Uralic studies, languages, and researchers

Edited by Sándor Szeverényi

Uralic studies, languages, and researchers

Proceedings of the 5th Mikola Conference
19–20, September 2019

Edited by Sándor Szeverényi

Szeged, 2021

Table of contents

Foreword

On September 19–20, 2019, the Department of Finno-Ugric Linguistics, University of Szeged, held the 5th Tibor Mikola Memorial Conference, to celebrate the memory of its late professor and department head. This volume presents papers from this conference, to commemorate the 20th anniversary of his untimely death in 2000. The special theme of our 2019 conference was the history of (Uralic) linguistics, and the two invited talks, by Beáta Wagner-Nagy and Rogier Blokland, are the opening papers of this volume, on *Undiscovered treasures: From the field research archive to the digital database*, and *Notes on Nicolaes Witsen and his Noord en Oost Tartarye*, respectively. The rest of the papers address topical issues in several subfields of Uralic linguistics, such as Ob-Ugric languages, Samoyed languages, and Permic languages – that is, exactly those branches of the Uralic family of languages, research into which in Szeged began under Tibor Mikola's guidance and is still carried out at our department. In addition, the volume also includes a paper on Mari, and two papers overviewing research into several languages.

It is our sincere hope that, similarly to earlier commemorative Mikola volumes, the present collection will also contribute to Tibor Mikola's legacy.

Sándor Szeverényi

Notes on Nicolaes Witsen
and his *Noord en Oost Tartarye*

Rogier Blokland

Uppsala University

1. Introduction

Over the last years there has been a quiet upsurge in research on Nicolaes Witsen (1641–1717), intermittent mayor of Amsterdam and author of *Noord en Oost Tartarye* (henceforth: *NOT*), a chaotic but fascinating and invaluable tome on the geography, topography, history and ethnography of that vaguely defined region previously known as 'Tartary'. A voluminous Dutch-language biography aimed at the general public was published in 2010 (Peters 2010), nearly 130 years after Gebhard's 1881 biography, which, however, focused on his political career and paid little attention to his 'hobby', i.e. his research on Eurasia. In 2010 a Russian translation of the 1705 edition of *NOT* was published. Some years before Wim Lucassen (1945–2006), the noted Dutch Abkhazologist and member of the team that had prepared the Russian edition, had suggested that publishing a separate volume on the linguistic material in *NOT* would be a worthwhile endeavour (cf. Naarden 2018: 2); this resulted in a book (Naarden 2018) with 30 articles of varying length on topics ranging from Witsen's Dutch–Georgian wordlist to Witsen as a western pioneer of Koreanology. It must be pointed out, however, that older sources such as Witsen's diary, published in three volumes in the 1960s (Locher and de Buck 1966–67), though not unknown in our field, are also still able to occasionally yield interesting nuggets of information. The time is therefore ripe, in my opinion, for a new brief overview of Witsen and his *NOT*.

2. Witsen's background

Nicolaas ('Nicolaes' in older Dutch spelling) Witsen came from a well-to-do, powerful, Amsterdam family. His father, Cornelis Janszoon Witsen (1605–1669), had been councillor in the Amsterdam city council, mayor of the city a number of times, and one of the managing directors of the Dutch East India Company (often considered to be the world's first multinational). Due to his important functions Cornelis was immortalized in Bartholomeus van der Helst's painting *Banquet of the Amsterdam Civic Guard in Celebration of the Peace of Münster* (1648), which can be admired the

Rijksmuseum in Amsterdam. He was perhaps not a pleasant man, being described as 'a drunk, … a bloodsucker and a hypocrite' (Schama 1999: 593), and is perhaps best (but unfairly) known for having caused Rembrandt's bankruptcy in 1658 (thus e.g. Schama 1999: 613; Crenshaw 2006: 78).[1] Growing up in an affluent household, Nicolaes had his own manservant, accompanied his father to England when Cornelis went there to meet Oliver Cromwell (1599–1658), and was given private lessons in 1657 by Jan Comenius (1592–1670), the renowned theologian, philosopher and pedagogue. He also studied law in Leiden where he attended lectures by Jacob Golius (1596–1667), the famous orientalist.

Nicolaas was thus used to power and money from the beginning, and it is then no surprise that he too followed in the Witsen family footsteps and held during his lifetime a number of important positions: he was (like his father Cornelis, nephew Jonas, and his second cousin Johannes Hudde) mayor of Amsterdam (13 times between 1682 and 1706, as at the time one could be mayor for two years within a three-year period), ambassador extraordinary in England, political commissar, delegated councillor, artillery commissar of Amstelveen, Urk en Emmeloord, and, ultimately of significance with regard to his famous opus *Noord and Oost Tartarye*, board member of the Dutch East India Company (a life-long appointment). This powerful position, in addition to the mayoralty of Amsterdam, was especially useful for him, as through it he gained contacts all over the world who were often to some extent obligated to him and whom he could thus pester for information for his research projects. Thus, for instance, the Dutch sailor Willem de Vlamingh (1640–1698), in an unashamed attempt to flatter his patron, named an island northeast of Kolguyev Island (part of the present-day Nenets Autonomous Okrug) after Witsen, which Witsen immediately added to his 1690 map of Tartary. However, the island later turned out to be non-existent, though it occurs on some other maps of northern Russia (cf. Plate 1); Peters (2010: 92) suggests Vlamingh may have played a practical joke on Witsen.

---

[1] It has recently been shown, however, that this is actually not true, as the loan Rembrandt had contracted for from Witsen was paid out in full out of the sale of movable property, which was sold before Rembrandt was forced to sell his house (Bosman 2019: 118–119).

Plate 1. Witsen Island northeast of Kolguyev Island on the 1742 Covens &
Mortier map 'Magnae Tartariae Tabula',
based on Guillaume de l'Isle's 1706 map of Tartary.

Witsen is well known for having had extremely varied interests, including history, cartography, astronomy, geography, geology, zoology, botany, epigraphy, and linguistics. These diverse fields may at first sight seem to be completely unrelated to his political pursuits, but in fact they were all connected to his one main interest: the prosperity of the city of Amsterdam. As this was based mainly on trade with other countries, so the more one knew about these, including their history, their physical features, and their inhabitants, the more successful one could be. This had already been advocated by the humanist Casparus Barlaeus (1584–1648), professor of philosophy and rhetoric at the Athenaeum Illustre (the precursor of the University of Amsterdam), who in his 1632 inaugural oration *Mercator sapiens* asserted that a knowledge of geography, cosmography, ethnology and navigation is of use for merchants and commerce.[2] The idea, therefore, that merchants and traders could benefit from delving into fields outside of their own purview was already current at the time. Incidentally, the catalogue of Witsen's library (Schouten 1747), which was

---

[2] Cf. *Illud ostendam: Optimum esse Mercaturæ cum Sapientiæ ac litterarum studiis commercium, nec augendæ rei curam mentis contemplationibus, nec has illi obesse, verum optimis rationibus inter se conspirare, mercandi & philosophandi facultatem, ut tanto mihi felicior sit futurus mercator, quanto philosophari poterit luculentius.* (Barlaeus 1632: 8).

put on sale in 1747 after his nephew Nicolaes Lambertszoon had died the previous year (Peters 2010: 341), showed Witsen had owned a copy of Barlaeus' oration (Schouten 1747: 14).

Witsen also had a specific interest in Asia in general, again mainly for commercial reasons: was there a northern route to India (by which other nations trading with India could be outmanoeuvred)? Could one draw up a map of Asia with all possible trading possibilities? Closely related to these matters were his religious interests, both practical (the proselytization of 'pagans' in Asia) and theoretical (how did humanity spread from Paradise to such faraway places as Australia?). Logically related to the question about the spread of humanity is the question of the relationship of the world's other languages to Hebrew; in the 17th century it was still generally accepted in Western Europe that Hebrew was the source of all languages.

3. The mission to Muscovy

In the late 16th century European trade with Muscovy was mostly in English hands, but due to their access to goods from Asia, Dutch trade with Muscovy started to increase, and by 1600 the Dutch had overtaken the English. There was another increase in Dutch–Russian trade after 1650, and as there was no permanent representative of the Netherlands in Muscovy before 1678, Dutch diplomatic missions were sent there from the 1600s onwards, especially in order to resolve commercial disputes. It was one of these missions that the young Witsen, 23 at the time, was able to join due to his background as son of the (erstwhile) mayor of Amsterdam. He joined the mission led by ambassador Jacob Boreel (1630–1697) as a 'state nobleman', a title he was given merely to increase the mission's prestige. The mission lasted from 17 September 1664 to 12 August 1665, and comprised 47 members, including also a secretary, an interpreter, a surgeon, assorted noblemen and officers, stewards, a quartermaster, an equerry, grooms, drummers, trumpeters, and pages. The equerry would not return to the Netherlands, as he had his throat cut in Russia (Peters 2010: 37).

Witsen kept a diary during the mission to Moscow, which has its own interesting history. The original is now lost, but in 1668, after his return from Moscow, Witsen gave a copy of it to the French author and scientist Melchisédec Thévenot (1620–1692) when Thévenot was visiting Amsterdam. Amongst his many other activities Thévenot also published travelogues; this was probably the reason Witsen gave him a copy of the diary. Thévenot later bequeathed it to the Royal Library in Paris, and the diary, which Witsen had never meant to have published, lay there undiscovered[3] for

---

[3] E.g. Gebhard, in his biography of Witsen, writes that the diary had been lost (Gebhard 1881: 33).

nearly two centuries, and it was not until 1886, when the library published a catalogue of its Dutch manuscripts (Huet 1886: 49, nos 47–49), that it was (re-)discovered. It was then published in 1966–67 by Theodor Locher and Piet de Buck, but this Dutch-language publication received little attention in Uralic circles; Stipa (1990: 46–47) may have been the first to have made use of it. A Latin translation had been planned in the early 18th century, but nothing came of it (Peters 2010: 189), and an edition was planned by the Finnish historian Erkki Kuujo (1921–2004) in the 1950s, but when Kuujo heard of the prospective Dutch edition he selflessly forwent his own project (Locher and de Buck 1966/I: LXXII). A Russian translation[4] of the diary was published in 1996.

The diary, though it does not describe everything Witsen saw during the voyage to and from Moscow, as it was more meant to be an aide-memoire for himself (cf. Peters 2010: 38), is not unimportant for Uralic studies in that it notes a number of minor facts and occurrences relating to Uralic peoples, which are often either not identical with the information in *NOT* or not mentioned there at all. For instance, on the 15th of February 1665 Witsen was invited to dinner at an Englishman's residence, where he saw a dance by a group of 'Samoyeds' who had recently arrived in Moscow and describes their features, dress, behaviour, and language;[5] this description is not featured in *NOT*. Two weeks later, on the 1st of March 1665, a group of Samoyeds also visited ambassador Boreel; Witsen notes they worship 'images' but know of God's existence.[6]

In addition to the diary, among the manuscripts in Paris there are also brief notes[7] Witsen made in Moscow about Russia in general; this can be seen as the starting point for his subsequent pursuit of collecting information about Eastern Europe, and North and Central Asia. In a description of 'Mordvin Tatars', which Witsen obtained in Moscow in conversation with the Dutch trader Jan van Zweden (d. 1668), he notes

---

[4] The Russian translation was carried out by Wilhelmina Triesman (1901–1982), a Dutch woman who emigrated to Soviet Russia in 1925 and worked as an assistant at the Museum of Anthropology and Ethnology (cf. Jager 2012: 239–240, 248–249, 272).

[5] *Dese dan dansten op een been, klapten in hunne handen, sloegen 't hooft op een sonderlinge wyse en songen op haer tael voor ons, dat een belachelyk aanschouwen gaf*. 'They danced then on one leg, clapped their hands, hit their heads in a strange manner and sang in their language for us, which was an absurd spectacle' (Locher and de Buck 1966/II: 139).

[6] *Se bidden beelden aen, echter weeten van Godt te spreeken; als men wilde sy eens op haer wys soude bidden, wygerde dat: 't selvige geschiet sonts op een vreemde wyse*. 'They worship images but know to speak of God; when asked to pray in their own manner they refused: it usually occurs in a strange manner.' (Locher and de Buck 1966/II: 148).

[7] *Aentekeningen van saeken mij voorgekomen op myn Moscovische reyse soo van de Russen selve als harens landtsaert* 'Notes on matters which occurred during my voyage to Muscovy about the Russians and their character' (Locher and de Buck 1967/III).

how they, despite being farmers, have 10 to 12 wives each (Locher and De Buck 1967/III: 381); in *NOT* (1705: 624), however, the text says 'six or seven' wives. There is also a brief description of the Russian Saami (Locher and De Buck 1967/III: 332–334), which is completely missing from *NOT*. A careful comparison of the information in the diary and the notes with the information found in *NOT* could therefore be a worthwhile undertaking.

In Moscow Witsen also met Andreas Winius (Russian: Андрей Андреевич Виниус; 1641–1717), the son of a Dutch merchant who had emigrated to Russia and (unbeknownst to either) a distant relation; Winius would turn out to be probably the most important of Witsen's contacts in Moscow, as he moved in the highest circles, was rich, spoke both Dutch and Russian natively, and held a number of influential positions, including that of first postmaster general (1675–1701) of Russia, director of the Apothecary Chancellery, and, from 1697 to 1703, chancellor of the Siberian Chancellery. Winius also had the Russian cartographer Semyon Remezov (1642–after 1721) compile a map of Siberia, a copy of which Witsen also obtained. A number of letters sent by Witsen to Winius have been preserved (cf. Peters 2010: 106-108), and it is generally assumed that Winius functioned as intermediary who sent Witsen linguistic material on the languages of Russia collected by others. He is never mentioned in *NOT*, however, probably so that there would be no danger of him being potentially accused of treason (cf. Boterbloem 2013: 72).

It is unlikely that Witsen had already developed all his above mentioned interests during the mission; it is often forgotten that there were exactly 25 years between his return from Moscow and the printing (sic!) of the first edition of *NOT* (on which more below). He did, however, receive a list of miscellaneous questions to ask in Moscow from Jacob Golius (1596–1667), the abovementioned Orientalist from Leiden (e.g. How far did the Tsar's realm reach to the east of the River Ob? Were there precious metals to be found in Muscovy? How did the Russians mate? Etc.; cf. Peters 2010: 34). The notes accompanying the diary that Witsen gave to Thévenot in 1668 also prove that Witsen had already commenced with his own research during his stay in Moscow: they include a description of 'Samoyeds' that he had received from an Englishman trader from Archangel; this is also included in *NOT* (1705/II: 895–897), but the inclusion of the description in the notes shows that Witsen already had it by 1668, and, as he travelled back via Novgorod and Riga and did not visit Archangel (cf. Locher and de Buck 1966/I: LXIV–LXV), he must have received it either in Moscow or soon afterwards by post in Amsterdam (Witsen already received his first post from Moscow in 1666; Peters 2010: 102). In contrast to what is occasionally thought,[8] Witsen made no further trips to Russia or Asia (Peters 2010: 45).

---

[8] '... voyagea plus loin et pendant de longues années en Russie' (Muller 1859: 58).

4. Noord en Oost Tartarye

After his return from Muscovy he again dedicated himself to his political and civic duties, which he always considered the most significant tasks bestowed on him. He also aimed to compile a map of 'Tartary' and write a commentary to this map: these were supposed to form a cartographic and written overview of trade possibilities in Russia and Asia. Nowadays Witsen is also best known for these printed works, which were in fact the result of his leisure pursuits, and, in a way, mere distractions from the main business of statecraft. In total he wrote five important works: in 1671 *Aeloude en hedendaegsche scheeps-bouw en bestier* 'Ancient and modern shipbuilding and management', a 556-page richly illustrated description of the history of shipbuilding from the Greeks to the 17th century. It consists of two sections: the first section recites what classical and modern authors have written about the topic, whilst the second section describes the present-day situation. This structure is apparent in *NOT* too. *Aeloude en hedendaegsche scheeps-bouw en bestier* was the only book which was actually published in Witsen's lifetime, quickly gaining renown outside the borders of the Netherlands as well: soon after its publication there were already copies in the libraries of well-heeled intellectuals in Sweden, Italy, Russia, and even the Dutch Indies (Peters 2010: 152); a second edition was published in 1690.

Witsen's second important publication is a 1687 map of 'Tartary' (= Siberia and Mongolia). This map, the *Nieuwe Lantkaarte Van het Noorder en Ooster-deel van Asia en Europa strekkende van Nova Zemla tot China*, has its own intrinsic importance as one of the best foreign maps of Siberia at the time; it is extant today in only 13 copies. It is officially dated to 1687, but most of the extant copies in the Netherlands are inscribed with a dedication to the tsar 'Peter Alexowitz', who only became tsar in 1696. This is typical for Witsen's working method: his achievements, be they books or maps, would be printed, but not actually published, and he would keep them at home (there being no mundane need for him to earn money through their sale) and continuously improve and refine them, making minute changes whilst constantly acquiring new information. The importance of the map for us lies in the fact that *NOT* was in fact planned merely as a commentary on this map.

Witsen's third important publication and, undoubtedly, the best known to our readership, is *NOT*, first published in 1692. As shown by the diary, already in Moscow Witsen had started to establish ties with people with whom he would later correspond, and later this would only increase. The ceaseless flow of information from his global network of informants also had an effect on the structure of *NOT*: what was at first conceptualized as a supplement and guide to the map of Siberia grew gradually, as Witsen kept adapting and rewriting it, into a fragmented and disorganized work. At the time, however, this was not necessarily seen as a flaw; rather, it was thought that a text was learnèd only if it contained an unending accumulation of facts.

If we want to be more accurate, we have to say that the 1692 edition was printed but not published, i.e. it was not commercially available. As mentioned above, there was no need for Witsen to sell any of his work, and so his usual mode d'emploi was to print but then keep the actual copies of his work at home. There was also a specific reason for Witsen to not distribute the book, which had to do with the visit in 1697 to the Netherlands by Tsar Peter the Great (incognito, because, according to an ecclesiastical edict, the Tsar was not allowed to leave holy Russian soil). In the Netherlands the Tsar of course met Witsen, mayor of Amsterdam at the time, as it was already known then that Witsen knew a great deal both about Russia and about shipbuilding; this last fact was important as the tsar wanted to make Russia into a naval power. Witsen set himself up as the Tsar's friend, guide, and intermediary, and used the occasion to gather more information about distant lands and exotic peoples from the Tsar's entourage with its hundreds of members and which included both important people such as e.g. Fëdor Alekseevich Golovin (1650–1706), governor of Siberia, and James Daniel Bruce (1669–1735), statesman and general, as well as e.g. Georgian royalty, 'Tartar' slaves, Circassians, and others. The Tsar asked Witsen to continue with his geographical and ethnographical research, which was to have a momentous consequence: Witsen decided then to make many additions to the map, which he had kept at home anyway, also deciding to completely rewrite his book. It was now no longer supposed to be a guide to the map, but rather a compendium of everything known about the lands east of the Urals, though in fact Witsen never restricts himself areally to Asia. This second, much expanded, edition was printed in 1705, in two volumes. Again, the book was printed but not published, and as with the 1692 edition, Witsen kept the actual copies of the book at home and continuously made changes to it; the first edition which was actually published and sold was the second printing of the second edition in 1785, 68 years after Witsen's death, i.e. the first 'public' edition was posthumous. This also explains why there are so few extant copies of the 1692 edition: Witsen did not think it good enough and did not want people to have it. The print run of the 1692 edition is unknown, and Peters (2010: 185) claims that there are only four extant copies (one each in the university libraries of Amsterdam and Utrecht, and two in Moscow: one in the Russian National Library and one in the library of the Academy of Sciences), but a quick online library search has turned up five more: three copies in Germany (one each in Munich in the Bavarian State Library, in Göttingen in the State and University Library, and in Stuttgart in the State Library of Württemberg), one in Paris (in the National Library of France), as well as one in Denmark (in the Royal Library). However, the same fate was to befall the second edition, as here too Witsen never had it sold, and as a matter of fact it was not actually even totally ready in 1705: Witsen was still making changes to the text and to the plates. The date '1705' therefore does not mean that anybody could have

possessed a copy in 1705. There were a number of reasons why Witsen did not finish the book: the engraver was slow; Witsen had promised the Tsar a new version of the book but could not send him a copy before a new version of his map was ready (as the book was initially meant to be an accompaniment to the book); as he kept receiving more information from his world-wide network, Witsen just could not stop trying to improve the text, and finally, Witsen felt his strength ebbing in his old age.[9] It was then only after his death that the existence of a second edition of *NOT* came to light (cf. Peters 2010: 151).

Because there were so few physical copies of the 1692 and the 1705 editions, the book was in fact not very well known, and it was not completely translated into another language until the 1950s. Jurčenkov (1995: 160), citing Adelung (1864: 206–207), claims it was translated into French, German and English, but Adelung in fact observes that it was *not* translated because it was (a) very big, and (b) written in Dutch.[10] Tibor Mikola (1936–2000) made a valuable contribution by translating the sections on the Uralic peoples from Dutch into German and publishing these in 1975.

In 2010 a Russian translation of the complete *NOT* was published in Amsterdam, which also has an interesting history. It was translated from the Dutch into Russian by the aforementioned Wilhelmina Triesman. She was working in the Museum of Anthropology and Ethnography in St. Petersburg in 1945 when she, as a native speaker of Dutch, was asked to translate *NOT* into Russian in order to make it available for Soviet specialists on Siberia and Asia, using a copy of the book from the library in Leningrad; the very same copy of the book had been used often in the 19th century by German-Russian scientists such as e.g. (the aforementioned) Friedrich von Adelung (1768–1843) and Leopold von Schrenck (1826–1894), whose knowledge of German made it easier to use the book. The (hand-written) translation was finished by 1950, but in the 1950s the political and ideological climate changed, and publication of a Russian version of *NOT* was out of the question. However, it could be used by interested scholars,[11] as the manuscript was registered and available in the library, and

---

[9] In 1713 Witsen writes to his friend Gijsbert Cuper (Peters 2010: 195): *De lust is mij hier ook so seer benomen tot de studie door oneyndelijke klijnigheden, die men mij aen heeft gedaen, dat ik er schier onder buck. Echter geeft Godt mij sterkte en tijt van leven, verhope nog iets te doen.* 'My desire to continue with my work has been greatly curbed by the infliction of unending trivialities that it weighs heavily upon me. But if God gives me strength and time, I still hope to do something.'

[10] 'То обстоятельство, что сочинеіе Витсена, столь важное и любопытное, не было переведено ни на один язык, по истинѣ возбуждаетъ удивленіе, и можетъ быть объяснено тѣмъ только, что оно очень велико, равно и тѣмъ, что писано на Голландскомъ языкъ.' (Adelung 1864/II: 207).

[11] Mokšin (1993: 24) is therefore mistaken when he claims in 1993 that there was no Russian translation; the existence of the translation, however, was not generally known.

there were still attempts to have it published in the 1970s, whilst Triesman, Witsen-like, kept on improving the text, doing so even on her deathbed. It was not until 2010, however, 22 years after Triesman's death, that the Russian translation was finally published (cf. Jager 2012: 164–170, 271–274).

5. The collection of the Uralic material in *Noord en Oost Tartarye*

We know very little about when or from whom Witsen received his linguistic material, and, regrettably, the contents of three boxes with all his notes (30 volumes) and practically all letters were mislaid or lost in 1817 (Peters 2010: 13, 19–20). Naarden (2018b: 7) states that Witsen used more than 700 printed, handwritten, and oral sources, ranging from classical authors to letters sent to him from Russia. In *NOT* he often mentions that he obtained material from a 'friend', and as pointed out above, even the name of Andreas Winius, the friend he had made in Moscow, does not occur in it. Already in Moscow he met, to the extent that it was possible, people he considered interesting (including the Patriarch Nikon of Moscow; cf. Locher and de Buck 1966/II: 267–289), but there is no mention in the diary of any linguistic material collected in Moscow, and it has also been shown with the aid of Witsen's correspondence with Leibniz (cf. Blokland 2018a: 285) that Witsen very likely did not come into the possession of any material in Uralic languages during his stay in Moscow, and that he received the vast majority (if not all) the linguistic material present in *NOT* by post long after his return to the Netherlands; he received the first letters from Moscow already in 1666 (Peters 2010: 102). None of the Uralic linguistic material in *NOT* had been previously published, so Witsen must have acquired all of it through his personal contacts. It is sometimes thought that Witsen personally visited areas where Uralic languages were spoken (e.g. Turkin 1993: 278; Mokšin 1993: 26–27; Ivanov 2003: 33; Fournet 2008: 13), but Witsen's diary shows that he did not travel further than Moscow.

It is also generally known that the 1692 and the 1705 editions differ as to the Uralic linguistic material they contain: the 1692 edition already contains the Dutch-Moksha wordlist (which has been analysed already a number of times; cf. e.g. Feoktistov 1959; Feoktistov 1963; Fournet 2008; Maticsák 2012: 39–56; van Pareren and Blokland 2018), but the translations of the Lord's Prayer in Mari, Zyrian Komi, Mansi, Nenets, Enets, Nganasan and Selkup are only found in the second, 1705, edition, and have in general, probably due to their size, received less attention (for recent analyses, see the corresponding articles in Naarden 2018). Here too we do not know who sent these to Witsen; again, we have to be satisfied with typical vague comments such as the one in a letter from Witsen to Leibniz, dated the 9th of April 1699: 'mes amis de Mosco m'écrivent d'avoir reçû quelques autres Oraisons Dominicales, mais qu'à cause des troubles passez & et de plusieurs autres affaires, ils

n'avoient pas en du tems pour les faire copier' (Leibniz 1717: 367).

It has also long been recognized that Witsen's interest in wordlists and translations of the Lord's Prayer in the languages of Russia and Asia ensued at least partially from his correspondence with Gottfried Wilhelm Leibniz (1646–1716) (cf. Müller 1955: 6, 17; Stipa 1990: 161; van Hal 2016; Blokland 2018a: 285), and with the German Orientalist Andreas Müller (1630–1694), who had published (under the pseudonym 'Barnimus Hagius') a collection of translations of the Lord's Prayer in 85 languages (Müller 1680; Peters 2010: 261–262); Leibniz[12] was specifically interested in translations of the Lord's Prayer as comparative material because he thought a comparison of the world's languages could shed light on early human history. The correspondence between Witsen and Leibniz commenced in 1694 (Müller 1955: 6), but we do not know when his correspondence with Müller started; as Müller dedicated a book to Witsen in 1680 (Peters 2010: 331), we can surmise that they must have become acquainted much earlier. We can thus presuppose that Witsen's interest in acquiring such translations was probably inspired by Müller, also because Leibniz' interest in languages in general is reliably assumed to have started only in 1685 (Vermeulen 2015: 68). Interestingly, for a number of the translations of the Lord's Prayer we can narrow down the date Witsen acquired them: it has been shown that Witsen received the translation of the Lord's Prayer in Mari between the 16th of October 1697 and the 5th of July 1699 (Blokland 2018b: 307); for Zyrian Komi and Mansi it could be shown more specifically that Witsen received the translation between the 22nd of May and the 5th of June 1698 (Blokland 2018c: 322).

Similarly, for a translation in a Samoyed language ('l'Oraison Dominicale en Langue Samoyede'), Witsen writes to Leibniz in a letter dated the 21st of July 1698 that he has just received it (Leibniz 1717: 365), and in another letter dated the 5th of July 1699 he writes that he has received two others in 'Samoyed', where Witsen already says that they are different in comparison to the one he already sent, and to each other (Leibniz 1717: 367–368).[13]

Considine (2017: 154) makes the interesting observation that Witsen's Kalmyk wordlist (*NOT* 1705: 297–304; for a recent analysis, see Badagarov 2018) and the Moksha wordlist were based on the same prompt-list; the Kalmyk wordlist is nearly twice as long as the Moksha one, but many sections have an identical order of Dutch lexical items. Considine therefore assumes that the interpretation of local languages in Russia was in some way centrally coordinated, but it may simply mean that one of

---

[12] Leibniz' interest in language samples, especially from Russia, is well known (cf. Vermeulen 2015: 49; van Hal 2016), although from the 1690s onwards he realized that wordlists are not sufficient for proper language comparison (Müller 1955: 28; Vermeulen 2015: 78).

[13] '... les deux autres en langue *Samojede*, toutes deux differentes de celle, que je vous ay ci-devant envoiée, & d'aussi differentes entr'elles, ...'

Witsen's Russian contacts sent the same list to different local collectors. Another interesting fact with regard to the Kalmyk and Moksha wordlists is that the translation *amidu* for Dutch *leeft* ('is alive') in the Moksha wordlist (cf. *NOT* 1705: 625) is in fact Kalmyk (cf. Kalmyk *emd* 'alive', Written Mongolian *amidu* 'id.', Written Oirat *amidu* 'id.'; Badagarov 2018: 489). The mistake may have been made by Witsen, or the mixup may already have occurred before the letters with the Moksha and Kalmyk material reached him. It has already been shown (cf. Badagarov 2018 for Kalmyk) that much of the material was mediated via Russian, but despite his interest in languages,[14] Witsen's own language knowledge was not very impressive: he read Greek with difficulty, could write some Latin, spoke French and English, but not very well (Peters 2010: 253; Naarden 2018b: 9).[15] In Moscow he attempted to learn Russian but was not allowed to do so, and could only attain some elementary knowledge of the language (cf. Locher and de Buck 1966/I: LXI). It is thus not surprising, both due to the convoluted route the material was sent to Witsen and because of his modest knowledge of Russian and its alphabet, that there are many inaccuracies and misinterpretations in the translations.

6. Conclusion

Noord en Oost Tartarye is a fascinating book, though it is not an easy read. As Müller (1733: 216–217) states, it seems as if Witsen added bits and pieces to his magnum opus as and when he received them.[16] For a number of reasons (his wealth, his worry that he might be plagiarized, his promise to Tsar Peter the Great, his unending efforts to improve his book, the language it was written in, and ultimately his failing health) led to *NOT* not being made commercially available and therefore not as well-known as it could have been. The Russian translation, carried out more than 50 years ago by Wilhelmina Triesman in the Soviet Union in extremely difficult circumstances, has now finally seen the light of day, and its availability both as a printed book (Witsen 2010) as well as an online Dutch–Russian bilingual edition (Witsen 2015) will hopefully cause a renewed interest in Witsen's work, as there still remain many unanswered questions.

---

[14] His library included more than 50 dictionaries and more than 40 grammars of languages ranging from Chinese to Syriac (Schouten 1747; Peters 2010: 360).

[15] Jorink (2010: 327) claims that Witsen 'studied Greek, Coptic, Arabic, Turkish, Aramaic and cuneiform, as well as more contemporary languages such as Hottentot, Kalmuk (= Kalmyk) and Samoyed', but this is undoubtedly a misunderstanding.

[16] 'Ausser dieser General-Eintheilung ist im Wercke fernerhin keine Ordnung anzutreffen. Die weitläuffigsten Beschreibungen und Rapporten, worin offtmahls von gantz unterschiedenen Materien gehandelt worden, sind in Forma eingerücket, und es scheinet nicht selten, als habe sich der Verfaszer in Mittheilung derselben blos nach der Zeit des Empfangs gerichtet.'

References

Adelung, Friedrich (von) 1864. *Kritiko-literaturnoe obozrenie putešestvennikov po Rossii do 1700 goda i ikh sočinenij*. Čast' II. Moskva.

Badagarov, Jargal 2018. Witsen's Kalmyk glossary. In: Naarden 2018: 461–538.

Barlaeus, Casparus 1632. *Mercator sapiens, sive oratio*. Amsterdam: Guillielmus Blaeu.

Blokland, Rogier 2018a. IV. Finno-Ugric languages. Introduction. In: Naarden 2018: 283–288.

Blokland, Rogier 2018b. The Lord's Prayer in Mari. In: Naarden 2018: 305–317.

Blokland, Rogier 2018c. The Lord's Prayer in Zyrian Komi. In: Naarden 2018: 319–334.

Bosman, Machiel 2019. *Rembrandts plan. De ware geschiedenis van zijn faillissement*. Amsterdam: Athenaeum-Polak & Van Gennep.

Boterbloem, Kees 2013. *Moderniser of Russia. Andrei Vinius 1641–1716*. Houndmills: Palgrave Macmillan.

Considine, John 2017. *Small dictionaries and curiosity. Lexicography and fieldwork in post-medieval Europe*. Oxford: Oxford University Press.

Crenshaw, Paul 2006. *Rembrandt's bankruptcy. The artist, his patrons, and the art market in seventeenth-century Netherlands*. Cambridge: Cambridge University Press.

Feoktistov, Aleksandr P. 1959. Esimene mordva sõnastik. *Emakeele Seltsi aastaraamat* 5: 268–287.

Feoktistov, Aleksandr P. 1963. Mordovskie jazyki i dialekty v istoriko-ėtnogra-fičeskoj literature XVII–XVIII vv. Koljadënkov, M.N. (ed). *Očerki mordovskikh dialektov*. Tom II. Saransk: Mordovskoe knižnoe izdatel'stvo. 3–36.

Fournet, Arnaud 2008. Le vocabulaire mordve de Witsen. Une forme ancienne du dialecte Zubu-Mokša. *Études Finno-Ougriennes* 40: 1–26.

Gebhard, Johan Fredrik 1881. *Het leven van Mr. Nicolaas Cornelisz. Witsen (1641-1717). I. Levensbeschrijving*. Utrecht: J.W. Leeflang.

van Hal, Toon 2016. Leibniz, das Vaterunser und die Sprachvielfalt. In: Grötschel, Martin et al. (eds). *Vision als Aufgabe: das Leibniz-Universum im 21. Jahrhundert.* Berlin: Berlin-Brandenburgische Akademie der Wissenschaften. 255–264.

Huet, Gédéon 1886. *Catalogue des manuscrits néerlandais de la Bibliotheque Nationale*. Paris.

Ivanov, Ivan G. 2003. *Marij literatur jộlme istorij*. Joškar-Ola: Marij kugộžanộš universitet.

Jager, Janine 2012. *Wilhelmina Triesman 1901-1982. Een Nederlandse in Leningrad*. Amsterdam: Pegasus.

Jorink, Eric 2010. *Reading the book of nature in the Dutch Golden Age, 1575–1715*. Brill's Studies in Intellectual History, vol. 191. Leiden and Boston: Brill.

Jurčenkov, Valerij 1995. *Vzgljad so storony. Mordovskij narod i kraj v sočinenijakh zapadno-evropejskikh avtorov VI-XVIII stoletij*. Saransk: Mordovskoe knižnoe izdatel'stvo.

Leibniz 1717 = von Eckhart, Johann Georg (ed) 1717. *Illvstris Viri Godofr. Gvilielmi Leibnitii Collectanea Etymologica, Illvstrationi Lingvarvm, Veteris Celticæ, Germanicæ, Gallicæ, Aliarvmqve Inservientia.* Pars II. Hanoverae: Foerster.

Locher, Th. J. G. and Buck, P. de (eds.) 1966-1967. *Nicolaas Witsen. Moscovische reyse 1664–1665: journaal en aantekeningen.* I–III. Werken der Linschoten Vereeniging LXVl–LXVIJI. 's-Gravenhage: Martinus Nijhoff.

Maticsák, Sándor 2012. *A mordvin írásbeliség kezdetei* (XVII-XVIII. század). Debrecen: Debreceni Egyetemi Kiadó.

Mikola, Tibor 1975. N. *Witsens Berichte über die uralischen Völker.* Studia Uralo-Altaica VII. Szeged: Universitatis Szegediensis de Attila József nominata.

Mokšin, Nikolaj F. 1993. *Mordva glazami zarubežnykh i rossijskikh putešestvennikov*. Saransk: Mordovskoe knižnoe izdatel'stvo.

Muller, Frederik 1859. *Essai d'une bibliographie neerlando-russe. Catalogue d'une collection remarquable de livres, atlas, cartes, portraits, planches, manuscrits hollandais, et de plusieurs livres étrangers, tous concernant la Russie et la Pologne.* Amsterdam: Frederik Muller.

Müller 1680 = *Orationis Dominicae versiones fermè centum. Oratio Orationum. S S. Oratiois Dominicae Versiones præter Authenticam ferè Centum eæque longé emendatiùs quàm antehàc et è probatissimis Auctoribus potius quàm prioribus Collectionibus, Jamque singulæ genuinis Linguæ suæ characteribus adeòque magnam partem ex ære ad editionem à Barnimô Hagiô traditæ, editæque à Thoma Ludekenio, Solqv.* March. Berolini, Ex Officina Rungiana, Anno 1680.

Müller, Gerhard Friedrich 1733. Nachricht von einem raren Wercke betitult: Noord- en Oost-Tartarye durch Nicolaes Witsen. In: Müller, Gerhard Friedrich. *Sammlung Russischer Geschichte* 1(3). Sankt-Petersburg: Kayserl. Academie der Wissenschaften. 196–221.

Müller, Kurt 1955. *Gottfried Wilhelm Leibniz und Nicolaas Witsen. Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin. Klasse für Philosophie,* Geschichte, Staats-, Rechts- und Wirtschaftswissenschaften. Jahrgang 1955, nr. 1. Berlin: Akademie-Verlag.

Naarden, Bruno et al. (eds) 2018. *The fascination with Inner Eurasian languages in the 17th century. The Amsterdam mayor Nicolaas Witsen and his collection of 'Tartarian' glossaries and scripts*. Amsterdam: Pegasus.

Naarden, Bruno 2018b. Historical introduction: In: Naarden 2018: 5–60.

van Pareren, Remco and Blokland, Rogier 2018. The Dutch-Mordvin glossary. In: Naarden 2018: 289–303.

Peters, Marion 2010. *De wijze koopman. Het wereldwijde onderzoek van Nicolaes Witsen (1641–1717), burgemeester en VOC-bewindhebber van Amsterdam.* Amsterdam: Bert Bakker.

Schama, Simon 1999. *Rembrandt's eyes*. London: Allen Lane.

Schouten, Salomon 1747. *Bibliotheca nitidissima sive catalogus librorum, theologicorum, juridicorum, historicorum, philosophicorum, de hist. naturali scriptorum, numismaticorum, antiquariorum, & miscellaneorum [...]. Quibus, dum vixit, Usus est Vir Nobilissimus atque Amplissimus Nicolaus Witsen, Senator & Judex Reipublicae Amstaelodamensis. Quorum omnium publica Auctio habebitur in Aedibus Salomonis Schouten, in platea vulgo de Kalverstraat, ad Diem Lunae 20 Martii & seqq. Anni 1747.* Amsterdam: Salomon & Petrus Schouten.

Stipa, Günther Johannes 1990. *Finnisch-ugrische Sprachforschung. Von der Renaissance bis zum Neupositivismus*. Mémoires de la Société Finno-Ougrienne 206. Helsinki: Suomalais-Ugrilainen Seura.

Turkin, Adol'f I. 1993. Ob odnom maloizvestnom pamjatnike drevnepermskom pis'mennosti. In: Festschrift für Raija Bartens zum 25.10.1993. *Mémoires de la Société Finno-Ougrienne* 215. Helsinki: Suomalais-Ugrilainen Seura. 277–281.

Vermeulen, Han F. 2015. *Before Boas. The genesis of ethnography and ethnology in the German enlightenment.* Lincoln & London: University of Nebraska Press.

Witsen, Nicolaes 1705. *Noord en Oost Tartarye, ofte Bondig Ontwerp van eenige dier Landen en Volken, welke voormaels bekent zijn geweest, beneffens verscheide tot noch toe onbekende, en meest nooit voorheen beschreve Tartersche en Nabuurige Gewesten, Landstreeken, Steden, Rivieren, en Plaetzen, in de Noorder en Oostelykste Gedeelten van Asia en Europa* (2 vols.). Amsterdam: François Halma.

Witsen 2010 = Vitsen, Nikolas 2010. *Severnaja i vostočnaja Tatarija* I–III. Amsterdam: Pegasus.

Witsen 2015 = http://resources.huygens.knaw.nl/retroboeken/witsen/

Undiscovered treasures: From the field research archive to the digital database

Beáta Wagner-Nagy, Chris Lasse Däbritz, and Timm Lehmberg

Universität Hamburg

1. Introduction[1]

Finno-Ugric Studies, or to be more precise, Uralic Studies traditionally belongs to the field of empirical linguistics, even if —especially in the past— the collected data were often processed in the context of historical linguistics. It must be noted that apart from some languages, such as Hungarian, Komi, or Finnish, which were written relatively early, most Uralic languages have no —or only very few— historical written sources. Of course, this does not mean that we do not have any archival data from these languages that are older than the written form in these languages. Even if the data are recent in comparison to e.g. Hungarian, Greek, or Latin, we still have data. Nevertheless, these data were not created naturalistically, that is, they are not written thoughts or literary pieces, and they are not even translations from other languages, etc. but were created as researchers' field records. For example, the older records by Gerhard Friedrich Müller, or, somewhat later, by Matthias Castrén or Kai Donner, are records that the researchers wrote down, and not records of the respective speakers studied by them. In the records made later, such as the manuscript materials of Angeline Kuzmina in the 1960s, one can also find recordings that were not made by the researcher but by a speaker themselves. All these materials were created during or after field research and, usually, only a small portion of the collected materials have ever been published. On the one hand, researchers usually collect much more data than they then use for their publications, such as the grammars and dictionaries of the Samoyedic languages in the case of Castrén (1854, 1855). On the other hand, in some cases, as in the case of Kuzmina's notes, very few records were published (see Tuchkova and Helimski 2010), or even as in the case of Müller's notes, only small parts of the work were used by other researchers. However, access to primary data

may also be important for future generations of researchers. The processing of unpublished materials can bring new knowledge about languages or peoples, but as Bowern (2008: 185) has pointed out, studying the archive materials can also bring knowledge about the way researchers work.

This study aims to throw some light on the special nature and potential of these resources and to open the discussion for further issues of, beyond other things, information modeling, and visualization. Section 2 focuses on what archives are and how to find archive materials, such as field notes of researchers. The first thing to think about in this topic is, however, what field notes are. This topic is discussed in section 3. In section 4, two examples are provided, specifically about how we can open field notes in a modern way.

2. Archives and archive materials

It is common in science and scholarship, especially in the early times when the main part of the work was analog rather than digital, that after the death of a researcher their (scientific and/or scholarly) material was preserved by their successors or family. But one question arises: where? The simplest answer is: in an archive. But if this is the answer to this question, we must first clarify what an archive is. The easiest answer is that an archive is a physical place where letters, reports, notes, memos, photographs, drafts, and final manuscripts as well as other primary sources are kept. Assuming that we accept this simple definition, we can name many places archives. And so we do! For example, an archive is often referred to as an individual cupboard or drawer in which, for example, field note manuscripts are kept, without this place ever being an official archive. This can be called a private archive. Some archives are at least in an institution, others are actually in private households. Some of them may be cataloged and ordered, others are not. Only a few field notes are preserved in official archives. However, archives of private companies or laboratory archives as well as those of associations, unions, and political parties are also private archives. In contrast, public archives are usually maintained by public services, such as administrative institutions, governments, etc. But the main task is the same: "to systematically take over, record, order, permanently store and index written, image and sound carriers and electronic storage media from public services, other institutions (associations, companies) or individuals" (Reimann 2004: 20). Examples of storing field research materials in a public archive can also be found, e.g. the field notes by Gerhard Friedrich Müller (e.g. section 4.1 below).

Everything that is kept in an archive is called archive material (German *Archivgut*). However, typical archive materials are documents of administrative activities, letters, etc. These materials are usually unique, their copies or reproductions can never replace the original. This also applies to the field notes, even if despite their variation and

uniqueness, they are extremely rarely recognized as research objects but as a helpful tool to gain relevant data. Hence, they are not curated and preserved as other empirical material is, not even if they can be considered a manuscript.

3. What are field notes and where do field notes come from?

Every researcher who does fieldwork, be it linguistic or ethnographic-anthropological, takes memos, notes, etc. We can call these notes *field notes*. Field notes do not represent the object of research itself, they represent the researcher's reception of the object of research. This holds also for drawings of artifacts in archeology and descriptions of something observed, whereby they usually contain manifold and multi-layered information of interdisciplinary relevance. Due to the individual character of field notes, there is no standard format and content for these records. Depending on the objective or the goal of the target work, these data may be primary language materials, but they may be notes about the people or their culture. If we look at it from the perspective of a linguist, we can and must conclude that beyond the primary data, the data contain secondary linguistic materials as well. Such information may include, for example, information about the target language, translations, or analyses of the language data and geographical information as well as metadata of individuals and personal notes. Thus, field notes always contain multi-layered information of interdisciplinary relevance, and often in multilingual form.

