://lingvodoc.ispras.ru/dictionary/660/7/perspective/660/8/view (accessed 06.10.2022)

Kannisto, A. 1925. Die tatarischen lehnwörter im Wogulischen. *Finnisch-Ugrische Forschungen*. 1925. Band XVII. Heft 1-3. pp. 1─264.

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

Koryakov, Yu. 2017. *Problema "yazyk ili dialect" i popytka leksikostatisticheskogo podhoda* ["Language or dialect" problem and an attempt at lexicostatistical approach]. *Voprosy yazykoznaniya* [Issues in Linguistics].

Koryakov, Yu. 2018 *Problema "yazyk ili dialect" i samodiyskie yazyki* ["Language or dialect" problem and the Samoyed languages]. *Uralo-altajskie issledovaniya* [Ural-Altaic Studies], no. 4 (31), pp. 156–217

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

Munkácsi B. & Kálmán, B. 1986. *Wogulisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Normanskaya, Yu. 2015. *Novye polevye i arhivnye dannye po mansijskim dialektam i ikh znachenie dl'a pramansiyskoy rekonstruktsii sistemy vokalizma pervogo sloga* [New field and archival data of the Mansi dialects and their significance for the reconstruction of the first syllable vocalism in proto-Mansi]. *Uralo-altajskie issledovaniya* [Ural-Altaic Studies], no. 4 (19), pp. 40–59.

Normanskaya, Yu. 2020. *Slovari permskikh mansi "aborigenov Sibiri", sobrannye P. S. Pallasom v 18 veke* [Dictionaries of Perm Mansi "native speakers of Siberia", collected by P. S. Pallas in the 18th century]. *Uralo-altajskie issledovaniya* [Ural-Altaic Studies], 2020, no.3 (38), pp. 71─80.

Normanskaya, Yu.& Bezenova, M. Classification of Udmurt dialects based on digital methods. In print.

Normanskaya, Yu. & Koshel'uk, N. 2020. *Neopublikovannyj mansiyskiy slovar' P. S. Pallasa - ranee neizvestnyy mansiyskiy dialekt?* [The unpublished Mansi dictionary of P. S. Pallas - an earlier unknown Mansi dialect?]. *Uralo-altajskie issledovaniya* [Ural-Altaic Studies], no. 1 (36), pp. 92‒100.

Rédei, K. 1970 *Die Syrjänischen Lehnwörter im wogulischen*. The Hague: Mouton & Co.

Rozhanskiy, F. & Zhivlov, M. 2019. Votic and Ingrian core lexicon in the Finnic context: Swadesh lists of five related varieties. *Linguistica Uralica*, 2019, no. 2, pp. 81–108. https://dx.doi.org/10.3176/lu.2019.2.01

Slovar' verkhoturskogo dialekta mansiyskogo yazyka (iz arhiva A. M. Shogrena, perepisannogo iz materialov P. S. Pallasa) [Dictionary of Verhotur'e dialect of Mansi (from A. J. Sjögren's archive rewritten from the materials of P. S. Pallas)]. URL: http://lingvodoc.ispras.ru/dictionary/2685/1653/perspective/2685/1654/view (accessed 06.10.2022)

Slovar' dialekta mansiyskogo yazyka (iz arhiva A. M. Shogrena, 19 в.) [Dictionary of a Mansi dialect (from A. Sjögren's archive, 19th century]. URL: http://lingvodoc.ispras.ru/dictionary/2654/9320/perspective/2654/9324/edit?page=6 (accessed 06.10.2022)

Slovar' kungurskogo dialekta mansiyskogo yazyka (iz arhiva A.M.Shogrena, perepisannogo iz materialov P.S.Pallasa) [Dictionary of Kungur dialect of Mansi (from A. J. Sjögren's archive rewritten from the materials of P. S. Pallas)] URL: http://lingvodoc.ispras.ru/dictionary/2685/6/perspective/2685/7/view (accessed 06.10.2022)

Slovar' mansiyskogo yazyka (sobran P. S. Pallasom v 18 v., iz arhiva A. M. Shogrena) [Dictionary of Mansi (compiled by P. S. Pallas) in the 18th century, from A. J. Sjögren's] URL: http://lingvodoc.ispras.ru/dictionary/1393/29131/perspective/1393/29132/view?page=6 (accessed 06.10.2022)

Slovar' mansiyskogo yazyka, sostavlennyy protoiyereem S. Cherkalovym (g. Solikamsk, 1783) [Dictionary of Mansi compiled by archpriest S.Cherkalov (Solikamsk, 1783)]. URL: http://lingvodoc.ispras.ru/dictionary/867/9/perspective/867/10/view ((accessed 06.10.2022)

Slovar' severnogo dialekta mansiyskogo yazyka (s.Berezovo, sobran P. S. Pallasom v 18 v.) [Dictionary of a northern dialect of Mansi (village Berezovo, compiled by P. S. Pallas in 18th century]. URL: http://lingvodoc.ispras.ru/dictionary/3066/11/perspective/3066/14/view (accessed 06.10.2022)

Slovar' cherdynskogo dialekta mansiyskogo yazyka (iz arhiva A. M. Shogrena, perepisannogo iz materialov P. S. Pallasa) [Dictionary of Cherdyn' dialect of Mansi (from A. J. Sjögren's archive rewritten from the materials of P. S. Pallas)] URL: http://lingvodoc.ispras.ru/dictionary/2685/846/perspective/2685/847/view (accessed 06.10.2022)

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/ (accessed 08.08.2020). (In English)

Stenin, I. & Normanskaya, Yu. *Slovar' sredneobskogo dialekta mansiyskogo yazyka* [Dictionary of Middle-Ob dialect of Mansi]. URL: http://lingvodoc.ispras.ru/dictionary/656/2/perspective/656/3/view (accessed 06.10.2022)

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

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)