The primacy of the language data can already be debated in so far as it always is a written representation of *eo ipso* oral material. This has two important consequences: first, field notes always and necessarily contain an interpretation of the researcher who wrote them down, and, second, field notes contain more information than just the object of research itself. To illustrate this, the image below (cf. Figure 1) shows an instance of linguistic field notes, coming from the Kuzmina Archive, which is stored at the Institute of Finno-Ugric/Uralic Studies at the University of Hamburg. It can be seen at first glance that the researcher uses at least two languages (Selkup and Russian). The word-by-word glosses are given in Russian. In addition to the Selkup words, one can find corrections written in a different pen. It is not indicated who made these corrections and when, but from the handwriting, we can deduce that it is the same person who made the field notes.
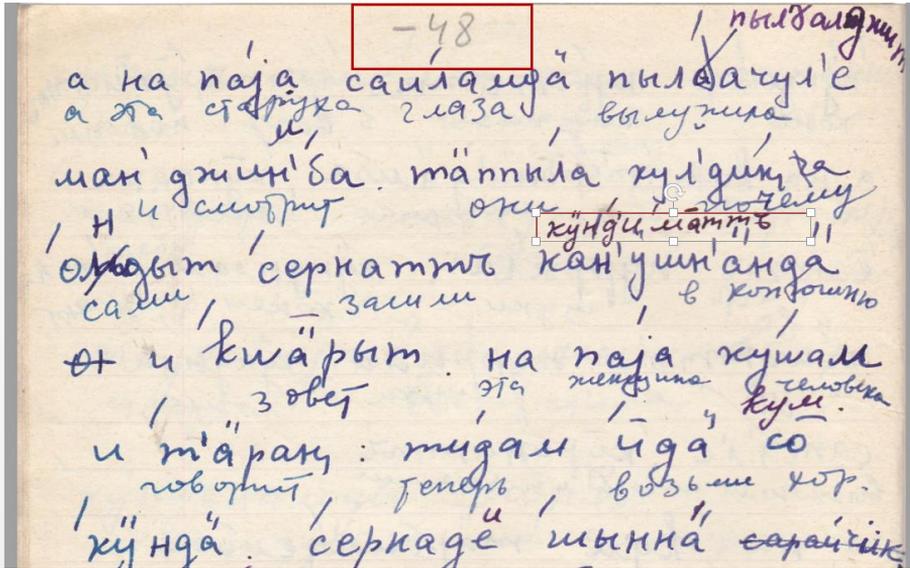
Figure 1. Manuscript field note page from Kuzmina Archive
(volume 26, book 2, page 48)

As mentioned above, many of these records are never researched. There are several reasons for this. One of them is the fact that field notes are hardly ever considered as a research topic; however, they are unique written artifacts. Another reason is that researchers always work on only a part of the collected material. A linguist publishes the linguistic data, such as texts or elicited sentences, lexical materials, etc., and probably the associated metadata as well as describes the language based on the data collected, therefore interpreting the linguistic material (cf. Himmelmann 2006, Woodbury 2011, among others). An anthropologist looks at the data from his or her point of view. A further aspect is that many field notes are either partially or not at all accessible to other researchers or the broader audience.

From the history of Uralic studies, we know very well that the collected materials did not always find their way to the scientific community. There are, again, several reasons for this. Often the researchers were not able to publish their materials, for instance, because of their untimely death (e.g. Castrén or Donner, among others) or for other, sometimes political reasons (e.g. Gerhard Friedrich Müller). Other researchers then edited the materials from their estates as in the case of Donner, whose Kamas materials were compiled and published by Aulis Joki, or in the case of Castrén, whose fieldwork materials were compiled, edited, and published by his friend Anton Schiefner. Nevertheless, the question of what happens to the original recordings keeps arising. This issue is considered in section 4 below. And as noted above, it would

often be useful to be able to look at the original materials as well, and not just the final product such as a published text collection or grammar.

4. Publishing old fieldwork materials

As already mentioned, the goal of the practice of working on manuscript field notes seems to be to extract the data of the given researcher's interest and publish it. This is what happens with many field note manuscripts. Either the linguistic data is extracted and, for example, a grammar is compiled or only the collected texts are published. However, ignoring manuscript field notes as research objects can lead to misunderstandings and even a loss of quality in any further research based on them. An example may illustrate this: the Finnish scholar Matthias Alexander Castrén conducted several field trips to numerous places in Northern Eurasia in the first half of the 19th century. Passing away at the age of only 38, he left a lot of materials and data from his trips, e.g. detailed notes on the grammar of all Samoyedic languages. In 1854, Anton Schiefner compiled a comparative grammar of the Samoyedic languages on the basis of Castrén's field notes (Castrén 1854). Without denying the merits of Schiefner's work, it must nevertheless be noted that Castrén's field notes are much more detailed and provide much more information (cf. Däbritz unpublished). However, Schiefner's work never aimed at researching the field notes as manuscripts or as written artifacts, instead, he wanted to publish the first results on these languages. With today's technical capabilities, however, it is possible to make the manuscripts accessible and available for the research community and the general audience so they can be explored in themselves as manuscripts.

However, this raises the question, how? It must be borne in mind that the general audience and the specialists sometimes have completely different needs. The options will be explained via two examples. The first is Gerhard Friedrich Müller's collection, the second is Angelina Kuzmina's field materials.

4.1. Müller's materials

Gerhard Friedrich Müller (1705–1783) was a historian and ethnographer who was invited to Sankt Petersburg to co-found the Russian Imperial Academy of Sciences. He not only participated in the Second Kamchatka (also known as Great Siberian) Expedition (1733–1743), but he was the leader of the "scientific group". During the field trip, he described the people he met and collected not only geographical and historical but extensive linguistic materials as well. He has collected words of semantic fields from various languages. These included numerals, religious vocabularies, such as the names of god and devil. Müller used a word list with approximately 240 lexemes. The source language was mostly Latin, sometimes

Russian. The notation of what he used can be considered phonetic. After returning from the expedition, Müller published only some ethnographic materials, among other things the book titled *Report on three pagan peoples, the Cheremis, Chuvash, and Votiak, living in the vicinity of the city of Kazan* (1759). He devoted his time to the study of the history of Siberia. Thus, not only a good portion of his ethnographic descriptions but also his linguistic materials remained unpublished for centuries. From the 1980s onwards his ethnographic records, letters, and other materials were gradually published (cf. Bucher 2002, Hoffmann 1995, Müller 2010, 2018 among other publications; for the Kamchatka Expedition, see Dahlmann 2009). But this cannot be said about the linguistic materials: they remain unpublished still. However, this does not mean that these materials were never seen by anybody. Helimski has dealt with Müller's estate several times. He not only described Müller's manuscripts preserved in the Russian State Archives of Ancient Acts (RGADA)[2] in Moscow and other archive materials from the 18th century (Helimski 1993, 1987) but also used Müller's materials in his several publications. However, Helimski never published the whole material.

Before we consider how these legacy materials could be published or presented, let us take a look at Müller's manuscripts. What materials do we have? How is the data organized, how is it written? Above all, where exactly are they located? This latter question can be answered quite simply. All those who are familiar with the history of research or who know the works on Müller know where these materials are preserved or know in which article or book the corresponding information can be found. As mentioned above, these materials are in the Russian State Archives of Ancient Acts. And here they are located in archival collection 199, archival inventory 2, portfolio 513, files 2 and 7.[3] Finding this collection is not difficult. The archive has an electronic catalog where one can easily find the collection. Figure 2 shows the search results.

---

[2] http://www.rgada.info/
[3] фонд архивного собрания 199, опись архивного фонда 2, портфель 513, дело 2, 7.

Номер фонда: 199 Номер описи: Название фонда: Название описи: Найти

Выводить по 30 результатов

Описи и указатели

Обнаружено 6 записей

№ фонда	Название фонда	№ описи	Название описи	Годы	Кол-во дел	Аннотация	Электронная версия
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	1		1265 - 1783	1477	Материалы (гл. образом в копиях) по истории России XIII - XVIII вв. Копии и выписки из дел Московских и сибирских архивов. Летописи и выписки из них. Сочинения В.Н.Татищева, материалы по истории русского дворянства, русской церкви. Личные и служебные документы А.Д.Меншикова, Б.П.Шереметева, А.И.Остермана, М.И.Воронцова.	Смотреть
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	1m	Машинописная опись.				Смотреть
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	2		1509 - 1783	1185	Ученая, личная и служебная переписка Г.Ф.Миллера. Дела по Академии наук, Московскому университету. Документы (подлинники и копии) местных учреждений Сибири XVII - XVIII вв. Дела об экспедициях и открытиях в Сибири "Сибирская история" Г.Ф.Миллера, описание сибирских народов, городов, словари языков народов Сибири.	Смотреть
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	2m	Машинописная опись.				Смотреть
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	3	Справочная опись "Краткий реестр портфеля историографа Г.Миллера"			199	Смотреть
199	Миллер Герард-Фридрих (1705-1783), историк, член Российской Академии наук, управляющий Московским архивом Коллегии иностранных дел.	4	Справочная опись "Портфели историографа Г.Ф.Миллера"			199	Смотреть

Figure 2. Search results of collection 1994

If we look at these results, we notice that we have not received any results for file 7. This means that the cataloging has not yet been completed. Some manuscripts, however, are already available on the site. If we click on inventory 2, we get to the digitized manuscripts. Unfortunately, the digitized version does not contain portfolio 513. However, for linguistic research, this portfolio would be interesting. The researcher then has to visit the archive.

Let's look at the materials anyway. What can be said about the physical material? We know that at Müller's time, the paper was naturally hand-made, from textile waste (rags). This paper is much thicker than the conventional paper used today and has various sizes. We can assume that Müller used loose pieces of paper, which were bound together with adhesive tape at a later point in time. When exactly this happened is not known, but from its appearance, we can assume that it did not happen in Müller's lifetime but much later. Figure 3 shows a piece of the manuscript. In the middle of the picture, the tape is visible.
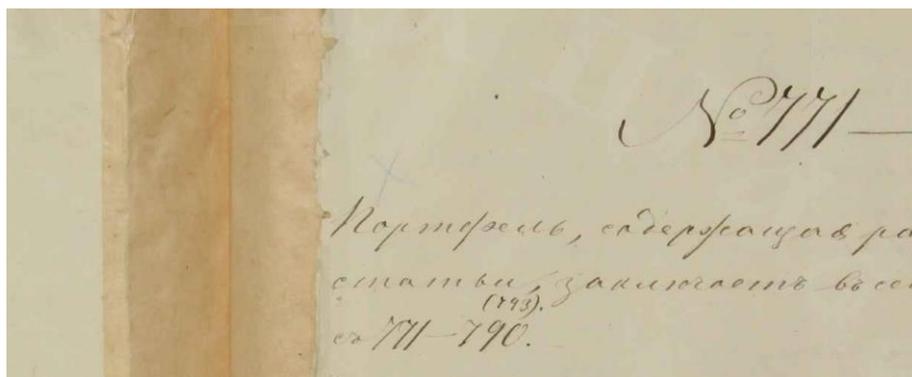
Figure 3. A piece of the manuscript
(Source: http://rgada.info/opisi/199-opis_2/0260.jpg; Accessed on 18 September 2019)

As Helimski already described (1987: 2), specialists know significant parts of these materials. In various works various parts were published, e.g. Helimski published the Mator and Kamas wordlists in the work mentioned above. He not only published the list but also commented on the materials and attached a photocopy of the manuscript. Would it be worth working with these manuscripts again? Most likely, yes. When we look at these manuscripts, we immediately see that some parts were already prepared for publication. The words are already organized by language and semantic fields. If we decide to publish the whole of the manuscript, we would have to clarify in what form it should be published. Should it be in paper form, as in Helimski's work, or in digital form? One graph-based approach to be mentioned in this context that takes into

account both the multi-layered nature and the individual character of field note collections is described by Jettka and Lehmberg (2020). In this case, connections between the semantic categories could be established relatively quickly. To do this, a set of prerequisite sub-objectives is to be accomplished, such as (a) the digitization of the wealth of analog archived materials, (b) the conversion of the digitized material into a searchable electronic database that would allow multiple modes of analysis and information extraction, and (c) the development of the capacity for dynamic visualizations of analytical results. A possible result would be, for example, as illustrated in figure 4. Here the search query was Kamas, and only Kamas items were searched for.[5]
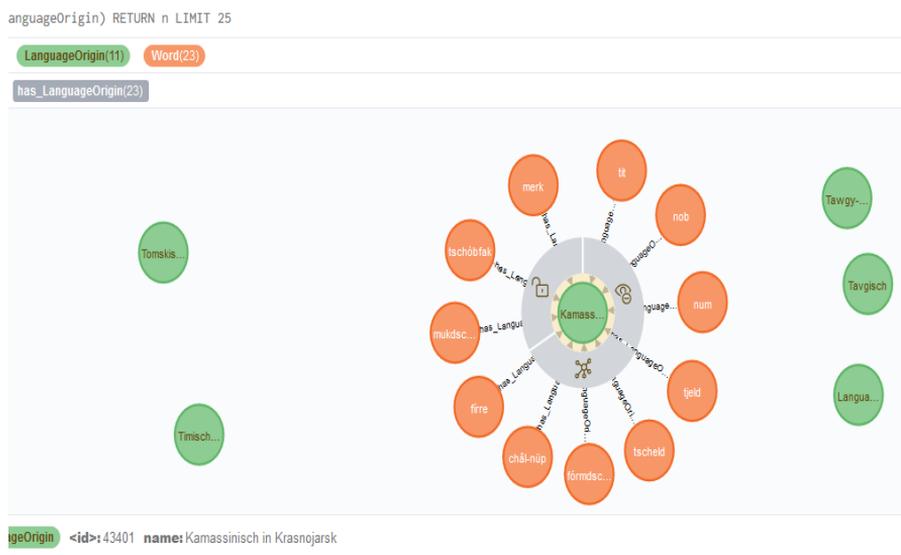
Figure 4. Searching for Kamas

A somewhat more complex search is shown in Figure 5. The search term 'windstorm' (blue ball) was used. The green units (balls) show the languages in the Müller Archive (red ball), the brown ones are the common names. The results show that the searched word is *tschóbfak* in Kamas. The graph also shows the relations. The word comes from Müller's archive, in the same archive, other units can be found, which are also presented and connected to another language. One can also see that there are more Kamas words to be found in Müller's collection.

---

[5] The search results do not represent the entire material, but only show a working version.

Figure 5. Searching for 'windstorm'

Of course, beyond simply searching for lexemes and the links between them, one can also create statistics. If one sorts Müller's language data according to today's names or puts the names together, one can visualize a relationship. This is shown in Figure 6. Note that the figure does not show the actual relation in Müller's manuscript, but the current processing status in the database.[6]

---

[6] The data for graphic search and for the statistics were prepared by Daniel Jettka, for which the authors are very grateful.

Figure 6. Visualization of the statistics of Müller's wordlist

Such a publication of the data shows well the relations between the collected lexemes and allows for a quick generating of statistical analyses. In this format, it is difficult to write, for instance, comments on the individual lemmas. For this reason, it cannot necessarily replace the traditional paper-based publication, but it can complement it. However, the original manuscript is still not accessible.

4.2. Kuzmina's materials

Angelina Kuzmina's fieldwork manuscripts are preserved at the Institute of Finno-Ugric/Uralic Studies[7] in Hamburg. Kuzmina was a student of Dulson, who worked in Tomsk on Siberian languages. She researched the German dialects of Siberia before turning to the study of the Selkup language. Between 1962 and 1969 and later in 1976 and 1977 she visited all regions where the Selkups lived and collected valuable materials. Her estate consists of 35 handwritten volumes, 30 of which are preserved in Hamburg. The volumes contain 357 books, with a total of 8,554 pages. When working with these materials, several details must be taken into account. Although an initial review of the materials took place and a temporary catalog was published (Tuchkova and Helimski 2010), the materials were not fully cataloged. Consequently,

---

[7] https://inel.corpora.uni-hamburg.de/portal/kuzmina (Accessed on 26 November 2020).

cataloging is necessary so that the wider audience and the research community can find out what is in the field notes. A very simple solution for this is a tabular view, but this offers few options. For the first search, however, it is ideal.[8] The second issue, however, is the original question of how to make the manuscripts and content available as research objects. Figure 7 shows the cover of one field notebook.
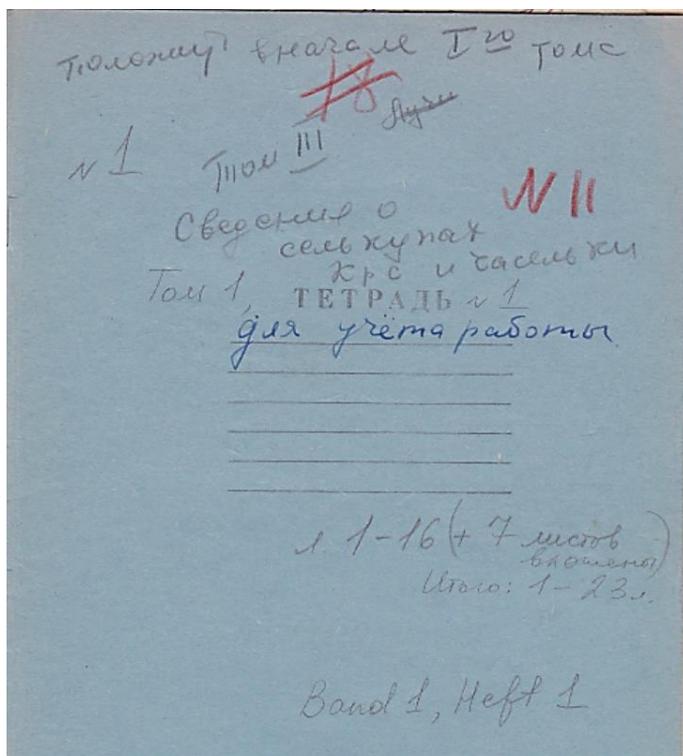
Figure 7. Cover of the first notebook in volume 1

At first glance, one can see that there is a lot of information available here. It is indicated which notebook of which volume this is (notebook 1 in volume 3). There is additional information, which was most likely created later and contains instructions and remarks. It is also visible that a later numbering with a red pencil was also done (Nr 11). And a third numbering occurred even later. This is at the bottom of the cover and was made by Eugen Helimski. It is not only his handwriting that indicates this,

---

[8] See Lehmberg 2020 for the potential that arises from the application of data analysis and visualization routines to a linear structured information resource derived from the Kuzmina Archive.

but that this numbering is in German. Nevertheless, the question arises as to who the red numbering comes from. It would be easy to say that they come from Kuzmina. This red numbering is not only on the cover, but the pages are numbered in some booklets also with a red pencil. However, this holds true only for the booklets that originate from the earlier research trips, but not for those that were written in the late 60s or the 70s: there the numbering is written in grey pencil. Compare the two pages below in Figures 8a and 8b.
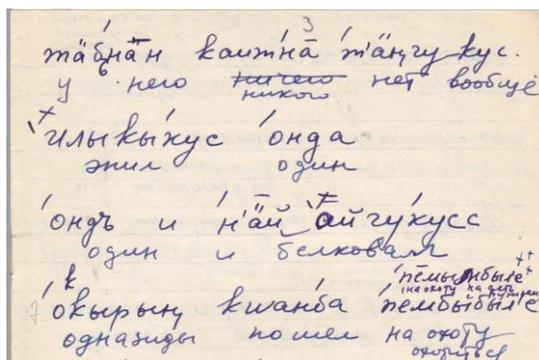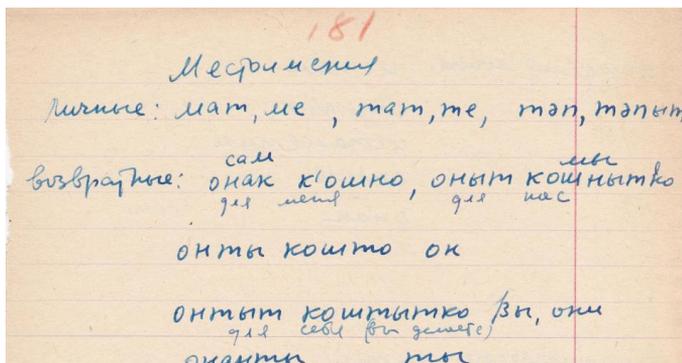
Figures 8a and 8b. Volume 2, book 9, and volume 26, book 1.

We know that Kuzmina left Tomsk in 1964 and moved to Novosibirsk, where she worked until the end of her life. If we open the field research manuscripts which are preserved in Tomsk and were written in Dulson's lifetime, we find exactly such red numbers as in Kuzmina's manuscript; in the notebooks written by Dulson as well as in the notebooks prepared by his students. All these facts indicate that these red numbers were made by Dulson.

As mentioned earlier, the field notes contain not only linguistic data but also other materials, such as metadata, sketches, and maps. This is true of Kuzmina's materials as well, which contain maps, metadata about the speakers, elicited sentences, conjugated verb forms, etc. During the processing of the fieldwork materials, the collected texts are usually published. It can be done as an electronic resource, such as a transcript in a corpus. But it can also simply be published in paper form. During Kuzmina's lifetime, only the latter option existed, so she published some texts in text collections. Regardless of which publication method we choose, some of the information contained in the field notes or which derivable from them remains hidden. The *INEL Selkup Corpus* (Brykina et al. 2020)[9] contains a text (AGS_1964_HeroAndDragon1_flk) which was published by Kuzmina herself. This information is given in the corpus, in the metadata of the text. The information in which volume this text can be found is also given. While editing the text, a transcript is generated and added to the corpus. Even if the INEL Selkup Corpus contains a relatively large amount of information, such as the original Cyrillic spelling, some corrections, etc., some properties remain hidden. The reason is that a corpus or a transcript in a corpus can never reproduce everything from a manuscript. Figure 1 above shows a part of the manuscript. Figure 9 shows the beginning of the text.
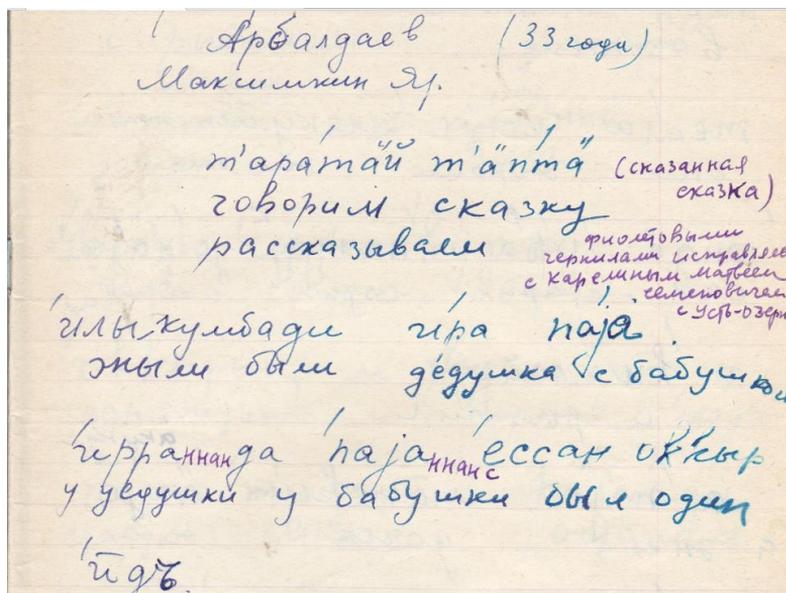
Figure 9. First sentences of the text in the manuscript

---

[9] http://hdl.handle.net/11022/0000-0007-E1D5-A (Accessed on 26 November 2020).

Here we get the information about not only who was recorded providing the text, but we will also explain who made the purple corrections. Kuzmina checked the text with another Selkup speaker. Looking at the metadata, it can be seen that the speakers come from two different villages. The storyteller (Arbaldaev) lived in Maksimin Yar, the other speaker, Karelin Matveev Semenovich, lived in Ust-Ozyornoe. Even if they speak the same dialect (Middle Ket), they have pronunciation variants and possible grammatical differences as well. Right at the beginning of the text, we see an important correction by Karelin. The second sentence is a possessive construction, in which normally the possessor is marked with the adessive suffix, but in the field notes, the possessor is unmarked. Karelin corrected it, of course. Well, two questions must be asked: is the possessor unmarked in the variety spoken by Arbaldaev? In this case, there is a dialect difference, or at least an idiolectal difference. This question can be answered very quickly by examining the other texts supplied by Arbaldaev. In all other texts the possessor was marked by him also with a postposition, so, obviously, here there is a mistake. It is not possible to find out who the error came from. It can be a recording error by Kuzmina, but it can also be an error by Arbaldaev. But the second question is, how should the text be published: with or without this grammatical mistake. From the metadata, one can know that this text was already published by Kuzmina (1967: 294–296, 317–318). The published version is shown in Figure 10.
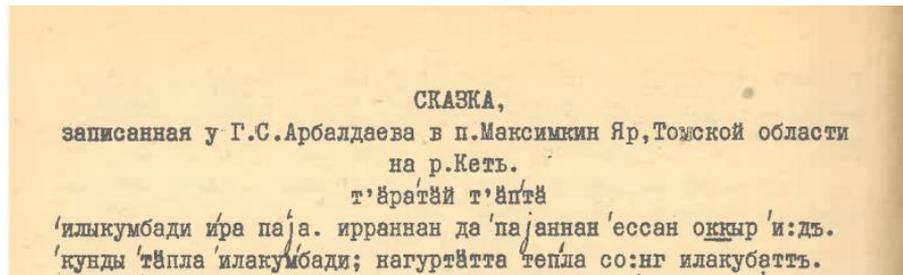
Figure 10. First sentences in the published text

Here we can see that Kuzmina has incorporated the corrections and corrected the ending. Concerning this text, there are two more records among the manuscripts. The second version of the text is in book 4 (88–96) in the same volume. This version was written by Arbaldaev himself, not by Kuzmina. The text is not identical to the previous one (book 2, page 25), but the content is the same. As figure 11 shows, the correction is still missing. Accordingly, another variant must be considered.
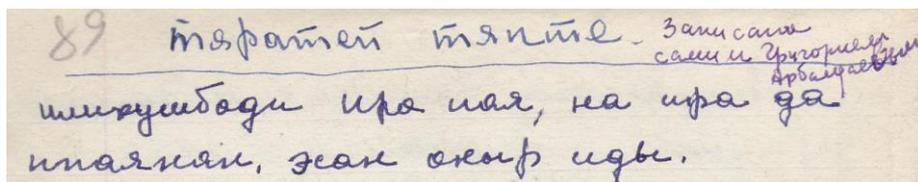
Figure 11. The first sentences of the text written by Arbaldaev

The third variant is an envelope in volume 26, which contains a typed copy of the text from 1964. As figure 12 shows, the correction is already incorporated. So it must have been made between 1963 and 1964. It is not known when exactly, but with a closer examination of the data in the archive, it could be found out. For this, it would be necessary to evaluate when Kuzmina worked with Kalinin. Most of his material is from 1963, but obviously, there was a later point in time when Kuzmina consulted with him. This still requires further research and evaluation of the materials.
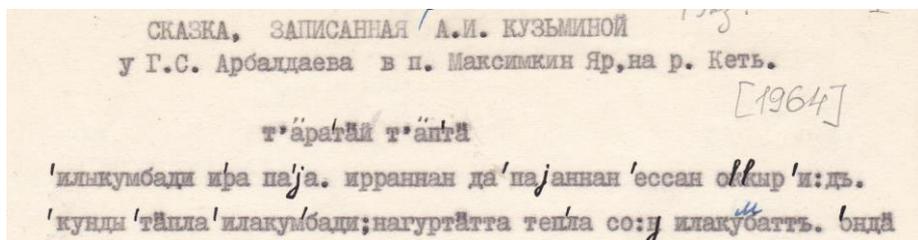
Figure 12. The first sentences of the text in the emvelop

5. Summary

From the examples discussed above, it can be seen that a closer look at the field research materials can provide a lot of information that can explain not only the origin of the texts but also the working methods of the researchers or the speakers they worked with. All this information is part of the text that is likely to be published. For this reason, when working with such archives as the Kuzmina archival materials, one has to think about how to make what is derived from the manuscripts accessible. A good approach is what the project chose: the manuscripts and the published versions of all texts, if they are available, are published. But one can go one step further and build the whole archive catalog in such a way that one links the geodata, the catalog units, and the transcripts that are in the corpus. This would make it easier to find the information. Future approaches that aim at the sustainable (digital) availability of all information contained in these resources will have to find both resource-specific and generic way of information modeling and visualization.

References

Bowern, Claire 2008. *Linguistic fieldwork: A practical guide*. London: Palgrave Macmillan Limited.

Brykina, Maria, Svetlana Orlova, and Beáta Wagner-Nagy. 2020. INEL Selkup Corpus, Version 1.0. Publication date 2020-06-30. Archived in Hamburger Zentrum für Sprachkorpora. http://hdl.handle.net/11022/0000-0007-E1D5-A. In: Wagner-Nagy, Beáta, Alexandre Arkhipov, Anne Ferger, Daniel Jettka, and Timm Lehmberg (eds.). *The INEL corpora of indigenous Northern Eurasian languages.*

Bucher, Gudrun 2002. Von Beschreibung der Sitten und Gebräuche der Völcker: die Instruktionen Gerhard Friedrich Müllers und ihre Bedeutung für die Geschichte der Ethnologie und der Geschichtswissenschaft. Stuttgart: Steiner.

Castrén, M. A. 1854. *Grammatik der samojedischen Sprachen*. Herausgegeben von Anton Schiefner. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Castrén, M. A. 1855. *Wörterverzeichnisse aus den samojedischen Sprachen*. Bearbeitet von Anton Schiefner. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Dahlmann, Dittmar 2009. *Sibirien vom 16. Jahrhundert bis zur Gegenwart*. Paderborn – München – Wien – Zürich: Ferdinand Schöningh.

Däbritz, Chris Lasse. *Manuscript fieldnotes as multilayered written artefacts: An interdisciplinary approach*. Unpublished manuscript.

Helimski, Eugen 1987. Two Mator-Taigi-Karagas vocabularies from the 18th century. *Journal de la Société Finno-Ougrienne,* 81: 49–132.

Helimski, Eugen 1993. Samoyedic vocabularies from the 18. century: A list of archive manuscripts. *Ural-Altaische Jahrbücher: NeueFolge,* 12: 249–265.

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In: Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel (eds.): *Essentials of language documentation*. Berlin: de Gruyter, 1–31.

Hoffmann, Peter (ed.) 1995. *Geographie, Geschichte und Bildungswesen in Rußland und Deutschland im 18. Jahrhundert: Briefwechsel Anton Friedrich Büsching – Gerhard Friedrich Müller 1751 bis 1783.* Berlin: Akademie-Verlag

Jettka, Daniel and Timm Lehmberg 2020. Towards flexible cross-resource exploitation of heterogenous language documentation data. In: *Proceedings of the 12th Conference on Language Resources and Evaluation* (LREC 2020), 2901–2905. available at https://www.aclweb.org/anthology/2020.lrec-1.354.pdf (Accessed on 17 July, 2020)

Kuzmina, Angelina [Кузьмина, А.И.] 1967. Диалектологические материалы по селькупскому языку. In *Исследования по языку и фольклору*. Новосибирск, 267–329.

Lehmberg, Timm 2020. Digitale Edition des Kuzmina Archivs. *Finnisch-Ugrische Mitteilungen* 44: 121–131.

Müller, Gerhard Fridrich 2010. *Ethnographische Schriften* 1. Edited by Wieland Hitzsche. Halle: Verlag der Frankeschen Stiftungen.

Müller, Gerhard Fridrich 2018. *Ethnographische Schriften* 2. Edited by Wieland Hitzsche and Aleksandr Elart. Halle: Verlag der Frankeschen Stiftungen.

Reinmann, Norbert 2004. Grundfragen und Organisation des Archivwesens. In: Reinmann, Norbert (ed). *Praktische Archivkunde*. Münster: Andrey Verlag.

Tuchkova, Natalia and Eugen Helimski [Тучкова, Н.А. and Хелимский, Е.А.] 2010. *О материалах А.И. Кузьминой по селькупскому языку = Über die selkupischen Sprachmaterialien von Angelina I. Kuz'mina*. Institut für Finnougristik/Uralistik der Universität Hamburg.

Woodbury, Anthony C. 2011. Language documentation. In: Austin, Peter K. and Julia Sallabank (eds.): *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press. 159–186.

On the language use of the first Finnish medical text

Meri Juhos

University of Debrecen

1. Introduction

The present paper analyzes the language use of the first Finnish medical text, its sound representations, spelling rules, and foreign origin words in the corpus. It was in 1788 that the first written work of medical subject matter was published in Finnish titled *Maan-Miehen Huone- ja Koti-Aptheeki* [Farmers' home remedies]. The author of this book was Kristfrid Ganander (1741–1790), who made field trips to collect popular medicine practices among the Finnish people. He complemented these with medical formulae earlier published in other languages in medicine books, translating them into Finnish. He also included an index in his book.

Ganander graduated as a theologian and served as a chaplain in Rantsila, near Oulu, from 1755 to his death. He is best known for the first Finnish comprehensive dictionary titled *Nytt Finskt Lexicon* [New Finnish Encyclopaedia] (1787), and for *Mythologia Fennica* (1789). He published eleven works altogether on a variety of subject matters, including collections of folk poetry and folk tales, glossaries, medical writings, and also poems of his own. He also collected riddles and proverbs that he published in the volume titled *Aenigmata Fennica, Suomalaiset Arwotuxet Wastauksen kansa* [Finnish riddles with answers] in 1783. This book contained 337 riddles with answers and was the first of its kind. Although most of Ganander's work is in the humanities, some of his writings cover topics of medical nature, including *Maan-Miehen Huone- ja Koti-Aptheeki* and *Eläinden Tauti-Kirja* [Farm animal diseases], the latter of which is about the diseases of farm animals and their treatment (for more about Ganander's life and work, see Gustaffson 1995: 203–219; Häkkinen 1995: 161–180; Pentikäinen 1995a: 133–160).

To understand the importance of the collection of formulae, it is important to know a little about the medical science of the time.

2. Contemporary medical science

In 18th century Finland, medical treatment was the privilege of a few. Although the

Royal Academy of Turku, founded in 1640, provided an opportunity for the training of medical doctors, this did not mean at all that everybody in need of medical care received sufficient care. The country was divided into eight districts that all had a doctor of its own, and if there was no licensed doctor available, priests and folk healers provided treatment for sick people. People living far from big cities and working in agriculture could not afford to visit doctors.

It is not very surprising, given these circumstances, that the mortality rate of Finland was particularly high compared to neighboring countries. The main reasons for this were widespread poverty, the lack of basic hygiene, and the absence of doctors available at a reasonable distance. There were outbreaks of epidemics that could have been prevented by following basic rules of hygiene, for example, tuberculosis, dysentery, typhoid fever, pneumonia, pox, smallpox, chickenpox and rubeola (Forsius 2005: 81–99; Haarala 1999; Harjula 2007: 16; Ignatius 2000: 508–510).

The appearance of pharmacies in Finland sometime in the 18th century had a positive effect on this situation by providing ingredients not available at home or imported from abroad (Forsius 1995: 36–38).

During his field trips, Ganander became aware of the lack of health care and observed a number of practices used by folk healers. By publishing the descriptions of popular medicine practices he collected, his intention was to provide simple farmers living far away from doctors with a book that was easy to understand and that described formulae using local, easily accessible ingredients.

The structure of the book

Maan-Miehen Huone- ja Koti-Aptheeki contains 101 formulae that describe the preparation, dosage, and availability of ingredients of the given medicine in great detail. The intended audience was the community of Finnish-speaking farmers, so it was evident that the book had to be written in Finnish. However, the scientific and academic language of the time, including medical language, was dominated by Latin and Swedish, which meant that a number of phrases did not have Finnish equivalents, and some of the ingredients were put down by Ganander during his field trips in another language (primarily Swedish). The translation and simplification of the earlier Swedish formulae was probably not an easy task either, because they had to be understandable and still unambiguous and complete for simple people without any medical expertise.

Excerpt from the volume. Formula 1:

Hikiä ajawa Tincturi eli öljy. Ota Walerianan (Ruton juuren) juurta ja putken juurta (Angelika) kumpiain 1 luodi, Liwerstikan juurta 1/2 luodi, leikkaa hienoxi, sekoita; lyö Palowijnaa päälle, sido kalwolla, pane lämpymään poroon, wetäymään, sijwihte

ja pane tallelle. Ota 20 eli 40 tippua – ajaa hikeen ja weren lijkkumaan. On myös matoja ja rewäsintä wasten.

In today's Finnish:

Hiostava tinktuura eli öljy. Ota Valerianan (Ruttojuuren) juurta ja putken juurta (Angelika) kumpaakin 1 luoti, liverstikan juurta 1/2 luotia, leikkaa hienoksi, sekoita, laita paloviinaa päälle, sido kalvoilla, pane lämpimään poroon, vetäytymään, suodata ja pane talteen. Ota 20 tai 40 tippaa – hiostaa ja stimuloi verenkiertoa. Sopii myös matoja ja vatsatautia vastaan.

English translation:

"Sweat tincture or oil. Take some valerian (butterbur) and angelica root, a lot[1] of each, and half a lot of lovage. Chop them up finely, mix them, pour some spirit on the mix, tie it up in a piece of thin leather, put it in hot ash until it shrinks, filter it and put it away. Use 20 or 40 drops a time – it makes you sweat and stimulates blood circulation. It is also good for worms and colic."

3. Vocabulary in the medicine book

The main text is in Finnish, but some Latin and Swedish expressions also occur in the vocabulary. The present study examines medical vocabulary in the book. Medical vocabulary is a collective name for the following: pharmaceutical ingredients; diseases and symptoms; technical tools and methods; adjectives for the description of the consistency or other quality attributes of the given medical preparation; and units of measurement.

Latin expressions

The Latin words of the text are usually names of ingredients. They mostly refer to components that were only available in pharmacies at the time: various types of minerals, resins, and plant parts not available in nature in Finland, e.g.:

> No. 4: *Ota 1 naula Waxia, (...) ja 4 luodia **Storax Calamita*** 'take a pound of wax (…) and four lots of resin' < Lat. *storax* 'resin', *calamita* 'residue'.
>
> No. 5: *ja 2 luodia **Myrrha**, molemmat hienoxi jauhoxi surwotut* 'take two lots of myrrh, both of them ground to fine powder' < Lat. *myrrha* 'myrrh'.
>
> No. 11: ***Camforti Spiritus**, ei ole muuta kuin **Camfortti** sulattu Palowijnasa hywin äkeäxi* 'camphor spirit is camphor dissolved in

---

[1] Translator's note: old unit of weight in many European countries, which equals about 13-14 grams.

spirit in a big concentration' < Lat. *camphorum* 'camphor', *spiritus* 'alcohol'.

In the case of ingredients only available in pharmacies, the Latin name remained unchanged as a loanword. The main reason for this was probably the fact that pharmacists knew these expressions and translation was not necessary.

As far as indigenous species of herbs are concerned, the Latin name was sometimes used to complement the Finnish name, in brackets. This was probably intended to make the identification of the ingredient easier. E.g.:

No. 19: *Oxennus juurta (**ipekacuanha**)* 'ipecac (Ipecacuanha)'.

In some places, the Latin name is primary and the Finnish name is given only for the sake of clarification:

No. 4: *Ota 1 naula Waxia, ja saman werta **Colophoniumia** eli Hartsia* 'take a pound of wax and the same amount of resin'.

It is perfectly reasonable to use multiple names at the same time for the identification of a certain plant, because in the case of collecting herbs from the natural environment, it was very important to be able to recognize similar but poisonous or useless plants compared to the one needed for a given medicine. Still, there are instances in the book where the names given for an herb as synonymous in fact refer to two different plants. This is the case with *waleriana* 'valerian' (Fi. *rohtovirmajuuri*), where the bracketed expression for "clarification" refers to the plant called *butterbur* (Lat. *Petasites*, Fi. *ruttojuuri*).

No. 1: *Ota **Walerianan** (Ruton juuren) juurta* 'Take some valerian (butterbur) root'.

The explanation lies in the fact that folk healers did not necessarily differentiate between different species of herbs, they often used a collective name for certain medicinal plant parts. For example, a name like 'laxative root' could refer to a number of plants that had the same effect, or of which the same part was used as medicine. Thus, it could happen that the healer used different kinds of roots mixed together as if they had been one single plant.

The names of drugs and some of the materials necessary for their preparation are also used in Latin, e.g.:

Preface: *Lääkitysten, **Metikamentein** ja Woidetten nimeä* 'the names of medicine, drugs and ointments' < Lat. *medicamentum* 'drug, medicine' (the Latin loan is already suffixed in Finnish here, in the genitive plural).

No. 1: *Hikiä ajawa **Tincturi** eli öljy* 'Sweat tincture or oil' < Lat. *tinctura* 'alcoholic solution'. This form is also partly Finnicized, ending in an *-i*, characteristic of loanwords in Finnish.

No. 11: *on tämä **Spiritus** hywä saipuan kansa* 'this spirit is good for that, applied together with soap' < Lat. *spiritus* 'spirit, alcohol'.

No. 16: *Että tehdä **Essentia Witae** (Elämän öljyä) ja mihenkä se kelpaa ja on hywä* 'preparation of the water of life, what it is favorable and good for' < Lat. *essentia vitae* 'elixir of life'.

Apparently, Latin expressions refer to ingredients or technical information that could be useful first and foremost for a professional preparing the drug. Therefore, these words are used in the text without (or with only minimal) modification (Finnicization). Body parts or symptoms, however, are never referred to in Latin, because Finnish readers would not have been able to understand the formula then.

Ganander left the spelling of Latin words essentially unchanged. The intended audience did not understand Latin, and those few who did (doctors and pharmacists) were able to read the formula with the original spelling as well. However, it is interesting that certain words in formulae collected from various territories often differ in spelling, e.g. *camphoratus* 'camphor' is the original Latin form, but the variants *camfortti*, *comfortti*, *camfertti*, *camforti* are Finnicized to a certain extent. The fully Latin spelling is more characteristic of formulae that the author adopted from other books of medicine, while the other variants were probably collected and used on the basis of the dialect of the given area.

4. Swedish loan elements

The book contains a number of Swedish loan elements as well. It is hard to assess whether these were still foreign elements at the time of collection in the areas visited by Ganander or had already been integrated into the language as loans. The majority of Swedish words in the text are ingredients and particularly herbs or parts of plants:

No. 2: *ota **Waxia**, ja Suutarin pikiä* 'take some wax and cobbler's pitch' < Sw. *vax* 'wax'.

No. 33: *ota keltaisia **Reenfanan** ja Camillin kukkia, warsista nykityt* 'take some tansy and chamomile flowers, plucked from the plant' < Sw. *renfana* 'tansy'.

No. 17: *Malin-nuppia ja **Lawendeli**-kukkia, kourallinen kumpiain, hakataan hienoxi* 'chop a handful of wormwood buds and a handful of lavender flowers finely' < Sw. *lavendel* 'lavender'.

No. 5: *ja 2 luodia walkiata **Lilje**-öljyä, ja wihdoin 3 luodia (muutamalla palowijna pisaralla) hienoxi jauhettua Camforttia* '2 lots of lily oil

and 3 lots of powdered camphor (mixed with some drops of distilled spirits)' < Sw. *lilja* 'lily'.

The author frequently provides us with the names of herbs in more than one language and sometimes uses both Swedish and Finnish names for the same plant within the same formula:

> No. 7: *pakkaa Wanhoja Tuohia ja **Newa kanerwia (get pors)** isompaan pataan* 'put some old birch bark and heather (ling) in a larger bowl' < Sw. *getpors* 'heather' (literally: 'goat + shrub').

This method had probably been intended to make the identification of the given herb easier. However, in some cases, it turned out to be counterproductive: the bracketed name does not always refer to the same plant as the main expression (as it is the case with the Latin names as well).

It seems probable that after finishing his collecting work, Ganander did not consolidate the various names for the same herb. Thus, if the informant used only a Swedish word for a plant, the Finnish counterpart was not put down next to it, but if the speaker used both Swedish and Finnish expressions, the author recorded both.

Another large group of Swedish-origin words are expressions referring to utensils and methods for preparing drugs, and other words describing certain qualities of a medicine (type, color, time period, and consistency):

> No. 30: *on kuumentawa lääkitys, **pruukataan** ainoastans nijlle, joilla on kinonen ja kylmä ruumis* 'warming medicine applied if one's body is stiff or cold'. The *pruukata* form first occurred in Agricola's work. Its precedent was Sw. *bruka* 'to use, to apply' (SSA 2: 413).

> No. 16: *yhteen korttelijn **Franskt** (...) wijnan sekaan* 'mixed with one kortteli[2] of French spirit' < Sw. *fransk* 'French'; the neuter gender form is *franskt,* which clearly shows that it is still the Swedish word that is used here (SSA 3: 47). (The *kortteli* is a unit of volume, one-quarter of a pint.)

> No. 15: *sixi kuin se tulee **klaarixi*** 'until it is clean'. The first written occurrence of *klaari* in Finnish is from 1638, and from 1731 as a verb *(klaarata)*. It is an adoption of the Swedish adjective *klar* 'bright, clear' (cf. Lat. *clarus*) (SSA 1: 378).

---

[2] Translator's note: old Finnish unit of measurement that equals about 330 ml.

No. 7: *ja se reikä kansi wäliin, sawella **klijsteröity**, tee walkia päälle* '[and insert] the punctured cover plate, seal it with clay and make a fire on it' < Sw. *klister* 'glue', *klistra* 'to glue'.

No. 3: **klasi** *kaataan sitten täyteen wettä* 'fill the bottle with water'. Agricola used the form *clasi*. Today's Finnish form is *lasi.* It is a Swedish loan, cf. Sw. *glas* 'bottle, glass' (SSA 2: 49).

No. 4: *sixi kuin se tulee **pruunixi*** 'until it turns brown' < Sw. *brun* 'brown'.

Apparently, these Swedish loanwords were adopted in everyday language as well, not only in the language of medicine. Ganander probably collected the formulae containing these in areas of strong Swedish influence.

Certain diseases and symptoms are given in Swedish as well, in brackets, for the sake of disambiguation, e.g.:

No. 15: *Lasten nawan puhkemiseen itkusta **(bråck)*** 'When a child's abdominal wall ruptures from crying (hernia)' < Sw. *brock* 'hernia'.

No. 18: *Jolla on (...) ja jalka-tauti **(podager)*** 'those who have a foot disease (podagra, gout)' < Sw. *podager* < Lat. *podagra* 'gout'.

It is generally true of the Swedish expressions occurring in the formulae that the author tried to use a way of spelling that is as close to Finnish as possible. E.g. the characteristically Swedish *b* and *g* sounds are usually substituted with *p* and *k* (*bruka > pruukata, brun > pruun*; *glas > klasi*), and words are affixed in Finnish. All of this shows that these Swedish loans had already been integrated (or were in the process of integration) into Finnish language in the region where Ganander collected this material (cf. Juhos 2019: 109–115).

5. Spelling

Ganander's orthography is rather inconsistent. One of the reasons may be that formulae collected from various places and from a number of different written sources applied different ways of spelling for the same word. Also, variation may result from the fact that the informants spoke several different dialects. Today's standard *-ts-* sound combination for example is also marked with the dialectal *-ht-* (*sijwihte = siivitse* 'to filter', *paihti = paitsi* 'except') and *-tt-* (*watta = vatsa* 'abdomen, belly') in the book. In any case, Ganander did not check and consolidate the spelling of the finished book.

It is a clear development in Ganander's way of spelling compared to Agricola, that the *ä* and *e* sounds show an almost completely regular pattern. Latinized forms are still frequent in his work as well, including letters that are not used in Finnish. This

demonstrates that although Ganander lived and worked almost two hundred years after Agricola, certain sounds and especially sound combinations still did not have a fixed written representation.

Sometimes, Ganander uses the voiced counterpart to mark voiceless stops: *luodi* (in today's Finnish: *luoti* 'lot [measurement]'), *tygö* (*tykö* 'to [suffix]'). After nasals, he sometimes uses *d* and *g*, although very rarely: *pandu* (*pantu* 'put [past tense, 3rd person singular]'), *jonga* (*jonka* 'who [relative pronoun]'), *kuitengin* (*kuitenkin* 'still, however'). But: *kypsentää* 'to become ripe, to mature', *monta* 'a lot of', *parantaa* 'to heal', *henkiwedon* (*henkiveto* 'breath'), *johonka* 'to somewhere'. The *mb* letter combination does not occur anywhere, only *mp*: *isompaan* (< *isompi* 'larger'), *kumpiaan* (< *kumpi* 'which'), *pienempi* 'smaller', *suurempi* 'larger', *ympäri* 'around'.

Similarly to many others before, he always and consistently marks the *ks* sound combination with *x*: *yxi* (*yksi* 'one'), *kaxi* (*kaksi* 'two'); *jauhoxi* (*jauhoksi* < *jauho* 'flour'), *hienoxi* (*hienoksi* < *hieno* 'fine, tiny'), *kowaxi* (*kovaksi* < *kova* 'hard'), *päiväxi* (*päiväksi* < *päivä* 'day'), *pienexi* (*pieneksi* < *pieni* 'small, little') etc.

The sound *v* is only once marked with *v* in the corpus examined (*päivä* 'day'), and in every other case it is marked with *w*, e.g. *ajawa* (*ajava* 'to lead, to carry'), *arwet* (*arvet* 'scars'), *haawa* (*haava* 'wound'), *hywä* (*hyvä* 'good'), *päiwä* (*päivä* 'day'), *kanerwia* (*kanervia* < *kanerva* 'heather'), *kowaxi* (*kovaksi* < *kova* 'hard'), *nawan* (*navan* < *napa* 'navel') etc.

The long *i* (*ii*) is usually marked with *ij*, e.g. *lijna* (*liina* 'fabric'), *tijma* (*tiima* 'hour'), *nijn* (*niin* 'this way'), *sijhen* (*siihen* 'there'), *sijtä* (*sitä* 'that'), *wiinaa* (*viina* 'vodka, spirit'), but rarely the *niin* 'like that', *siihen* 'there', *wiina* 'vodka' forms are used as well (cf. Juhos 2018: 92–94).

6. Morphological issues

The inessive case is very regular in Ganander's work: in one- or two-syllable roots it is *-sa*, e.g. *josa* (in today's standard Finnish: *jossa* < *joka* 'something'), *kiehuesa* (*kiehuessa* < *kiehua* 'to boil'), *taudisa* (*taudissa* < *tauti* 'disease'), *kuumisa* (*kuumiissa* < *kuuma* 'hot'), *wedesä* (*vedessä* < *vesi* 'water'); in longer roots it is *-s*: *ihmisis* (*ihmisissä* < *ihminen* 'person, man'), *hewoisis* (*hevoisissa* < *hevonen* 'horse').

The illative case is marked as follows: *astiaan* (< *astia* 'vessel, bowl'), *isompaan* (< *isompi* 'larger'), *kattilaan* (< *kattila* 'pot'). Ganander uses both the *ij* and *ii* letter combinations in illative forms: *kaikkijn* (*kaikkiin* 'to/for everything'), *kellarijn* (*kellariin* 'to the cellar'), *klasijn* (*lasiin* 'to a bottle, to a glass'), *tautijn* (*tautiin* 'into a disease') – *kaikkiin*, *kuppiin* (< *kuppi* 'cup'), *muottiin* (< *muotti* 'shape'), *paperiin* (< *paperi* 'paper'), *selkiin* (< *selkä* 'back'), *tautiin* (< *tauti* 'disease, ailment').

Genitive plural: *jalkain*, *marjain*, *patain*. There is some variation in today's Finnish as well: next to the *jalkojen*, *marjojen*, *patojen* forms, the variants used by Ganander exist as well (*jalka* 'foot, leg', *marja* 'berry', *pata* 'pot').

7. Summary

In Ganander's time, the book titled *Maan-Miehen Huone- ja Koti-Aptheeki* filled a void long present, since this was the first Finnish language book of medical subject matter. The fact, however, that the book was to be published in Finnish, made it necessary to create medical vocabulary understandable to the Finnish lay person, as a significant part of the medical language was in Swedish and Latin earlier. The base text of the book is Finnish, complemented with some Latin and Swedish origin expressions.

The Latin elements usually refer to resins, minerals and chemicals only available in pharmacies, and also plant parts not accessible in nature in Finland. The author did not find it necessary to translate or Finnicize these expressions, because pharmacists knew this vocabulary anyway.

In the case of Swedish loan elements, it is not certain whether the given expression was still a loan or had completely integrated into everyday Finnish language at the time of collection. It is characteristic of these words that they do not necessarily belong to the vocabulary of medicine and can refer to general attributes like consistency or color. Swedish elements are usually Finnicized in the text, both from the point of sound and letter correspondences and affixation. This also indicates that these loanwords were probably integrated into the Finnish language at the time. It is especially frequent in the case of plant names where only the Swedish name is given, which makes it probable that the informant knew only this one name for the herb. In contrast, body parts and general symptoms (e.g. pain, fever) are all in Finnish.

Spelling cannot be considered as standard and established in the book. This can be attributed to the different places of birth and dialects of the informants and also to the influence of Latin on the orthography of the time. Ganander did not consolidate the use of spelling or the names recorded in several different variants.

Ganander's book represents an important milestone in the development of Finnish medicine. With the detailed listing of the ingredients and the simple and clear way of phrasing, his goal was to enable simple Finnish people to use these healing practices and to facilitate that the collected formulae reach as wide an audience as possible.

References

Forsius, Arno 1995. Lääkärit apteekkareina. *Suomen Lääkärilehti,* 33: 3638.

Forsius, Arno 2005. Kuolinsyyt vuosien 1749–1877 väkilukutauluissa. In: Epidemioista, väestötapahtumista ja kuolinsyistä. *Suomen sukututkimusseuran julkaisuja* 40: 81–97. Pieksämäki.

Ganander, Kristfrid 1788. *Maan-Miehen Huone- ja Koti-Aptheeki*. Rantsila.

Gustaffson, Philip 1995. Kristfrid Ganander – 1700-luvun monipuolinen valistusmies. In: Pentikäinen 1995: 203–219.

Haarala, Risto 1999. Miten lääketieteen suomea huolletaan? *Kielikello* 1999/3. https://www.kielikello.fi/ /miten-laaketieteen-suomea-huolletaan

Häkkinen, Kaisa 1995. Kristfrid Ganander etymologina ja leksikografina. In: Pentikäinen 1995: 161–180.

Harjula, Minna 2007. *Terveyden jäljillä. Suomalainen terveyspolitiikka 1900-luvulla*. Tampere: Tampere University Press.

Ignatius, Jaakko 2000. Lääketieteet. In: Tommila, Päiviö (toim.), *Suomen tieteen historia* 3. Helsinki: Luonnontieteet, lääketieteet ja tekniset tieteet. WSOY.

Juhos, Meri 2018. Kristfrid Ganander orvosságoskönyve (1788) [Kristfrid Ganander's Medical Guide (1788)]. *Folia Uralica Debreceniensia,* 25: 73–98.

Juhos, Meri 2019. Jövevényelemek Kristfrid Ganander orvosi nyelvezetében [Loan elements in Kristfrid Ganander's medical terminology]. *Folia Uralica Debreceniensia,* 26: 107–122.

Lääkäriliitto 2008. *Lääkärien määrän kehitys vuosina 1900–2008.* http://www.laakariliitto.fi/tilastot/laakaritilastot/kehitys.html

Pentikäinen, Juha (toim.) 1995. *Kristfrid Ganander, Mythologia Fennica. Suomalainen Mytologia*. Tampere: Recallmed Oy.

Pentikäinen, Juha 1995a. Kristfrid Ganander mytologina. In: Pentikäinen 1995: 133–160.

Pesonen, Niilo 1973. *Lääkärinä ja virkamiehenä. Lääkintöhallituksen pääjohtaja muistelee*. Helsinki: WSOY.

SSA = Itkonen, Erkki and Kulonen, Ulla-Maija (toim.): *Suomen sanojen alkuperä* 1–3. Helsinki: Suomalaisen Kirjallisuuden Seura, 1992–2000.

Sajnovics, the responsible fieldworker

Sándor Szeverényi

University of Szeged

1. Introduction

The birth and reception of the *Demonstratio* has been already examined from many points of view, but it has largely remained within the linguistic discourse. In particular, recent works (e.g. Aspaas and Kontler 2015 and 2019, Kontler 2011 and 2013, S. Varga 2005) have explored the origins and reception history in a broader context, not only in the context of historical linguistics, but also the political, cultural and social background of the 18th century, and, on the other hand, by analyzing new sources, the genesis story has become more reconstrable (cf. Éder 2014, C. Vladár 2016, 2017, and relevant parts of Aspaas's work cited above); even the history of science is discussed from the astronomical point of view as well (e.g. Hansen and Aspaas 2005, Aspaas 2012).

But, there is still a question that arose at most peripherally regarding the reception of Sajnovics's work, and the simplest way to put it is what Sajnovics (and Miksa Hell) thought about the Lapp/Saami people. We do not know much about Sajnovics's relationship with them, how he worked with them, how many informants he worked with, and how exactly he collected his linguistic data from them. Did he consider his direct, human relationships important, or did he only consider Lapps as his research subjects? In brief, what type of fieldwork did Sajnovics do – if he did any at all? All these questions are hardly answered by Sajnovics and Hell themselves – and that is why there is no exact answers in the literature, but this silence can be telling and suggestive. The researcher inevitably moves to more unfirm ground: s/he is forced to reconstruct.

Why are these questions so relevant? Because 18th century ethnographic, anthropological, and linguistic fieldwork was in its infancy in the mid-18th century (or even non-existent), the description of the circumstances and the evolution of methodology can be traced back to the mid-19th century, and even more so to the last decades of the 19th century.

The afterlife of *Demonstratio* is notoriously noisy, because of the fact that the Lapps are relatives of the Hungarians – although Sajnovics limited the genetic relatedness to the language (or the starting point was the language in his explanation). This northern relationship was very far from the origin of its supposed Hungarian

historical consciousness and folk tradition. We can also say that Sajnovics did not take into account factors outside the language, because he could not do it – even those that might have been considered national characteristics in the mid-18th century.

Sajnovics's work is pioneering in many ways. A forerunner of comparative historical linguistics (Hegedűs 2003: 14), he has long been known and acknowledged in the history of linguistics (e.g. Robins 1999: 189, Campbell 2002: 86, 90, Campbell 2016, most recently Klein et al. 2017: 1, 172). Every pioneering work is characterized by its opposition to the norm of its age: something unexpected, unusual, new is associated with a discovery. For Sajnovics, this meant that in order for the *Demonstratio* to be born, he had to do things (and at the same time ignore many factors) that were unusual by the standards of the age. The most important of these (apart from the narrower methodological observations now) are as follows:

(a) Sajnovics was the first Hungarian who visited an area of a "primitive" or "exotic" people that is not supposed to be a relative of the Hungarians. Before Sajnovics, as far as I know, there was no work in which the author compared language material collected from an indigenous people with his mother tongue for the sake of ignoring their contemporary civilization and culture. Campbell also notes (Campbell 2016: 252) that linguistic fieldwork and the comparative historical method were first combined in Sajnovics' work – he was the first scholar to do fieldwork for linguistic and comparative research.

(b) The aim of Sajnovics's work – as opposed to, for instance, that of the missionary Jesuits and his comrades – was not to map the language but to compare it with his own native language. But his analysis could not be independent of his own identity. Contemporary scholars, missionaries, or those whose work served as antecedent either worked from written materials or by collecting materials "in the field" through learning the language, did not have to deal with defining their own (linguistic or ethnic) relationship with the people they studied.

(c) In Sajnovics's work, a language, a culture, and a people appear to be related to the Hungarians, which in every respect is perceived and appreciated as *foreign* when compared to the Hungarian folk traditions, historical traditions, and in fact, the language. And Sajnovics does not consider it as a problem.

Beyond that, it must be mentioned that these questions and findings cannot be independent of the activity of Maximilian (Miksa) Hell. Recently, more and more works have been and are being published (e.g. Aspaas 2012, Vladár C. 2016, 2017), which prove that Hell played a more significant role in the development of the *Demonstratio*'s concept than had been previously thought. So now, when I mention Sajnovics, Miksa Hell's name could be added almost always, with just a little exaggeration.

2. What, how and from whom did Sajnovics and Hell collect materials?

The names of Sajnovics and Hell appear in hardly any history of anthropological research, which is not a coincidence, since it basically dates from the Enlightenment, and also justifies the fact that their ethnographic collections were not published in the end. Sajnovics, a pioneer in linguistic fieldwork, is hardly ever mentioned. We know that Miksa Hell, in his planned but unpublished three-volume work, *Expeditio Litteraria*, would have devoted a separate section to the ethnographic description of the Lapps (Aspaas published its summary in English in 2012; in Hungarian, see Sajnovics 1994). At the same time, the lack of information about the Lapp people in the *Demonstratio* is in line with the purpose of their work: they focused mainly on the language, not to describe it primarily, but to compare it with Hungarian. But Sajnovics also provides very little information on the methodology and how he proceeded, and only in Chapter 3 does he write about it in a few paragraphs, and in the remainder of the subsequent chapters we can only make inferences about working conditions. According to these remarks, Sajnovics had his own collected language material, but in Chapter 4 he only writes about Porsanger's use of Leem's dictionary, not what happened to his own notes, how extensive they were, and what they were about. In Chapter 3, he also describes how he listened to the Lapps talk. Exactly when, how many times, and under what circumstances he did so is not clear from his diary. He mentions two specific cases there, but one of them was a Karelian speaker he met in Mauersund. The other case, when he came from the mountains with Lapps and asked for words in the presence of Hell and with the help of an interpreter, was essentially that part of the vocabulary that we call basic vocabulary. He also mentions that "I repeated this often", and that there were "speakers of different dialects on the island", meaning that meetings of this kind took place several times. Although he does not elaborate on the details of this, many have already stated that he was also a pioneer in his method (e.g. Stipa 1990: 210).

As I mentioned before, in the history of comparative linguistics in Hungary, there is no activity similar to Sajnovics's. Until the end of the 18th century, language comparisons were typically not based own collected language materials but on the vocabularies and dictionaries collected and published by *other* scholars and travelers. This was applied in particular to the "holy languages", especially the Hebrew language. Before Hell and Sajnovics's expedition in 1768, no one had traveled to the "East" or North to document largely unknown languages with the aim of investigating linguistic "kinship". Furthermore, it is also well known that the Hungarian language was dealt with mostly by foreigners rather than Hungarians (Hegedűs 2003). In the Hungarians' consciousness of their origins, as reflected in Hungarian chronicles,

language plays an inherently central role: it was not even mentioned that the languages of the Hungarians and the Huns are different. Julianus found Hungarians and not the relatives of Hungarians called Hungarians, who spoke different languages and had different lifestyles. The concept of the Caucasus homeland in the mid-18th century was also based on the similarity or identical nature of the languages spoken. The place called "Magyar(vár)" (Madschar) was assumed, which meant that there had to be Hungarians there. But the search for the Caucasian and other eastern homelands did not really begin until a few decades later, in the first half of the 19th century. Thus, attempts to find languages related to Hungarian the language purely until Sajnovics's time were based solely on sources collected and published by others, who were typically not Hungarians. We can mention here Martin Vogel of Germany, Markus Wöldike of Denmark, Philipp von Strahlenberg of Sweden – whose works were also used by Sajnovics.

It is clear from the *Demonstratio* that Sajnovics and Hell also had extensive knowledge of the research into the Lapp language. However, in addition to these passages, Sajnovics hardly mentions anything about the Lapp people in his *Demonstratio*, not even his own personal relations with any of them. He is not very wordy anywhere else either. His diary's description of the period in Lapland is recitant, focusing primarily on his work in astronomy and his local, rather official, human relationships. He rarely writes about the Lapps in his diary (Lakó 1973: 37–38, Kisbán 1942: 34). In his entries of September 28 and October 6, 1768, he mentions only concrete experiences with the Lapp people. Shortly after his arrival in Vardø, he gives an objective description of their appearance, and his later entry similarly describes the living conditions of the Lapps (and/or Finns) (Sajnovics 1990: 67–68, 70). In addition, he gives a brief description of the Lapp people in a few letters. No new sources are mentioned in recent literature – that is, in fact, he writes about very few personal experiences in his writings. In his letter to Miklós Benkő, dated April 5, 1769, he gives the most detailed description of the Lapp people: "I have not yet mentioned the Lapps, let us say a few words about them" (Sajnovics 1990: 223), and then we get some information about the appearance and lifestyle of the Lapps. All this is assessed by future generations as "Sajnovics's interest in the Lapp language [which] arose at the same time as the ethnographic interest in the Lapps" (Lakó 1973: 47), for which, however, no direct evidence is found in *Demonstratio*. Knowing what Sajnovics's task was, the direction of the progression may have been exactly the opposite.

All in all, it is not clear what kind of relationship Sajnovics had with the local Lapps. Anders (Biret-Ánde) Porsanger (1735–1780) was undoubtedly the most important Saami person, but Sajnovics worked with him only on his journey back to Copenhagen. Porsanger was not a typical Saami, in the sense that he was no longer

simply the son of a people: he was a Saami native speaker and the first educated Saami, who studied theology. Porsanger became assistant to Knud Leem – he did not have anyone to follow in his footsteps for a long time. At the same time, Sajnovics was not surprised by this:

> "The memory of the tiring journey was immediately swept away by the sincerest sight that filled him, realizing how close he was to his Lapp people and the Hungarians, though he had long known this excellent belief, and other educated people throughout Denmark and Norway, even from historians. In addition, he exalted the magnificent Divine Providence in his cause, and he alone attributed it, unlike the domestic custom, educated and ultimately capable of taking on this great task and earning immortal merit **in his country**" (D34–35, emphasis added).

He met and worked with Porsanger in the spring of 1770 after returning from the north. According to his diary, they sometimes had lunch together (he mentioned two such events) and once mentioned that Porsanger was ill. Porsanger was also influenced in his work by Sajnovics: later he wrote a Saami grammar himself, although it was not published and the manuscript was lost (Martinussen 1992: 43-59). The question is how much Porsanger's personality influenced Sajnovics' relationship with the Lapps. Little is known about this. Sajnovics's comment "unlike the domestic custom" indicates that Sajnovics was aware that Porsanger had an unusual career path. It is probable, however, that for Sajnovics this was still natural: the notion of nation did not depend on the language and/or the language and culture of the various ethnicities living in the nation, but on the structure of the nation.

At the same time, there is no doubt that Porsanger faced many difficulties in his studies and work until his death, which is basically the result of discrimination against the Saami (Holdvangen et al. 2000: 35, Kelemen in this volume). The question is whether these discrimination phenomena may have been encountered by Sajnovics during their journey, and if so, whether they had been detected and appreciated at all.

In the *Demonstratio*, Sajnovics does not analyze his collected material but rewrites Leem's Danish–Lapp dictionary by adding Hungarian equivalents to it (cf. Campbell 2016: 252). Porsanger also participated in the work on the dictionary as Leem's assistant, and this was the main reason that he became Sajnovics's assistant. Sajnovics in the *Demonstratio* does not give details of what the material he collected from the Lapps was exactly like and what it consisted of. From this point of view, we could even say that because of the language comparison, he did not need to travel so far north and probably would have achieved the same results if he had worked with Porsanger in Copenhagen. Also, the word queries described at the beginning of the

Demonstratio seem more like getting acquainted with the language than a real collection of material.

Although Sajnovics does not say anything about his Lapp materials, he explains why he does not use it, namely, because of the lack of its verifiability (D55, Sajnovics 1994: 59): "in the Elenchus I included no Lappish words that are absent from Leem's *Nomenclator*, and no Hungarian words that are not in the Molnár's Hungarian dictionary either. All this in order to prevent anyone from believing that I am quoting Hungarian or Lappish words in bad faith to increase the catalog." This projects the image of the responsible fieldworker far into the future: he wants to prove a kinship between two languages that can be verified by others, but there was no other way to do this at the time. However, posterity regarded it as a shortcoming, for example, Zsirai (1994: 497) commented as follows: "It is far more regrettable that Sajnovics, by leaving his Hungarian language unexploited and providing opportunities for field studies, limited himself to written sources". What is more typical is that Sajnovics's picture of fieldwork in Vardø – why else would he go there? – proves the kinship, and in Copenhagen th e work is "only" written. This is confirmed by quotes from early literature, such as the following:

> "He came in contact with mountain ranges from a distance of ten miles, via Sajnovics Daass, and asked them about the names of the so-called ancient concepts common to all peoples. This process added valuable elements to his vocabulary. This way, once again, he became possessed of many word similarities. His comparative vocabulary was now quite authoritative, with the help of his mountain, he could continue to observe phonetic, morphological and later syntactic differences. In fact, he has already been through the exhausting work of collecting material. Now he felt he could begin to summarize the results of his mountain research." (Kisbán 1942: 37)

It is quite clear that in reality the causal relationships were different.

3. Sajnovics's method

Based on the *Demonstratio* and Sajnovics's diary, the method of language documentation was as follows:

(1) Comparison of sounds. It is not clear to what extent this meant learning the language and how many notes it involved. As he writes, "while I was listening to their conversation", "we didn't understand the words", but by "sound" they were sure that the two languages were "identical" (D22, 35).

(2) Using Leem's *Nomenclator*. He used Leem's work as a point of reference. Hell sat beside him to help, they browsed through the Lapp words in *Nomenclator*,

matching the Danish and Latin meanings, and checking to find Hungarian words with similar meanings and sounds. Sajnovics writes that this was a very slow and tiring task (Sajnovics 1994: 36).

(3) "Interviews". In addition to this, Sajnovics and Hell constantly gather material from Lapp speakers, but the amount of material and the frequency of interviews can only be guessed. Basically words (and to a lesser extent texts) were queried, thematically grouped, focusing on the basic vocabulary. As mentioned above, their quantity and quality are hard to guess. For example, the remark that "I myself have experienced and Mr LEEM explains in detail that the Lapps of Finnmark have different dialects" (D52) may indicate a very rich, diverse set of direct language skills. This is also referred to in the note "I learned from living Lappish speech" (D40).

There are two notes regarding the size of the material: "And while they are large in number, it will be much larger if I publish the rest of my notes..." and " [e]very time I scroll through my manuscripts, I always find new and new Lappish words" (D80), but "I did not understand the Lapp speakers" (D13).

(4) Compiling *Nomenclator* with Kauriing. He had to make *Nomenclator* self-explanatory, since it was published in Danish, and the Latin index was incomplete.

(5) Using Leem's *Grammar* with Kauriing. Following the glossary, he became familiar with grammar.

(6) Porsanger in Copenhagen, Leem's dictionary. Clarifying all of these, but only comparing them with verifiable – that is, previously published – Lapp materials.

4. The Lapps of the 18th century: Situation, documentation, information

Sajnovics's image of Lapps was also influenced by what kind of knowledge he had in advance. It is clear that the Lapps were a well-known population in Europe by the mid-18th century, and much of the available literature was accessible to Sajnovics and Hell, and he listed it himself in the *Demonstratio*.

The Saami lexicography was fundamentally different from that of other European languages, and it was much more "international" than the lexicography of Finnish (Considine 2017: 169): the most northerly living nation with no cities or universities. Northern Saami literacy began in the mid-17th century. As it is the case with many indigenous peoples, the first texts were made available thanks to missionary activity for the Saami. The earliest Saami texts are from Nicolaus Andreae from 1619 (Piteå, northern Sweden). The first book (*Swenske och Lappeske ABC-book*) was published in 1638, and then again in 1640, followed by the *Manuale Lapponicum* in 1648. In 1643 Johan Tornaeus received from the Swedish state the task of translating Swedish church texts into Saami. His work was based on the Swedish versions of North Saami and aimed to create a standard. Tornaeus was helped by speakers of various Saami

dialects to set this certain standard. The collection of religious texts with nearly a thousand pages was based on the Torneå Saami, of course. A major change took place in the early 18th century, mainly due to the activities of Knud Leem. Most of Leem's works were already known to Sajnovics: Leem published the first grammar in 1748, and the first dictionary in 1756 and 1768. This was the basis of Sajnovics's comparative work. Leem's activities were made possible by the influence of German Pietism. from the early 1700s, and he founded his school in 1716, where he also taught and where later Porsanger became his student.

The first Saami texts (Røros Saami) were written around 1750 in Denmark, with which Norway formed a union at the time. At the same time, it was enacted in 1774 that the language of instruction for Saami children should be Norwegian (Korhonen 1981: 56). The views of Norwegians on the Saami language were reflected in their views on the Saami people and culture. Thus, in the introduction to Fjellström's *Grammar* (1738a), where the Swedish author states that the Saami language is spoken by an unfriendly and uneducated people who have fallen into dialects due to irregular commuting and have lost their regularity with other languages (Fiellström 1738a: 8, cited in Hovdhaugen et al. 2000: 35). Sajnovics used Fjellström's work a lot, but he did not refer to this passage, but quoted strictly phonetic and grammatical descriptions only. It is not clear, however, which of Fjellström's works was published in 1738: a dictionary or a grammar. In the 1994 edition of the *Demonstratio*, it is assumed that he used the dictionary, at least the grammar is not mentioned in the appendix (Sajnovics 1994: 131), but due to the high number of grammatical references, the *Grammar* is also mentioned by Sajnovics (D13, Sajnovics 1994: 29).

Sajnovics might have known the views of other scientists, who, however, were less negative and took other aspects into account. For example, the Swedish Ganandrus / Ganander *Grammar* (1643), also used by Sajnovics, in which the author believed that Saami is a very ancient language, like Hebrew. According to Ganander, Saami, along with Finnish and Estonian, was one of 70 languages created after the confusion of the tongues at the Tower of Babel. Leem, the author of Sajnovics's most important source, also speaks of Saami in his preface to *Grammar* (1748), summarizing what was thought of Saami at that time: "an absurd, wild, and confused language with no rules for vocabulary, inflexion and the like. Others imagine this is a random mixture of many languages" (Hovdhaugen et al. 2000: 35). However, Leem disagreed, saying that Saami was not just a very old language that had long been used, by people "who had the same faith and religion as us, the same God and they serve a king with us, we live in the same kingdom" and it is "an interesting, accurate and rich language" (Leem 1748, *Preface*, quoted by Hovdhaugen et al. 2000: 35).

There is little work reflecting on the 18th century European awareness of the Lapps. There is no doubt that the Lapps, as indigenous native peoples, were

considered as exotic as other native peoples outside Europe. However, due to their geographical proximity, relatively few modern myths were woven around them (cf. Vértes O. 1938). What was the image of Lapland in 18th century Europe? In addition to the ancient authors (Tacitus, Procopius), the works of Swedish Catholic bishop Olaus Magnus (1492–1557), *Carta marina* (1539) and *Historia de gentibus septentrionalibus* (1555), and the influence of Johannes Schefferus' *Lapponica* (1673) made the Lapps known in Europe. Schefferus' work was immediately translated into English (1674), German (1675), French (1678), and Dutch (1682). It should be added that Schefferus himself never went to Lapland. He met most Saami at the University of Uppsala (Rydving 2010, cited by Nordin and Ojala 2015: 117) and in the urban market (Löw 1956: 16, cited by Nordin and Ojala 2015: 117), and it is known that in the 17th century the Saami lived in greater numbers in central Sweden, meaning people did not have to travel to northern Lapland meet them. Schefferus acquired geographic, religious, economic, and cultural knowledge from them. Illustrations of the volume became well-known throughout Europe, defining the image of Lapland in Europe. (Thon (2016) gives an excellent interpretation of 17th-century Saami iconography.)

If we are to draw an 18th century European image of Lapland, even though we go back to the beginnings of modern anthropology, we find few handholds, and Sajnovics makes no mention of general works in this field. In relation to native peoples, "until the birth of anthropology in the nineteenth century, two stereotypes of opposite ends emerged" in philosophical thinking, dating back to the 16th century (Vargyas 2009: 3). Citing Vargyas' paper (2009), one view is related to Hobbes, Hobbes saw the "state of nature" as the rule of brutal violence, "where every man is Enemy to every man", which he called for the eradication of tyranny, the symbol of the biblical Leviathan. "The life of man, solitary, poore, nasty, brutish, and short", and under these circumstances there is no place for diligence, because its fruit is uncertain; there are no means of transporting powerful objects to and fro, no knowledge of the earth, no time, no arts, no literature, no social contact, and worst of all, eternal fear, the danger of violent death (Hobbes 1970 [1651]: 109). Rousseau and, generally, the "good, noble savage" of the French philosophers, argued that "nature made man happy and good, but that society depraves him and makes him miserable", and "born free and everywhere he is on the chain" (Rousseau 1972 [1762]: 6, 22). These idealized "wild people" provided "ammunition" to pillory western society. Whether Sajnovics read or indirectly influenced the work of Rousseau or Voltaire and Hobbes is unknown. Sajnovics was primarily a naturalist and, moreover, a Jesuit, but in the case of Hell, the issue was more complicated because of his active relationship with Jesuit historians. And the French philosophy of history began to have an impact after the birth of the *Demonstratio*, even more so in Hungary, and this will be particularly

important in judging the kinship, of languages and nations especially as a result of the work of the prominent poet and playwright of the Englisghtenment, György Bessenyei. The first wave of the influence of French world history dates from 1760 to 1770, the second wave from 1770s can be hallmarked by Bessenyei's historical works (see e.g. Penke 2000).

Vértes O. wrote in her classic 1938 work about the appearance of Finno-Ugric peoples, including the Lapps, in French literature. As she mentioned, among the Finno-Ugric peoples, the literature on the Lapps was the most extensive, and the interest was general. However, they could not compete with the popularity of non-European peoples in terms of "being fashionable".

The anthropological peculiarity of the Lapps appeared already in the 17th century. The term "race" was applied to human groups by François Bernier, a French physician and traveler. He was the first who visited Poland, Egypt, and India (Vermeulen 2015: 367). In his short article in the *Journal des Sçavans*, he (Bernier 1684: 148) distinguished four or five races: "the first being Europeans, Egyptians, the Hindus, and American Indians; the second is Africans; the third is the Chinese and the Japanese; and the fourth is the Lappish. The classification was geographical and somatic: based not only on skin color, but also on the features of the face, such as the shape of the nose, lips, teeth, and hair. Bernier admitted that he knew almost nothing about the Lapps (his experiences with the African people in the Turkish and Arabic slave markets). Its species classification was a judgment of value" (Vermeulen 2015: 367). This example also shows that the Lapps were known in Europe, even if they had not been seen.

4. Identity and language

At the turn of the 18th and 19th century, György Bessenyei – and many of his contemporaries – simply overlooked the question of language and kinship. Language was not a "sign of nation". The instrumental conception of language was general: the goal was the common good, which could be achieved through the dissemination of science, and the key to the dissemination of science was language, that is, language was a means of achieving the common good and not the nation. This is illustrated very well by, for example, the reception of Joseph II's Language Decree of 1784: in 1784, 37 counties voted for Latin, and only 20 for Hungarian (Soós 2005), for example, Ferenc Kazinczy also supported the decree. Let there be no misunderstanding, at that time Kazinczy was "extremely interested in the Hungarian language as a writer, but not as interested in the language of the nation" (Bíró 2010: 69). So what did one need to do with one's northern relatives to be acceptable? To confer glorious qualities on them. Just as András Dugonics did in his novel *Etelka*. If they are to be depicted in a

negative color, the negative "moral" qualities must be emphasized, as László Perecsenyi Nagy did in 1804 in his work, *Orithia* (see Szeverényi 2002).

Previously, little attention was paid to the role played by the mother tongue and its history in the formation of the national identity. The situation of the Hungarian language and the development of the concept of nation can be traced most closely to the evolution of literature, because "there is also a significant chapter of another great story, the history of the concept of nation" (Bíró 2005: 582). From this point of view, the time of the *Demonstratio* is key. In the second half of the 18th century, earlier Hungarian concepts of community underwent a transformation. In pre-18th century Hungary, three concepts of nation coexisted (e.g. Bíró 2005; Bíró 2010): on the one hand, Hungarians' consciousness as a territorial community concept was linked to the institution of the kingdom, the community of Hungarians, regardless of the language. The *natio Hungariae*, which is organized in the corporate sense of the Order, provides conceptual equality and communication. This was conditional on belonging to the country's privilegedand "representative" political community (*communitas regni*) and "body", but not to language. The order's solidarity with the court was for the most part, and for a relatively long time, stronger than the other bonds. The third concept is language- and culture-based identities. The latter intensifies in the second half of the 18th century, at the expense of the previous two but in fact only appears in the generation following Bessenyei's, relatively suddenly. This is also evident in Kazinczy's change of mind in just ten years: "In 1780 he looked at the language of Prešov/Eperjes with no particular emotion, and was not annoyed by the miserable fate of the Hungarian language, but by the fact that he had no opportunity to practice French. In ten years, however, in 1790, non-native speakers of Hungarian who would not want to learn Hungarian are threatened with hunger" (Bíró 2005: 592). S. Varga's (2005) paradigm division in the evolution of Hungarian literature corresponds to this three-tiered concept of nation: community of traditions, union of states, and community of origin. Bessenyei and his contemporaries, according to the community of origin, reject linguistic affinities, because if the language is added to character traits, it also expresses the national character: "the affinity with the inferior character of the speakers of related languages is inevitable" (S. Varga 2005: 244).

Sajnovics's identity was far from the linguistic-cultural concept, he was a real *Hungarus*. The way he treated the Lapps and the Lappish relatives is entirely consistent with this. So, not surprisingly, the most part of the negative reviews of the Demonstratio was born later, from 1790's.

5. Conclusion

The *Demonstratio* could not be the product of any other time – at least in this form – before the third quarter of the 18th century. If the Venus retreated either sooner or

later before the Sun, it is unlikely that the expedition would have taken place. Certainly not a few decades earlier, because, in the absence of relevant language data, the idea of a northern kinship had not yet arisen. Furthermore, in Josefinism, linguistic-cultural identity was so powerful that it might not be a coincidence that a few decades later, the indigenous and relatives were searched in the East rather than in the North. So we return to the original question: how did Sajnovics look at the Saamis? Let's see what we know:

(1) Sajnovics certainly worked with Saami speakers, typically in the presence of Hell and with the help of an interpreter. For this reason, no deeper personal contact with the locals could be established. The exception was Porsanger, who was not a typical member of the Lapp community. His comments suggest little about this. In his diary, well-known letters, and in the text of the *Demonstratio*, it is more likely that he did not maintain an active relationship with the Lapps. He does not detail his relationship with Porsanger, he sees a European worker, a member of the Danish state and not just a native person. Sajnovics has no subjective comment about the Lapps. This neutrality corresponds to the consciousness of the *Hungarus*.

(2) The only Saami to qualify was Porsanger, who is considered to be the first university educated Saami person and lived in Copenhagen. The work was done on the way back from the expedition and in Copenhagen – when Sajnovics, more or less, no doubt – had his own direct experience of the Saamis.

(3) His method of documentation demonstrates a sense of responsibility.

References

Aspaas, Per Pippin 2008. Maximilian Hell's invitation to Norway. *Comm. in Asteroseismology,* 49: 10–20.

Aspaas, Per Pippin 2012a. Maximilian Hellin ja Johannes Sajnovicsin "Expeditio litteraria ad Polum arcticum" ja suomalais-ugrilaisen kielentutkimuksen synty. In: Osmo Pekonen and Johan Stén, eds. *Lapin tuhat tarinaa*. Ranua: Mäntykustannus 65–88.

Aspaas, Per Pippin 2012b. Maximilianus Hell (1720–1792) and the eighteenth-century transits of Venus. https://munin.uit.no/handle/10037/4178. Date of access: August 15, 2020

Aspaas, Per Pippin and László Kontler 2015. Before and after 1773: Central European Jesuits, the politics of language and discourses of identity in the late eighteenth century Habsburg Monarchy. In: Gábor Almási and Lav Šubarić (eds.), *Latin at the crossroads of identity: The evolution of linguistic nationalism in the Kingdom of Hungary*. Leiden and Boston: Brill, 95–120.

Aspass, Per Pippin and Kontler, László 2019. *Maximilian Hell (1720–92) and the ends of Jesuit science in enlightenment Europe*. Jesuit Studies, Volume 27. Leiden and Boston: Brill

Bíró, Ferenc 2005. Nyelv, „tudományok", nemzet [Language, "sciences", and nation]. *Holmi,* 17: 580–594.

Bernier, François 1684. Nouvelle Division de la terre par les différentes espèces ou races d'hommes qui l'habitent. *Journal des Sçavans*, April 24: 85– 89.

C. Vladár, Zsuzsa 2017. Hell mint nyelvész: A Kar-jelia etimológia és a kínai hasonlítás példája [Hell as a linguist: the Kar-jelia etymology and the example of Chinese comparisons]. In: Forgács, Tamás, Németh, Miklós and Sinkovics, Balázs (eds*.). A nyelvtörténeti kutatások legújabb eredményei IX* [The most recent findings of historical linguistic research, vol. 9]. Szeged: Szegedi Tudományegyetem, Magyar Nyelvészeti Tanszék, 337–350.

C. Vladár, Zsuzsa 2016. Valójában ki a szerzője a Demonstrationak? [Who is the author of *Demonstratio* really?] *Magyar Nyelv*, 112: 316–324. DOI: 10.18349/MagyarNyelv.2016.3.316

Campbell, Lyle 2002. The history of linguistics: In: Aronoff, Mark and Rees-Miller, Janie (eds.). *The handbook of linguistics*. Oxford: Wiley-Blackwell, 81–104.

Campbell, Lyle 2016. Language documentation and historical linguistics. In: Andrea L. Berez-Kroeker, Diane M. Hintz and Carmen Jany (eds.), *Language contact and change in the Americas: Studies in honor of Marianne Mithun*. Studies in Language Companion Series 173. Amsterdam and Philadelphia: John Benjamins, 247–272. DOI: 10.1075/slcs.173.11cam

Considine, John 2017. Small dictionaries and curiosity: Lexicography and fieldwork in post-Medieval Europe. Oxford: Oxford University Press.

Éder, Zoltán 1999/1976. Újabb szempontok a „Demonstratio" hazai fogadtatásának kérdéséhez [New aspects of the reception of *Demonstratio* in Hungary]. In: Éder, Zoltán. *Túl a Duna-tájon* [Beyond the Danube region]. Budapest: Mundus Magyar Egyetemi Kiadó, 47–61. (First edition: Beiträge zum IV. Internationalen Finnougristen kongress II. Istituto Universitario Orientale, Napoli 1975. 3–21.)

Éder, Zoltán 2014. Sajnovics Demonstratiójának első recenziója: „Viennensis recensitor opusculi mei Hafniae editi" [The first review of Sajnovics's *Demonstratio*: "Viennensis recensitor opusculi mei Hafniae editi"]. *Magyar Nyelv,* 110: 85–94.

Fjellström, Petrus 1738. *Dictionarium Sueco-Lapponicum, eller en Orda-Bok*. Stockholm: Literis Joh. Laurentii Horn.

Fjellström, Petrus 1738. *Grammatica Lapponica*. Stockholm: Literis Joh.

Hám Sándor 1889. Sajnovics János élete és Demonstratioja [János Sajnovics's life and *Demonstratio*]. Esztergom.

Hansen, Truls Lynne and Aspaas, Per Pippin 2005. *Maximilian Hell's geomagnetic observations in Norway 1769*. Tromsø Geophysical Observatory Reports. No. 2. Tromsø: University of Tromsø.

Hegedűs, József 2003. *Hiedelem és valóság* [Beliefs and reality]. Budapest: Akadémai Kiadó.

Hobbes, Thomas 1970[1651]. Leviatán, vagy az egyházi és világi állam anyaga, formája és hatalma. Budapest: Helikon Könyvkiadó

Hovdhaugen, Even, Karlsson, Fred, Henriksen, Carol and Sigurd, Bengt 2000. *The history of linguistics in the Nordic countries*. Helsinki: Societas Scientiarum Fennica.

Kelemen, Ivett 2018. Knud Leem, a 18. századi számi lexikográfia úttörője: Vázlatok a Lexicon Lapponicum Bipartitum történetéről [Knud Leem, a pioneer of 18th century Sami lexicography: Notes regarding the history of *Lexicon Lapponicum Bipartitum*]. *FUD* 25: 99–108.

Kisbán, Emil 1942. Tordai és kálózi Sajnovics János, 1733-1785 [János Sajnovics of Torda and Káló z, 1733–1785]. Debrecen: [s.n.]; Budapest: Csáthy.

Klein, Jared, Joseph, Brian, and Fritz, Matthias (eds.) 2017. *Handbook of comparative and historical Indo-European linguistics*. Handbücher zur Sprach- und Kommunikationswissenschaft 41.1. Berlin and Boston: De Gruyter Mouton.

Kontler, László 2011. The Lappon, the Scythian and the Hungarian, or our (former) selves as 'others': Philosophical history in eighteenth-century Hungary. Encountering Otherness. In: Abbatista, Guido ed. *Diversities and transcultural experiences in early modern European culture*. EUT, Edizioni Università di Trieste, 131–145.

Kontler, László 2013. Politicians, patriots and plotters: Unlikely debates occasioned by Maximilian Hell's Venus Transit Expedition of 1769. *The Journal of Astronomical Data*, 19(1): 83–93.

Korhonen, Mikko 1981. *Johdatus lapin kielen historiaan*. Helsinki: SKS.

Lakó, György 1973. *Sajnovics János.* Budapest: Akadémiai Kiadó.

Löw, Bengt 1956. Johannes Schefferus och hans Lapponia. In Johannes Schefferus, *Lappland*. Uppsala: Almqvist & Wiksell, 9–23.

Martinussen, Bente 1992. Anders Porsanger: Teolog og språkforsker fra 1700-tallets Finnmark. *Nordlyd: Tromsø University Working Papers on Language and Linguistics* 18: 15–59.

Nordin, Jonas M. and Ojala, Carl-Gösta 2015. Collecting Sápmi. Early modern collecting of Sámi material culture. *Nordisk Museologi* 2015/2, 114–122.

Penke, Olga 2000. *Filozofikus világtörténetek és történetfilozófiák* [Philosophical histories of the world and philosophies of history]. Budapest: Balassi.

Robins, Henry Robert 1999. *A nyelvészet rövid története* [A short history of linguistics]. Budapest: Osiris – Tinta.

Rousseau, Jean-Jacques. 1972 [1762] A társadalmi szerződés. Bukarest: Kriterion („Téka").

Rydving, Håkan 2010. Samiska överhetspersoner i Sverige-Finland under 1600-talet." In Else Mundal and Håkan Rydving (eds.). Samer som "de andra", samer om "de andra. *Identitet och etnicitet i nordiska kulturmöten*. Umeå: Umeå universitet, 259–265.

S. Varga, Pál 2005. *A nemzeti költészet csarnokai* [Halls of national poetry]. Budapest: Balassi.

Sajnovics, János 1990. *Sajnovics János naplója* 1768–1769–1770 [János Sajnovics's diary, 1768–1769–1770]. Budapest: ELTE.

Sajnovics, János 1994. *Demonstratio*. Budapest: ELTE.

Soós, István 2005. II. József nyelvrendelete és a „hivatalos Magyarország" [Joseph II's decree on language and the "official Hungary"]. In: Bíró, Ferenc (ed.): *Tanulmányok a magyar nyelv ügyének 18. századi történetéből* [Studies on the 18th c. history of the status of the Hungarian language]. Budapest: Argumentum, 261–301.

Storm, Dikka 2009. Network of missionaries and the establishment knowledge: Creating space. In: Charlotte Damm and Janne Saarikivi (eds.) Networks, interaction and emerging identities in Fennoscandia and beyond. Tromsø, Norway, October 13–16 2009. *Suomalais-Ugrilaisen Seuran Toimituksia/ Mémoires de la Société Finno-Ougrienne,* 265: 263–284.

Stipa, Günter 1990. Finnisch-ugrische Sprachforschung. *Suomalais-Ugrilaisen Seuran Toimituksia/ Mémoires de la Société Finno-Ougrienne* 206.

Szeverényi, Sándor 2002. *Perecsenyi Nagy László*. Szombathely: BDTF.

Thon, Jahn Holljen 2016. Knud Leem's illustration of the unfamiliar. *Studia IBB Philologica* 61(3): 143–162.

Vargyas, Gábor 2009. A zsákmányolók képe a kulturális antropológiában [The image of the plunderer in cultural anthropology]. In: Berta Péter, ed. *Ethno-Lore,* 26: 207–238.
http://nti.btk.mta.hu/images/evkonyv/2009/vargyasgabor.pdf

Vermeulen, Hans F. 2015. *Before Boas: The genesis of ethnography and ethnology in the German enlightenment*. Lincoln & London: University of Nebraska Press.

Vértes, O. Auguszta 1938. A finnugorság felfedezése a francia irodalomban [The discovery of Finno-Ugric peoples in French literature]. Budapest: Kókai Lajos kiadása.

Zsirai, Miklós 1937/1994. *Finnugor rokonságunk* [Our Finno-Ugric relations]. Budapest: Trezor Kiadó, 474–582.

The life and work of the Saami theologian and linguist: Anders Porsanger

Ivett Kelemen

University of Debrecen

1. Introduction

The unique life of Norwegian Saami priest Anders Porsanger (1735–1780) and his relation to Hungary has not yet, or only superficially, been discussed in the Hungarian linguistic literature. Porsanger's collaboration with János Sajnovics at Copenhagen was decisive in the career of both: their suggestion of introducing Hungarian orthographic rules and their efforts to transliterate Knud Leem's trilingual dictionary titled *Lexicon Lapponicum Bipartitum* applying Hungarian period-orthography could have brought fame and professional advancement for Porsanger and further (foreign) achievements for Sajnovics (subsequent to the publication of the *Demonstratio*). Thus, it is not surprising at all that the first and most detailed portray of Porsanger comes from Sajnovics:

> "This man was a Lapp, came from Porsanger and studied in Nidrosia (i.e. Trondheim) as a young man. He received excellent education in the field of humanities there. At the University of Copenhagen, he studied philosophy and theology. He spent three years as a missionary among his fellow Lappish people, and then was appointed as a priest at the church of the Trondheim hospital. […] This excellent scholar spoke not only his Lappisch mother tongue but also Danish, German, Latin, Greek and Hebrew, to the glory of the talented Lappish people" (Szíj 1994: 43).

Sajnovics was aware of how "unique" a Saami man Porsanger was, having been raised "in a way different from the ways of his homeland" and having chosen to become a scholar at a time when the majority of Saami people were undereducated or outright illiterate.

2. Historical background

In 18th century Denmark–Norway, ecclesiastical activity was closely intertwined with political interests. Although the Swedish expansion has already ceased by then, in the North the border issue remained unresolved (Kristiansen 2016: 1055–1056).

The church also played a role in the definition of the borders. In addition to the church building efforts, the increasingly popular movement of pietism[1] introduced by Thomas von Westen[2] and the Saami mission (*Finnemisjonen*) established under this movement also provided new ways to reach the Saami people. A crucial point of the mission's strategy was the education of Saami-speaking missionaries. The future missionaries were trained in Trondheim Seminary (*Seminarium Scholasticum*),[3] which belonged to the Cathedral School (*Trondheim katedralskole*).[4] Von Westen's aim was to train Saami-speaking missionaries and Saamis who could help missionary work as school teachers. Some of them indeed became teachers, but nobody started theological studies (unlike in 17th century Sweden). This is why Anders Porsanger is a prominent "exception", the first Saami linguist and priest of Norway.

3. The early years

Anders (Biret-Ánde) Andersen[5] was born in September 1735 in the village of Olderfjord, Porsanger municipality, Finnmark[6] county. He was born to Anders Henriksen (Heandaraga-Ánde 1701–1768) and Berit Mortensdatter (Mortte-Biret, ca. 1695–1768). The local vicar, Danish Ole Hierild (1708–1788), was the first to recognize Porsanger's great talent and started to teach him. He saw such promise in him that he wanted to bring the boy with him to Denmark when he returned home, but his parents, although sensing the importance of this offer, did not agree to this. Thus,

---

[1] Pietism (< Lat. *pieta* 'piety, tenderness') was a revivalist movement established within the Protestant church in the 17th and 18th centuries. It looked upon reformation as an ongoing process instead of seeing it as a past and finished event. Its focus was on the importance of practical Christian life instead of dogmatic theology (Oftestad et al. 2005: 144−148).

[2] Thomas von Westen (1682–1727) was a Norwegian missionary and minister who managed to organize the Saami mission in just a few years and had several chapels and schools built (Steen 1954: 150−197).

[3] This institution (established in 1717) provided Saami linguistic education for future priests and missionaries before they began their studies in Copenhagen. It was also intended to impart Bible knowledge and to train teachers of Norwegian language. Seminarium Scholasticum was closed upon Westen's death.

[4] For more on the school's history, see Eriksen 1897: 168−187.

[5] Andersen assumed the name Porsanger as a sign of respect for his homeland.

[6] The northeasternmost and largest (48,618 km^2) county of Norway with a population of 75,000, it constitutes a part of Lapland, and a quarter of the population are Saami. The most important Saami-populated municipalities are Kautokeino, Karasjok, Tana, Nesseby, and Porsanger. Finnmark was part of the Trondheim Diocese until 1803.

their son stayed in Porsanger and started to work with missionary Hans Hagerup Krogh (1723–1785).[7]

At the request of the College of Missions, Copenhagen[8] the Bishop of Trondheim, Frederik Nannestad[9] established an institute called *Seminarium Lapponicum Fredericianum*[10] in 1752 to promote the Saami linguistic education of Norwegian missionaries (in fact, continuing the tradition of the earlier Seminarium Scholasticum). This new institution was part of the efforts the public authorities made to strengthen the connection of the Saamis with the (Danish-) Norwegian nation state. The king himself provided financial support for its establishment and as Bishop Nannestad also found it important to use the Saami language on Saami territories, the activity of the institution would not be hindered by anything. Knud Leem (1697–1774),[11] the prominent Norwegian linguist of Saami language and culture, was appointed as head of the institution.

4. Studies

Professor Knud Leem was promised funding from the College to hire young "language masters" from Finnmark. Johan Falch (1707–1758), a priest from Talvik, was appointed to choose the most suitable candidate (speaking both Saami and Norwegian), and he chose Porsanger, 17 years of age at the time. The job of the language master was to act as a native interviewee in linguistic studies and to assist in such studies (Kristiansen 2016: 1060). The most prominent argument for Porsanger

---

[7] The mission was in charge of informing the public authorities of conditions at the borders from the beginning. In 1750, Krogh (together with Porsanger) was appointed to define the borders situated furthest from the northern coastline (Kristiansen 2016: 1059).

[8] In 1714, Frederick IV established the College of Missions (*Missionskollegiet*) in Copenhagen to finance and manage Protestant missions under royal patronage. The principal goal of the institution was to organise a mission to India, but from 1716 it also managed the Saami missions.

[9] Frederik Nannestad (1693–1774) was Bishop of the Trondheim Diocese from 1748 (for further reference, see Hansen and Olsen 2006: 305−324).

[10] The institution, in fact, continued the tradition started by the Seminarium Scholasticum, and operated until Leem's death in 1774.

[11] He had already finished his theological studies when, in 1715, his interest turned towards the Saami language and culture. In 1723 he applied to the Missionary College to become a missionary among the Saamis of Finnmark. Around this time, he also began his Saami language studies in order to be able to preach to the Saamis in their own language. He went to Trondheim, where he studied at the Seminarium Scholasticum between 1723 and 1725. From 1725 he served as a missionary in Finnmark for 10 years and accumulated a wide range of knowledge about the Saami language, earning the title of professor for this in 1752 (about Leem, see Kelemen 2018).

was probably the fact that he spoke the same (Porsanger[12]) dialect as Leem, which Leem also used in his work.[13] In 1752, Porsanger arrived in Trondheim to work as a language master, and at the same time he enrolled in the second grade of the Trondheim Cathedral (Latin) School.

Porsanger ended up in the same class as the son of the influential headmaster Gerhard Schøning (1722–1780).[14] This led to a conflict between Bishop Nannestad and the headmaster, since Schøning was of the opinion that Saamis had to learn to read and write in Danish, and that the bishop exceeded his powers by this move. Nevertheless, he had to accept the bishop's decision, but he pulled his son out of the class in a demonstrative manner.[15] Porsanger was an excellent student and received a number of scholarships.[16]

From 1758, Porsanger studied theology at the University of Copenhagen and graduated successfully in 1761. His studies in Copenhagen were financed by the College of Missions. During his stay, in 1760, he and Michael Baade (1739–1799)[17] translated certain parts of the Bible (three epistles of John's) from Danish to Saami upon the request of the college[18] (Hansen and Olsen 2006: 143).

---

[12] It belongs to the eastern subdialects of Northern Saami's Finnmark dialect in Norway (cf. Sammallahti 1998: 11).

[13] *En lappisk Grammatica* (1748)*; En lappisk Nomenclator* (1756)*; Lexicon lapponicum bipartitum* (1768–1781).

[14] The first historian professor in Norway, headmaster of the Cathedral School from 1751.

[15] It is useful to be acquainted with Schøning's idea of the Saamis to understand the nature of the conflict between the bishop and the headmaster. In his 1751 paper (*Forsøg til de nordiske landes, særdeles Norges, gamle Geografi* [Inquiry into the ancient geography of the Nordic countries, particularly that of Norway]), he suggested that the Saamis are in fact immigrants, having arrived and settled in Finnmark county of Norway at a later point of time than the Norwegians. In addition, Schøning also claimed that the Saami people, who lived close to nature, were characterised by a lower cultural level than the Norwegians, which made it necessary to bring them to a higher level, that is, to make them learn to speak and write in Danish (Kristiansen 2016: 1061–1062).

[16] Further proof of his exceptional talent is the fact that the majority of scholarship holders were the sons of well-known priests.

[17] Michael Baade was a Danish missionary, Knud Leem's student. He also translated the first six chapters of the Gospel of Matthew into Saami (Kjølaas 1995: 49)

[18] Porsanger also translated the First Book of Moses and the Acts of the Apostles, but these translations were destroyed. Porsanger himself burned them in anger after not receiving the position of parson of Trondheim Cathedral (cf. Apelseth 2004: 183).

5. Missionary in Varanger

After finishing his studies, Porsanger was commissioned as a missionary to Varanger[19] by the new bishop, Johan Ernst Gunnerus (1758–1773).[20] On his way to East-Finnmark, he arrived in Trondheim at Christmas, 1761, where he got stuck because of foul weather. He was ordained as a priest in the Trondheim Cathedral in April 1762 and continued his way to Varanger later in the same year (Martinussen 1992: 18). Bishop Gunnerus intended to make him stay in Trondheim to help the elderly Leem finish his trilingual dictionary. The Seminary also planned to employ him as a senior lecturer, but no salary would have been paid for this position at the time. Therefore, Gunnerus had to let him continue his journey to the north. In 1762, Porsanger arrived in Varanger and settled down in Nesseby. There is little information about his activity as a missionary there. It can be assumed that Saami language teaching and preaching were important to him. This is supported by a 1764 letter of his to the missionary board about the immediate need for a Saami alphabet book and catechism (Martinussen 1992: 19).

6. Hospital priest and senior lecturer in Trondheim

During Porsanger's stay in Varanger, Bishop Gunnerus (together with Knud Leem) was trying to solve the problem of his employment (which was first and foremost of financial nature). Finally, Porsanger was appointed as a hospital priest[21] in the Trondheim Hospital and as a senior lecturer in the Seminarium Lapponicum (1896: 6). This meant that he worked full time as a hospital priest and also received a salary from the College of Missions as a lecturer. In the same year, he married Anna Catharina Hagerup (?–1780), and they lived together in humble circumstances despite the two salaries. Porsanger had numerous duties in addition to dictionary writing, since he was obliged to undertake further jobs to earn a living. This hindered work on the dictionary, which heightened the tensions between Leem and him, present from the beginning.

Leem taught only a few classes in the Seminary and did little work on the dictionary (as his hands trembled so much by then that even a signature required serious effort on his part; cf. Martinussen 1992: 20–21). The majority of the work was Porsanger's responsibility, still he did not dare complain. The conflict between them reached its peak in 1769–70 when he was summoned to Copenhagen by the College

---

[19] The easternmost municipal unit of Finnmark.

[20] About Biskop Gunnerus, see Hansen and Olsen (2006: 324−358).

[21] Hospital priests were among the lowest ranking members of the church hierarchy, because they worked with poor and sick people and received a small salary if they received any salary at all.

to take a stand in the issue raised by János Sajnovics (and Miksa Hell): whether the Hungarian way of spelling should be adopted in Leem's dictionary (already in press) or not.

7. Collaboration with János Sajnovics

János Sajnovics and Miksa Hell participated in an astronomy expedition to the island of Vardø [1768–1769], where they also started studying the Saami language. Sajnovics had Leem's grammar book and the vocabulary list of the *Nomenclator* as help, but he had great difficulty in using these because of the "strangeness" of their spelling and his lack of Danish knowledge. However, when he listened to how spoken Saami sounded, he concluded that the Hungarian way of spelling would be more suitable for the transcription of the language, because of the linguistic relatedness of the two languages, among other things. Sajnovics and Hell shared this theory with Otto Thott,[22] their patron and current president of the The Royal Danish Academy of Sciences and Letters (*Kongelige Danske Videnskabernes Selskab*), who was enthusiastic about the idea. At Thott's request, Bishop Gunnerus appointed Porsanger as Sajnovics's assistant, and Sajnovics had a good opinion of Porsanger (Solbakk 1998: 13–14).

Porsanger was quick to get acquainted with Hungarian spelling, and in 1770 he and Sajnovics undertook to transcribe the Leem dictionary[23] according to Hungarian orthography, which they finished as early as May 1770.[24] In June, Porsanger submitted his 17-page report[25] to the College of Missions, in which he mostly agreed with Sajnovics's suggestions, but he also complemented them with his own views. Thus, his recommendation on Saami orthography was a different, third version. With this, the conflict between Leem and Porsanger reached its climax, as the professor was against the new way of spelling from the beginning.

8. Fall from grace

Although Porsanger saw the opportunity of professional advancement and fame in the Copenhagen assignment, it turned out quite differently. Leem fiercely opposed the

---

[22] Danish count and statesman. He held several positions at the time, e.g. he was a member of the Privy Council, Highest Chancellor of the Danish Chancery and President of the Royal Danish Academy of Sciences and Letters.

[23] For more on the dictionary, see Kelemen (2018: 99−108).

[24] Sajnovics wrote the following in his Diary on May 15, 1770: "I have presented the Lappish dictionary written with Hungarian spelling to Count Thott" (Szíj 1995: 145).

[25] *Betænkning om den lappiske Orthographies Forandring efter den Ungarske* [Report on the modification of Lappish orthography on the basis of Hungarian spelling].

new spelling throughout their cooperation and probably considered the suggestion as a personal attack. It might be legitimate to consider why Porsanger and not Leem was asked to examine the applicability of Hungarian spelling in the dictionary. The answer probably lies in the fact that a native speaker of Saami was regarded as more authentic as far as linguistic competence was considered than a linguist professor who had learnt Saami "only" as a foreign language. Also, Porsanger might have been appointed for this task as a potential successor of Leem.[26]

Whatever the case may be, Porsanger was a very ambitious person, and he became well-known within church circles outside Trondheim as well. It was reasonable for him to hope for advancement in his career, as during his Trondheim stay he was promised a higher position, the parson of Trondheim Cathedral.

The position became vacant in 1771, and Porsanger officially applied for it, but his application caused great "discontent" in Trondheim, as he clearly had bad reputation by this time. He was considered a selfish, defiant man characterized by rude comments and disrespectful behavior. He was rumored to be drinking and playing cards with ordinary people regularly. The higher circles of the church presumably became aware of Porsanger's ambitions, and the elite could not agree with such a high ranking position being filled by a common man who was a Saami as well. Gunnerus had been staying in Copenhagen, and only after returning home had he confronted the atmosphere around Porsanger. As a consequence, in a letter written on October 25, 1771, Gunnerus also turned against his protégée, asking the king not to approve of Porsanger's appointment on the grounds that he lacked a work ethic and the critical observations described above. Thus, the position was offered to Marcus Frederic Bang[27] (1711–1789) instead of Porsanger. Bang also served as Bishop of the Trondheim Diocese from 1773 to 1787.

9. On a mission again – The last attempt

It became clear that the government of the church did not want Porsanger to serve as a priest in Trondheim. He became the parson of Vadsø in 1772 and served as a priest in East-Finnmark from 1777. In fact, he had been removed from the center of religious authority and lost the chance to restart his career (Wåle 1975: 2).

---

[26] In his 1769 letter to the College of Missions, Bishop Gunnerus requested 200 rigsdaler (the currency in the personal union of Denmark–Norway) for Porsanger until "Mr. Leem's departure" (Dahl 1906: 45). This phrasing supports the theory that Porsanger could have been Leem's successor.

[27] Bang despised the Saami people. He was against the Saami mission and the use of the Saami language.

The successor of Gunnerus, Bishop Bang also played a part in the ending of Porsanger's academic career with his "anti-Saami" ideology. After Leem's death, he made sure that the Seminarium Lapponicum ceased to exist in the same year, 1774. He was strictly against Saami language teaching and insisted that everyone had to learn Danish.

Hoping for a better position and salary, Porsanger and his family set out to Copenhagen in 1780, but their ship ran aground at Risør and the whole family died. Porsanger was only 45 years old (Hanssen 1986: 126).

10. Summary

Kristiansen (2016: 1071) raised an important issue in his paper about Porsanger as to what extent (if at all) Porsanger's Saami origin played a part in his tragedy. It is more likely, however, that the collective impact of several factors caused his work to be fruitless. He had ground-breaking potential, but still could not achieve a leading position, even though he had everything needed for a career: talent, ambition, and, initially, support from the higher circles of the church. He lived in a class-conscious time when it was unacceptable for a common, ordinary man to be appointed as a high ranking member of the church. Although he was a revolutionary of his time, he has not been recognized as a great Saami linguist. This lack of recognition is also due to the fact that all of his work had been consumed by fire: some of his work perished in the great fire of Copenhagen, while his translations were burnt by himself.

After his 1762 ordination, almost 160 years passed before another priest of Saami origin was ordained in the Norwegian church.

References

Apelseth, Arne 2004. Anders Porsanger: eit vekkskusla språktalent; Ein jesuitt, ein sameprest og ein terskelkunnskap. In: Apelseth, Arne, ed. *Den låge danninga : skriftmeistring, diskursintegrering og tekstlege deltakingsformer 1760-1840 Bergen: Universitetet i Bergen, Det historisk-filosofiske fakultetet, Nordisk institutt*, 180−186.

Dahl, Ove 1906. Biskop Gunnerus' virksomhed: fornemmelig som botaniker tilligemed en oversigt over botanikens tilstand i Danmark og Norge indtil hans død. *Det Kongelige Norske Videnskabers Selskabs Skrifter*. No.4. 1–102. Trondhjem: Aktietrykkeriet I Trondhjem.

Eriksen, A. E. 1897. Træk af Trondhjems katedralskoles historie. In: Heggtveit, H.G. ed. *Trondhjem i fortid og nutid 997−1897*. Horten: C. Andersens forlag.

Hansen, Lars Ivar and Olsen, Bjørnar 2006. *Samernas historia fram till 1750*. Stockholm: Liber AB.

Hanssen, Einar Richter 1986. Anders Porsanger. In: *Porsanger bygdebok* 1, 121−126.

Kelemen, Ivett 2018: A 18. századi számi lexikográfia úttörője: Knud Leem. Vázlatok a Lexicon Lapponicum Bipartitum történetéről [Knud Leem, the pioneer of Saami lexicography in the 18th century Tracing the history of Lexicon Lapponicum Bipartitum]. *Folia Uralica Debreceniensia*, 25: 99–108.

Kjølaas, Per Oskar 1995. *Bibelen på samisk: En bok om samisk bibeloversettelse*. Oslo: Det Norske Bibelselskap.

Kristiansen, Roald E. 2016. Den norske kirken og det samiske. In: Lindmark, Daniel and Sundström, Olle, eds. *De historiska relationerna mellan Svenska kyrkan och samerna. En vetenskaplig antologi*. Band 2. Uppsala: Svenska kyrkans forskningsenhet. 1053–1083.

Martinussen, Bente 1992. Anders Porsanger – teolog og språkforsker fra 1770-tallets Finnmark. *Nordlyd* 18: 15–59.

Norsk biografisk leksikon. https://nbl.snl.no/. Date of access: August 20, 2019.

Oftestad, Bernt T., Rasmussen, Tarald, and Schumacher, Jan 2005. *Norsk kirkehistorie*. Oslo: Universitetsforlaget.

Sammallahti, Pekka 1998. *The Saami languages: An introduction*. Kárášjohka: Davvi Girji.

Solbakk, Aage 1998. Anders Porsanger – Biret Ánde. In: Solbakk, John Trygve – Gaski, Harald, ed. *Cafe Boddu: essay-čoakkáldat 3.* Kárášjohka: Davvi Girji, 7–19.

Steen, Adolf 1954. *Samenes kristning og Finnemisjonen til 1888*. Avhandlinger utgitt av Egede Instituttet 5. Oslo: Egede Istituttet 5.

Szíj, Enikő (ed.) 1990. Sajnovics naplója 1768–1769–1770. [The diary of Sajnovics 1768–1769–1770] ELTE. Finnugor Nyelvtudományi Tanszék, Budapest.

Szíj, Enikő (ed.) 1994. *Sajnovics János: Demonstratio. Bizonyítás. A magyar és a lapp nyelv azonos.* [János Sajnovics: Demonstratio. Demonstration. The Hungarian and the Lappish languages are the same]. Bibliotheca Regulyana 2. Budapest: ELTE.

Thrap, Daniel 1896. Finne-Præsten Anders Porsanger. In: *Norsk Folkekalender. Separat af Barmanns Folkekalender.* Kristiania: Oscar Andersen Bogtrykkeri.

Wåle, Helge 1975. Anders Porsanger – den første presten av samisk ætt i Norge. *Nordkalotten*, 10(25/26): 2.

The use and semantics of the Northern Mansi diminutive -riś~rəś[1]

Bernadett Bíró

University of Szeged

1. Introduction

In my paper I discuss the use and the various meanings of the Northern Mansi diminutive suffix *-riś~rəś* and propose a structure for its semantics applying Jurafsky's Radial Category Theory (1996).

I restrict my investigation to the Northern Mansi dialect, which is the only remaining dialect of Mansi[2] and a seriously endangered language with cca. 1000 speakers. The data used for this research are taken from written sources dated between the 1890's and 2019.

In what follows, at first I discuss the notion of the diminutive as well as present the various and often contradictory meanings it can express cross-linguistically (Section 2). Section 3 deals with the Mansi diminutives while in Section 4 the use and the different meanings of the Northern-Mansi diminutive suffix *-riś~rəś* are presented. Finally, after introducing Jurafsky's Radial Category Theory, I propose a structure for the semantics of the diminutive suffix *-riś~rəś* within Jurafsky's theory (Section 5).

2. The diminutive

The diminutive function seems to appear universally, diminutives are used in many languages and they can express a wide range of semantic (and also pragmatic) content cross-linguistically (cf. Jurafsky 1996, Schneider 2013). As Schneider points it out (Schneider 2013: 140) "A standard description of diminutive meaning is that the meaning of the base word is essentially retained, and that the semantic component

---

[2] Mansi is a Uralic/Finno-Ugric language, its closest related languages are Khanty and Hungarian. It is spoken in Western-Siberia, along the river Ob and its tributaries (Keresztes 1998). It is a highly endangered language, almost all of its speakers are Russian-Mansi bilinguals (cf. Bíró and Sipőcz 2006, 2009, Pusztay 2006: 45).

small is added through the diminutive marker. This additional component does not change the meaning of the base word, but merely modifies it." Thus, *cubelets* and *droplets* are still cubes and drops, however smaller than the "normal" or prototypic cubes and drops; and their meanings can be glossed as 'small cubes' and 'small drops'. So this standard description may well apply to cases like those of *cubelet* and *droplet*, but it is not always adequate, let us think of words like *wifelet* and *princelet*, for example. The meanings of these words cannot be glossed as 'small wife' and 'small prince'. Here, instead of the component small, the diminutive adds a negative component to the base word, expressing the negative attitude of the speaker towards the referent (Schneider 2013: 140–141). Thus, beside smallness, the diminutive can express an extreme variety of meanings – as it has been mentioned before –, some of them being even contradictory. Among these meanings are imitation, exactness, approximation, individuation, partitive, intensification, attenuation as well as such pragmatic senses like affection, contempt, playfulness, politeness etc. (cf. Jurafsky 1996). Let us see some examples of these various meanings of the diminutives cross-linguistically:

(1)

imitation:	Mandarin *fo zhur* 'monk's beads' (< *zhu* 'pearl');
exactness:	Mexican Spanish *llegandito* 'immediately after arriving' (< *llegando* 'arriving' < *llegar* 'arrive');
approximation:	Mexican Spanish *altillo* 'tallish, rather tall' (< *alto* 'tall');
individuation/partitive:	Yiddish *dos zemdl* 'grain of sand' (< *der zamd* 'sand');
intensification:	Mexican Spanish *ahorita* 'immediately, right now' (< *ahora* 'now');
attenuation:	Cuban and Dominican Spanish *ahorita* 'soon, in a little while' (< *ahora* 'now');
affection:	*Jimmy;*
playfulness:	Colombian Spanish *monstrico* (< 'monster') (used jokingly by a woman when addressing her husband);
politeness:	Spanish *¿Gusta un **cafecito**?* 'Would you like some coffee?' (< *café* 'coffee');
contempt/derogative:	Cantonese *sek*[22] *nui*[35] 'frigid woman' (< *nui*[25] 'woman').

(The examples are taken from Gladkova 2015: 38, Jurafsky 1996: 534, 536, 548 and Mendoza 2011: 137, 141, 150, 153.)

I will discuss the problems related to the various meanings of the diminutive cross-linguistically in more detail in Section 5. For the purposes of this paper, following

Jurafsky, I define the diminutive as a morphological **device which means at least 'small'** (Jurafsky 1996: 534).

Regarding the formation of diminutives, prototypical diminutives are considered to be created by morphological devices, namely by derivation, cf. e.g. Schneider (2013: 137–138): "Prototypical diminutives, i.e. diminutives generally considered to be the "best" examples of this category, are nouns derived from nouns by attaching a suffix which functions as the diminutive marker (or 'diminutivizer') […]." Beside derivation, however, there are other morphological and also non-morphological devices and processes as well to form diminutives, such as prefixation, reduplication, compounding, truncation, inflection and periphrastic constructions (Štekauer et al. 2012: 237–303, Schneider 2003: 7–10).

3. Diminutives in Northern Mansi

In Mansi two diminutive suffixes are used: *-riś~-rəś* and *-kwe* (*-ke*)[3]. Both diminutive suffixes are very productive and can be likely attached to any word with the exception of conjunctions (Rombandeeva 1973: 76)[4]. These diminutive suffixes occur also in verb conjugation attaching to any verb stem, with the possibility that then voice, tense and mood markers can also attach to them. When attached to verbs, they serve to express subjectivity and speaker's stance. The diminutive *-kwe* (*-ke*) expresses the speaker's positive stance and adds a meaning of affection and politeness to the verb, e.g.

(2)   *toti-ke-n!* (bring-DIM-2SG.IMP) 'Bring some, dear!',

(3)   *toti-ke-m* (bring-DIM-1SG) 'I bring willingly, gladly, cordially'.

While *-riś~rəś* expresses regret and scorn, e.g.

(4)   *tājə-riś-en!* (eat-DIM-2SG.IMP) 'Eat (you poor, pitiable hungry wretch)!',

(5)   *toti-riś-əm* (bring-DIM-1SG) 'I bring, poor me (it is hard for me)'.

        (Examples are taken from Kálmán 1989: 61.)[5]

---

[3] The suffix *-kwe* (*-ke*) has several allomorphs: *-ka, -kä, -kə.* (cf. Riese 2001: 73)

[4] According to Rombandeeva (1973: 77), however, the suffix *-riś/-rəś* can hardly be attached to the names of villages, towns, rivers and lakes.

[5] Munkácsi (1894) and Kálmán (1989) describe these verb forms as a distinct mood in the Mansi verbal system, namely as the "precative" mood. According to other linguists (e.g. Rombandeeva 1973, Rombandeeva and Vahruseva 1989, Riese 2001), these verb forms are considered simple derivative forms. The fact that the verbs containing these derivative suffixes can also take mood markers definitely supports the view that these are "simply" diminutive suffixes and not mood markers. Thus, I agree with those authors who claim that there is no precative mood in Mansi.

In the case of nouns derived with the two diminutive suffixes, the positive and negative meanings are not conveyed necessarily, cf. *piγriś* 'little boy', *piγkwe* 'little boy'.[6]

Beside these two diminutive suffixes, there are so called "diminutive words" in Mansi as well[7], for example:

ōwl 'beginning, end' (e.g. *tuľōwl* 'small finger' < *tuľ* 'finger'),

sup 'piece, part, half' (e.g. *χāpsup* 'small boat' < *χāp* 'boat'),

sow 'fur, leather, skin, bark, crust, skin, dress' (e.g. *pūtsow* 'small kettle/cauldron' < *pūt* 'kettle, cauldron'). (Cf. Riese 2001: 147–149)

4. The Northern Mansi diminutive suffix *-riś/-rəś*

This suffix originates from an independent word probably with the meaning 'piece' which is of Finno-Ugric origin: FU **räćɜ* 'piece, bite, part' (Uralonet, UEW No. 854[8], Riese 2001: 107). This word is no longer traceable in Mansi. The development of this word into a suffix began early, possibly already in Proto-Mansi.[9] (Riese 2001: 107)

The suffix occurs mostly in the form of *-riś* although there are less common variants like *-rəś/-räś*. In the Northern parts it can appear also as *-ľiś*. This diminutive suffix is most common in the Northern dialect and absent from the Southern (or Tawda) dialect. As Riese states it (Riese 2001: 108), in contrast to the other diminutive suffix (*-kwe/-ke*), *-riś/-rəś* often has a pejorative, contemptuous undertone. This semantic development is presumably secondary and has led to a semantic differentiation regarding the use of the two diminutive suffixes. Riese also notices, however, that this differentiation has never gone as far as it is sometimes described, especially in the non-Northern dialects. He also points it out that in case of some words, the diminutive suffix does not have a pejorative meaning at all, e.g. *āγiriś* 'little girl', *piγriś* 'little boy', *kūťuwriś* 'little dog, doggy'. (Cf. Riese 2001: 108)

---

[6] In this paper I examine the semantics and use of only the diminutive suffix *-riś/-rəś*. It is noteworthy, though, that the exact rules guiding the use of the other diminutive suffix need further research, too.

[7] Riese has a more restrictive view considering the diminutive words. Thus, contrary to other researchers he does not consider such words like *sam* 'eye; grain, berry', *lōmt* 'piece, part, stain', *pāl* 'side, area, half etc.', *śiśkwe* 'mother, darling (loved one)', *tārəs* 'hair, vein, thin root' diminutive words because the original meaning of these words are still clearly recognizable in the compounds formed with them. (Cf. Riese 2001: 147)

[8] http://www.uralonet.nytud.hu/eintrag.cgi?id_eintrag=854&locale=hu_HU

[9] This change of meaning is also supported by the diminutive function of the Komi cognate, e.g.: *ńań-ręć* 'a piece of bread'.

(http://www.uralonet.nytud.hu/ eintrag.cgi?id_eintrag=854&locale=hu_HU)

The suffix can express several semantic and also pragmatic contents, as it has been mentioned before. These contents are the following:

a)   'small'
    e.g. *pukiriśe* 'his small belly' (belly-DIM-PX3SG)
    *kwol **sisriśəmnə*** 'to the back of my little house'
    (*sis-riś-əm-nə* house back-DIM-PX1SG-LAT)[10]

b)   'child'/'young'/'offspring'
    āɣiriś 'little girl' (*āɣi* 'girl')
    pāsiɣriś 'reindeer calf' (*pāsiɣ* 'id.')
    *Mān iŋ **māńriśit** ōlsūw.* 'We were still very small/young.'
    (LS 2014/13: 14) (*māń-riś-it* small-DIM-PL)

c)   affection
    apiɣriśəm 'my dear grandson/granddaughter'
    (grandson/granddaughter-DIM-PX1SG)
    *Am saka **āɣiriś** ōńśuŋkwe taχsum.*
    'I wanted to have a daughter very much.'
    (LS 2014/5: 7)

d)   indefiniteness, approximation
    *kwoťliŋ ēt **kwoťəlriśət*** 'around midnight'
    (*kwoťəl-riś-ət* middle-DIM-LOC)
    *taw minəmä **jui-pālriśət*** 'not long after s/he has left'
    (*jui-pāl-riś-ət* back-part-DIM-LOC)[11]
    (cf. Szabó 1904: 77)

e)   The suffix appears also in some lexicalized forms:
    ōpariś nam 'patronymic' (*ōpa, ōpa-riś* 'grandfather' + *nam* 'name')
    ūjriś 'bird' (*ūj* 'animal')
    mātāpriś 'mouse, mole'.

Since the pejorative or negative meaning of the suffix raises several questions, I discuss this subject in more detail. As it has been mentioned before, the pejorative or negative undertone is clearer in the case of verbs. When attached to verbs, the suffix can express regret and pity (and in connection with these feelings, sometimes also

---

[10] In the postpositional phrase sometimes the postposition takes the diminutive suffix instead of the noun.

[11] The postposition *jui-pālt* 'after' originates from the word *jui-pāl* 'back side' (*jui-pālt:* back side-LOC).

sympathy) or scorn and disdain. Regret and pity can be seen in the following examples:

(6)   *wāγtal*   **pat-əm-riś-əm**
        weak   become-PTCP.PST-DIM-PX1SG
        'poor me, I have become weak'
        (Kálmán 1989: 61)

(7)   *jā*   *naŋ*   *kuńər*   *rūpiγtə-nə*   *χum,*   *mēt-χum,*
        well   you   poor   work-PTCP.PRS   man   wage-worker
        naŋ   **χuj-əriś-en***!*[12]
        you   lie-DIM-IMP.2SG
        'Well, you poor, working man, wage-worker, you just lie down!'
        (VNGy IV: 334)

(8)   *ań*   **min-əriś-en**   *jūw!*
        now   go-DIM-IMP.2SG   inside
        'And now, just go inside, poor you!'
        (VNGy IV: 329)

In case of example (8) the field-worker who collected the tale translated this sentence like this: 'Just go inside now, my dear!'. In my opinion both translations can be correct since this sentence is addressed to a poor man who has been beaten up and is in a very bad condition. (In the previous sentence he is addressed as *naŋ kuńer* (you poor) 'poor you'.) Thus, the speaker can feel both regret and sympathy towards this man who is the protagonist of the tale, actually.[13]

In the following examples (9–10) *-riś/-rəś* clearly expresses scorn and disdain:

(9)   *Vōľiŋ*   *nē*   **āmp-ərəś** *woss*   **sus-s-nū-rəś.**
        [proper noun] woman   dog-DIM   let   watch-PST-COND-DIM
        'The woman of Vōľiŋ-village, the dog, just would watch it.'
        (VNGy IV: 46)

---

[12] The suffix *-riś/-rəś* is often attached with an epenthetic vowel (*-ə*) to stems ending with a consonant.

[13] It is noteworthy that in the tale this poor man's belongings are also referred to by using the diminutive suffix *-riś/-rəś,* e.g.: *kwolriś* '[his] wretched little house', *ēntəpriśä* 'his wretched little belt', while otherwise the diminutive suffix *-kwe* is used: *ujkwe* 'little animal', *nājkwe* 'small fire' etc. Here the differentiation between the two diminutive suffixes seems intentional, *-riś/-rəś* has a negative undertone and it expresses bad quality and, in addition, perhaps sympathy and regret as well.

Here the negative, pejorative undertone of the suffix is underlined by its double use, once with a verb and once with a noun. The woman (the woman of Vōľiŋ-village) is referred to as an *āmpərəś* (dog-DIM) and here it is clearly not a nickname or an endearing expression since in the song this woman – who is most probably the singer's wife – is the one who keeps the singer from obtaining the desired and loved other woman (who is referred to as *āɣikwə* (girl-DIM) throughout the song).[14]

Another example for the pejorative use of *-riś/-rəś* is also taken from a faith-song, here the singer speaks about her sisters-in-law who are of a very bad nature, lazy and mean with her:

(10) *tanänəlnə*   *at*   **tēli-riś**   *wārnut*   *wār-uŋkwə.*
      they.DAT   NEG   arise-DIM   work   do-INF
      'They are unable to do any work.'
      (VNGy IV: 58)

In the case of nouns the negative or pejorative undertone of the diminutive suffix is not as dominant as it is in the case of verbs. There are some examples, however where *-riś/-rəś* clearly expresses bad quality or scorn and disdain. This meaning of the suffix is the most obvious in those cases where *-riś/-rəś* and *-kwe* are used in contrast with each-other, the former expressing the negative while the latter expressing the positive content. A good example of this is a Mansi tale, namely *The Tale of the Mōś-Woman and Por-Woman* (*Mōś-nēɣ, por-nēɣ*). In this tale the protagonist (at first the Mōś-Woman and after her death her daughter) is always marked by the diminutive suffix *-kwe* while the antagonist is marked by *-riś/-rəś* (however, she is marked by the suffix only once). Then there is a scene where the daughters of the two women meet and there is a clear difference in the quality and beauty of their clothes consequently marked by the two different diminutive suffixes: *māstər śaχikwet, māstər wājikwet* [the Mōś-Woman's] 'masterly [made] fur-coat, masterly [made] boots'; *matər-mat sūlnəl wārim śaχiriśt, sāltnəl wārim wājriśt* [the Por-Woman's] 'some shabby fur-coat made of bark, shabby boots made of shuck'.

There is another good example of the contrastive use of the two diminutive suffixes in a faith-song where the singer talks scornfully about the man who would like to marry her: *ti māń Xal-pauliŋ χumriś* 'this little man from Xal-village' (VNGy

---

[14] The example is taken from a so-called "faith-song" which is a typical genre of the Mansi folk poetry. In a faith-song the singer speaks about his or her life, about the things which are important to them, and the subject of these faith-songs is often the marriage or the difficulties for men to get a wife or for women to be given to a stranger or to an unkind man as a wife. (In older times Mansis used to maintain a strict exogamy so men had to get a wife from another tribe, from another village. Thus, for women marriage was not always a happy event since it meant that they were taken away from their family and brought to an unfamiliar place.)

IV: 60) while the man who she loves is marked by the diminutive suffix *-kwe*. In another faith-song when the singer talks about herself and her things, she often uses the diminutive suffix *-kwe* while when she talks about that after her death who will replace her, she uses the suffix *-riś/-rəś:*

(11) *ti*   *pasän*   *woipi*   *χūrəm*   **ńōr-kə-m**
      this   table   similar   three   hill-DIM-PX1SG
      ti   *uläs*   *ľūľit*   *χūrəm*   **ur-kə-m**
      this   chair   high   three   mountain-DIM-PX1SG
      χoti   **nē-riś-ən**   *jōma-we-t?*
      what.sort.of   woman-DIM-LAT   walk-PASS-3PL
      'These three dear little hills of mine, similar to tables these three dear little mountains of mine, tall as a chair, what kind of a woman will walk on them?' (VNGy IV: 2)

Beside the contrastive use of the two suffixes, there are also some cases where only *-riś/-rəś* is present and it clearly has a negative meaning, see (12) for example:

(12) *χonal*   *woss*   *joli-oul*   *māń*
      some.day   most   lower.end   small
      pant-kä-nəm   *mat-χūrip*   *sajim*   *χul*
      brother.in.law-DIM-PXPL.1SG   what.kind.of   rotten   fish

      sān-riś   *wiγ-ət;*   *ēriŋ*   *tanki*
      birch.bark.box-DIM   take-3PL   maybe   they.EMP
      as-riś-änl   *ut*   *wērm-eγət.*
      hole-DIM-PX3PL   thing   be.able-3PL
      'Some day my dear youngest (lit. last) brothers-in-law will marry some birch bark box of rotten fish (worthless woman) who perhaps will not be able to [sew] their own clothes (the holes in their clothes).' (VNGy IV: 15) (Note that the word 'brothers-in law' is marked by the diminutive suffix *-kwe* (*-kä*).)

Other examples for the pejorative or scornful meaning of the suffix are when a woman calls her lover's wife *kuľriś* (devil-DIM) (accompanied by a very vulgar adjective) or when a man calls the woman he does not love *āmpərəś* (dog-DIM) (see example (9)) and also when a woman calls her husband who she does not love and wants to leave *χum-piγriś* (man boy-DIM).

In some cases when the suffix has a negative meaning it expresses rather regret and pity or bad quality, not scorn or disdain. See some examples in Footnote 8 and also in example (13) where a woman speaks about herself and her bad life with her husband's sisters and family:

(13) *ti   sāt   χum   **ōńiɣ-riś**   aj-elāl-eɣəm*
      this seven man   sister.in.law-DIM   drink-FREQ-1SG
      '[As] these seven men's (poor) sister-in-law, I am just drinking
      [brandy].'
      (VNGy IV: 59)

In example (13) the diminutive suffix clearly expresses self-pity, similarly to example (14) where the Moś-woman's daughter is moaning after her little brother's death:

(14) ***amť'-ť'e-riś-əm***   *tit*   *jōm-iɣt-ēɣəm*
      me.EMP-DIM-DIM-PX1SG   here   walk-FREQ-1SG
      'Poor me, I am just wondering here, on my own.'

Here both diminutive suffixes (*-ke/-ť'e/*[15] and *-riś*) are added to the word *amki* (*amťi*) 'I myself'. This is not a unique case, there are some other examples of words marked by both diminutive suffixes, e.g. *piɣ-riś-akwē-ɣ* 'two little boys', *āɣi-riś-kwe* 'little girl'. According to Rombandeeva (1973: 78) if both diminutive suffixes are attached to the word then *-kwe* normally follows *-riś/-rəś* and in this case the meaning of both diminutive suffixes is positive. This statement concerning the order of the two diminutives seems to be supported by my examination as well. If the suffixes are attached to the word in the reverse order, however, then they have the following meaning: 'something what we used to like has become a negative thing' (Rombandeeva 1973: 78). In example (14) we can find the reverse order of the diminutives but in my opinion they do not have the meaning proposed by Rombandeeva. It also has to be noted that in the case of personal pronouns almost only the first person pronouns (mostly in singular, i.e. 'I') take the diminutive suffix and they rather take *-riś/-rəś* than *-kwe (-ke)*. (This may due to the fact that one usually talks about oneself with modesty and not with affection.) There are some examples with the second person pronouns as well (again, mostly in singular) but examples containing the third person pronouns are very scarce. The vast majority of the examples thus contain the first person singular personal pronoun and if both the diminutive suffixes are present then the order of them seems to be always *-kwe (-ke)* and *-riś/-rəś* (like in example 14). The reason for this together with the case of personal pronouns marked by the diminutives, however, needs further research as do the possible differences between the different orders of the two diminutives.

---

[15] In the Sygva subdialect of the Northern dialect there is a *ť'* instead of *k* before palatal vowels (*ē, i*), e.g. *ť'ēr* 'iron', *ť'iwər* 'inner part' vs. *kēr, kiwər* in the other subdialects (Kálmán 1989: 12).

In the newer texts there are only a very few examples where *-riś/-rəś* has a negative meaning, one of them is example (15) (taken from the Mansi newspaper) where the meaning of the suffix is 'bad quality':

(15) *Māk   māńśi   ākań   jūnt-uŋkw   χosa,*
      proper   Mansi   doll   sew-INF   long
      *moľaχ   māγəs   mān   kūstər   **ākań-riś-ət***
      quick   for   we   quick   doll-DIM-PL
      wār-s-ūw.
      prepare-PST-1PL
      'It takes long to sew a proper Mansi doll, so in order to be quick, we made quick Mansi dolls.'
      (LS 2015/1: 14)

However, in example (15) the meaning 'bad quality' can be attributed to the diminutive suffix only on the basis of the context, in my opinion. A similar situation appears in a tale (published also in the newspaper), where Nenets people are referred to as *jōrnriśət* (Nenets-DIM-PL). Here only the context tells us that Nenets people act as bad persons in this story (as it is often the case in Mansi tales but not always), thus here the meaning of the diminutive cannot be 'small' or 'dear' or it cannot have any other interpretation than a negative one.

Beside those examples where *-riś/-rəś* has a negative meaning, there are a lot more examples where it definitely expresses a positive content, mostly affection.[16] In (16) I list a few of these examples. (These examples are taken mostly from older texts collected at the end of the 19th and at the beginning of the 20th century since in the newer texts diminutives appear mostly with kinship-terms only.)

(16) *Tūpəl ūsəriśəmnə*   'to my beloved town of Tobolsk',
      Pētərrəśəm   'my dear Peter',
      kwolriś   'dear little house',
      jiwriś   'dear little tree',
      jāmsəkrəśəm   'my dear wagoner',
      āwiriśəm   'my dear little door',
      tōrriśəm   'my dear little kerchief',
      jänriśəm   'my dear little bow',
      māń χumriś   'dear little man',
      nanriś   'you (PL), dear'.

---

[16] Rombandeeva (1973: 77–78) mentions that there are such constructions where the subject is marked by the suffix *-riś/rəś* while the verb is marked by *-kwe*. According to her in these constructions *-riś/rəś* has a positive meaning, too, it expresses pride or respect.

In addition, we can mention the kinship terms which can take either *-riś/-rəś* or *-kwe* for expressing affection (or the meaning 'small'), see some examples in (17):

(17) *āɣiriś*   'dear daughter' or 'little daughter/girl',

  piɣriś   'dear son' or 'little son/boy',

  jaɣriś   'dear father',

  jiɣriś   'dear (younger) sister',

  apiɣriś   'dear grandson/granddaughter', 'dear nephew'

  mańriś   'dear daughter-in-law'.

Concerning the use of the diminutive suffix *-riś/-rəś* in conclusion we can say the followings:

In the case of verbs, the suffix expresses regret and pity (and in connection with these feelings, sometimes also sympathy) or scorn and disdain. When attached to nouns it can have both positive and negative meanings. In older texts it is clear that when the intention was to express scorn and disdain or regret and pity or 'bad quality' then the suffix *-riś/-rəś* was used. The other diminutive suffix (*-kwe*) does not have this function at all. In newer texts, however, there are only a few examples where *-riś/-rəś* has a negative meaning and in these examples the role of the context supporting the negative undertone is much stronger than in the older texts. There are a lot more examples, however, where *-riś/-rəś* has the meaning 'small' or it expresses affection, in older texts it attaches to various base words while in newer texts it is mostly used with kinship terms. (This is also true for the other diminutive suffix.) It is noteworthy that in older texts (tales, songs, riddles etc. collected from native speakers in fieldworks at the end of the 19th and at the beginning of the 20th century) it is clearly observable that some informant most probably favoured *-riś/-rəś* over *-kwe* so they used this diminutive suffix expressing affection (or the meaning 'small') throughout their texts instead of *-kwe*. But in the vast majority of the texts *-kwe* is used far more often than *-riś/-rəś*. The only exceptions are the words *piɣ* 'boy' and *āɣi* 'girl' which are used almost equally with both diminutive suffixes. (The words *piɣriś* and *āɣiriś* may have been lexicalized forms, actually.) In newer texts *-riś/-rəś* appears almost only with kinship terms and cca. 70% of its use consists of the words *piɣriś* and *āɣiriś* (boy/son-DIM, girl/daughter-DIM).

5. The Radial Category Theory

As Jurafsky (1996: 534–535) points it out, the description of the semantics of the diminutive causes several problems both synchronically and diachronically. From a synchronic perspective, it is difficult to explain the varied and often contradictory meanings of the diminutive (like the fact that it can express both intensification and attenuation, for example). The wide scale of meanings expressed by the diminutives

cannot be considered language-specific idiosyncracy because these meanings appear with surprising frequency cross-linguistically. Beside the various semantic contents, the diminutive can express a number of pragmatic senses as well, as it has been mentioned in Section 2. These pragmatic senses usually co-occur with the semantic ones. Diachronically the problem lies in that no consistent proposal for a semantic reconstruction of the diminutive has been made so far, although there are proposed reconstructions for individual languages. These reconstructions lead back the semantics of the diminutive to small/child, approximation, resemblance and various emotional connotations. On the other hand, even if there was a consistent reconstruction, it still would prove difficult to explain the development of the original meaning into the astonishingly varied senses of the diminutive. As Jurafsky states it (1996: 535): "Previously proposed mechanisms of semantic change (metaphor, conventionalized implicature, generalization) can explain the development of some of the senses […] But for some senses, such as 'approximation' or 'exactness', previous methods are insufficient."

In order to solve these problems, Jurafsky (1996) proposes a structured polysemy model based on the radial category (George Lakoff 1987). Radial categories are networks of senses which arise due to the metaphorical conceptualization involved in the structuring of a particular domain or a grammatical category. The central sense in this network is connected to other senses through metaphorical and metonymical extensions, image schematic transfer and inference. The prototypes of senses are represented by nods and the metaphorical extensions, inferences etc. are represented by links. (Jurafsky 1996, Mendoza 2011: 140) "Thus when interpreted as a synchronic object, the radial category describes the motivated relations between senses of a polysemous category. When interpreted as a historical object, the radial category captures the generalizations of various mechanisms of semantic change." (Jurafsky 1996: 542)

Figure 1 shows Jurafsky's (1996: 542) proposal for a universal radial category for the diminutive.
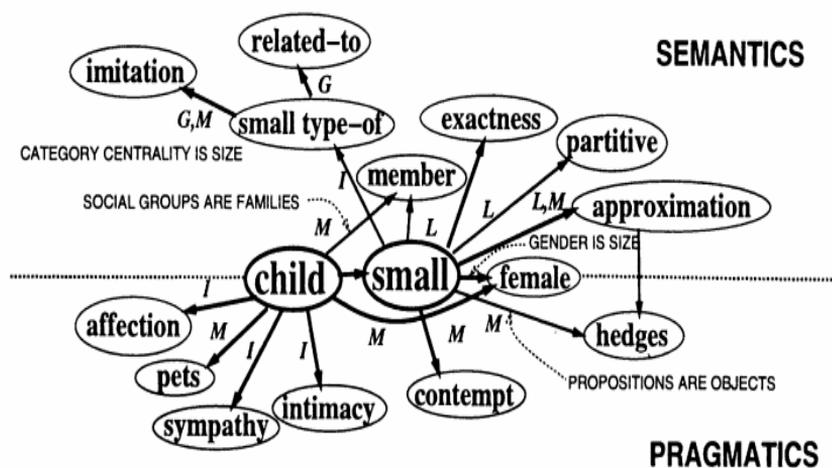
Figure 1. Jurafsky's proposed universal structure for the semantics of the
diminutive (Jurafsky 1996: 542)

As it can be seen in Figure 1, Jurafsky (1996) proposes that the diminutive cross-linguistically originates from words which semantically or pragmatically are linked to children. Other meanings of the diminutive in a given language will develop diachronically from the central senses towards the senses of the edge through the mechanisms of semantic change, i.e. metaphor (M), inference (I), generalization (G)[17] and lambda-abstraction (L) (Jurafsky 1996: 542–543).

Jurafsky introduces a new mechanism for semantic change related to generalization, namely lambda-abstraction-specification. This mechanism creates quantificational and second order meanings from propositional ones by taking a predicate in a form and replacing it with a variable. "For the diminutive, this process takes the original concept 'small(x)', which has the meaning 'smaller than the prototypical exemplar x on the scale of size', and lambda-abstracting it to 'lambda(y)

---

[17] Metaphor: "A meaning shifts to a new domain, based on a general metaphor which maps between the old and new domains. The mapping will preserve certain features of the old domain." (Jurafsky 1996: 544)

Inference: "A morpheme acquires a new meaning which had been an inference or implicature of its old meaning. The historically earlier meaning of a morpheme causes the listener to naturally draw some inference; this inference gradually becomes conventionalized as the literal meaning of the morpheme." (Jurafsky 1996: 544)

Generalization: "A new sense is created from an old one by abstracting away specific features of meaning. The new meaning is more general and less informative than the old one." (Jurafsky 1996: 544)

(smaller than the prototypical exemplar x on the scale y)'." (Jurafsky 1996: 555) The approximation sense of the diminutive, for example, developed from the sense 'small' by lambda-abstraction. Thus, the semantics of the diminutive form *reddish* can be captured as 'dim (point x, scale y) = lower than x on y' where the scale is the scale of redness and the point is the prototype of red. In other words, a reddish object is a marginal member of the category of red objects, and the predicate *red* does not fully apply to it. (Jurafsky 1996: 555, Mendoza 2011)

On the basis of the Radial Category Theory, I propose the following structure for the semantics of the diminutive suffix *-riś~rəś:*
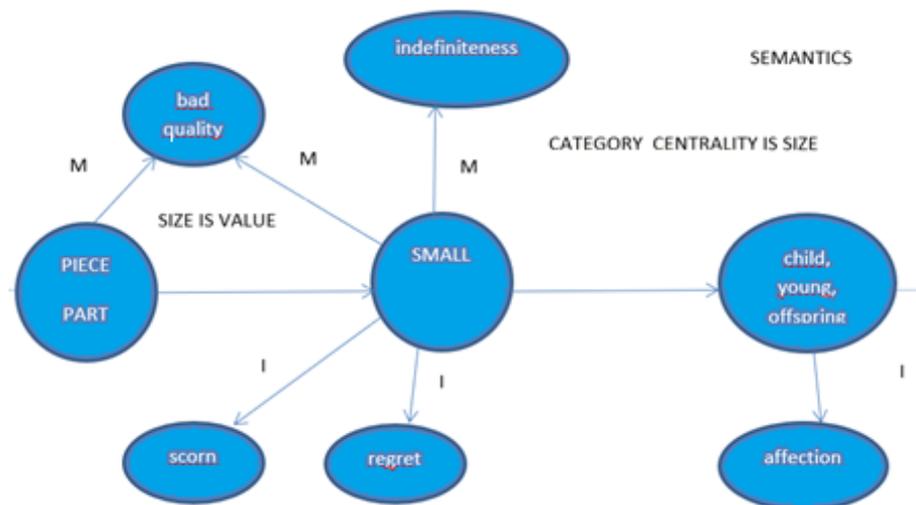
Figure 2. Proposed structure for the semantics of the diminutive *-riś/-rəś* within the Radial Category Theory

The original meaning of the diminutive was 'piece, part' which gave rise to the meaning 'small' and thus the prototypical diminutive function. From the meaning 'small' originate the meanings 'child, young, offspring'. The meaning 'bad quality' can originate both from the meanings 'piece, part' and 'small' through the conceptual metaphor SIZE IS VALUE. (Small things are of a worse quality while bigger things are of a better quality.) The 'indefiniteness' sense could develop from the meaning 'small' via the metaphor CATEGORY CENTRALITY IS SIZE. This metaphor links the marginal members of a category to small size while the central or prototypical members are associated with big size.

Regarding the pragmatic senses, the 'affection' sense developed from the meaning 'child, young, offspring' via inference. It is a natural tendency that we feel affection

towards children. Thus, hearing a diminutive referring to children causes the hearer to draw the natural inference that the speaker feels affection towards the diminutivized object (child). (Cf. Jurafsky 1996: 551) Similarly, the senses 'scorn' and 'regret' both can originate from the meaning 'small' via inference. Towards small objects (which often can be of a bad quality, cf. the metaphor SIZE IS VALUE) we often feel scorn and/or regret thus the hearer can again draw the inference that the speaker feels scorn and/or regret towards the person or object marked by the diminutive.[18]

Abbreviations

COND	conditional
DAT	dative
DIM	diminutive
EMP	emphasis
FREQ	frequentative
IMP	imperative
INF	infinitive
KM	Konda Mansi dialect
LAT	lative
LOC	locative
NEG	negative
PASS	passive
PM	Pelym Mansi dialect
PL	plural
PST	past
PTCP.PRS	present participle
PTCP.PST	past participle
PX	possessive suffix
SG	singular
ULM	Upper Lozva Mansi dialect

---

[18] It must be mentioned that according to Rombandeeva (1973: 78–79) the diminutive suffix *-riś/rəś* originates from the Mansi word *rus* [*ris*] 'weak, fragile, soft' (a Komi loanword, cf. Munkácsi and Kálmán 1986), cf. *ris jiw* 'a weak tree', *ris* 'something weak'. I consider Riese's and the Uralonet's etymology more convincing, in the case of the other etymology there are phonetic problems as well. (How and why has changed the original *-s* into a palatalized *-ś*?) But if the original meaning of the suffix was 'weak, fragile, soft' then the meanings 'small' and 'bad quality' could develop from this meaning as well.

References

Bíró Bernadett and Sipőcz Katalin 2006. Manysiul ma? *Nyelvtudomány II. Új folyam*. XLIII: 21–35. Szeged.

Bíró Bernadett and Sipőcz Katalin 2009. Language shift among the Mansi. In: Stanford, James N. and Preston, Dennis R. (eds.), *Variation in Indigenous Minority Languages*. Amsterdam and Philadelphia: John Benjamins Publishing Company. 321–346.

Gladkova, Anna 2015. Ethnosyntax. In: Farzad Sharifian (ed.), *The Routledge Handbook of Language and Culture*. Abingdon – New York: Routledge.

Jurafsky, Daniel 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72(3): 533–578.

Kálmán, Béla 1989. *Chrestomathia Vogulica*. (3rd edition) Budapest: Tankönyvkiadó.

Keresztes, László 1998. Mansi. In: Abondolo, Daniel (ed.), *The Uralic Languages*. London: Routledge. 387–427.

Lakoff, George 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

LS = *Luima Seripos* (Mansi newspaper). Khanty-Mansiysk. http://www.khanty-yasang.ru/luima-seripos/archive

Mendoza, Martha 2011. Size matters: grammaticalization, metaphor, and the Spanish diminutive. Estudios de Lingüística Aplicada, año 29, número 54, diciembre de 2011, 135–157.

Munkácsi, Bernát 1894. *A vogul nyelvjárások szóragozásukban ismertetve*. [Declension of the Mansi dialects.] Budapest: MTA.

Munkácsi, Bernát and Kálmán, Béla 1986. *Wogulisches Wörterbuch*. Akadémiai Kiadó: Budapest.

Pusztay János 2006. *Nyelvével hal a nemzet. Az oroszországi finnugor népek jelene és jövője 11 pontban*. Budapest: Teleki László Alapítvány.

Riese, Timothy 2001. *Historische Nominalderivation des Wogulischen*. Wiesbaden: Harrassowitz Verlag.

Rombandeeva, E. I. 1973. *Mansijskij (vogulskij) jazyk*. [Мансийский (вогульский) язык.] Москва: Издательство «Наука».

Rombandeeva, E. I. and Vahruseva, M. P. 1989. *Mansijskij jazyk*. [Мансийский язык.] Ленинград: Просвещение Ленинградское отделение.

Schneider, Klaus P 2003. *Diminutives in English*. Tubingen: Max Niemeyer Verlag.

Schneider, Klaus P. 2013. The truth about diminutives, and how we can find it: Some theoretical and methodological considerations. *SKASE Journal of Theoretical Linguistics* [online]. 2013, vol. 10, no. 1 [cit. 2013-02-04]. http://www.skase.sk/Volumes/JTL22/pdf_doc/08.pdf

Štekauer, Pavol,Valera, Salvador, and Körtvélyessy, Lívia 2012. *Word-formation in the world's languages: a typological survey.* Cambridge: Cambridge University Press.

Szabó, Dezső 1904. A vogul szóképzés. [Derivation in Mansi.] *Nyelvtudományi Közlemények* 34: 417–443.

Uralonet (Uralic Etymological Database). http://www.uralonet.nytud.hu/ (Accessed 24.6.2020.)

VNGy = Munkácsi Bernát 1892–1896. *Vogul Népköltési Gyűjtemény* I–IV. [A Collection of Vogul (Mansi) Folkloric texts.] Budapest: MTA.

The event of "giving" and "getting" in Siberian Uralic languages[1]

Katalin Sipőcz

University of Szeged

1. Introduction

This paper investigates the linguistic expression of the event of "giving" and "getting" in Siberian Uralic languages. The description is based on more aspects, it applies the criteria of language typology on the one hand, and also takes into account aspects of cognitive linguistics and cultural linguistics on the other hand. As we will see, these approaches partly complement and partly explain each other. Data from the Mansi, Khanty, and Nganasan languages are included in this paper. All three languages belong to the Uralic language family, Mansi and Khanty represent the Ob-Ugric languages of the Finno-Ugric branch, they are closely related languages, while Nganasan is one of the Northern Samoyedic languages. The Ob-Ugric and Samoyedic languages are distant relatives, but areally they all belong to the Western Siberian language area. The language data discussed in the paper are taken from digital databases and partly from native informants (Brykina et al. 2018, TDUL, DDML).

The three participants of the event of giving are the giver, the recipient and the thing transferred, its most typical verb is the verb 'give'. In this paper I investigate the recipient side of the event, too, thus the event of getting is also included. These two processes complement each other, there is no giving without getting and vice versa.[2] And the two events are connected also by the fact that they are expressed by the same verb – the verb 'give' – in numerous languages (see sections 2 and 3). Both events are characterised by the same Thematic Roles: Agent (A), Recipient (R), and Theme (T). The most common verbs of the events are 'give' ang 'get', but they can be expressed by other verbs, too (e.g. 'pass', 'hand', 'present', 'sell'; 'accept', 'receive' etc.). This paper investigates only the verbs meaning 'give' and 'get', which are basic verbs in this lexical group (see section 3). Cf.:

---

[2] Research focusing on the event of giving and the verb 'give' is significantly wider. The event of getting and the verb 'get' received much less attention in the linguistics literature.

(1)   *Mary*   *gives*   *a book*   *to John.*
         A   T   R

         John   *gets*   *a book*   *from Mary.*
         R   T   (Source = A)

2. The aspect of syntactic typology

In linguistic typology, ditransitive constructions and their arguments have recently become a popular research topic. Ditransitive verbs are three-argument verbs which typically express physical transfer (e.g. *give*, *sell*, *bring* etc.), but also other, semantically close verbs can behave syntactically in a similar way. In this case they are also listed as ditransitive verbs, e.g. verbs of communication expressing mental transfer between the speaker and the addressee (e.g. *say*, *tell*, *sing*, *show*), verbs of caused motion (e.g. *send*, *post*), ballistic verbs (e.g. *throw*, *roll*, *drop*), and verbs of creation typically with a Benefactive argument (e.g. *build*, *make*, *cook* etc.).[3] The most typical ditransitive verb is 'give'. There are many papers focusing on the morphosyntax of this verb (Haspelmath 2013, Newman 1996). Ditransitive verbs are used in a type of construction, the ditransitive construction, and there are languages with more than one ditransitive construction. There have been many descriptions of ditransitive structures in many languages in the past few decades, and due to this the typology of the ditransitive morphosyntax is thoroughly elaborated. The main issue of the typology is which argument of the construction appears in the same syntactic position as the Patient of the monotransitive construction (cf. Haspelmath 2005, 2013, Wunderlich 2006, Margetts and Austin 2007 Malchukov et al. 2010). Several papers examining ditransitivity in Ob-Ugric languages (mainly Mansi) have been published recently, too (Bíró 2015, Bíró and Sipőcz 2017a, 2017b, 2018, Sipőcz 2011, 2013, 2015b, 2015c, 2016, 2017, Virtanen 2011).

Two ditransitive constructions are used in the Khanty and Mansi languages.[4] One is the indirective type in which the Theme argument is the syntactic object, thus it is

---

[3] It is language specific onto which verb groups the use of ditransitive structure is spread beyond the physical transfer verbs (Sipőcz 2015, 2016).

[4] This phenomenon is called ditransitive alternation. It is well-known from English, where the indirective and the neutral constructions alternate. (Neutral alignment: the Theme and the Recepient are coded in the same way as the Patient of the monotransitive construction.) The alternation used in Ob-Ugric languages is more common than the alternation used in English (cf. Malchukov et al. 2010: 18).

coded as the Patient of the monotransitive construction,[5] and the Recipient is marked
with the lative-dative suffix or with a postposition of a similar function. Cf.:

(2) Khanty

āśe-l	*pox-al-a*	*uli*	*ma-l.*
father-3SG	boy-3SG-LAT/DAT	reindeer	give-PRS.3SG

'The father gives a reindeer to his son.'

(3) Mansi

tōrəm	*naŋən*	*matər*	*mi-s*
God	you.LAT/DAT	sth.	give-PST.3SG

'God gave you something.'

The other is the secundative type, in which the Recipient is coded as the Patient
of the monotransitive construction and the Theme argument gets an oblique marker
(the instrumental case suffix in Mansi, and the locative or instrumental suffix in the
Khanty dialects). Cf.:

(4) Khanty

āśe-l	*pox-al*	*uli-jn*	*ma-l-li.*
father-3SG	boy-3SG	reindeer-LOC	give-PRS-OBJ.3SG

'The father gives a reindeer to his son.'

(5) Mansi

uwśi-m	*tor-əl*	*mi-s-lum.*
sister-1SG	kerchief-INSTR	give-PST-OBJ.1SG

'I gave a kerchief to my elder sister.'

These constructions can passivize: the passivization of the secundative alignment
(moving the Recipient to the subject position) is more frequent (6-7), and it is more
common typologically, too. The Recipient passivization of the construction
containing the verb 'give', in fact, is used to express the meaning 'get'. This is the
general way of expressing the notion of getting, since there is no basic verb meaning
'get' in the Ob-Ugric languages.

5 There is no accusative case suffix in Khanty or in Northern Mansi (as opposed to other Mansi
dialects), the object of the clause is in the nominative case, except the personal pronouns, which
have an accusative form in all Ob-Ugric dialects.

(6) Khanty

āŋki-nə[6]	ńēwrem	ńāń-a	mə-ʌ-i
mother-LOC	child	bread-INSTR	give-PRS-PASS.3SG

'The child gets bread from the mother.'

(7) Mansi

tōnt tax	ōs	akw	Buran-əl	mi-w-et
then PTCL	PTCL	one	Buran-INSTR	give-PASS-3PL

'They get (lit. they are given) one more new Buran (snowmobile).'

The passivization of indirective alignment (moving the Theme into subject position) is rarer, but there are examples of this, too.

(8) Khanty

mā-nə	nüŋati	järnas	jɔnt-ʌ-i
I-LOC	you.LAT	shirt	sew-PRS-PASS.3SG

'A shirt was sewn for you (by me).'

(9) Mansi

ti	rupata	war-ne-nəl	mayəs	tananeln	tax	man
this	work	do-AN-3PL	for	they.LAT/DAT	then	we

okrug-uw-nəl	akw	Buran	mi-w-e
district-1PL-ABL	one	Buran	give-PASS-3SG

'Then they were given a Buran for their work from our district.'

As we have seen in the examples above, in the Ob-Ugric languages the ditansitive alternation is a device for differentiation between the events of giving and getting through the single verb 'give' assigning the participants of the event different grammatical roles. In the case of the active ditransitive constructions containing the verb 'give', the event expressed by the verb is 'giving'. The subject of the clause is the giver, who is at the same time the Agent of the event and usually the topic of the discourse. The event in which the giver is the Agent can be only the giving, irrespective of the fact whether the Theme or the Recipient is in the object position.

However, the passive constructions containing the verb 'give' allow the interpretation of both giving and getting. In the case of Theme passivization the subject of the construction is the Theme argument and, regarding the connection of information structure and clause structure, the Theme is the topic of the discourse.[7] The event in which the Theme appears as the topical element can be either the giving

---

[6] The Agent of the passive construction is marked with the locative suffix in Khanty.

[7] By topic I mean a previously mentioned contextually or situationally given information, cf. Dalrymple and Nikolaeva (2014: 48–57).

or the getting, thus both can be stated about the Theme: it was given / it was got. In case of Recipient-passivisation the subject of the construction (and, thus, the topic of the utterance) is the recipient. The event whose topic is the Recipient is the getting and not the giving. In other words, from the perspective of the Recipient, the primary aspect of the event is the getting. From this it follows that the verb 'give' in the secundative passive construction is used to express the notion 'get'.

In the Nganasan language a possessive ditransitive construction is used, which is typologically rather rare. In the example below the Recipient argument of the construction is coded on the Theme with a possessive suffix following a destinative suffix. (As if we were saying 'I gave his/her book.'). This construction can be passivized, and as it can be seen in example (11), the passive ditransitive construction containing the verb 'give' expresses the meaning 'get', as in the Ob-Ugric languages.

(10) Nganasan

mənə	*kńiga-ðə-mtu*	*mi-śiə-m*
I	book-DST-ACC.3SG	give-PST-1SG

'I gave him/her the book.' (Wagner-Nagy and Szeverényi 2013: 28)

(11) Nganasan

kiribaʔkü-ðə-mə	*tətu-ru-bata-ðə*	*ŋuəntəə-tə*
bread-DST-NOM.1SG	give-PASS-INFER-SG.R	boss-LAT.SG

'I got bread from the boss.' (MVL_080226_TwoHorses_flks.426, Brykina et al. 2018)

3. The aspect of lexical typology

The frame for my lexical typological description is provided in a paper by Viberg (2010), which is a contrastive description of the verbs of possession in Swedish, English, German, French, and Finnish. The verbs of languages typically form an open word class with thousands of members, and with hundreds of semantic fields (e.g. motion verbs, verbs of communication, verbs of emotion, perception verbs etc.), also verbs of possession represent a semantic field. The verbs of possession express getting into possession, giving into possession, and being in possession, e.g. English *have*, *get*, *take*, *give*, *need*, *pay*, *keep*, *buy*, *provide*, *sell*, *pay*, *lack*, *own*, *send*, *reach*, and *hand*. The verbs 'give' and 'get' investigated in this paper belong to this verb class.

Viberg selected the basic verbs of possession on the basis of their frequency. Regarding their frequency, verbs can be divided into a small number of basic verbs and a large number of non-basic verbs. Basic verbs are the most common and most extensively used verbs of the given semantic fields. On the basis of frequency investigations it can be stated that the usage of basic verbs makes up the majority of verb uses. In European languages, the twenty most frequent verbs tend to cover almost

half of all the occurrences of the verbs in running text, and these verbs show great cross-linguistic similarities with respect to their meaning. Within the basic verbs Viberg distinguishes the nuclear verbs, which tend to be realized as basic verbs in all languages, thus they are universals. But at the same time the nuclear verbs can have language specific features, too. Nuclear verbs are usually polysemic, and their polysemic patterns can vary in different languages. According to Viberg, the group of nuclear verbs is an important category for L2 learners since they tend to favor them (Viberg 2002, 2010). Similar tendencies can be found regarding language acquisition, too (De Villiers and De Villiers 1978: 143).

Based on word frequency tests, Viberg listed the following verbs as the basic verbs of possession: 'have', 'get', 'take', and 'give', of which 'give' and 'take' are nuclear verbs. Further possession verbs – even the most common ones – show significantly rarer usage. The basic verbs of possession can be characterized by polysemy in addition to their frequency, they are typically short lexemes, often with irregular inflection. Furthermore, these verbs are often source verbs for several grammaticalization processes. Historically the basic verbs of possession have been usually derived from physical action verbs related to movements and manipulations with the hands, e.g. English *have* < PIE **kap* 'grasp', Swedish *få* 'get' < 'catch' (Viberg 2010). Regarding their semantics, the meaning of basic verbs is simpler than the meaning of non-basic ones. E.g. the verbs 'give' and 'take' differ from the verbs 'sell', 'pay', and 'buy' in the respect that 'give' and 'take' have only the component transfer of possession where the transfer is initiated by one of the interactors, and only the direction of the transfer differs depending on who the initiator of the transfer is; the verbs 'sell', 'pay', 'buy' etc. have other semantic components, too (e.g. notion of obligation to transfer money)[8] (De Villiers and De Villiers 1978: 143).

Table 1 summarizes the semantics of the basic verbs of possession: regarding the role of the subject and the direction of the transfer, 'give' is opposed to 'take' and 'get'. The dynamic system distinguishes Causative, Inchoative, and State verbs.[9] The source-based pair of 'get' in the empty space would be a verb meaning 'lose', but verbs with this meaning cannot be considered as basic according to the criteria of basicness. ("The space is left empty because those verbs do not reach a very high frequency and do not have characteristics which justify calling them basic to the same degree as the four verbs included in the table. Each of the basic verbs serve as the

---

[8] This is parallel to Dixon's notion of nuclear and non-nuclear verbs. Nuclear verbs cannot be defined in terms of other verbs, but non-nuclear verbs are semantically more complex and can be defined referring to other verbs (Dixon 1972: 293).

[9] This kind of semantic characterization applies to other verbs, too. Verbs are stative or dynamic types (cf. *have – get* or *know – realize*), and dynamic verbs can be inchoative or causative (cf. *lose – steal, die – kill*). (Viberg 2002:129)

superordinates of a number of hyponyms, whereas a superordinate term is lacking for the empty space in the grid." Viberg 2010) Because of the focus of this paper and the languages investigated, the meaning 'have' is not relevant, only the description of the verbs 'give', 'take', and 'get' is needed.

Dynamic meaning	**Source based**	**Goal based**
causative	'give'	'take'
inchoative		'get'
state	'have'	

Table 1. The basic verbs of possession (Viberg 2010)

We have seen above (in section 2) that in Khanty, Mansi, and Nganasan the meaning 'get' is expressed by the verb 'give' in a passive construction (examples 6, 7, and 11) (cf. Table 2). This phenomenon is not rare as it can be seen from the following Swahili example and its English translation:

(12) Swahili
a.

Halima	*a-li-m-pa*		*zawadi*	*Fatuma.*
Halima	she-PST-HER-give		gift	Fatuma

'Halima gave a gift to Fatuma.'
b.

Fatuma	*a-li-p-ew-a*		*zawadi*	*na*	*Halima.*
Fatuma	she-PST-GIVE-pass		gift	by	Halima

'Fatuma was given a gift by Halima.'
(Malchukov et al. 2010: 22)

The system of the verbs of possession in the Uralic languages examined in this paper can be seen in Table 2. (The verbs 'have' are listed for completeness' sake only.) The verb 'get' is absent in all three languages, this meaning is expressed by the verb 'give'. In the Nganasan language the verb 'take' is also absent. This meaning is represented by several verbs, but they are semantically specific and thus cannot be considered basic, cf. *ńakələsi* 'take, take away, obtain', *mintələsi* 'take, take away; drag', *kəmə-* 'catch' (Sipőcz and Szeverényi 2019).

a) Mansi

Dynamic meaning	**Source based**	**Goal based**
causative	*mi-*	*wi-*
inchoative		
state	*ōńś-*	

b) Khanty

Dynamic meaning	Source based	Goal based
causative	mă-	wu-
inchoative		
state	taj-	

c) Nganasan

Dinamic meaning	Source based	Goal based
causative	misji tətuďa[10]	-
inchoative		
state	honsɨ	

Table 2. The basic verbs of possession in Mansi, Khanty and Nganasan

Languages represented in Viberg's paper showed the greatest variety regarding the verbs meaning 'get'. In languages having a basic verb with this meaning usually the verb 'get' is etymologically old, and in connection with its oldness this verb has complex polysemy and several grammaticalized usage, and it is typically one of the most frequent verbs of the given language (cf. English *get*, Swedish *få* or Finnish *saada* as modal or causative auxiliaries). In languages without a basic verb 'get', this meaning can be expressed by the passivization of the verb 'give' or by other basic verbs of possession. This phenomenon can be seen in Viberg's following tables (Tables 3–4).

---

[10] In Nganasan there are two basic verbs meaning 'give' whose distribution depends on the person of the Recipient: *tətuďa* is used if the Recipient is 1st or 2nd person, *mis^ji* is used if the Recipient is 3rd person (cf. examples 10 and 11) (Wagner-Nagy and Szeverényi 2013). A similar suppletive split appears also in other languages, e.g. Saliba, Kolyma Yukaghir (Margetts and Austin 2007).

English		
Dynamic meaning:	**Source-based**	**Goal-based**
Causative	give	take
Inchoative		get
State	have	

German		
Source-based		**Goal-based**
geben		nehmen
		bekommen/ kriegen
haben		

French		
Dynamic meaning:	**Source-based**	**Goal-based**
Causative	donner	prendre
Inchoative		
State	avoir	

Finnish		
Source-based		**Goal-based**
antaa		ottaa
		saada
Locational possessive		

Table 3. The basic verbs of possession in English, German, French, and Finnish
(Viberg 2010, Table 11.)

Even these genetically and/or areally related five languages show differences (the 5th language in Viberg's description is Swedish): while in Finnish (and also in Swedish) there are four basic verbs of possession, in English 'give' also functions as 'get' in a passive construction, in German the verbs *bekommen* and *kriegen* share the meaning 'get', and in French *donner* and *avoir* appear in this function.

In non-European languages the differences can be greater. While in Chipewyan there are no basic verbs of possession, in Sango the direction of transfer is not relevant and 'get' is expressed by a verb meaning 'find', in Turkish the causative possession verbs express the meaning 'get', and in Swahili 'get' is expressed by the passive use of the verb 'give' (see also example 11).

	Chipewyan (Dene Sųłiné)	
Dynamic meaning:	Source-based	Goal-based
Causative	No basic verbs of possession	
Inchoative		
State		

	Sango	
Dynamic meaning:	Source-based	Goal-based
Causative	mû 'give=take, transfer'	
Inchoative		wara 'find'
State	With-Possessive	

	Turkish	
Dynamic meaning:	Source-based	Goal-based
Causative	vermek	almak
Inchoative		
State	Genitive schema	

	Swahili	
Dynamic meaning:	Source-based	Goal-based
Causative	-pa	-chukua
Inchoative		-pata
State	With-Possessive	

Table 4. Basic verbs of possession in Chipewyan, Sango, Turkish and Swahili
(Viberg 2010: Table 13)

In the investigated Uralic languages, it can be seen – mainly in newer sources – that in addition to the passive use of the verb 'give' the verb 'take' is also acquiring the function of 'get' (cf. section 5 below). I intend to display the cognitive and sociocultural background of this phenomenon in the following sections.

4. The perspective of cognitive linguistics

'Giving'/'getting' is one of our most common activity, it is a "basic act occurring between humans" (Newman 2002: 79). According to Viberg, the verb 'give' is a nuclear verb, this meaning tends to be lexicalized in all languages. The verb 'get' is a basic verb "only", and in many languages its meaning is not lexicalized as a basic verb

(and has thus no specialized meaning) (see the examples in 3). Khanty, Mansi, and Nganasan also belong to this latter group of languages.

The phenomenon is worth examining from a cognitive perspective. The figure-ground organization of clause structure widely investigated in Cognitive Grammar can be detected in give clauses, the grammatical subject of the clause is in connection with the figure-ground perspectives. All three participants of the event may function as grammatical subject (cf. *Mary gave a book to John. John got a book from Mary. This book has changed hands*.). Nonetheless it is very common that speakers tend to describe the event of giving so that the source of the action occurs as the grammatical subject. In an experiment where speakers had to describe different physical transfers, they preferred to put the giver into the subject position. Other clause structures are also possible, since the use of the different structures in an utterance is always determined by the given situation and by which constituent is pragmatically emphasized, but from a cognitive perspective the most basic way is featuring the giver as a grammatical subject (Newman 2002: 79–81).

The notion 'get' is a kind of abstraction. It is supported by several conditions. The lexicalization of this notion is not as universal as the lexicalization of 'give' or 'take' (see section 3 above). If a language has a basic verb with this meaning, etymologically this verb originates in more concrete meanings denoting actions done with hands, e.g. Swedish *få* 'get' < 'catch' (Viberg 2010), Finnish *saada* 'get' < **saγe* 'come, catch, reach' (UEW 429), Hungarian *kap* 'get' < an onomatopoeic verb with the meaning 'snap' (EWUng 684–5). In addition, the syntax and semantics of the verb 'get' display a kind of controversy. The event of getting involves an actor who is not the subject of the verb (in active voice), and the subject is not an actor. Thus, in the case of the verb 'get' the syntactic subject is the Recipient. Cognitively this may explain the "unstable" status of this verb within the group of the basic verbs of possession, and the use of other, cognitively "more stable" verbs to express this event.

5. The perspective of ethnosyntax[11]

It is a well-known fact that languages reflect on the world around most visibly in the lexicon, but also the grammar of a language can display features reflecting the aspects of the speakers' environment and culture. A good example of it is the nominal classification system found in some languages which represents both cultural and grammatical features. According to Wierzbicka (1973: 313) "every language embodies in its very structure a certain world view, a certain philosophy".

Dixon writes about the possible connections between culture and the linguistic expression of the event of giving in Dyrbal (1973). He claims that while several 'give'

---

[11] For the term ethnosyntax and its "narrow and broad sense", see Enfield (2002: 7-8).

type verbs are used in English with specialized commercially oriented meaning (like *sell*, *rent*, *lend*, *pay*, *award* etc.), the verbs belonging to the same lexical field have different semantic characteristics in Dyirbal: they have semantic components related to position, movement and kinship obligation relations. (For instance, a different verb is used depending on what movement is involved in the transfer, or whether the transfer involves relatives etc., cf. Dixon 1973: 206-210.)

Newman investigated the ethnosyntactic correlations of give-type clauses in one of his papers (2002). He distinguishes between languages in which the linguistic expression of the 'give' event reflects cultural circumstances on the one hand, and languages in which the structure of the give clause shows language-specific features, but these features do not reflect greater cultural, non-linguistic values. The Japanese language provides a good example of the connection between the linguistic expression of transfer and certain cultural and social circumstances. Politeness phenomena are relatively well-known features of the Japanese language, since respect is an inherent characteristic of Japanese society. Japanese language use is fundamentally determined by the relative status of the interlocutors and the subordinate/superordinate status of the speakers in the hierarchy of rank. This is also evident in the expression of 'give' events: the choice of the verb is dependent on the relative status of the giver and the recipient in the hierarchy as well as on the given situation, that is, whether the speaker is the giver, the recipient, or neither (Newman 2002: 82–84).

Newman describes as an example of the second type, that in some languages the recipient is expressed as the syntactic object, in contrast with other languages where the transferred thing occupies this position. (In the typological framework presented in section 2, these construction types correspond to the indirect and secundative alignments.) In Tuggy's 1998 terminology, the two construction types correspond to the Human Interaction and Object Manipulation perspectives: the syntactic object is the recipient in the former, and the transferred thing in the latter. These perspectives can also be manifested in other aspects of grammar: for instance, in the Nahuatl and Zulu languages they are represented in the causative and applicative constructions also, in addition to ditransitives. But there is no reason to suppose that these features of grammar have any cultural motivations (Newman 2002: 91-93).

As we have seen in section 2, the Ob-Ugric languages employ both the Human Interaction and Object Manipulation perspectives for expressing 'give' events. This can be regarded a language specific feature, but without being associated with any cultural or social characteristic. From the typological perspective it is important to add that the use of both of these construction types is found in many languages of the world. Another language specific phenomenon is the absence of the basic verb meaning 'get' in the investigated Uralic languages. As we have seen, this is not a unique phenomenon either, it is typical of several languages (see section 3), and

presumably is in connection with the cognitive background of the notion 'get' (see section 4). No cultural or social characteristic is associated with it, since there is no reason to believe that Siberian Uralic people, for instance, regard 'giving' as a more important thing than 'receiving' – this, again, is just a language specific phenomenon.

However, we can also observe in the case of the Ob-Ugric languages that recently – typically in the newer sources – beside the passive use of the verb 'give' also the verb 'take' appears with the meaning 'get'. As we could see above (section 3), it is not unprecedented, the verb 'take' can have this function also in other languages. When I asked my Khanty informant to translate the sentence 'I got this book from you' into Khanty, she said "It is impossible to say in Khanty that I got something. It is in Russian". Finally, after some hesitation she chose the verb *wu-* 'take' (12). I had a similar experience with my Mansi informant who used the verb 'take' when she tried to translate Russian sentences with the verb 'get'.

(13) Khanty

ma	*năŋ*	*ełten*	*tăm*	*nepek*	*u-s-em*
I	you	PP	this	book	take-PST-1SG

'I got this book from you.'

This lexical change – namely, the extension of the meaning of the verb 'take' – can be traced in the dictionaries. "Modern" Russian–Mansi dictionaries contain the verb *wiγ* 'take' as the Mansi equivalent of the Russian verb *получать* 'get' (Rombandeeva 2005, Rombandeeva and Kuzakova 1982). Dictionaries based on earlier collections do not mention this meaning of the verb *wiγ* (Munkácsi and Kálmán 725–726, Kannisto 137–139). Nor does the Khanty dictionary based on older texts and collections mention the meaning 'get' of the verb *wu-* 'take' (DEWOS 1549). In modern Russian–Khanty dictionaries either the verb *получать* 'get' is translated with a non-basic verb (like 'catch', 'capture' etc.) or they do not contain the entry 'get' (Syazi and Skameyko 1992, Solovar 2006, Tereshkin and Solovar 1981, Volkova and Solovar 2016). In some Khanty–Russian dictionaries the verb *wu-* is translated by the Russian word *получать* 'get' in addition to the lexemes *взять* and *брать* 'take' (Syazi and Skameyko 1992). The following Mansi sentence pair is from the same newspaper article, the sentences express the same situation, but the verbs are different. In (14) the passive construction containing the verb 'give' is used, in sentence (15) the same meaning is expressed by the verb 'take'.

(14) Mansi

Kit-it	*mesta-l*	*Nižnewartowskij*	*ūs-t*	*ōl-ne*
two-dx	place-INSTR	Nizhnevartovsk	town-LOC	live-PTCP.PRS
xantə-t	*maj-we-s-ət.*			
Khanty-PL	give-PASS-PST-3PL			

'Khanty people from Nizhnevartovsk got second place.'

(15) Mansi

Os   xūrmit   mesta   ńeft'ejuganskij   rajon-t
and   third   place   Nefteyugansk   district-LOC
ōl-ne   xōtpa-t   wi-s-ət
live-PTCP.PRS   person-PL   take-PST-3PL
'And people from Nefteyugansk took/got 3rd place.'
(LS 2016/13: 9)

In recent Nganasan sources we can observe a similar phenomenon with a different lexical solution. In example (16) we can see an active clause structure to express the situation of getting by using the Russian verb *получать* 'get':

(16) Nganasan

*pr'emə   **polut'i-ir-ü**   kəmüðü-tü*
award.ACC   get-FRQ-PRS.[3SG.S]   fut[STAT]-PRS[3SG.S]
'(We were competing, who arrives first to Volochanka), that person will get the award.' (KTD_MyLife_nar.exb, Brykina et al. 2018)

The examples above demonstrate that probably due to the influence of bilingualism in Russian among the speakers of Siberian Uralic languages today, a new clause structure has emerged for expressing the notion 'get'. In addition to the previous passive construction with the verb 'give', the speakers have also started to apply an active construction for expressing the event of getting which was not previously used in these languages. And this change is connected to the social and cultural circumstances of the speakers, namely, to the predominant bilingualism of Ob-Ugric and Samoyedic people appearing rapidly in the last few decades. This is an example of the phenomenon where language contact is manifested not in the lexicon but in a structural change. It cannot be named unambiguously an ethnosyntactic phenomenon (not in the narrow sense of ethnosyntax), but it is undoubtedly in connection with the speakers' conditions. Table 4 demonstrates this modified system of the basic verbs of possession in Mansi, Khanty, and Nganasan.

a) Mansi

Dynamic meaning	**Source based**	**Goal based**
causative	*mi-*	*wi-*
inchoative		
state	*ōńś-*	

b) Khanty*

Dynamic meaning	Source based	Goal based
causative	*mă-*	*wu-*
inchoative		
state	*taj-*	

c) Nganasan

Dynamic meaning	Source based	Goal based
causative	*misji* *tətuďa*	-
inchoative		(+ Russian *получать*)
state	*honsɨ*	

Table 4. The basic verbs of possession in recent Mansi, Khanty and Nganasan

6. Conclusion

In this paper I have investigated the event of giving and getting from different theoretical perspectives in three Uralic languages spoken in Siberia, namely, Khanty, Mansi, and Nganasan. From a syntactic perspective we could see that the verbs meaning 'give' and 'get' are used in constructions which are typologically attested in other languages, too, and in all three languages the notion 'get' is expressed by a passive construction containing the verb 'give'. This is a common phenomenon from the perspective of lexical typology, too. The meaning 'get' does not lexicalize as a basic verb in numerous languages, and very often other basic verbs of possession express it. Probably this phenomenon is in connection with some cognitive features of the event of getting. On the one hand, the syntax and the semantics of the verb 'get' display a kind of controversy: in clauses with the verb 'get' the syntactic subject – regarding its Thematic role – is the Recipient, not the Agent. On the other hand, the meaning 'get' is a result of abstraction, which is confirmed among other things by the etymology of 'get' verbs, which originally denoted concrete actions done with hands (grabbing, grasping, catching etc.). In recent language usage a new structure has appeared in the investigated languages which is absent in earlier sources. In this structure the event of getting is expressed by an active clause, in the Ob-Ugric languages the verb 'take' is getting acquire the function of 'get', and in the Nganasan language the Russian verb meaning 'get' appears in this function. Behind this change

we can see the intensive influence of Russian and the rapidly developed bilingualism of the speakers.

With this complex analysis of the event of giving and getting I have aimed to demonstrate that in analysis of a given language phenomenon different approaches (syntactic, lexical, cognitive, and cultural linguistic) may be interconnected in complex ways and can provide a more comprehensive description.

Abbreviations

A	agent of a (di)transitive clause
ABL	ablative
ACC	accusative
AN	action nominal
CAUS	causative marker
DU	dual
DEST	destinative suffix
DX	derivational suffix
FRQ	frequentative DX
FUT	future
INFER	inferential
INSTR	instrumental
LAT/DAT	lative-dative
LOC	locative
NEG	negative particle
PASS	passive
PL	plural
PP	postposition
PRS	present
PST	past
PTCL	particle
PTCP	participle
R	recipient
SG	singular
T	theme

References

Bíró, Bernadett 2015. Az északi és déli manysi ditranzitív szerkezetek szórendje [The word order of ditransitive constructions in Northern and Southern Mansi dialects]. In: É. Kiss, Katalin, Hegedűs, Attila, and Pintér Lilla (eds.). *Nyelvmélet és dialektológia* 3 [Linguistic theory and dialectology 3]. Piliscsaba: PPKE, 43–58.

Bíró, Bernadett and Sipőcz, Katalin 2017a. The Mansi ditransitive constructions. *Finno-Ugric Languages and Linguistics* 6: (1).
http://full.btk.ppke.hu/index.php/FULL/article/view/53/59

Bíró, Bernadett and Sipőcz, Katalin 2017b. Ditransitivity in the Mansi language from typological perspective. *Finnisch-Ugrische Mitteilungen,* 41: 25-43.

Bíró, Bernadett and Sipőcz, Katalin 2018. Ditransitive constructions in Mansi dialects. In: Korn, Agnes and Malchukov, Andrej (eds). *Ditransitive constructions in a cross-linguistic perspective*. Wiesbaden: Reichert Verlag, 187–207.

Brykina et al. 2018: Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor and Wagner-Nagy, Beáta. 2016. "Nganasan Spoken Language Corpus (NSLC)." Archived in Hamburger Zentrum für Sprachkorpora. Version 0.1. Publication date 2016-12-23. http://hdl.handle.net/11022/0000-0001-B36C-C.

Dalrymple, Mary and Nikolaeva, Irina 2014. *Objects and information structure*. Cambridge: Cambridge University Press.

DDML: Bíró, Bernadett and Sipőcz Katalin. Ditransitive database of the Mansi language. Unpublished manuscript.

de Villiers, J. G. and de Villiers, P. A. 1978. *Language acquisition*. Cambridge, Mass: Harvard University Press.

DEWOS: Steinitz, Wolfgang 1967–1991. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. 1–14. Lieferungen. Berlin: Akademie Verlag.

Enfield, N.J. (ed.). 2002. *Ethnosyntax: Explorations in Grammar and Culture*. Oxford: Oxford University Press.

EWUng: Benkő, Loránd (Hrsg.) 1993–5. *Etymologisches Wörterbuch des Ungarischen I–II*. Budapest: Akadémiai Kiadó.

Haspelmath, Martin 2005. Argument marking in ditransitive alignment types: Linguistic discovery. 3/1. https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/280?htmlOnce=yes. Date of access: November 24, 2020.

Haspelmath, Martin 2013. Ditransitive constructions: The verb 'give'. In: Dryer, Matthew S. and Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/105

Kannisto, Artturi, Vuokko Eiras, and Arto Moisio 2013. *Wogulisches Wörterbuch*. Helsinki: Société Finno-Ougrienne.

Malchukov, Andrej, Haspelmath, Martin, and Comrie, Bernard 2010. Ditransitive constructions: A typological overview. In: Malchukov, Andrej, Haspelmath, Martin, and Comrie, Bernard (eds.). *Studies in ditransitive constructions: A comparative handbook*. Berlin: Mouton de Gruyter, 1–65.

Margetts, A. and Austin Peter K. 2007 Three-participant events in the languages of the world: towards a crosslinguistic typology. *Linguistics.* 45(3): 393–451.

LS: Lujima Seripos. Mansi newspaper. Khanty-Mansiysk. (http://www.khanty-yasang.ru/)

Munkácsi, Bernát and Kálmán Béla 1986. *Wogulisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Newman, John 1996. *Give: A cognitive linguistic study*. Berlin: Mouton de Gruyter.

Newman, John 2002. Culture, cognition, and the grammar of 'give' clauses. In: N. J. Enfield (ed.). *Ethnosyntax: Explorations in grammar and culture*. Oxford: Oxford University Press.

Rombandeeva, E. I. 2005. *Russko–mansiyskiy slovar*. Sankt-Peterburg: Mirall.

Rombandeeva, E. I. and Kuzakova, E. A. 1982. *Mansiysko–russkiy i russko–mansiyskiy slovar.* Leningrad: Prosveshenie.

Sipőcz, Katalin 2011. Az obi-ugor ditranzitív szerkezetek történeti háttere [The historical background of the Ob-Ugric ditransitive constructions] In: Bakró-Nagy, Marianne and Forgács, Tamás (eds.), *A nyelvtörténeti kutatások újabb eredményei* VI [The most recent results of historical linguistic investigations, vol. 6]. Szeged: Szegedi Tudományegyetem, 221–229.

Sipőcz, Katalin 2013. Ditranzitív igék a manysiban [Ditransitive verbs in Mansi]. *Nyelvtudományi Közlemények,* 109: 123–137.

Sipőcz, Katalin 2015a. Tárgyjelölés a manysi ditranzitív szerkezetekben [Object marking in Mansi ditransitive constructions]. In: É. Kiss, Katalin, Hegedűs, Attila, and Pintér, Lilla (eds.), *Nyelvelmélet és dialektológia* 3 [Linguistic theory and dialectology, vol. 3]. Piliscsaba: PPKE, 234–246.

Sipőcz, Katalin 2015b. Ditransitivity in the Ob-Ugric languages. In: Harri Mantila, Kaisa Leinonen, Sisko Brunni, Jari Sivonen, and Santeri Palviainen (eds.), *Congressus Duodecimus Internationalis Fenno-Ugristarum. Plenary Papers*. Oulu: University of Oulu, 133-159.

Sipőcz, Katalin 2016. A manysi ditranzitív igék szintaxisa szemantikai megközelítésből [The syntax of ditransitive verbs from a semantic perspective] *Jelentés és Nyelvhasználat,* 3: 49–61.

Sipőcz, Katalin 2017. A language without 'get'? In: Khabtagaeva, Bayarma and Szeverényi, Sándor (eds.), *Uralic and Siberian lexicology and lexicography*. Studia Uralo-Altaica 51. Szeged: Szegedi Tudományegyetem, 173–186.

Sipőcz, Katalin 2018. A birtoklás alapigéi a manysiban lexikális tipológiai megközelítésből [The basic verbs of possession in Mansi from a typological point of view]. In: Kenesei, István, and Bakró-Nagy, Marianne (eds.), *Általános Nyelvészeti Tanulmányok* XXX: Uralisztikai tanulmányok [Studies in general linguistics vol. 30: Papers in Uralic studies]. Budapest. Akadémiai Kiadó, 233–248.

Sipőcz, Katalin and Szeverényi, Sándor 2018. A birtoklás alapigéi a manysiban és a nganaszanban lexikális tipológiai szempontból [The basic verbs of possession in Mansi and Nganasan from a lexical typology perspective]. Presentation at "Össznyelvész Konferencia" 2018. Budapest, ELTE.

Syazi, Z. I. and Skameyko, R. R. 1992. *Slovar' khantyjsko-russkij i russko-khantyjskij (shuryshkarskij dialekt)*. Sankt-Peterburg: Prosveshenie.

Solovar, V. N. 2006. *Hantyjsko-russkij slovar' (kazymskij dialekt)*. Sankt-Peterburg: Mirall.

TDUL: Typological database of the Ugric languages. http://finnugor.elte.hu/index.php?q=ugadat. Date of access: November 24, 2020.

Tereshkin N. I. 1981. *Slovar' vostochnohantyjskih dialektov.* Leningrad: Nauka.

Tuggy, David 1998. Giving in Nawatl. In: John Newman (ed.). *The linguistics of giving.* Amsterdam: John Benjamins, 35–66.

UEW: Rédei, Károly (Hrsg.) 1988, 1991. *Uralisches Etymologisches Wörterbuch* I–III. Budapest – Wiesbaden: Akadémiai Kiadó – Otto Harrassowitz

Viberg, Åke 2002. Polysemy and disambiguation cues across languages: The case of Swedish få and English get. In: Altenberg, B. and S. Granger (eds.), *Lexis in contrast corpus-based approaches*. Amsterdam: Benjamins, 119–150.

Viberg, Åke 2010. Basic verbs of possession: A contrastive and typological study. In: M. Lemmens (ed.), *Unison in multiplicity: Cognitive and typological perspectives on grammar and lexis*. CogniTextes 4. https://journals.openedition.org/cognitextes/308?lang=en. Date of access: November 24, 2020.

Virtanen, Susanna 2011. Ditransitiivien vaihtelua mansin itämurteissa [Ditransitive alternation in Eastern Mansi]. *Folia Uralica Debreceniensia,* 18: 165–176.

Volkova, A. N. and Solovar, V. N. 2016. *Kratkij russko-xantyjskij slovar' (surgutskij dialekt)* [A short Russian-Khanty dictionary (Surgut dialect)]. Xanty-Mansijsk: Obsko-ugorskij institut prikladnyx issledovanij i razrabotok

Wagner-Nagy, Beáta and Szeverényi, Sándor 2013. On the argument structure of the 'give' verbs in Nganasan and Selkup. *Tomskyj zhurnal lingvisticheskih i antropologicheskih issledovanij,* 1: 27–36.

Wierzbicka, Anna 1979. Ethno-syntax and the philosophy of grammar. *Studies in Language,* 3(3): 313–383.

Wunderlich, Dieter 2006. Towards a structural typology of verb classes. In: Wunderlich, Dieter, (ed.). *Advances in the theory of the lexicon*. Berlin: Mouton de Gruyter. 57–167.

A word-formational approach to neologisms in modern Northern Mansi: Bilingual compound names of professions, documents and institutions

Susanna Virtanen

Helsinki University

1. Introduction

Neologism are new words recently adapted to the language. Usually neologisms are created by borrowing from other languages or by deriving and compounding already existing words, or by bringing up wholly new coinages. This article is one part of a more extensive study on Mansi neologism. In this article, I will outline one word-formational group of neologisms in modern Northern Mansi: bilingual (Mansi-Russian) compounds. I will concentrate especially on three thematic groups: names of professions, official documents and institutions. These thematic groups were chosen, because they also form the most frequent word-formational types of bilingual compounds appearing in my data. In other words, there is a connection between thematic groups and word-formational types.

Mansi language belongs to the Ob-Ugric branch of the Finno-Ugrian language family. It is spoken in Western Siberia, in the Khanty-Mansi Autonomous Area of the Russian Federation. Mansi is an urgently endangered language: there are still approx. 1,000 speakers alive (see the Russian Census 2010). Only one of the four main dialects is still spoken. The speakers of Mansi live under the Russian laws and inside the Russian society. In elementary schools, Mansi is often taught as a second language only, and it is not used as a teaching language at all. Mansi is not used in everyday situations by authorities or public offices. Urbanization and oil drilling speed up the death of the language. For this reason, there are also a lot of Russian loanwords in Mansi language.

The data for my study is gathered from up-to-date newspaper articles. Methodologically this study combines together word-formational analysis and semantics: the results of word-formational and thematic classification processes are compared with each other. The aim of this study is to present a formal description of

bilingual compounds (Russian-Mansi combinations), and to describe, what kind of connections or correlations there are between word-formational types and thematic groups. Bilingual compounds include either a Russian generic part and a Mansi specific part, or a Mansi generic part and Russian specific part. The semantic properties of both types will be observed.

2. Language contacts of Mansi

During the centuries, Mansi has been under influence of several surrounding languages. Proto-Iranian and Proto-Turkic influence started already in the common Ugric era and continued after Hungarian broke away (Keresztes 1989: 422). Among the Uralic languages, Komi, Khanty and Nenets have influenced Mansi. Contacts with Komi people started in the 900's (Keresztes 1989: 423), and also the first Russian loanwords were adapted with Komi mediation. Also Samoyedic influence has been observed: especially the Mansi reindeer-breeding terminology has been influenced by the Nenets language (Steinitz 1959). Southern dialects of Mansi have also been in contact with Tatar languages from the 14th Century (Keresztes 1989: 422). With the Russian population, there have been direct contacts since the 17th Century (see e.g. Kálmán 1961, Bakró-Nagy 2018).

Later on, a great deal of new words have been adapted to Mansi from Russian when adapting the whole phenomenon from the Russians. This covers a large amount of political, technological, scientific, cultural and administrational vocabulary, and vocabulary on other thematic groups as well. The current sociolinguistic situation is that Mansi is used only in a restricted area of everyday life: many situations are excluded from Mansi language using area, because for example administration and education are organized mainly in Russian. Practically all speakers of Mansi are Russian-Mansi bilinguals, and for a big majority of them, Russian is the language they speak best. This means that Russian language strongly influences Mansi language and makes maintaining the minority language much more difficult. One of the aims of this study is to map out the circumstances where new phenomena are named with Mansi-origin words are used in neologisms.

3. Research on neologisms

As for example Häkkinen (1997: 85–87) states, new words come to a language first and foremost by compounding and deriving old words, by borrowing words from other languages and sometimes by inventing completely new coinages. Neologisms are generally defined as words that have recently come into the language or have not been established yet. For example, Bolganbaev (1988: 112) defines neologism as follows: "Neologisms are words that have appeared in a language in connection with

new phenomena, new concepts… but which have not yet entered into the active vocabularies of a significant portion of the native speakers of the language." Fischer's (1998: 3) definition is basing on similar grounds: "A neologism is a word which has lost its status of a nonce-formation but is still one which is considered new by the majority of the members of a speech community."

The definitions above emphasize that neologisms are connected to new cultural concepts and phenomena, and also that neologisms are not yet established in the language. The connection to new cultural concepts is the starting point of gathering my own data as well: my data focuses on the modern society and culture of the 20th and 21st centuries. This kind of cultural features have come to the Mansi community almost exclusively through Russian society. In this study, such lexical items are presented, which are connected to modern society and culture: technology, health/medicine, cultural life, modern professions and administration/society.

Similar to other types of lexicology, also neologisms can be classified according to their functions, coinage processes, formation, etymological sources, or structural and organizational feature. Further, a formational classification includes several categories. Smyk-Bhattacharjee (2009: 36), referring to Helfrich (1993) classifies neologisms into four groups:

1. morphological neologisms
2. semantic neologisms
3. loan neologisms
4. coinage neologisms

My classification is quite similar to that of Smyk-Bhattacharjee, but my terminology and way of classification differs slightly from hers: I have ended up using partly more punctual terms *derivation* (morphological neologism), *metaphor* (semantic neologism), *loanword* and *compound*. Shortenings or coinages are not included in this study.

This article is one part of a larger lexical study on neologisms and loanwords in modern Northern Mansi. The more extensive study is a descriptive study on the formation and semantics of Mansi neologisms, involving mainly formational perspective. The study includes three sub-studies: Russian loanwords, Mansi-origin words and bilingual compounds (see Figure 1 in Section 5.1). As described in detail in Section 4, this partial study is a more detailed approach to one word-formational part of my data.

4. Research data

My data are gathered from the volumes 2014, 2018 and 2019 of *Lūimaa Sēripos*, the only regularly published Mansi-language newspaper in the world. It is published twice

a month in Khanty-Mansijsk and is also readable online[1]. The data have been gathered in a traditional way by reading texts and picking up words belonging to particular thematic groups: all groups are connected to technological, political and cultural innovations of the 20th and 21st centuries. The data have been classified according to two criteria: word-formational category and thematic category. The whole data include a corpus of 500 lexical items, including the following thematic groups:

1. Modern professions
2. Technology, Computing, Internet
3. Modern Society, Administration
4. Education, Cultural life, Science
5. Health care, Medicine

Further, the data have been classified into Russian loanwords, Mansi-origin words and bilingual compounds (see Figure 1 in Section 5.1). The need for this kind a classification originated from the very nature of the data: the details were decided just after observing the content of the data first.

In this case, the demarcation of the data is not straightforward, and the age of a single word cannot always be defined exactly, especially in case of a less documented language. The starting point of this study is to collect words belonging to thematic groups that have only emerged in human culture during the last hundred years. At this stage (2021), not all of the words in my data necessarily meet the definitions above, but the starting point is that the words have at least been neologisms once within the past 100 years.

5. Neologisms in my data

In this section, the above described data is first presented as a whole in 5.1. Further, the different bilingual compound types are presented in 5.2 and 5.3.

5.1 Overview of my data

The data for this study have been classified by two different perspectives: the thematic and the word-formational one. The thematic groups were already presented in Section 4. This article concentrates only on one part of the data, but first it is appropriate to present the ford-formational classification of my whole data. The classification is presented in Figure 1:

---

[1] http://www.khanty-yasang.ru/luima-seripos
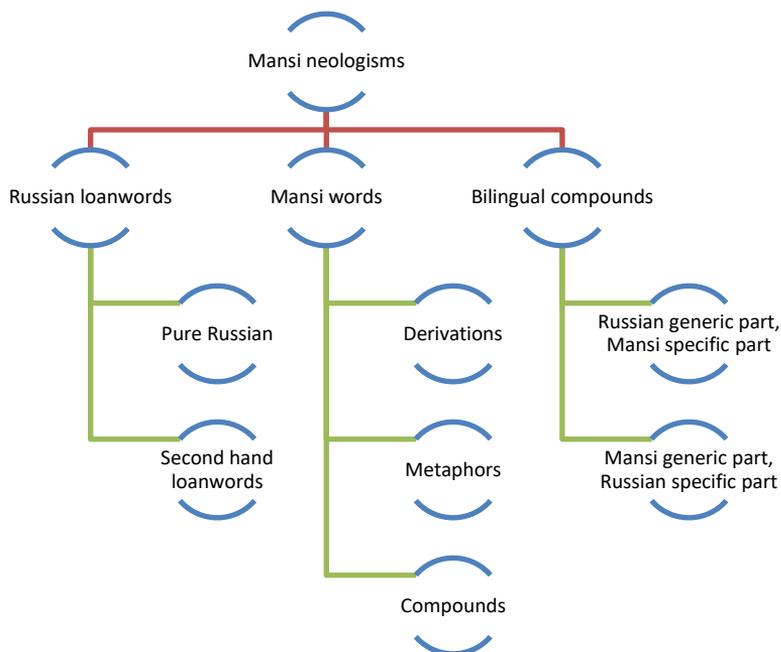
Figure 1. Word-formational classification of my data.

This study is an analysis of one part of the above mentioned data: I will concentrate on bilingual compounds (the right periphery of Figure 1). Bilingual compounds make up 39% of the whole data. They are mostly used in profession names, institution names, document names, other bureaucratical terms and medical terminology. I have divided the bilingual compounds into two types: the 'Russian specific part + Mansi generic part' type (see Section 5.2) appears much more often than the 'Mansi specific part + Russian general part' type (see Section 5.3). The first mentioned type can also be further divided into thematic sub-types.

The two types differ from each other both word-formationally and semantically. First, they differ in thematic groups connected to them. Secondly, there are clear differences between these two types especially in the way the generic parts behave in compounds. Both types are, however, productive. In the following subsections, I will present the most productive types and sub-types of bilingual compound in my data from both thematic and word-formational perspective. The similarities and differences between the two main types of compounds will be described in detail. Compounds with a Mansi generic part are presented in Section 5.2, and those with a Russian generic part in 5.3.

5.2. Russian specific part, Mansi general parts

In the following subsections 5.2.1–5.2.3 I will present three thematic groups (sub-types of one of the main types) that are connected to the 'Russian specific part + Mansi generic part' types compounds. These are compound document names (5.2.1), compound institution and building names (5.2.2) and compound profession names (5.2.3). The subsections are named by the thematic groups, and under these titles, also the word-formational properties are observed in detail.

5.2.1 Compound document names with a Russian specific part and the Mansi generic part *нэпак* 'book; letter'

One of the most productive bilingual compound types in my data are the document names created with a Russian specific part and the Mansi generic part *нэпак* 'book; letter'. The Russian specific part is always a Russian document name, and it remains unmodified.

All of the examples in this group are created in the same way, for example: *документ-нэпак* 'document', *полис-нэпак* 'insurance document', *виза-нэпак* 'visa', *диплом-нэпак* 'diploma, examination certificate', *договор-нэпак* 'contract', *заявка-нэпак* 'application', *заявление-нэпак* 'application', *квитанции-нэпак* 'receipt', *реквизит-нэпак* 'personal information form'.

Also some book names are created in the same way: *букварь-нэпак* 'ABC Book', *диссертация-нэпак* 'PhD Thesis', *атлас-нэпак* 'atlas'. Still, the formation methodology is exactly the same: a Russian term is adapted to Mansi language by turning it into a compound and adding a Mansi word as a generic part. Further, just as a side note and for comparison, also more typical bilingual compounds are found in the data, like *программа-нэпак* 'program booklet', *сāккон-нэпак* 'statute book' (Russ. *закон* > *сāккон*), where the specific part is specifying the generic part in a normal way.

A common feature for all examples in this group is that they include a Russian document (or book) name, which is turned into the specific part of a compound by adding the Mansi term *нэпак*. *Нэпак* serves as a generic part: from this point of view the specific part's function is to tell, what kind of document is in question. However, the meaning of the original Russian document name remains the meaning of the compound as well: semantically it already includes the meaning 'document'. In other words, the function of the Mansi generic part is more likely to adapt the term to Mansi language. The Mansi generic part gives us the information that there is talk about a document, while the same information is included in the specific part as well.

5.2.2 Compound names of institutions and buildings, with the Mansi generic part *кол* 'house'

In the same way as 'нэпак' in 5.2.1, also the Mansi word *кол* 'house' functions as a generic part of bilingual compounds: compounds produced with *кол* are names of institutions or concrete buildings. Names in this group include a Russian institution name or a name of a building as a specific part, and the Mansi term *кол* as a generic part. Further, both appellatives and proper names of institutions are included as specific part.

Appellatives appearing as specific parts typically refer to institutions maintained by society: *садик-кол* 'kindergarten' *библиотека-кол* 'library', *интернат-кол* 'dormitory', *театр-кол* 'theatre', *клуб-кол* 'club', *консерватория-кол* 'conservatory' *архив-кол* 'archive' *профсоюзов кол* 'trade union office' *радио кол* 'radio house', *спортивный яныг кол* 'sport center',

In addition to appellatives presented above, also many proper names are turned into compounds in the same way: *Соссаӈ мир культура кол* 'Indigenouse Peoples' Cultural Center', *Центр искусств кол* 'Art Centre', «*Институт развития образования*» *кол* 'Center for Development of Education', *Центр профессиональной патологии кол* 'Center for Occupational Pathology'. All of these are Russian proper names just refined with the Mansi word *кол*.

Again, the meaning of the Russian generic part remains the meaning of the whole compound. Technically the Mansi part is a generic part, and it includes the information that there is talk about a building or an institution, but the same meaning is included already in the specific part. Again, the Mansi generic part functions as an "adapter" to Mansi language.

5.2.3 Compound profession names with Mansi generic part *хӯтпа* 'person', *ӧйка* 'old man', *мāхум* 'people', *хум* 'man' or *нӭ* 'woman'

Third similar group to 5.2.1 and 5.2.2 are the compound profession names. They are produced with a Russian profession name as a specific part, and a Mansi human-related word as a generic part. Typical generic parts are *хӯтпа* 'person', *ӧйка* 'old man', *мāхум* 'people', *хум* 'man' and *нӭ* 'woman'. In other words, the generic part varies depending on the gender and the age of the professional. Naturally, there may be different variants of one profession name according to gender and age.

Some of the names are written with a dash, some not: there seems not to be any consistent practice for using or not using the dash. For example: *юрист-хӯтпа*

'lawyer', *учёный ōйка* 'male researcher', *жюри хōтпа* 'jury member', *яныг л̄эккар-хōтпа* 'senior physician' ['a big doctor-person'], *лэ̄ккар-ма̄хум* 'medical staff'.

This thematic group includes also compounds, where the specific part consists of a Russian noun and a participle form of a Mansi verb: Russian *кина ва̄рнэ ма̄хум* 'film industry people' ['people making films'].

5.3. Mansi specific part, Russian generic part

Compound words with a Mansi specific part and a Russian generic part differ structurally from the examples presented in 5.2.1–5.2.3. Further, they do not form as consistent and homogenous group as the three types in the previous subsections. A common feature for all examples also in this group is that the Russian generic parts of the compounds represent phenomena of modern society, adapted from the Russian culture.

The Russian generic parts are specified by Mansi terms. For example *сым а̄гм пусмалтан пӯльница* 'hospital for heart diseases', *ма̄-вōй нōх-винэ компаниян* 'oil drilling company', *щāнь лāтӊут ӯргалан Фонд* 'fund for protecting minority languages', *ханищтан программа* 'teaching programme', *ханищтап департамент* 'department of education', *сым пусмалтан лē̄ккар* 'oculist'.

The generic parts of the compounds presented in this subsection represent cultural phenomena adapted from Russians to Mansi society. Differently from the examples presented in the previous subsections, the roles of specific and generic parts of the compounds' follow the normal semantic structure of compounds. The Mansi specific part is really specifying the generic one: Russian general terms are specified by Mansi expressions. Old words in both old and new meanings are used.

5.4. The most productive compound types in brief:

To sum up, the Russian profession, institution or document names build up a special type of compounds (see 5.2.1–5.2.3). Russian names of professions, institutions and documents are turned into specific parts of compounds. All of them are added a Mansi generic part, which is semantically a quite simple and concrete phenomenon. The Russian terms are added a semantically simple Mansi word (generic part) to produce a compound, while the meaning of the compound remains the same as the original term's meaning in Russian. The function of the generic part is not to bring new information on the target but to adapt the term into Mansi language.

Is this a special word-formation type or a simple compound? Technically it is a very typical compound: the generic part expresses the general nature of the phenomenon, and the specific part gives more specific information on it. However, word-formationally and semantically the question is more complicated: the specific

part already includes the same information as the generic part. Also Hauel (1993) has observed similar lexical compounds in Khanty, when investigation Russian-origin loan translations in modern Khanty. Hauel's perspective and starting point are different from those in this study, but her data include examples similar to those in my data.

Hauel mentions Khanty profession names very similar to the Mansi examples presented above: also in Khanty, profession names are created with a human-related word like *нэ* 'woman', *хоят* person', *хе* 'man' and *ёт* 'people' as a generic part (Hauel 1993: 122). Also words *вер* 'work', *ут* 'thing' and *сир* 'manner' are used in compounds in the same way. Differently from the Mansi data presented here, they usually include a Khanty-origin specific part, or a Russian-origin word adapted to Khanty voice system. Hauel compares the appearance of loan translations (Lehnübersetzung), loan renditions (Lehnübertragung) and loan creations (Lehnschöpfung) (Hauel 1993: 123). All of these categories are not comparable with Mansi bilingual compounds, especially not with those with a Russian specific part and a Mansi generic part. In other words, many of them are nearer to the Mansi-origin derivations or compounds, which are not discussed in this partial study (but many appear in my data). However, the Khanty generic parts and their nature are interesting from the point of you of my study as well.

Hauel calls these generic parts of compounds 'half-suffixes' (Hauel 1993: 123). This refers to the generic parts' derivational nature: according to this point of view, they are not clearly parts of compounds but more likely derivational suffixes. The same idea can be applied to Mansi data as well: from one perspective, the Russian terms are adapted to Mansi language by a special way of derivation, by creating compounds, which actually are Mansi derivations of Russian-origin words.

It is good to remember that beside bilingual compounds, also monolingual Russian or Mansi synonyms are used. If we turn back to Figure 1, some phenomena may have equivalents in every branch of the figure, or at least in two of them. For example all three variants *мāнь няврамыт кол*, *садик* and *садик-кол* appear in the meaning of 'kindergarten'. Both *договор* and *договор-нӭпак* appear for 'contract'. Basing on this information, we can say that providing a derivation-like bilingual compound is not obligatory from any perspective. It is one way to bring new words to language, and get them better adapted, but not the only one.

Further, these data were gathered from newspapers. Newspaper language is more official and more formal than some other registers of communication. It is possible that some different kind of data would show different results. By investigating newspaper language we cannot find out, whether people use loanwords or bilingual compounds or something else for example in oral discussions or private letters.

6. Conclusions and further questions

The data presented in this study show that bilingual compounds appear especially in the field of society, professions and bureaucracy. The productive types of bilingual compounds in my data are:

1) "Mansi derivations" of Russian neologisms (69%), including document names with a Russian specific part and the Mansi generic part *нэпак*, institution names with a Russian specific part and the Mansi generic part *кол* and profession names with a Russian specific part and a Mansi generic part referring to human beings.

2) Institution and organization names with a Mansi specific part and a Russian generic part (21%).

3) Other types 10%.[2]

A common feature for all is that they are semantically related to phenomena adapted from the Russian society. Types 1 and 2 differ from each other word-formationally. In type 1, the function of a Mansi generic part is to adapt the Russian term into Mansi language: the meaning of the specific part becomes the meaning of the new compound as well. In type 2, a Russian generic part is specified by a Mansi specific part in a normal way.

Finally, to present some further questions and perspectives, the topic could be further elaborated by qualitative, quantitative or typological aspect. A qualitative approach includes the question of productivity of the compound types with a Mansi generic part like *нэпак*, *кол* or *хōтпа*: which other generic parts can this model be extended to? Also a quantitative approach on the topic will be useful: with how many specific parts do these generic parts appear?

Finally, a typological survey by comparing the Mansi data with other languages, also those with a different sociolinguistic status, would give many answers. Are there parallel situations in other languages of the world? Is this a feature of an endangered minority language only? Is this a feature of a typologically particular language? Do compounds like this appear in any other sociolinguistic contexts? These questions are topics of further studies, and investigating them will perhaps enlighten the motivation for creating untypical compounds in Mansi as well.

---

[2] These types or cases are not discussed here, because they are quantitatively marginal cases.

References

Bakró-Nagy, Marianne 2018. Mansi Loanword Phonology: a Historical Approach to the Typology of Repair Strategies in Russian Loanwords in Mansi. In: Bartos, Huba, den Dikken, Marcel, Bánréti, Zoltán and Váradi, Tamás (eds.), *Boundaries Crossed, at the Interfaces of Morphosyntax, Phonology, Pragmatics and Semantics.* Studies on Natural Language and Linguistic Theory 94.

Bolganbaev, A. 1988. Lexicology in Kazakh language. *Almaty Mektep.* 14–23.

Fischer, A. 1989. Aspects of Historical Lexicology. In: Fries, U., and Heusser, M. (eds.), *Meaning and Beyond: Ernst Leisi zum 70 Geburtstag*. Tübingen: Gunter Narr. 71–91.

Hauel, Petra 1993. Ostjakische Lehnübersetzung im Bereich der Nominalen Zusammensetzung. In: Bakró-Nagy, Marianne and Szíj, Enikő (eds.), *Hajdú Péter 70 éves*. Linguistica Series A. Studia et Dissertationes 15. Budapest: MTA Nyelvtudományi Intézet.

Helfrich, U. 1993. *Neologismen auf dem Prüfstand. Ein Modell zur Ermittlung der Akzeptanz französischer Neologismen.* Wilhelmsfeld: Gottfried Egert.

Häkkinen, Kaisa 1997. *Mistä sanat tulevat?* Tietolipas 117. Second edition. Helsinki: Finnish Literature Society.

Kálmán, Béla 1961. *Die russischen Lehnwörter im Wogulischen*. Budapest: Akadémiai kiadó.

Keresztes, László 1998. Mansi. In: Daniel Abondolo (ed.), *The Uralic languages*. London/New York: Routledge. 387–427.

Smyk-Bhattacharjee, Dorota 2009. *Lexical Innovation on the Internet – Neologisms in Blogs*. Doctoral Thesis. Faculty of Arts, University of Zurich. Steinitz, Wolfgang 1959. Zu den samojedischen Lehnwörtern im Ob-Ugrischen. *Ural-Altaische Jahrbücher* 31: 426–453.

Word and stem repetitions in the heroic epic songs collected by Antal Reguly

Mária Sipos

Hungarian Research Centre for Linguistics, Budapest

"Repetition, therefore, serves to conjure up a world
in which nothing is hurried:
actions are elaborated at a deliberate pace;
even in the heat of battle
the words are exchanged with a degree of formality.
[...]
This audience enjoys the poet's play on the patterns
they have come to know;
it is essential to their experience of oral song.
They find pleasure in traditional forms and traditional strategies;
and it gives them pleasure
to observe the poet working inventively with this same material."

(Minchin 2016: 27, 28)

1. Introduction

In November and December 1844, Antal Reguly collected 12 heroic epic songs from an old Khanty singer, Nikilov Maxim, and this collection constitutes the only documentation of the Sygva dialect. Of this folklore material, the heroic epic songs amounting to about 17,000 lines were processed, translated, and published by Hungarian scholars (ONGy, OH I, OH II, OH III/I, OHIII/II). Since publication, the epic songs have been investigated from various aspects, e.g. metrics (Austerlitz 1958, Schmidt 1990), linguistic issues (the grammaticalization of present and past participles, A. Jászó 1976), historical perspectives, typological subgroups of the genre (Demény 1977, 1978), cognitive mechanisms in parallel lines (couplets) (Bakró-Nagy 1985), and formulas (Widmer 2001), and were also used as a source for historical studies and cultural anthropology (Hatto 2017). But despite all this research, Nikilov

Maxim's heroic epic songs are still not widely known internationally in folklore research.

The heroic epic has been in the focus of attention for a long time because of what has been termed the Homeric question, which in fact consisted of several sub-questions, namely, how the *Iliad* and *Odyssey* came into being, whether they are works of a single poet or a number of contributors, how singers memorized long texts etc. Investigations carried out by Milman Parry and Albert B. Lord revealed that long poems can be transmitted orally, and the composition of folklore texts of this type takes place during the performance itself. These findings generated interest in the specifics of oral poetry, or rather, in verbal improvisation. Most research in the field has concentrated on the mnemotechnic function of repeated structural units, i.e. subformulas, formulas, type scenes, and rhetorical figures. Later this one-sided approach was criticized, claiming that numerous important aspects had been neglected, e.g. poetic and aesthetic functions, the significance of cultural background etc. (Delić 2015).

These days, there are investigations concentrating on, among other things, the cooperation of singer and audience. Minchin (2016) analyzes types of repetition in the Homeric epic, laying great emphasis on the interests of listeners processing long poems. She highlights that "oral song imposes far greater demands on both the poet and his audience than everyday talk imposes on a group of speakers" (Minchin 2016: 14). Regarding the singer, fluency is a natural requirement, the performance is monologic, and lasts typically for a long time. As far as the listeners are concerned, limitations of memory and attention constrain them also as they process the poem (Minchin 2016: 15). Minchin also emphasizes that the specific language of epic poetry using formulas and many other manifestations of repetition slow the presentation of new information and allow listeners to absorb it. "....[I]ndeed, capacity in short term memory is freed up so that the listener can focus on the storyline and absorb the poet's evaluative cues" (Minchin 2016: 16). These claims explain the importance of repetition techniques also in Ob-Ugric poetry.

Repetition is the most salient feature of the language of Ob-Ugric folklore, manifesting itself in various forms. Its significance was recognized already at the beginning of processing Khanty and Mansi folklore texts. Austerlitz began the summary of his findings in his fundamental monograph on Ob-Ugric metrics with the sentence "[t]he over-all feature of Ob-Ugric metrics is varied repetition" (Austerlitz 1958: 123). Although he concluded claiming that parallelism is the most common and characteristic type of repetition, in his work he gives a careful analysis also of etymological figures (Austerlitz 1958: 108–119).

Decades later, Schulze claimed that the technique of building etymological figures must have been vivid, and places an emphasis on the dichotomy of two mechanisms

during the process that is called composition-in-performance by folklorists, namely, invoking invariable *figurae etymologicae*, as well as the singers' freedom in creating and using them (Schulze 1975: 684).

The objective of the present paper is to describe some characteristic features of the *figura etymologica* used in northern Khanty folklore on the basis of the heroic epic songs collected by Antal Reguly. The investigation does not include a comprehensive description of all types of the *figura etymologica* as such classifications have already been written.

In this paper, the *figura etymologica*, along with other similar rhetoric schemes, is not considered to be a stylistic device aiming to produce a poetic effect. From a metrical point of view, the role of this type of repetition in the building up of lines has been discussed in the above mentioned monograph by Austerlitz (1958). In what follows, items of intra-sentential repetition are regarded as devices that can support singers in creating texts totaling hundreds of sentences, i.e. during composition-in-performance, on the one hand, and as features of the given folklore material that can ease the task of listening and comprehending on the other hand.

The main questions are the following: to what extent are *figurae etymologicae* fixed and static elements of the text which are called up as invariable units? and are there signs of flexible usage, which gives significant scope for the singer's creativity?

1.1. Terminology

Among the numerous rhetoric schemes, it is the *figura etymologica* that is most often mentioned when discussing Ob-Ugric intra-sentential repetition. It is typically defined as words of the same etymological derivation that are used relatively close (Todd 2009b: i) or in close proximity to each other (DCC) (1):

(1) *So long <u>lives</u> this, and this gives <u>life</u> to thee'* (Sonnet XVIII, Shakespeare)

In addition, for identifying the rhetoric devices of the Sygva Khanty heroic epic, *polyptoton* is also needed. Polyptoton is the repetition of the same word with various inflection (DCC) (2):

(2) *Then know that I, one Snug the joiner, am/ A <u>lion-fell</u>, nor else no <u>lion's</u> dam.* (Midsummer night's dream, Shakespeare)

In short, the main difference between the two is morphological: the constituents of a *figura etymologica* are derived from the same root, while those of a polyptoton are inflected forms of the same word.

There is a further type of repetition in the folklore texts in question that hardly needs any terminological clarification. It is *word repetition*, i.e. when a word occurs within the clause twice or more in the very same form but in different phrases.

1.2. Ordering of discussion

In Section 2, a quantitative survey will be discussed, concentrating on the frequency of various kinds of repetitions realized within the boundaries of a line. In Section 3, those types of etymological figures will be introduced that have proven to be characteristic of the Synja Khanty material. In addition, it is emphasized that verbal phrases allow more flexibility for creating etymological figures than noun phrases. Then, in Section 4, it will be shown that, regarding the actual realizations of formulas, the *figurae etymologicae* are often optional. Furthermore, it will be illustrated that they are not necessarily fossilized phrases but are adjusted to the actual semantic requirements of the sentence.

Quantitative data are given on the basis of the survey of *Jeli us eri* (OH 1: 166–486), however, examples will be given from various epic songs.

Khanty data will be quoted in the original form recorded by Reguly. Stem and word repetitions will be indicated with bold letters, other details are highlighted with underlining. As parallelism and the realizations of intra-sentential repetition do not inevitably coincide, parallel lines are not taken into consideration, i.e. either the first or the relevant line will be quoted, depending on the actual position of the repetition in question.

2. The proportion of lines containing *figurae etymologicae*

On the basis of a heroic song from the archaic subtype, and another one from the new pieces, the proportion of lines exhibiting the *figura etymologica* is between 10–15%.

POLM,[1] belonging to the archaic heroic songs, consists of 617 lines. Nineteen lines express characters' names, which are regularly multi-word phrases occupying a whole line, so the *figura etymologica* is not likely to occur in them. As there are 60 lines containing the *figura etymologica*, it results in about 10%.

Regarding JELI,[2] a heroic song of the later types, the total number of lines is 2,795. Names of characters are given in 202 lines, and 521 lines bear the *figura etymologica*, which amounts to 18.6% of the lines.

---

[1] POLM = *Polm torom ar*. A pelimvidéki isten éneke [Song of the god of the Pelim region]. ONGy 69–99.

[2] JELI = *Jeli us eri*. Jeli város éneke [Song of the village of Jeli]. OH I. 166–487.

Song	POLM	JELI
Number of lines	617	2,795
Number of lines describing heroes' names	19	202
Number of lines apart from enumeration	598	2,593
Lines containing figurae etymologicae	60	521
Percent	10%	20%

Table 1. Data of lines containing the figura etymologica

According to Todd (2009a), same-clause repetition of the same stem occurs roughly once every 100 lines in the *Iliad*, and once every 50 lines in the *Odyssey*. Converting the above percentage, in Reguly's heroic songs, the *figura etymologica* appears once every 10 (archaic heroic song), or once every 5 to 6 (new type) lines, which is several times its respective proportion in the Homeric epic. As sentences are divided into several lines, this proportion would be even more impressive if we referred to sentences.

In conclusion, regarding quantitative aspects, all the above numbers prove that the *figura etymologica* plays an important role in the Khanty oral epic tradition.

3. Types of the *figura etymologica* in the northern Khanty heroic epic

Steinitz described the most important types of *figura etymologica* found in northern Khanty dialects (Steinitz 1976, first published in 1941), namely, two nominal groups and a verbal one, to which some further subtypes were added later by Schulze (1975), as a result of her survey of the Sygva Khanty folklore texts collected by Reguly. Steinitz and Schulze both differentiated nominal and verbal types, as well as some minor subgroups. Their classifications give similar overviews, i.e. *figurae etymologicae* belonging to a noun phrase show a greater extent of regularity and fewer subgroups compared to those belonging to the verbal part of the sentence.

Widmer, in her monograph on the formulas found in a heroic epic material very similar to Reguly's heroic songs, provides a thorough analysis of the structure of formulas containing a *figura etymologica* as well (Widmer 2001: 134, 144). Widmer processed the material mainly in relation to the metrical structure of the lines, and in her work only the *figurae etymologicae* belonging to the nominal group are described. This type of formulas can be described with the term subformula as they can combine with various verbal parts resulting in formulas differing in content (Sipos 2019).

Austerlitz (1958) also specified several types of *figurae etymologicae*, revealing their syntactic background when needed. In his analysis, the proximity of repeated elements played an important role, and he focused mainly on *figurae etymologicae*

within noun phrases, sometimes discussing ones covering also the verbal part of the sentence.

He differentiated *figurae etymologicae* from pure repetitions (Austerlitz 1958: 109–119).

In sum, the analyses typically concentrate on the *figurae etymologicae* operating within a noun phrase, and, thus, much less importance was attached to nominal-verbal and participial-verbal ones. However, in the heroic epic song processed in the present paper, the proportion of the purely nominal *figurae etymologicae* and the remaining types suggest that these two main groups are of similar importance.

3.1. The *figura etymologica* in noun phrases

The nominal etymological figures essentially cluster around two basic patterns. This is in accordance with the results of Steinitz and Austerlitz, who processed large corpora and enumerated several types of the *figurae etymologicae*, which are not of the same abundancy in their corpora (Steinitz 1976, Austerlitz 1958). Regarding the Sygva Khanty epic material, not all of Austerlitz' groups are represented in it. In JELI, the epic song used as a corpus of qualitative analysis, there are 294 nominal *figurae etymologicae* including 258 ones belonging to Type [*f' A f''*], and 32 to Type [*A f' f'' B*],[3] leaving 4 to further small groups (3) (4):

(3) URTE[4]-0440

    ***sänz**-ing*   *vuăs*   ***sass**[5]-l-ä*   *jogot-l-em*
    back-DER$_{ADJ}$ town   back-3SG-LAT   arrive-PRS-1SG
    'I arrive at the back of the town that has a back part'

(4) ASSP[6]-0543

    *jang **peng**-pi*   ***peng**-eng*   *ungel*
    ten   tooth-DER$_{ADJ}$[1]   tooth-DER$_{ADJ}$[2]   mouth
    '(his) toothy mouth with ten teeth'

---

[3] [*f' A f''*] and [*A f' f'' B*] were introduced by Austerlitz to symbolize the structure of the *figurae etymologicae*; *f'* and *f''* indicate the changing forms of the repeated item.

[4] URTE = *Urt enmem arl*. A fejedelem felnevelkedésének éneke [Song of the prince's growing up]. OH II. 180–255.

[5] The nasal element of the stem final consonant cluster in the noun 'back' is deleted or not depending on the syllable structure of the inflected or derived word form.

[6] ASSP = *Ass pogol eri*. Ob-falu éneke [Song of the village Ob]. OH II. 3–119.

3.2. The *figura etymologica* in verb phrases

In contrast, the verbal *figurae etymologicae* have more subgroups and show greater syntactic diversity.

3.2.1. Participle–verb figures

In this group, the first part of the figure is a participle (present or past), the second element, in turn, is the verbal predicate of the sentence. The etymological figures exhibiting this pattern occur in 133 lines of the corpus. Participle–verb figures are very diverse even within their own subgroup, i.e. the participles appear in various syntactic functions, and they can be attribute to the object or any adverbial of the sentence (5)–(7):

(5) JELI 0305

***untl*ti**	*kur-l*	*jirr*	*koti*	***untt-l-*em**
know.PTCP.PRS	foot-3SG	joint	how	know-PRS-1SG

'This is how I recognize his foot joint to be recognized...'

(6) ASSP-0686

in	*untä*	*măsing*	*ord*	***kongolt*i**	*numang*	*joch*
now	until	perhaps	hero	make.climb.PTCP.PRS	branch.DER_{ADJ}	tree

***kongolt*ä-s-ä**
make.climb-PST-PASS.3SG

'So far, he may have been executed on the tree used to execute heroes.'

(7) JELI-1940

koleut	*kăr*	***kill*-ti**	*aling*	*suj*	*aling*
tomorrow	reindeer.bull	get.up-PTCP.PRS	morning	noise	earl

***kil*-ä**
get.up-IMP.SG2

'Tomorrow, wake up to the sound of reindeer getting up.'

There are numerous examples of the use of archaic participles[7] known exclusively from folklore texts (8):

---

[7] In these participles the suffix *-ŋən* is attached to the participles ending in *-t, -m,* and *-la.*

(8) PAUT[8]-1087

jas-l	*ke*	*tăgărle-l*		
hand-3SG	if	be.stuck-PRS.3SG		

lui+mang	*joch*	***veleg+tangen***	***velech-l-em***
sprayey(?)	tree	snag-PTCP	snag-PRS-1SG

'When his hands get stuck I cut them like twigs.'

This subtype is represented in 22 *figurae etymologicae*, and they typically express similes whose syntactic function is adverbial of manner.

3.2.2. Polyptoton

In a further type, the finite verb of the sentence does not actually participate in the repetition, although the stem repetition itself exceeds a noun phrase. This type may be qualified as a *polyptoton* because the morphological difference between the two parts of the figure lies in only a case suffix.[9] The scheme of this repetition type is the following: an adjective or a numeral appears first in its basic form, and then with a lative case suffix. This type is represented in 42 lines in the heroic song in question and is illustrated in (9) and (10):

(9) JELI-1355

put- *ing*	*lant*	*jäm*	*sop*	***put-*** *ing-ä*	*le-l-eu*
pot-DER_ADJ	food	good	bit	pot-DER_ADJ-LAT	eat-PRS-1PL

'We are eating the delicious bit of cauldron food from the cauldron'

(10) JELI-0137

lung-en	*katl*	*kab-ung*	*ki*	***lăbed***	*moi*
summer-2SG	day	boat-DER_ADJ	if	seven	wedding.guests

labt- *ä*	*kus*	*jăch-mel*
seven-LAT	although	go-EVID.PST.3PL

'although on a summer day, seven separate groups of wedding guests with boats paid a visit'

In many cases, it seems to describe a change of state: however, the initial and final features, conditions etc. are the same, i.e. the semantics in this type can definitely have a tautologic character (11).

---

[8] PAUT = *Puling-aut eri*. Obdorszki ének [Song of Obdorsk]. OH I. 2–165

[9] This was also claimed by Schulze (1975: 682), although she does not use the technical term *polyptoton*.

(11) JELI-1294:

juh	*uni*	***kolim***	*schelm-el*	***kolm*-a**	*tag-os*
wood	big	three	spall-3SG	three-LAT	rip-PST.3SG

'Three big spalls of the wood ripped into three.'

Nevertheless, as is illustrated in (12), adverbials of mode are normally derived from adjectives with the derivative suffix -*a*, which is historically identical with the lative case suffix -*a*.

(12) JELI-2553

put	*vur-i*	*ălng-el*	***put*-ä**
black	blood-DER_ADJ	end-3SG	black-LAT
temes-tal			
flow-EVID.PRS.3SG			

'His black blood flows blackly.'

3.2.3. Noun–verb figures

There are examples of *figura etymologica* in which the stem of the sentence final finite verb is repeated as a noun, either as a stem or in derived form. There are 20 occurrences of this type (13)–(15):

(13) JELI-0295

unt	*koj-i*	***noms*-em**	***nomes*-l-em**
forest	man-DER_ADJ	brain-1SG	think-PRS-1SG

'I'm speculating with my forest man brain.'

(14) JELI-1598

jas-l	*pelek*	*jäm*	***amt*ep**	*schirtnet*	***amt*ellyi-ly**
hand-3SG	half	good	exultation	then	rejoice-PRS.3SG

'Then he is doing a one-hand jubilation.'

(15) JELI-0486

lănge	*ko*	***ălm*-en**	*iet*[10]	*schi*	***ăl*-l-en**
ancestral.idol	man	dream-2SG	?	PTC	sleep-PRS-2SG

'You sleep an idol's sleep.'

3.3. Pure repetition

The following type is classified as *pure repetition* by Austerlitz (1958: 116), which differs from the *figura etymologica* in that the repeated words are attributes to two

---

[10] Unknown meaning.

different nouns. The number of lines containing this type of repetition is 50. It is illustrated with (16) and (17):

(16) JELI-1316

ku	*poter*	***ku***	*kas*	*ămes-l-eu*
long	discourse	long	joy	sit-PRS-1PL

'We solemnize the long joy of long discourse.'

(17) JELI-1856

labed	*mu*	*zung*[11]	***labed***	*ort-en*
seven	land	region	seven	hero-2SG

'your seven heroes from seven regions'

3.4. Further types

In addition to the above mentioned types, there are some cases in which the basic viewpoints are not enough to classify the clauses containing the *figura etymologica*, and there are patterns of lower incidence rate.

3.4.1. Triple *figurae etymologicae*

Triple *figurae etymologicae* are created by combining a verbal type with a nominal one (Schulze 1975: 684). The following pattern should be regarded as a Type [*f' A f''*] applied together with a [Noun + verb] stem repetition (18):

(18) JELI-0049

mui	***nem-****ing*	*ord*	***nem-****em*	***nem****i-l-ä*
what	name-DER$_{ADJ}$	hero	name-1SG	name-PRS-PASS.3SG

'By what name is my hero's name named / By what famous name am I mentioned?'

3.4.2. Mixed types

The above tautologic construction containing a lative case construction and classified as a polyptoton (see 3.2.2 above; cf. example 11) can be expressed with participial constructions (19):

---

[11] Reguly: *muzung*.

(19) JELI-1646

säss	*belä*	*käi-*ti		*ar*	*kăt-em*
back	toward	leave-PTCP.PRS		many	house-1SG

sass	*belä*	*kai*-temen
back	toward	leave-EVID.PRS.1DU

'We leave the houses to leave behind us behind us.'

The lative case construction can further be enriched with word repetition, resulting in a combination of a polyptoton and a pure word repetition, cf. 3.4.1 (20):

(20) JELI-1603

labed	*mu*	*zung*	*seu-l*		*senhl-em*	*labed*
seven	region	corner	braid-3SG		grow-PTCP.PST	seven

nai	*labt-*a	*latlti-mem*
woman	seven-LAT	make.fly-EVID.1SG

'I made fly seven women with braids and coming from seven regions.'

4. The *figura etymologica*: static or dynamic?

There are some phenomena suggesting that, although the text of oral epic is generally based on formulaic language, the *figura etymologica* should not be considered automatically a part of a fossilized construction in every case. In what follows, I give some examples of two features of the use of stem repetition that illustrate the above observation.

4.1. The *figura etymologica* and variability

There are formulas that seem very similar to each other in the sense that the etymological figures in them are arranged around the same words as central elements, e.g. 'boat', 'town', or 'house.

Intrinsically, they are possessive constructions with the same possessor but with a different possessee, which latter, with the adjective of the possessee, actually constitute stem repetition. These formulas are intended to express direction or other detailed information about the place of the activity denoted by the verb, e.g. 'onto the roof of a house', 'in the corner of the house', 'from the aft end of a boat', 'in the spacious inner part of the village', 'to the dock of the town/village' etc. In other words, the representatives of this type of construction seem to be a result of combining various techniques, i.e. applying the [Adj – Possessor – Possessee] construction and generating *figura etymologica* at the same time, so they are not necessarily static, invariable, three-word formulas. The following examples exhibit constructions in which the central element is a boat (21a–f).

(21a) JELI-0328

langl-en   kăp   **lăngl**-el   eult[12]
cover-DER_ADJ boat   cover-3SG   from
'from the cover of the boat having a cover'

(21b) JELI-0746

lipe-n   kăp   **lipe**-l-nä
inner.part-DER_ADJ   boat   inner.part-3SG-LOC
iln   long-tel
down   enter-EVID.PRS.3PL
'they go down into the boat having an inner part'

(21c) JELI-0804

seuss-eng   kăp   **seuss**-el-nä
wheel-DER_ADJ   boat   wheel-3SG-LOC
'at the wheel of the boat having a wheel'

(21d) JELI-0672

pus-en   kăp   **pus**-el[13]
aft-DER_ADJ boat   aft
'from behind the boat having a back part'

(21e) JELI-1432

jelpi   katl   sai-l-na   **nyol**-eng
previous   day   hidden.place-3SG-LOC   nose-DER_ADJ
kăp   **nyăl**-en   katl-em   ko
boat   nose-2SG   catch-PTCP.PST   man
'the man who tied up the boat having a forepart by its forepart the previous day'

(21f) JELI-1490

näng   jelpe-n   **siing**   kăp   **sing-**em[14]
2SG   before-2SG   crook.DER_ADJ   boat   crook-1SG
ke   katl-s-em
if   catch-PST-1SG
'if I had tied up my boat having a crook earlier than you'

The very same observations can be made about the subformulas containing 'house' or 'town' as the possessor of the construction. In other words, such sets of nominal

---

[12] Reguly: *lăngle leult*.

[13] In the Hungarian translation it is indicated that the ablative postposition (Reguly: *eult*) is expected to follow the noun (OH I: 242).

[14] Reguly: *singemke*.

formulas prove that the singer not only invoked fixed phrases but also applied rules for building phrases with the etymological figure creatively, according to the semantic needs of the sentence.

4.2. The *figura etymologica* and optionality

There are both nominal and verbal formulas that can appear either in a form containing the *figura etymologica*, or in a form without this rhetoric scheme. In the following examples, in each pair, (a) and (b) formulas exhibit the same syntactic pattern but (a) sentences include a stem repetition while (b) sentences do not (22a–25b):

(22a) JELI-1040

kaltn*-*äng   *vuăs*   ***kaltn*-*el-na***   *katt-l-eu*
port-DER_ADJ   town   port-3SG-LOC   land-PRS-1PL
'We land at the port of the town that has a port.'

(22b) JELI-1179

täm   *kaltn-äng*   *vuăs*   *ilpe-l-na*[15]
this   port-DER_ADJ   town   bottom-3SG-LOC
'at the bottom of the town having a port'

(23a) JEMA[16]-0357

i   *pelek*   *vuet*   ***lui*-*pi***   ***luj*-*ing***   *jăs*   *katt-l-em*
one   side   five   finger-DER_ADJ   finger-DER_ADJ   hand   grab-PRS-1SG.O
'I grab it with one of my fingery hands having five fingers'

(23b) UORT[17]-1089

vuet   *lui-pi*   *elyang*   *kurr*   *vuatte-l-em*
five   finger-DER_ADJ   shiny   foot   fix-PRS-1SG
'I fix it on my five-finger shiny foot.'

(24a) MUNK[18]-1073

sas   *vuazi*   ***ămes*-*tangen***   *iln*   ***ămes*-*tal***[19]
ermine   duck   sit-PTCP   down   sit-EVID.PRS.3SG
'[The old man *śokor* from the underworld] has dropped down similarly to a duck'

---

[15] Reguly: *il pelna.*
[16] JEMA = *Jeming Ass mui ar*. A szentséges Ob görbületének éneke [Song of the gut of the holy Ob river]. OH II. 120–180.
[17] UORT = *Uort ar. Sarnyu uort arl*. Fejedelemének. Az Aranyfejedelemnek ő éneke [Song of prince. Song of the Golden Prince]. OH III/I. 14–170.
[18] MUNK = *Munkess Kănt Tărom eri*. A munkeszi hadisten éneke [Song of the war god of Munkes]. ONGy 1–68.
[19] Reguly: *il nămestal.*

(24b) JELI-2205

jii-l *vuăs-em* *kul-ing* *pull*
father-3SG fish-PTCP.PST fish-DERADJ fish.trap

vuăs-tangen[20] *ămselt-em*
fish-PTCP seat-PTCP.PST

'(he) has dropped down as if he took a seat on his father's fish trap'

(25a) UORT-0753

mărti *mi* *ko* *ar* *neng-en* *ass*
southern region man many woman-2SG Ob

kaleu *rohung* **nyăch** *konnzä* **nyăch-*tel***
seagull shrieking laugh on.one's.back laugh-EVID.PRS.3PL

'The numerous women of the man from the southern region are shrieking like Ob gulls.'

(25b) MUNK-0129

männz-em *käenzsi* *lăuat* *ord* *ake-m-pa*
pluck-PTCP.PST teal[21] size.of hero uncle-1SG-in.turn

lăr *kaleu* *rohung* *nyăch* *konnzä* *vuel-tal*
lake seagull shrieking laugh on.one's.back make-EVID.PRS.3SG

'My uncle, who is of the size of a plucked teal, is making the lake gulls' shrieking laugh.'

5. Conclusion

Repetition has a special importance during composition-in-performance. In addition, it plays an important role in the processing of texts from the point of view of the audience, which means that the use of the *figura etymologica*, as a typical kind of repetition in oral poetry, is motivated from both sides.

In the history of investigations of oral formulaic poetry, it was formulas that generally stood in the center of attention, as they are the components that can easily be mobilized either in a frozen form or creatively adjusted to the needs of the text.

However, in the investigation of Ob-Ugric folklore texts, great emphasis has been placed on rhetorical schemes because it had been recognized that manifestations of repetitions on various levels play a very important role in this oral tradition.

Besides parallelism, the most important way of repetition in the heroic songs in question is the *figura etymologica*. Polyptoton and pure word repetition, which have different morphological and syntactic characteristics, produce stylistic effects similar

---

[20] Reguly: *vuăs tangen*.

[21] Eurasian teal, cf. Steinitz (1966–1988: 0525): Krickente, Anas crecca.

to the *figura etymologica*. In addition, they have the same motivations (metrics, technique of performance, role in cognitive processes), and they are possible to combine, consequently, these three rhetoric schemes were discussed together.

These three rhetoric schemes together exhibit a high incidence rate. In an archaic heroic song and another one from the new layers, they occur in 10–20% of the lines. Compared to the data concerning the Homeric epic, these numbers represent a manifold greater occurrence of the above mentioned patterns, proving their significance in the Ob-Ugric folklore.

It has already been noticed that although the use of the *figura etymologica* is regulated by various rules, at the same time, singers can apply these rhetoric patterns creatively. The paper has focused on the details of the dual nature of handling and creating etymological figures in the course of the performance.

The results of the investigation show that the *figurae etymologicae* in verbal phrases, evidently for syntactic reasons, show greater variability. In other words, the main difference between nominal and verbal etymological figures lies in the number of main types and subgroups. Although nominal etymological figures seem to be more fixed, flexibility or variability occur in them. In the [Adj + Possessor + Possessed] construction, the phrases containing e.g. 'town', 'boat, 'house' as a central element proved to be diverse: they can be easily modulated according to the requirements of the sentence, i.e. they can express any place, direction or actuality that is necessary for describing the event, by means of modifying their lexical constituents.

Besides variability, another important feature of the use of the *figura etymologica* is optionality. Numerous examples prove that the etymological figures can be detached from other structural or stylistic units. Formulas expressing essentially the same content can occur in a form with or without a *figura etymologica*.

In fact, the individual formulas or any kind of a syntactic structure are not necessarily accompanied by the *figura etymologica*, in other words, the *figura etymologica* is an optional device that operates according to its own rules.

As the above investigations were carried out on the basis of one epic song, these results can be completed and specified by means of processing all the Sygva Khanty heroic songs collected by Antal Reguly.

Abbreviations

1	first person
2	second person
3	third person
ADJ	adjective
ADV	adverb
DER	derivative suffix

DU	dual
EVID	evidential
LAT	lative
LOC	locative
PL	plural
PTC	particle
PTCL	participle
PRS	present
PST	past
SG	singular

References

A. Jászó, Anna 1976. Megjegyzések a particípiumból alakult verbum finitumok mondattanához az északi osztjákban [Notes on the syntax of finite verbs formed from participles in northern Khanty]. *Nyelvtudományi Közlemények,* 78: 353–358.

Austerlitz, Robert 1958. *Ob-Ugric metrics: The metrical structure of Ostyak and Vogul folk-poetry*. FFC 174. Helsinki: Suomalainen Tiedeakatemia.

Bakró-Nagy, Marianne 1985. A paralelizmus szemantikájáról (kognitív mechanizmusok a paralelizmusokban) [On the semantics of parallelism: Cognitive methanisms in parallelisms]. Műhelymunkák III. Budapest: MTA Nyelvtudományi Intézet.

Delić, Lidija 2015: Poetic grounds of epic formulae. In: Detelić, Mirjana and Delić, Lidija (eds), *Epic formula: A Balkan perspective*. Belgrade: Institute for Balkan Studies of the Serbian Academy of Sciences and Arts, 13–42.

DCC = Gildenhard, Ingo, Hodgson, Louise et al. (eds) Cicero / De Imperio Cn. Pompei. Dickinson College Commentaries. http://dcc.dickinson.edu/cicero-de-imperio/rhetorical-terms. Date of access: November 15, 2019.

Demény, István Pál 1977. Az északi osztják hősi énekek tipológiai elemzése [A typological analysis of northern Khanty heroic epic songs]. *Nyelv- és irodalomtudományi Közlemények*, 21: 33–43.

Demény, István Pál 1978. Az északi osztják hősi énekek történeti fejlődése [The historical development of northern Khanty heroic epic songs]. *Nyelv- és irodalomtudományi Közlemények*, 22: 147–158.

Hatto, Arthur 2017. *The world of the Khanty epic hero-princes: An exploration of a Siberian oral tradition*. Cambridge: Cambridge University Press.

Minchin, Elizabeth 2016. Repetition in Homeric epic: Cognitive and linguistic perspectives. In: Mihailo Antovic and Cristóbal Pagán Cánovas (eds): *Oral poetics and cognitive science*. Berlin – Boston: Walter de Gruyter, 12–29.

OH I. = *Osztják hősénekek. Reguly A. és Pápay J. hagyatéka*. Közzéteszi: Zsirai Miklós [Khanty heroic epic songs. The estates of A. Reguly and J. Pápay. Published by Miklós Zsirai]. 1944. Budapest: Magyar Tudományos Akadémia.

OH II. = *Osztják (chanti) hősénekek II. Reguly A. és Pápay J. hagyatéka.* Közzéteszi: Zsirai Miklós [Khanty heroic epic songs, vol. 2. The estates of A. Reguly and J. Pápay. Published by Miklós Zsirai]. 1951 Budapest: Akadémiai Kiadó.

OH III/I. = *Osztják (chanti) hősénekek III/I. Reguly A. és Pápay J. hagyatéka.* Zsirai Miklós hagyatékából közzéteszi Fokos Dávid [Khanty heroic epic songs, vol. 3/1. The estates of A. Reguly and J. Pápay. From the estate of Miklós Zsirai, published by Dávid Fokos]. 1963. Budapest: Akadémiai Kiadó.

OH III/II. = *Osztják (chanti) hősénekek III/II. Reguly A. és Pápay J. hagyatéka.* Zsirai Miklós hagyatékának felhasználával közzéteszi Fokos Dávid [Khanty heroic epic songs, vol. 3/2. The estates of A. Reguly and J. Pápay. From the estate of Miklós Zsirai, published by Dávid Fokos]. 1965. Akadémiai Kiadó, Budapest.

ONGy = *Osztják népköltési gyűjtemény. Reguly Antal hagyatéka és saját gyűjtései alapján közzéteszi Pápay József* [A collection of Khanty folk poetry. From the estate of Antal Reguly and his own collection, published by József Pápay]. 1905. Budapest – Leipzig: Hornyánszky Viktor – Hiersemann, Karl.

Schmidt, Éva 1990: Az osztják metrika másik oldaláról [The other side of Khanty metric]. *Nyelvtudományi Közlemények*, 91: 181–194.

Schulze, Brigitte 1975. *Zur ostjakischen etymologischen Figur*. CIFU III. Pars 1. 681–684.

Sipos, Mária 2019. *Formula és típusjelenet a Reguly-hősénekekben* [Formulas and typican scenes in the Reguly heroic epic songs]. Manuscript submitted for publication.

Steinitz, Wolfgang 1966–1988. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. Berlin: Akademie Verlag.

Steinitz, Wolfgang 1976. Zur ostjakischen und wogulischen Volksdichtung. In: Sauer, Gert and Schulze, Brigitte (eds), *Wolfgang Steinitz: Ostjakologische Arbeiten*. Band II. Budapest: Akadémiai Kiadó, 1–62.

Todd, C. Clary 2009a. Solemnity, banality and sarcasm: Provenances of the figura etymologica in Homer. https://camws.org/meeting/2009/program/abstracts/11E3.Clary.pdf. Date of access: November 20, 2019.

Todd, C. Clary 2009b, Rhetoric and repetition: The figura etymologica in Homeric epic. https://pdfs.semanticscholar.org/f272/95c6bfcda71d29886c825fb7436434e10 1d9.pdf. Date of access: November 18, 2019.

Widmer, Anna 2001. *Die poetischen Formeln der nordostjakischen Heldendichtung*. Veröffentlichungen der Societas Uralo-Altaica 53. Wiesbaden: Harrasowitz.

The use of body part terms in expressing emotions in Udmurt

Rebeka Kubitsch

University of Szeged

1. Introduction

The paper discusses the role of body part terms in expressing emotions in Udmurt from a cognitive linguistic point of view and within the framework of the embodiment hypothesis in particular. The paper discusses the use of body part terms in the conceptualization of HAPPINESS, SADNESS, ANGER, FEAR, AND LOVE.

2. Conceptual metaphors and metonymies

A general tendency of human thinking is that more abstract domains are understood in terms of more concrete, tangible ones, for instance, metaphors. Conceptual metaphors are a set of system correspondences between two different conceptual domains (Kövecses 2010). The source domain is typically more concrete, and the target domain is more abstract, therefore the target domain is understood in terms of the source domain. There are three types of conceptual metaphors: structural, orientational, and ontological (Lakoff and Johnson 1980). In the case of structural metaphors, a concept is structured in terms of another (e.g. AN ARGUMENT IS A BUILDING). Orientational metaphors organize a system of concepts in relation to one another (e.g. HAPPY IS UP, while SAD IS DOWN). Ontological metaphors concern physical objects and substances representing abstract concepts as if they had physical properties (e.g. THE HUMAN BODY IS A CONTAINER).

Similarity is considered the basis of metaphor, but in addition to the similarity between the two conceptual domains, metaphors can be based on human experience and on various kinds of biological and cultural roots. For example, in AN ARGUMENT IS A BUILDING structural metaphor, the stability and structural similarity of a building is mapped on the foundation, strength and construction of an argument, cf. *She constructed a solid argument*, or *We have got a good foundation for the argument*. However, in the case of ANGER IS A HOT FLUID IN CONTAINER, the basis of this ontological metaphor is bodily experience, cf. *boiling with anger; simmer down; making one's blood boil* (Kövecses 2010: 54, 79, 92).

Conceptual metonymies provide mental access through one conceptual entity (vehicle) to another conceptual entity (target). The entities are a part of the same conceptual domain, or, in other words, the same idealized cognitive model (*icm*), hence metonymy is based on the relationship of contiguity (Kövecses 2010: 173–174). There are two main configurations which apply to different sets of *icm*: (i) THE WHOLE STANDS FOR A PART, or A PART STANDS FOR THE WHOLE; and (ii) A PART STANDS FOR ANOTHER PART. The first version typically gives rise to metonymies involving things (e.g. Thing-and-Part *icm*: THE WHOLE FOR THE PART metonymy), the second version applies to metonymies involving events (e.g. Causation *icm*: EFFECT FOR CAUSE) (Kövecses and Radden 2007).

3. The embodiment cognition thesis and emotions

The embodiment cognition thesis emphasizes that the human body has a substantial role in thinking. Our conceptual organization and structure of thinking are built on our physical, embodied experiences (cf. Evans 2007, Kövecses and Benczes 2010). In other words, according to the what has been called the embodiment premise:

> "Human language and thought emerge from recurring patterns of embodied activity that constrain ongoing intelligent behavior. We must not assume cognition to be purely internal, symbolic, computational, and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action." (Gibbs: 2005: 9)

Also, since the body in not culture independent, sociocultural factors also stimulate human language and cognition.

The human body often constitutes as a source domain for other, more abstract concepts (Kraska-Szlenk 2014: 8–12). One example can be the consistent mapping of body parts on the corresponding part of an object, e.g. Eng. *legs of the table*; Fi. *pullonsuu* 'mouth of the bottle'; Hu. *a cipő nyelve* 'tongue of a shoe'; Ru. *rukav rubashki* 'shirtsleeve' (lit. arm of the shirt). The human body is one of the most common source domains for conceptual metaphors and many metaphorical meanings derive from our bodily experiences (Kövecses 2010: 18).

The concept of emotion is a superior target domain (Kövecses 2010: 23). The cognitive model of emotion in folk theory (Kövecses 2000: 58–59) can be characterized as a five-stage scenario:

(1) cause → (2) emotion → (3) control → (4) loss of control → (5) (behavioral) response.

Thus, emotion is a caused change of state, from a non-emotional state to an emotional one. Emotion is assumed to affect the self, which may try to control emotional behavior. The self may respond to the emotion by undergoing emotional behavior,

rather than acting as a willful agent. Furthermore, emotions are subcategorized in the following way: states (cf. change of state), events (cf. behavioral response), actions (cf. control), and passions (cf. effect on the self).

The human body is often referred to when emotions are being described (Enfield and Wierzbicka 2002; Sharifian et al. 2008). In the description of emotions, a high degree of universality can be observed in the languages of the world. A reason for this universality is that the structure of the human body and the functions of the body parts are nearly the same everywhere. Another reason is the metonymic conceptualization and verbalization of the biological and physiological reactions to different kind of emotions (Kövecses 2000: 164). It is to be expected that cultural factors have great influence on the description and conceptualization of emotions as well.

Various forms of embodiment can be observed in respect to emotional concepts and emotional language (Kövecses 2000: 116–117). In the case of emotion metonymies, bodily responses are associated with emotion concepts, e.g. the physiological effects and expressive responses of an emotion stands for the emotion (more generally EFFECTS OF A STATE STAND FOR THE STATE) – BLUSHING FOR LOVE (*She blushed when she saw him*) (Kövecses 2000: 124, 134).

Considering emotional metaphors, two image schemas have significant roles: the CONTAINER schema and the FORCE schema. In the CONTAINER schema, emotions are events/states that happen inside the human body, which serves as a container (cf. the ontological metaphor THE HUMAN BODY IS A CONTAINER), e.g. HAPPINESS IS A FLUID IN A CONTAINER (*She was bursting with joy*) (Kövecses 2010: 111). In the FORCE schema it is considered that certain causes produce emotions, and emotions produce certain responses, e.g. SADNESS IS A NATURAL FORCE (*Waves of depression came over him*) (Kövecses 2000: 25) This schema roots in the generic level conceptual metaphor CAUSES ARE FORCES (cf. the folk theory of emotions).

In connection with the CONTAINER schema it can be concluded that emotions are located in the body (Kraska-Szlenk 2014: 40; Sharifian et al. 2008: 13), and a particular body part is considered the center of emotional activity in many languages (THE LOCUS OF EMOTIONS IS A PARTICULAR BODY PART). Not only the center of emotional, but the center of mental activity is also often associated with a certain body part. Typically, internal body organs are the source of such conceptualizations. There are three types of conceptualizations: abdomen-centering, holistic heart-centering and dualistic heart/head-centering conceptualizations. They can be linked to certain cultural models; consider the dual heart/head-centering and the religious (Christian) and philosophical (cf. Plato, Descartes) cultural models on distinguishing rational head and emotional heart in Western cultures (Sharifian et al. 2008).

4. Research material and methodology

I discuss conceptual metaphors and metonymies involving body part terms in respect to emotion categories, illustrating then with figures about the proportions of certain body-part terms occurring in the data.

I have examined five categories of emotion: HAPPINESS, SADNESS, ANGER, FEAR AND LOVE. These emotion categories can be considered general and perhaps universal (Kövecses 2000: 4). It is important to mention that emotions are not considered to be feelings exclusively. Emotion concepts have a complex structure, which consists of conceptual metaphors and metonymies associated with the emotion concept in question, related concepts and also cognitive models (Kövecses 1990: 40). This results in the "broader" interpretation of the abovementioned categories (cf. the folk theory of emotions, in which the cause and the behavioral and physiological responses are also part of the emotion). For example, the category of HAPPINESS will also include BEING CONTENT, GLAD, HAPPY, FEELING GOOD, DELIGHTED, ECSTATIC, etc.

The data used for this paper comprises 138 phrases and idioms with 142 occurrences of body part terms.[1] Expressions are taken from the Kozmács's 2002 dictionary, and from two dictionaries of phrases and idioms, Dzjuina (1967) and (1996). The material of the online Udmurt corpus[2] is also included in the analysis to some extent: collocational searches have been carried out based on the typology literature, particularly on Kövecses (1990) and (2000) and on Baş (2015). The classification of phrases and idioms is conducted according to the typology literature as well (Kövecses 1990, 2000, 2010).

5. Body part terms in conceptual metaphors and metonymies in expressing emotions

The present section reviews the conceptual metaphors and metonymies related to HAPPINESS, SADNESS, ANGER, FEAR and LOVE. Reflections on typological findings are made as well.

5.1. Happiness

In the data under consideration, 22 expressions stand for the emotion HAPPINESS. Of them, 14 (64%) are metaphors, and 8 (36%) are metonymies. The most frequent body part terms are *śulem* 'heart' (8 instances; 36%) and *śin* 'eye' (7; 31%) (cf. Figure 1).

---

[1] The occurrence of body part terms is higher than the number of phrases because there are some expressions that may refer to more than one emotion concept.

[2] http://web-corpora.net/UdmurtCorpus/search/ Date of access: January 19, 2020.

In conceptual metaphors the source domains are the following: FLUID/SUBSTANCE IN A CONTAINER (4 instances), FIRE (3), BEING OFF THE GROUND (2), LIGHT (1), WARM (1), UP (1), LIVING ORGANISM (1) and FORCE (1). The source domains BEING OFF THE GROUND and LIGHT seem to apply to HAPPINESS only, not just in Udmurt, but, for example, in English as well (cf. Kövecses 2000: 39). In expressions of HAPPINESS the CONTAINER metaphor is mostly connected to *śulem* 'heart', while the FIRE and LIGHT metaphors to *śin* 'eye'. Some examples of metaphorical expressions are presented below.

HAPPINESS IS A FLUID/SUBSTANCE IN A CONTAINER
(1) *šumpotonen śulemy paćyl tyrmiz* (lit. my heart **fills to the brim** with happiness) 'I am extremely happy'
(2) *śulem šumpotonen paćylme* (lit. the heart **overflows** with happiness) 'being extremely happy'
(3) *köt puškaz źeć esep ĺukaloz* (lit. one hoards up good mood to their stomach/belly) 'something good came up'

HAPPINESS IS FIRE
(4) *śinjosyz tyl kiśto* (lit. one's eyes **pour fire**) 'being euphoric'
(5) *šumpotysa ӡuaś śinjos* (lit. the eyes are **burning** with happiness) 'being extremely happy'

HAPPY IS LIGHT
(6) *ymnyryz pište šumpotonen* (lit. one's face **glows with happiness**) 'there is a glow of happiness on one's face'

HAPPY IS BEING OFF THE GROUND
(7) *syče šumpoti, pydjosy muzjemez ik öz šöde ńi* (lit. I was so happy; my feet **did not feel the ground**) 'I was very happy'
(8) *oj, šumpoti, pydy muzjeme ik ug jöty* (lit. oh, I was happy, my foot **did not touch the ground**) 'I was very happy'

From the conceptual metonymies, SMILING FOR HAPPINESS (2 instances) and BRIGHT EYES FOR HAPPINESS[3] (4) were the most frequent. Both can be traced back to the general type of conceptual metonymy where the EXPRESSIVE RESPONSES OF EMOTION STAND FOR THE EMOTION (7 instances altogether). The metonymy AGITATION/EXCITEMENT FOR HAPPINESS (1) can also be found, although only in one case in the data. This is a specific instance of THE PHYSIOLOGICAL EFFECTS OF EMOTION STANDING FOR THE EMOTION general metonymy. Such contiguity between expressive

---

[3] Instances of this conceptual metonymy are related to the metaphor HAPPY IS LIGHT as well.

responses/physiological effects and emotions is widespread, since these are substantial metonymic principles in our conceptual system (Kövecses 2000: 134) and can be encountered throughout the whole data.

THE EXPRESSIVE RESPONSES FOR AN EMOTION STAND FOR THE EMOTION

(9) **ymnyryz šaṅgi śiem bere kadʼ** (lit. one's **face looks like** they were eating *šaṅgi*[4]) 'glowing with happiness'

BRIGHT EYES FOR HAPPINESS

(10) **śinmyz ćilja, šude** (lit. one's eye is shining, playing) 'one's eyes are shining with happiness'

(11) **śinjosyz solen zoro** (lit. one's eyes are shiny) 'one's eyes are shining with happiness'

SMILING FOR HAPPINESS

(12) **ymyz peĺdoraz vuemyn** (lit. one's **mouth** got to one's temple) 'having a big smile'

Figure 1 shows the distribution of body part terms in the conceptual metaphors and metonymies considering HAPPINESS. The word *śulem* 'heart' takes a significant role in metaphoric expressions (seven out of 14; 50%), whereas the face and its parts (cf. *śin* 'eyes', *ym* 'mouth', *ymdur* 'lips') make up almost all instances (seven out of eight) of the metonymic ones. The term *śin* 'eye' appears not only in metonymies but it is exclusive for the HAPPINESS IS FIRE metaphor.
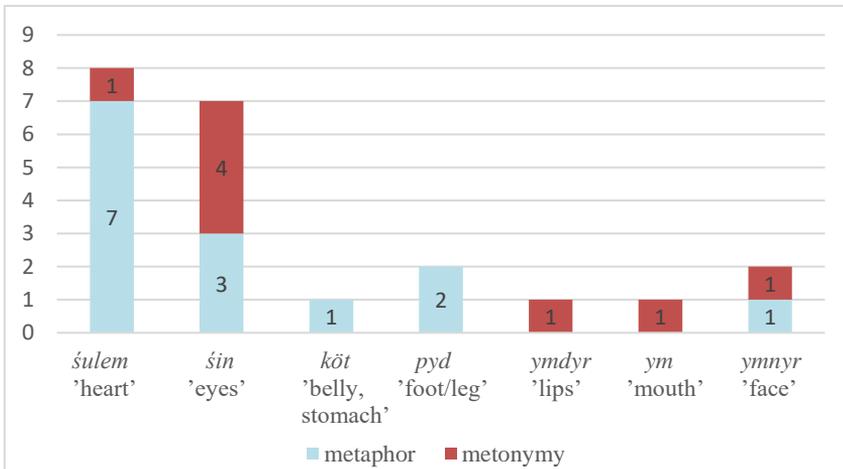

Figure 1. Body part terms in conceptual metaphors and metonymies for HAPPINESS

---

[4] *Šaṅgi* is an Udmurt pastry.

Figure 2 demonstrates the conceptual metaphors and metonymies in relation to the concept of HAPPINESS in the research material. The upper part of the figure shows the metaphors, the lower part the metonymies.

SUSBSTANCE IN A CONTAINER (4)	BEING OFF THE GROUND (2)	FIRE (3)	WARM (1)	LIVING ORGANISM (1)
		LIGHT (1)	FORCE (1)	UP (1)

HAPPINESS

OTHER EXPRESSIVE RESPONSE (1)	BRIGHT EYES (4)	SMILING (2)	AGITATION/EXCITEMENT EYES (1)
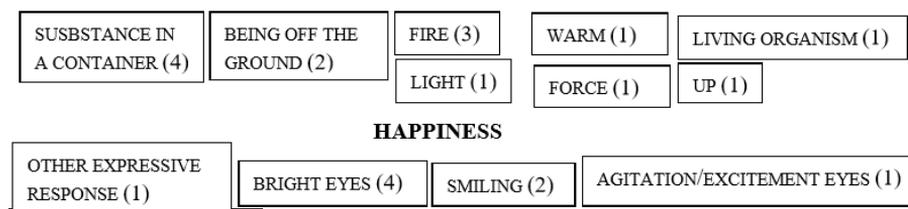
Figure 2. Conceptual metaphors and metonymies for HAPPINESS

5.2. Sadness

Phrases and idioms expressing SADNESS and its related concepts (like DISTRESS, WORRY, MISERY, FEELING BAD, UNHAPPINESS) make up the largest proportion of the research material (40 expressions). Just as in the case of HAPPINESS, the term *śulem* occurred in the highest numbers and proportions (25 instances; 62.5%). Of the 40 expressions, 25 are metaphors (62.5%), and 15 are metonymies (37.5%).

The source domains of metaphors for the concept of SADNESS are PHYSICAL AGITATION (9), PHYSICAL DAMAGE (4), FIRE/HEAT (3), LIVING ORGANISM (2), DOWN (2), CONTAINER (2), NATURAL FORCE (1), BLOCKING SUBSTANCE (1), and BURDEN (1). Although this paper does not intend to make a comparison between Udmurt and English metaphors and emotional concepts, it is noteworthy to mention the relatively high number of instances of SADNESS IS PHYSICAL AGITATION and SADNESS IS FIRE, since this type of conceptual metaphor is not mentioned in respect to English (cf. Kövecses 2000), but it can be well attested in Turkish[5] (cf. Baş 2015: 164, 168 SADNESS/DISTRESS IS PHYSICAL ANNOYANCE/AGITATION; SADNESS/DISTRESS IS FIRE). SADNESS IS PHYSICAL AGITATION can be linked to the generic level metaphors, which are EMOTION IS PHYSICAL AGITATION and EMOTIONAL DISTURBANCE IS PHYSICAL AGITATION, however, emotional disturbance stands metonymically for emotion. Considering the emotion scenario, this metaphor represents "cause → emotion" part: sadness is a disturbed state of mind that arises from some cause (Kövecses 2000: 80–81).

It is also important to mention the orientational metaphors HAPPY IS UP – SAD IS DOWN. Such oppositions like up vs. down, light vs. dark, and warm vs. cold focus on

---

[5] A crucial difference, however, between Turkish and Udmurt in respect to such metaphoric expressions is that in Turkish the physical agitation/annoyance affects the 'head', while in Udmurt it affects the 'heart' (cf. the metonymies HEAD FOR PERSON – HEART FOR PERSON).

the positive vs. negative evaluation of an emotion (like happiness as positive, sadness as negative) (Kövecses 2000: 44).

SADNESS IS PHYSICAL AGITATION

(13) *śulem ćutyrtyny* (lit. the heart **cramps**) 'feel sorrow, being really worried'

(14) *śulem syzla* (lit. the heart **stings**) 'being sad'

(15) *śulme mertćyny* (lit. to **grasp**, to **pierce** the heart) 'to make one sad'

SADNESS IS PHYSICAL DAMAGE (cf. EMOTIONAL DAMAGE IS PHYSICAL DAMAGE)

(17) *śulem putkoške* (lit. the heart **shatters**) 'feel great sadness'

(18) *śulem piliškoz* (lit. the heart will **break**) 'feel great sadness'

SADNESS IS FIRE/HEAT

(19) *śulem ƶua* (lit. the heart **ignites**) 'being sad'

(20) *śulem pyže* (lit. the heart **fries**) 'being sad'

SAD IS DOWN

(21) *jyr ošyny* (lit. **hang** the head) 'having a long face'

SADNESS IS A FLUID IN A CONTAINER

(22) *kurektonen paćylmem śinjos* (lit. the eyes **overflow** with misery) 'sad, miserable'

Emotion metonymies of the generic level metonymy EXPRESSIVE AND PHYSIOLOGICAL EFFECTS OF EMOTION STAND FOR THE EMOTION can be attested in the case of SADNESS as well. The specific level metonymies are PHYSICAL AGITATION FOR SADNESS (7), CRYING FOR SADNESS (5 instances), and INABILITY TO MOVE (2). There is one instance of expressive response that does not stand for SADNESS itself but rather for CRYING (which implies SADNESS, cf. example 28). The metonymy PHYSICAL AGITATION FOR SADNESS is also rooted in the generic level metaphor EMOTION IS PHYSICAL AGITATION. However, this metonymy represents the "emotion → response" part of the emotion scenario. Therefore, agitation does not come from a disturbed state of mind (as in the case of SADNESS IS PHYSICAL AGITATION), but it is a physiological response which arises from an emotion. As can be seen, almost as many instances of PHYSICAL AGITATION FOR SADNESS (7) are attested in the research material as SADNESS IS PHYSICAL AGITATION (9).[6]

---

[6] In the classification of certain Udmurt instances two strategies were applied. One was consultation with a native speaker and examining the use of phrases and idioms in the online Udmurt corpus. The other was analyzing the phrase: if it contained verbs meaning *lue, luiz* 'become, became', *pote* 'turn into something' (as the heart) or if it was explicit that the SADNESS

PHYSICAL AGITATION FOR SADNESS

(23) *mözmon śulemez **ćityr-ćutyr kare*** (lit. sadness **makes** the **heart crisscross**)
'being worried, distressed'

(24) *śulmyz **giźi-gaźi**[7] **luiz*** (lit. the heart **became constricted**) 'living in anxiety'

CRYING FOR SADNESS

(25) ***bördonaz** ymduraz* (lit. one's **crying** is on the lips) 'one is about to cry'

(26) *śinjos **vuźekto*** (lit. the eyes **fog up**) 'weeps, cries'

INABILITY TO MOVE

(27) *sujpyd **ug žutsky*** (lit. one's limbs **do not rise**) 'unable to move because of
sadness'

EXPRESSIVE RESPONSES FOR THE EMOTION STAND FOR THE EMOTION

(28) *śinkaš **ńebźemyn*** (lit. the eyebrow is **limp**) 'one gets teary-eyed'

In Figure 3, illustrating the distribution of body part terms in conceptualizing
SADNESS, *śulem* 'heart' stands out with its 25 occurrences. It is noteworthy to mention
that while in western cultures the heart is often associated with LOVE, in Udmurt it is
rather associated with SADNESS and DISTRESS. Consider the lexicalizations involving
śulem 'heart': *śulmany* 'to be sad, to dispear'; *śulmaśkyny* 'to take care, to worry';
śulemšug (lit. heart + trouble) 'worry, distress' (Kozmács 2002: 469).

"itself" caused the agitation (cf. example 25) the phrase was considered a metonymic
expression.

[7] Onomatopoetic.
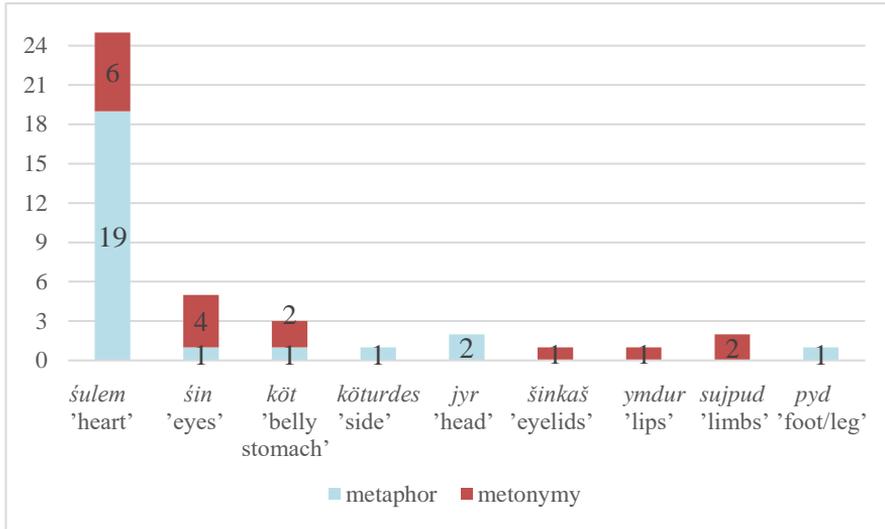
Figure 3. Body part terms in conceptual metaphors and metonymies for SADNESS

Figure 4 demonstrates the conceptual metaphors and metonymies in relation to the concept of SADNESS in the research material. The upper part of the figure shows the metaphors, the lower part the metonymies.
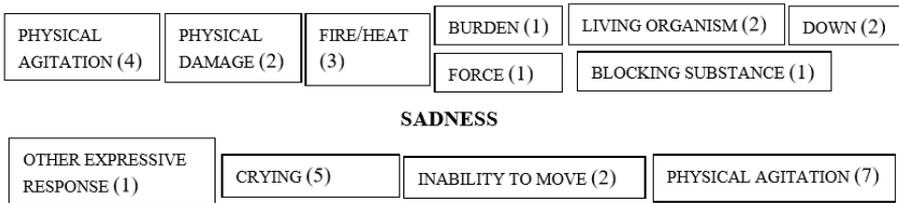
Figure 4. Conceptual metaphors and metonymies for SADNESS

5.3. Anger

In the data ANGER and related emotions (such as ANNOYANCE, IRRITATION, FURY, BEING UPSET etc.) were represented in 29 expressions. The most frequent body part terms were *śulem* 'heart' (7 instances; 24%), *jyr* 'head' (4; 14%), and *piń* 'tooth' (4; 14%). The number of metonymical (15; 52%) expressions was slightly higher than the metaphorical ones (14; 48%).

In conceptual metaphors the source domains were the following: EXPLOSION (3), LIVING ORGANISM (3), FIRE (3), INSANITY (2), OPPONENT (1), DARK (1), and FORCE (1).

All these source domains are typical for most emotion concepts. The ANGER IS A HOT SUBSTANCE/FLUID IN A CONTAINER metaphor is considered central in the conceptualization of ANGER in various languages and cultures (Kövecses 2000: 146, 154 Baş 2015: 177). The Udmurt expressions did not explicitly show this type of conceptualization, but ANGER IS EXPLOSION can be seen as a subtype of the CONTAINER metaphor, since when anger becomes too intense, the person explodes, as intense anger is putting pressure on the container (Kövecses 2000: 150). In this sense the ANGER IS A HOT SUBSTANCE/FLUID IN A CONTAINER and THE ANGRY PERSON IS A PRESSURIZED CONTAINER metaphors apply to Udmurt as well.[8] The cross-culturally similar conceptualization of ANGER can be traced back to the physiological responses given to anger (e.g. higher blood pressure).

ANGER IS EXPLOSION
(29) *sepez **pöśtem*** (lit. one's gall **erupts**) 'to bristle up'
(30) *vožde en vaj – pekĺajed **pöśtoz*** (lit. don't get angry – your kidney will **erupt**) 'don't get angry'
(31) *jyr **puštyny*** (lit. the head **explodes**) 'to get angry'

ANGER IS FIRE
(32) *köt **ǯua*** (lit. stomach/belly flames up) 'getting angry'
(33) *sinjosyz **ǯualskizy*** (lit. one's eyes lit up) 'getting mad'

ANGER IS INSANITY
(34) *jyr **yštyny*** (lit. **lose** the head) 'get angry'
(35) *jyrzy kuren **urmizy*** (lit. their head **got rabies** with anger) 'they got into rage'

The conceptualization of ANGER in relation to the head is also apparent in Turkish (Baş 2015: 106) and in Hungarian (Kövecses 2000: 149). Considering the compound verb *jyrkur* (lit. head + shame) 'anger' (Kozmács 2002: 172) the term is clearly associated with ANGER. Moreover, in Udmurt *jyr* 'head' is the locus of mental activity (Kubitsch 2019: 365) – its appearance in the conceptual metaphor ANGER IS INSANITY (and, in my opinion, the lexicalization denoting 'anger') illustrate the stage of loss of control in the cognitive model of emotion: the self attempts to control the emotion by the locus of mental activity but does not succeed (cf. LACK OF EMOTIONAL CONTROL IS INSANITY).

---

[8] In the data under consideration the only evidence of the substance being hot is the verb *pöśtyny* 'to erupt', which can be linked to the adjective *pöś* 'hot, occurring in 2 of the 3 EXPLOSION metaphors.

The relatively high proportion of metonymic expressions (52%) compared to other emotion concepts (36% for HAPPINESS and 37.5% for SADNESS) suggests that Udmurt conceptualizes ANGER through the responses given to the emotion, which is the final stage of the emotion model of the folk theory.

Subtypes of the general emotion metonymies PHYSIOLOGICAL/BEHAVIORAL/ EXPRESSIVE RESPONSE FOR THE EMOTION STANDS FOR THE EMOTION[9] can be attested. The specific-level metonymies are PHYSICAL AGITATION (1), VIOLENT, FRUSTRATED BEHAVIOR (10), AGGRESSIVE VERBAL BEHAVIOR (1) and DISAPPROVING GLAZE (2).

BEHAVIORAL RESPONSE FOR THE EMOTION STANDS FOR THE EMOTION

VIOLENT, FRUSTRATED BEHAVIOR STANDS FOR ANGER

(35) *pińjosses **jyrjyny*** (lit. **gnaw** the teeth) 'be angry'

(36) *ku vylyś kuze **ńijoz*** (lit. one will **strip** the skin from the skin) 'being very angry with sb'

(37) *vač́e jyrśi **vuyny*** (lit. **got to** each other's hair) 'to fall out with sb'

(38) *jyrde **kurtćozy*** (lit. they will bite your head off) 'speak to sb angrily'

AGGRESSIVE VERBAL BEHAVIOR STANDS FOR ANGER

(39) ***ym dyrti šuky** ponysa* (lit. putting **foam through the mouth**) 'shouting angrily'

EXPRESSIVE RESPONSE FOR THE EMOTION STANDS FOR THE EMOTION

DISAPPROVING GLAZE FOR ANGER

(40) ***kyryž śin** ućkyny* (lit. watching **crooked eye**) 'look angrily

(41) *śinmynyz **śie*** (lit. one **eats** with their eye) 'being very angry with sb'

The metonymy VIOLENT, FRUSTRATED BEHAVIOR STANDS FOR ANGER involves the body part term *piń* 'tooth' in four cases and *ku* 'skin' in three cases. The appearance of these terms is unique in the data for conceptualizing ANGER. Interestingly, in Mari (which is culturally close to Udmurt), the conceptualization of ANGER may also appear through the term *püj* 'tooth'.

(42) *iktaž-kö βaštareš **püjĝm šumaš*** (lit. one is **sharpening tooth** to sb) 'to be out to get sb; to bear a grudge against sb'[10]

Figure 5 illustrates the distribution of body part terms in relation to ANGER. As it shows, a great variety of body part terms is involved in the linguistic expressions of

---

[9] According to Kövecses (2000: 159), conceptualized physiology (conceptual metonymies) provides cognitive motivation for metaphoric conceptualizations of the angry person as a PRESSURIZED CONTAINER.

[10] The source is the online Mari–English Dictionary: https://www.univie.ac.at/maridict/site-2014/index.php?int=0 – Date of access: January 22, 2020.

the concept of ANGER. The occurrence of the term 'heart' is not as outstanding (24%) as it was in the case of SADNESS (62%) and as it is in the case of LOVE (cf. 4.5), below. In the conceptualization of ANGER, the terms *jyr* 'head' and *piń* 'tooth' also play a relatively great role.
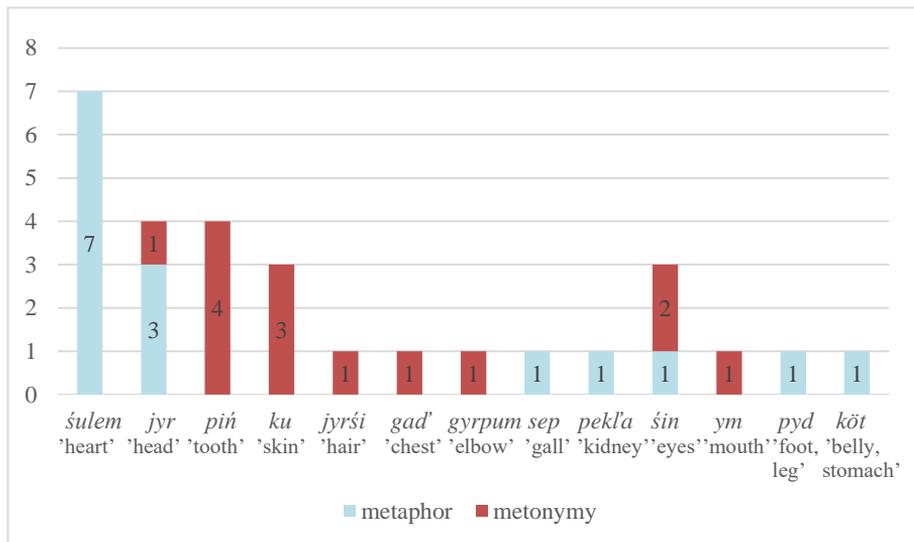
Figure 5. Body part terms in conceptual metaphors and metonymies for ANGER

Figure 6 shows the conceptual metaphors and metonymies in relation to the concept of ANGER in the research material. The upper part of the figure demonstrates the metaphors, the lower part the metonymies.
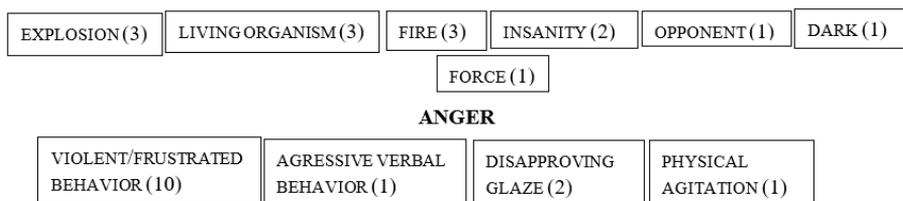
Figure 6. Conceptual metaphors and metonymies for ANGER

5.4. Fear

In the data 28 expressions referred to the emotion concept of FEAR. It is an emotion which seems to be characterized "by a rich system of its physiological effects and behavioral reactions" (Kövecses 1990: 70). This claim applies to Udmurt, since only 8 of 28 instances (29%) are metaphorical expressions, therefore 71% of the cases (20) are metonymic. In the conceptualization of FEAR the term *śulem* 'heart' has a

significant role (10 instances, 36%), but the term *pyd* 'foot' (5; 18%) also appears in a relatively large proportion.

As has been mentioned, only a handful of metaphoric expressions can be attested in the data under consideration. The source domains are the following: COLD (2), LIVING ORGANISM (2), FORCE (2) FOG (1), and RAPTURE/HIGH (1). The source domain of LIVING ORGANISM is considered to apply to all emotional concepts (cf. EMOTION IS A LIVING ORGANISM) and the domain of FORCE applies to most emotion concepts (Kövecses 2000: 36). These source domains appear in the conceptualization of all the other discussed emotion concepts in Udmurt. Therefore, they can be considered quite general for various emotions. The source domains COLD[11] and FOG, however, are unique for FEAR in Udmurt, and they can be grouped together as FEAR IS A NATURAL PHENOMENON. The motivation for FEAR IS COLD is probably the similarity in their physiological symptoms (trembling and shaking). The verb used in the expressions of FEAR IS FOG is *tölźyny* 'to dissolve' – this verb is used for the dissolution of fog as well. Also, other phrases, which are not included in the data, since they do not involve body part terms, refer to FEAR as fog. Consider the following expression: *kyškan bus kad' tölźoz* (lit. fear will dissolve like fog) 'one will not fear anymore'. According to Kövecses, FEAR appears to be characterized by both very general and specific metaphors (Kövecses 2000: 23). This claim seems to be true for Udmurt.

FEAR IS A NATURAL PHENOMENON

FEAR IS COLD

(43) *śulem ik **yrak-yrak** lue* (lit. the heart becomes **cool**) 'to gasp, to get scared'
(44) *śulemjosmy **dyrekjazy***[12] (lit. our heart **trembled**) 'we were afraid'

FEAR IS FOG

(45) *śulemyśtyz kyškan öz tölźy* (lit. fear **did not dissolve** from one's heart) 'one was still afraid'

FEAR IS A LIVING ORGANISM

(46) *śulemazy kyškan **karjaśkemyn*** (lit. fear **has nestled** in one's heart) 'one is afraid of sth'
(47) *śinjosazy kyškan **karjaśkemyn*** (lit. fear **has nestled** in their eyes) 'they are afraid of sth'

FEAR IS A FORCE DISLOCATING THE HEART

(48) *śulmy **tetćiśke** val* (lit. my heart **jumped**) 'I startled'

---

[11] The conceptual metaphor FEAR IS COLD is attested in Turkish as well (Baş 2015: 191).
[12] According to Kozmács's Udmurt dictionary (2002: 113), this verb specifically refers to trembling with cold.

In the data under consideration a great variety of specific metonymies of the PHYSIOLOGICAL EFFECTS OF EMOTION STANDS FOR THE EMOTION generic-level metonymy can be attested. These are the following: INABILITY TO STAND (6), LAPSES IN HEARTBEAT (3), HAIR STRAIGHTENS OUT (4), PHYSICAL AGITATION (2), WAY OF LOOKING (2), SWEATING (2), and BLOOD LEAVES FACE (1). The body part term *pyd* 'foot, leg' appeared in most conceptual metonymies (cf. INABILITY TO STAND). The term *śulem* 'heart' was encountered in four cases – the term rarely appears in metonymic expressions (in six cases of SADNESS and in one case of HAPPINESS), since terms of inner body organs tend to take part in metaphoric expressions.

INABILITY TO STAND

(49) **pydesśinjos** *kwalekjalo* (lit. the **kneecaps** tremble) 'trembles with fear'

(50) **pydjosyz** *žiĺnaśke* (lit. one's **legs** weaken) 'weaken by fear'

(51) **pydjos** *kaltyralo* (lit. the **legs** tremble) 'trembles by fear'

LAPSES IN HEARTBEAT STANDS FOR FEAR

(52) *śulem* **kwaĺaz** (lit. the heart **trembles**) 'one is startled'

(53) *śulem* **ńorkak** *kare* (lit. the heart **makes a large beat**) 'one is startled'

HAIR STRAIGHTENS OUT FOR FEAR

(54) **jyrśi** *vyllań* **žutske** (lit. **hair rises**) 'got scared'

(55) **jyrśiosyz** *ik* **žabyrskizy** (lit. one's **hair spears up**) 'one's got scared'

WAY OF LOOKING STANDS FOR FEAR

(56) **śinjosaz kyškan** *šödiśke* (lit. fear is felt in one's eyes) 'one is being afraid of sth'

BLOOD LEAVES FACE STANDS FOR FEAR

(57) **ymnyryz köd-köd** *luem* (lit. the **face** became **pale**) 'got scared'

Figure 7 illustrates the distribution of body part terms in relation to the concept of FEAR. Since FEAR typically conceptualizes through metonymies, body part terms appearing in the data do not denote inner organs but refer to the visible parts of the human body (i.e. face or hair), which function as a display of emotions (cf. section 5).
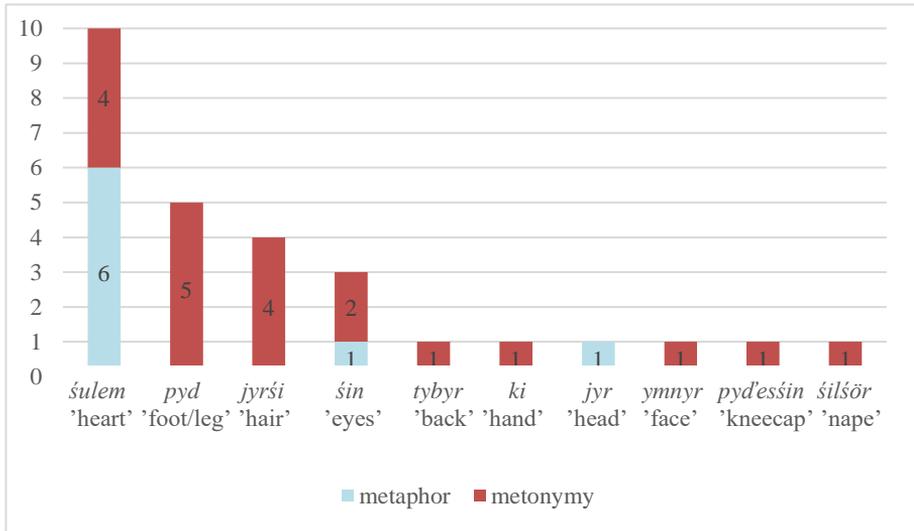
Figure 7. Body part terms in conceptual metaphors and metonymies for FEAR

Figure 8 illustrates the conceptual metaphors and metonymies in relation to the concept of FEAR in the research material. The upper part of the figure shows the metaphors, the lower part the metonymies.
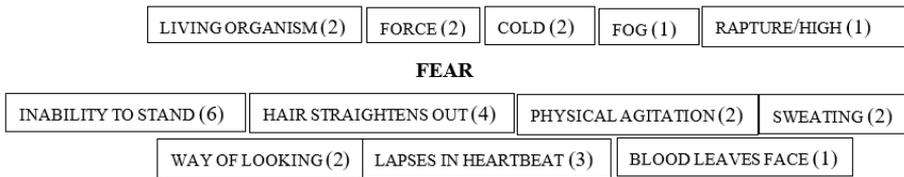
Figure 8. Conceptual metaphors and metonymies for FEAR

5.5. Love

LOVE and its related concepts (such as ROMANTIC LOVE, LIKING, and AFFECTION) are represented in 23 cases in the research material. As many as 21 out of 23 (91%) are

metaphorical expressions, and two (9%) are metonymical ones. According to Kövecses (2000: 27), "the concept of love is the most highly 'metaphorized' emotion concept".

A great variety of source domains are encountered in the 21 metaphoric expressions, which are the following: SUBSTANCE/FLUID IN A CONTAINER (6), FIRE (5), PHYSICAL FORCE (2), RAPTURE/HIGH (2), LIVING ORGANISM (1), WARMTH (1), INSANITY (1), THE OBJECT OF LOVE IS A VALUABLE OBJECT (1), THE OBJECT OF LOVE IS IN THE HEART (1), and PHYSICAL AGITATION (1). The source domains related to HEAT (WARMTH and FIRE) are especially common in the conceptualization of PASSIONS (Kövecses 2010: 21). The CONTAINER metaphor also has significant role in the conceptualization of LOVE (just like in the case of HAPPINESS and inexplicitly in the case of ANGER). It roots in the generic level metaphor BODY IS A CONTAINER FOR THE EMOTIONS (Kövecses 2010: 370). In Udmurt the container of LOVE (and the container of other emotions as well cf. section 5) is *śulem* 'heart'. The body part term *śulem* 'heart' appears in almost all of the conceptual metaphors with the exception of LOVE IS INSANITY and LOVE IS RAPTURE/HIGH – in these cases the term *jyr* 'head' is attested. The LOCUS OF MENTAL ACTIVITY is conceptualized in the head – the person under the influence of love is viewed as someone whose mental faculty functions improperly. Just like in the case of ANGER, these metaphors represent the stage of loss of control in the cognitive model of emotion; extreme anger or love can interfere with one's normal mental functioning.

LOVE IS A SUBSTANCE/FLUID IN A CONTAINER (IN THE HEART)

(58) *śulemyz jaratonen **paćylmemyn*** (lit. one's heart **overflows** with love) 'be in love'
(59) *śulemyz no lulyz **tyrmemyn** jaratonen* (lit. one's heart and soul is **filled** with love) 'be in love'
(60) *solen śulemaz jaraton **pyčam*** (lit. love **soaks** one's heart) 'be in love'

LOVE IS FIRE

(61) *śulem **ǯua** jaratonen* (lit. the **heart** burns with love) 'be crazy in love'
(62) *śulemam **tyl aratid** (lit.* you **lit fire** in my heart) 'I fell in love with you'
(63) *piosmurtlen śulemaz jaraton **kenǯytoz*** (lit. sb **ignites** love in the man's heart) 'sth made the man to fall in love'

LOVE IS PHYSICAL FORCE

(64) *śulem **kyskyny*** (lit. the heart **pulls**) 'to long for sth or sb, to prefer sth or sb'

LOVE IS RAPTURE/HIGH

(65) *jaratonen **poromem** jyr* (lit**.** the head is **dizzy** with love) 'be confused by love'

LOVE IS INSANITY

(66) *jyrze yštytoź šinmaśkem* (lit. **losing** one's **head** because of love) 'be crazy in love'

THE OBJECT OF LOVE IS A VALUABLE OBJECT

(67) *ki vylyn nullyny* (lit. **carry on** one's hand) 'to love, appreciate sb'

In the data under consideration there were only two metonymic expressions related to the concept of LOVE. These are LOVING VISUAL BEHAVIOR and INABILITY TO SPEAK, which are specific level metonymies of BEHAVIORAL REACTIONS OF THE EMOTION STAND FOR THE EMOTION.

LOVING VISUAL BEHAVIOR STANDS FOR LOVE

(68) *jaratiś murtlen šinmyz no veraśke* (lit. even the **eye speaks** of a person who is in love) 'love is seen in the eyes'

INABILITY TO SPEAK STANDS FOR LOVE

(69) *jaratiś murtlen kylyz myrk* (lit. the **tongue** is **dumb** of a person who is in love) 'unable to speak because of strong affection'

Figure 9 shows the distribution of body part terms in the expressions for LOVE. As it can be seen, despite the great variety of conceptual metaphors, only a few body part terms play role in the conceptualization of LOVE, and *śulem* 'heart' make up 70% (16 instances) of all the appearances.
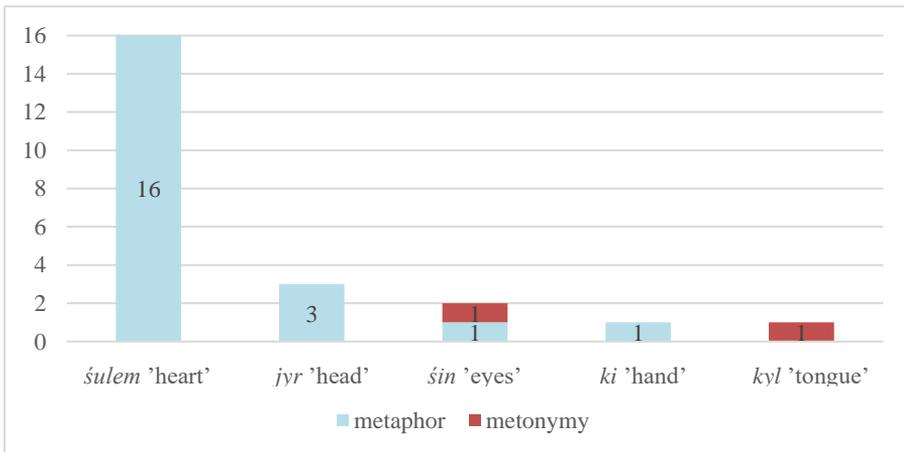

Figure 9. Body part terms in conceptual metaphors and metonymies for LOVE

Figure 10 shows the conceptual metaphors and metonymies in relation to the concept of LOVE in the research material. The upper part of the figure illustrates the metaphors, the lower part the metonymies.

SUBSTANCE IN A CONTAINER (6)	FIRE (5)	WARMTH (1)	OBJECT OF LOVE IS A VALUABLE OBJECT (1)	OBJECT OF LOVE IS IN THE HEART (1)
	RAPTURE (2)	FORCE (2)		
INSANITY (1)			LIVING ORGANISM (1)	PHYSICAL AGITATION (1)

LOVE

LOVING VISUAL BEHAVIOR (1)	INABILITY TO SPEAK (1)
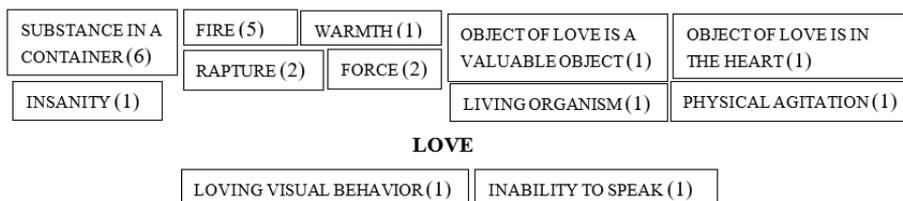
Figure 10. Conceptual metaphors and metonymies for LOVE

6. Distribution of body part terms in the data

Figure 11 shows the distribution of body part terms in all the examined emotion expressions. The total number of occurrences is 142. The term *śulem* 'heart' has a significant role in the conceptualization of HAPPINESS, SADNESS, ANGER, FEAR, and LOVE (66 instances, 46%). The term typically takes part in metaphoric mappings. In the discussed emotion concepts *śulem* 'heart' is either the container of the emotion in question (cf. happiness, love) or it is affected by the emotion (cf. sadness, fear). Based on the fact that *śulem* 'heart' makes up almost half of all the body part term appearances and that all the discussed emotion concepts conceptualize primarily through metaphors and metonymies which involve this term, it can be concluded that the center of emotional activity in Udmurt is in the heart: LOCUS OF EMOTIONS IS HEART.

The second most common body part term is *śin* 'eye' (19 instances, 13%). It principally takes part in the conceptualization of HAPPINESS (BRIGHT EYES FOR HAPPINESS and HAPPINESS IS LIGHT) and SADNESS (CRYING FOR SADNESS), but it appears in the expressions of all emotion concepts at a lower degree. As Yu (2004: 665) claims, although the face is the most distinctive part of a person, both physically and socially the focus is really on the eyes, which are commonly viewed as a "window to the soul" in many cultures (Kövecses 2010: 218).

The term *jyr* 'head' (10 instances, 7%) mainly appeared in metaphoric expressions related to ANGER and LOVE, since in Udmurt the LOCUS OF MENTAL ACTIVITY IS HEAD, which functions as a controller of emotions (Kubitsch 2019: 365). Considering the cognitive model of emotions implemented in the study, the self may try to control the emotion (cf. section 2). Expressions with the term *jyr* 'head' focus on lack of control (within the control aspect of the emotion concept), and this results in conceptualizations like LACK OF CONTROL IS INSANITY or LACK OF CONTROL IS RAPTURE/HIGH (Kövecses 2000: 43).

The term *pyd* 'foot' (9 instances, 6%) primarily takes part in the metaphoric conceptualization of HAPPINESS (cf. HAPPY IS BEING OFF THE GROUND) and in the metonymic conceptualization of FEAR (cf. INABILITY TO STAND FOR FEAR). It appeared twice altogether, once in relation to the concept of SADNESS (SADNESS IS NATURAL FORCE that sweeps sb off their feet) and once in relation to ANGER (ANGER IS AN OPPONENT which cannot be trodden down).

Although the body part term *köt* 'stomach, belly' does not have a high occurrence rate (5 instances, 3.5%), it is worthwhile to mention that dictionaries provide 'heart – locus of emotions' as a secondary meaning for this lexeme (Munkácsi 1896; Kozmács 2002). As can be seen, it has a role in the conceptualization of emotions, but it is not nearly as significant as *śulem* 'heart'. It appears in relation to SADNESS (SADNESS IS PHYSICAL AGITATION), in three instances in relation to ANGER (ANGER IS FIRE), and in relation to HAPPINESS (HAPPINESS IS A FLUID/SUBSTANCE IN A CONTAINER) in one instance. Considering the derivations and compound words (Kozmács 2002: 229–230) including *köt* 'stomach, belly', it is related to SADNESS and ANGER: *köttemmyny* (lit. become one without stomach, belly) 'losing hope'; *kötkajgu* (lit. stomach, belly + sadness) 'sadness'; *kötžož* (lit. stomach, belly + complaint) 'anger, being offended'; *kötšug* (lit. stomach, belly + difficulty) 'sadness, distress'.

From the other body parts the occurrence of *piń* 'tooth' (4 instances) is noteworthy on the one hand because it appeared exclusively in relation to the concept of ANGER, on the other hand because similar conceptualizations can be encountered in the Mari language as well.
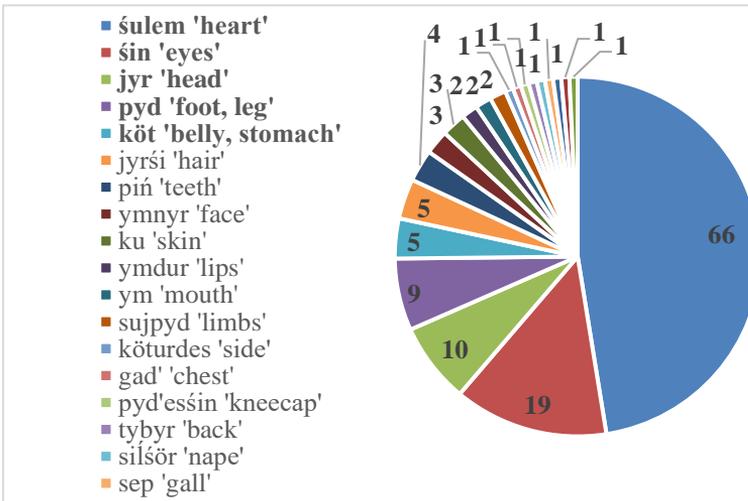

Figure 11. Distribution of body part terms in the entire data

Furthermore, it is important to mention that there are 20 metonymical expressions including the terms referring to the face or its parts (including the 12 metonymic uses

of *śin* 'eye'). This is rooted in the physical distinctiveness and social significance of face and eyes as it was highlighted before. The generic level metonymies PHYSIOLOGICAL EFFECTS OF EMOTION STANDS FOR THE EMOTION and EXPRESSIVE RESPONSES FOR EMOTION STANDS FOR THE EMOTION involves many phrases and idioms related to facial expressions. Because of this, the face can be considered the display of emotions.

7. Review of central metaphors in the research material

Reviewing the conceptual metaphors related to HAPPINESS, HAPPINESS IS A FLUID/SUBSTANCE IN A CONTAINER and HAPPY IS FIRE seem to be central (section 4.1, Figure 2). In the case of SADNESS frequent metaphors are SADNESS IS PHYSICAL AGITATION and SADNESS IS PHYSICAL DAMAGE, which primarily agitates/damages the heart. The metaphor SADNESS IS FIRE/HEAT also has a relatively high occurrence rate in the data under consideration (cf. section 4.2, Figure 4). Interestingly, FIRE does not seem to be a source domain either for HAPPINESS or for SADNESS in English, neither the source domains PHYSICAL AGITATION and PHYSICAL DAMAGE are in relation to SADNESS (cf. Kövecses 2000), but they are well attested in Turkish and seem to be just as substantial as they are in Udmurt (Baş 2015: 160).

Anger is mainly conceptualized as ANGER IS EXPLOSION (cf. ANGRY PERSON IS A PRESSURIZED CONTAINER) and ANGER IS A LIVING ORGANISM that nestles in the heart (cf. section 4.3, Figure 6). In the case of FEAR there was not such a conceptual metaphor which can be considered as central (at least considering metaphors including body part terms). FEAR IS COLD, FEAR IS A LIVING ORGANISM and FEAR IS FORCE appear equally (cf. section 4.4, Figure 8). FEAR is the only discussed emotion concept comprising more metonymic than metaphoric expressions. However, it is known that FEAR tends to be characterized rather by its physiological effects and behavioral reactions (Kövecses 1990: 70). In the case of LOVE, LOVE IS A SUBSTANCE/FLUID IN A CONTAINER (IN THE HEART) and LOVE IS FIRE are the central metaphors (cf. section 4.5, Figure 10).

The CONTAINER metaphor is central in the case of HAPPINESS, ANGER and LOVE, but it also appeared in the case of SADNESS as well. Conceptualizing emotions as being fluids or substances in a container roots in the generic level metaphor HUMAN BODY IS A CONTAINER FOR THE EMOTIONS (Kövecses 2000: 217). In the case of Udmurt, emotions are often viewed as fluids, but they are not necessarily so – generally, EMOTIONS ARE SUBSTANCES INSIDE A PERSON (HUMAN BODY), which serves as a container.

Another central metaphor in more emotion concepts is EMOTION IS FIRE. It takes part in the conceptualization of SADNESS, ANGER and LOVE and it highlights the intensity aspect of emotion concepts. This can be characterized by the INTENSITY OF

EMOTION IS HEAT metaphor. The metaphor has application beyond the domain of emotion as well (INTENSITY IS HEAT) (Kövecses 2000: 41–42). The significance of the intensity aspect of emotion concepts are seen in such metaphoric expressions which can refer to either LOVE, ANGER, SADNESS or HAPPINESS (cf. example 70), so basically to emotions that are intense.

(70) *śulem ǯua* (lit. the heart **ignites**) 'get angry/be in love/be sad/being euphoric'

The source domain FORCE also appeared with all emotion concepts. It is typical to view emotions as forces (cf. EMOTIONS ARE FORCES). The EMOTION IS A LIVING ORGANISM metaphor is attested in relation to all discussed emotion concepts to a higher (SADNESS, ANGER) or lower (HAPPINESS, LOVE, FEAR) degree. These source domains focus on the passivity of the self in relation to the emotion, since emotions are viewed as happening to us (Kövecses 2000: 42).

8. Conclusion

This paper has overviewed the role of body part terms in the conceptualization of HAPPINESS, SADNESS, ANGER, FEAR and LOVE in Udmurt. The results show that, similarly to other languages and cultures, body part terms (and the human body) play a significant role in the conceptualization of emotions.

Taking into account the metaphorical mappings, it can be concluded that the two central metaphors for emotions are EMOTIONS ARE SUBSTANCES IN A CONTAINER and EMOTION IS FIRE. Considering conceptual metonymies, all three types of generic level metonymies are attested in the data: PHYSIOLOGICAL EFFECTS OF EMOTION STANDS FOR THE EMOTION; BEHAVIORAL RESPONSES FOR EMOTION STANDS FOR EMOTION, and EXPRESSIVE RESPONSES FOR EMOTION STANDS FOR THE EMOTION. The proportion of metonymic expressions is not even, for instance, FEAR is rather conceptualized through metonymies, whereas LOVE through metaphors.

Considering body part terms in metaphoric and metonymic expressions in relation to the discussed emotion concepts it can be concluded that the term *śulem* 'heart' is substantial in the linguistic expressions of emotion concepts in Udmurt (cf. section 5), therefore LOCUS OF EMOTION IS HEART. This correlates with the typological findings, whereby typically inner body organs are conceptualized for the locus of emotional activity (Sharifian et al. 2008: 13). Inner body organs primarily take part in metaphoric mappings. The occurrence of *śin* 'eye' and other terms related to the face and its parts is frequent as well. The face is often viewed as a display of emotions (Kraska-Szlenk 2014: 43) and, because of this, it is usually involved in metonymic expressions.

Finally, although providing a systematic comparison of the conceptual metaphors and metonymies used for certain emotion concepts between Udmurt and other languages is beyond the scope of this paper, the results suggest that several emotion

concepts of Udmurt are characterized similarly to Turkish emotion concepts (cf. SADNESS IS PHYSICAL AGITATION/DAMAGE/FIRE; FEAR IS COLD or conceptualization of ANGER in relation to the head).

References

Baş, Melike. 2015. *Conceptualization of emotion through body part idioms in Turkish: A cognitive linguistic study*. Ankara: Hacettepe University Graduate School of Social Sciences PhD dissertation.

Dzjuina, Klavdia N. 1967. *Kratkij udmurtsko-russkij frazeologicheskij slovar*. Izhevsk: Udmurtia.

Dzjuina, Klavdia N. 1996. *Sredstva obraznovo vyrazhenyja v udmurtskom jazyke*. Izhevsk: Udmurtia.

Enfield, Nick. J. and Wierzbicka, Anna. 2002. The body in description of emotion. *Pragmatics and Cognition*, 10(1/2): 1–25.

Gibbs, Raymond W. 2005. *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Kraska-Szlenk, Iwona. 2014. *Semantics of body part terms: General trends and a case study of Swahili*. LINCOM Studies in Semantics 07. LINCOM, Muenchen.

Kozmács, István. 2002. *Udmurt–magyar szótár* [Udmurt–Hungarian dictionary]. Szombathely: Savaria University Press.

Kövecses, Zoltán. 1990. *Emotion concepts*. New York: Springer.

Kövecses, Zoltán. 2000. *Metaphor and emotion*. Cambridge: Cambridge University Press.

Kövecses, Zoltán. 2010. *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kubitsch, Rebeka. 2019. Testrésznevek szerepe az érzelmek és az értelem konceptualizálásában az udmurt nyelvben [The role of body part names in conceptualizing emotions and mind in the Udmurt language]. In: Makarov, A. M. ed. *Материалы XLVII итоговой студенческой конференции*. Izhevsk: Izd-vo «Udmurtski Universitet», 363–366.

Lakoff, George and Johnson, Mark. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Munkácsi, Bernát. 1896. *A votják nyelv szótára* [Votyak Doctionary]. Reprinted in 1990. Pécs: Az Uralisztikai Tanszék Kiadványai I.

Sharifian, Farzad; Dirven, Rene; Yu, Ning (eds.). 2008. *Culture, body and language: Conceptualizations of internal body organs across cultures and languages*. Berlin – New York: Mouton de Gruyter.

Yu, Ning. 2004: The eyes for sight and mind. *Journal of Pragmatics,* 36: 663–686.

The characteristics of responses given to compliments in Udmurt

Zoltán Németh

University of Szeged

1. Introduction

I am dedicating my study to the topic of linguistic politeness in connection with Udmurt, because, as was reported by Shirobokova (2011: 31–38), in case of the smaller Finno-Ugric languages the number of sociolinguistic studies is really low. This is true not only of sociolinguistics, but also of sociopragmatics, where politeness research also belongs. Before I started to do my research on this topic, basically there were no studies carried out on politeness in connection with the smaller Finno-Ugric languages spoken in the territory of the Russian Federation.

The question arises: is it necessary at all to do this kind of research on these languages? And the answer is yes, especially in the case of Udmurt, because as we can find it out from Pischlöger (2016), the Udmurt language is the most visible minority language on the social network sites among the minority languages spoken in Russia. It is important to highlight that this is not only true of the Finno-Ugric languages spoken there, but of every minority language spoken in Russia, outstripping languages like Tatar, which is the most widely spoken minority language in Russia. Because of this high visibility the speakers of this language have an higher level of probability of being exposed to intercultural communication, which is considered to be quite a common phenomenon in the western part of Europe, but a new challenge for the Udmurts. In the case of intercultural communication knowing the accepted norms of the other group is inevitable to being successful.

In this study I present my results regarding the use of Udmurt language, structuring it in the following way. After the introduction I briefly discuss the present-day situation of the Udmurt language, which is followed by the discussion of the data collection and my informants. After these the next step the theoretical background of the research, the analysis of the data, and finally the results.

2. The current status of the Udmurt language

The Udmurt language belongs to the Uralic language family and within that to the Finno-Ugric languages, more specifically to the Permic sub-branch. The closest affiliated language to Udmurt is Komi. As Pischlöger (2016: 109-110) also discusses, Udmurt is mainly spoken in the territory named after them, Udmutia, which is one of the republics of the Russian Federation. But the language is not spoken exclusively there, but also in the neighboring republics of Tatarstan and Bashkiria, the Kirov and Perm territories, and there are also speakers living abroad, e.g. in Estonia.

Unfortunately, like in the case of most Finno-Ugric minorities of Russia, the number of the people considering themselves Udmurt shows a declining tendency. In the results of the 2002 census their number was around 640,000, but in eight years it dropped to 550,000. And this is only the number of ethnic Udmurts, which does not equal the speakers of the language. That number is around 325,000 (Pischlöger 2006:110), which is about the 60% of all ethnic Udmurts. It also does not help the situation of the language that they form a minority even in their own republic. The Udmurt Republic is inhabited by approximately 1.5 million people. The largest ethnic group is the Russian, with about 912,000 people, or about 61% of the population. The second largest group is people who consider themselves Udmurt, totaling 410,000 people (27%), and the third largest is the Tatars, who number slightly less than 100,000 people (7%). The remaining 5% of people living within the borders of Udmurtia is made up by various smaller ethnic groups. One more detail about them that should definitely be mentioned is that the dominance of the Russian language is not only present at the level of the republic but also in the cities (Winkler 2001: 5).

As a result of all this, basically all Udmurts who speak the language are Udmurt–Russian bilinguals, because they have no chance to avoid the use of Russian language in their lives. There is speculation that in the oldest generation it could be possible to find Udmurt monolingual speakers, but even if this is true, their number would be so small that it is safe to say that it is negligible. But at the same time, mainly around the southern border of Udmurtia there are Udmurt–Russian–Tatar multilingual speakers as well. Because of this southern effect of the Tatar language it is also noticeable among the Udmurt speakers that the language use of the Northern Udmurts is mainly influenced by the Russian language, but that of the Southern Udmurts is heavily affected by the Tatar language (Edygarova 2014: 378).

3. Data collection and informants

For data collection I used the CCSARP (Cross-Cultural Study of Speech Act Realization Patterns) discourse completion tests (Blum-Kulka and Olshtain 1984), for four politeness situations, but in this study, I only discuss the results of the answers to

questions requiring compliment responses. I have chosen the CCSARP discourse completion test because it can be used cross-culturally (as the name suggests), so later my results can be compared to similar results in other languages. The exact questionnaire that mine is based on is the one used by Szili (2013) for her own research on the Hungarian language, with some modifications. First of all, it was translated into Udmurt and into Russian by myself, and the translations were checked by Udmurt lecturer Ekaterina Suncova of the Department of Finno-Ugric Studies, University of Szeged, Hungary, a native Udmurt speaker (who speaks both Udmurt and Russian), and who has a very high proficiency of Hungarian as well and is, therefore, able to compare the translations with the original texts. But the translation was not the only change to be made, there were other bigger or smaller modification to be done. By smaller changes, I mean localization: there were situations that would not have been familiar to the speakers, which had to be changed to something that is thematically the same but more realistic for the informants. I consider these smaller changes, because it is just the redressing of the same situation with no effect on the results. Bigger changes mainly affected situations where the speakers had to give negative answers. In these situations it was often explicitly in the description of the situation that it required a negative answer, but by stating this explicitly there was a fair chance of these being leading and hypothetical questions, and so the results would be unreliable.

In the part of the questionnaire referring to the speakers themselves, I asked them about their mother tongue(s), place of origin, age, and in what program they were studying at the moment of the completion of the task. I did not ask them about their gender specifically, because the questionnaire was structured in such a way that this information was acquired anyway (in the case of responses to compliments there was a separate set of situations for men vs. women).

I used my questionnaire to collect data from Udmurt speakers enabled by the short-term scholarship of the organization Campus Mundi which made it possible for me to carry out fieldwork at Udmurt State University in October, 2017. During the data collection I received a lot of help from the co-workers of the Institute of Udmurt Philology, Finno-Ugric Studies and Journalism, where I carried out my research. My target group was the Udmurt speaking students of the BA program of the University. I chose informants of university student age for multiple reasons. First, sociopragmatic studies are usually carried out with the participation of this age group, so this way my results on Udmurt could be compared later with results of studies on other languages. Second, if I had not been able to travel to Udmurtia to carry out my research, I still could have collected data from my subjects because this is the age group whose answers are the least affected by side effects of an online completion of the questionnaire. And the third reason is that, in my opinion, speakers of this age

group are the most endangered by language shift, because they usually come from villages which are usually Udmurt dominant, to the capital city which is heavily Russian dominant. So, it is important to do research on their language use because the more we know about it, the better strategies can be made to prevent them from shifting from their native language to Russian, and if we can reach that, there is a greater chance to stop the declining numbers or even turn them around.

The university students I asked were studying in the Udmurt language program of the Udmurt State University. There were 149 students enrolled in the Udmurt language program, of whom 94 were majored in Udmurt linguistics (28 first-year, 24 second-year, 22 third-year, and 20 fourth-year students) and 55 were teacher trainees in the Udmurt language (14 first-year, 11 second-year, 14 third-year, and 16 second-year students). I chose them because the language of instruction in their classes was mainly Udmurt, so to be able to complete their studies, they had to have a high enough proficiency in the language. Moreover, they used the Udmurt language on an everyday basis, which is not always true in the case of those students who may be proficient speakers of the language but less exposed to it, which can impact on their language use.

Although first I planned to analyze the answers of subjects of both genders, but in the end only the women's answers were taken into consideration. The reason is that among the students majoring in Udmurt at the Institute of Udmurt Philology, Finno-Ugric Studies and Journalism, the number of male students is very low. Of the 149 enrolled students only 10 were male, 3 of them in the first year, 2 of them in the second, 1 of them in the third, and four of them in the fourth. These were the numbers for potential male informants, which clearly shows the female dominancy in the numbers. The proportions were even more skewed when I was there doing fieldwork, because two of the ten potential male? informants were abroad in an exchange program. Because of these small numbers I decided to examine the language use of the female? speakers, because their number is high enough to get relevant answers, but the number of answers by the males would be too small for generalizations. In total I received 120 filled out questionnaires, which is about the 81% of the students enrolled, and after removing the answers of the males, I was still left with responses of 80% of all the students.

There was a problem that I encountered during the preparation of the study. As I mentioned above, those Udmurts who speak Udmurt also speak Russian. And not only do they speak both, but they often use mixed language (Pischlöger 2016:111). Because of this I had to find a way to get answers in Udmurt which would reflect the everyday use of the language, with Russian words mixed in, rather than purist and artificial answers. I decided to solve this problem by telling the subjects in the instructions that in the questionnaire all interlocutors are all assumed to speak Udmurt, even if they are

a person unknown to the subject, because in those cases they heard the other person speak on their phones in Udmurt. This implied to the subjects that Udmurt is the language expected for the answer but did not rule out the use of Russian language elements. This method was successful both during the pilot tests and in the main study as well.

4. The theoretical background of the study

Before starting the discussion of the research made by me, and analyzing the results, I need to define what is meant by politeness and by response to compliments.

There are two well-known theories that laid down the basics of politeness research: the first one is Leech's theory (1983), which is based on Grice's maxims (1975), and the other one is Brown and Levinson's (1987), based on Goffman's (1967) face theory. According to the former, we talk about politeness in those cases when the speaker violates one of the four maxims (of quantity, quality, mood, or manner) to avoid doing harm to the other person, e.g. the speaker provides less information than is needed, or composes their speech that way that the meaning is opaque.

The starting point for the latter theory is the two faces of people, discussed as positive and negative face. Positive face represents the speaker's desire to have their actions valued in a positive way. On the other hand, negative face expresses the desire of the speaker to not to be held back from carrying out their actions, and also their desire to carry out their actions autonomously. In those cases when someone is polite, theyr try to defend the other person's face from threatening even by doing damage to their own. This latter theory was later expanded by Foley (1997), who said that both participants of the communication should feel appreciated. This remark has an important role, for example, in the case of asking for forgiveness.

Searle (1975: 357) discusses response to politeness as a member of the group of expressives among speech acts, which expresses the attitude of the participant towards the current situation. Szili (2013: 156) claims that one of the most important characteristics of this speech act is that the compliment and the response to it cannot be separated from each other. The reason behind this is that in many cases in case of seeing the answer only we cannot reconstruct what the compliment could have been, e.g. in case of answers like *thanks*, *thanks, yours too* or *thank you, but you are exaggerating* we do not have any clue about what the compliment was. And according to Schlegoff and Sacks (1973: 296), the compliment and the response to it form an adjacency pair, because they are not only connected to each other functionally but temporally as well.

5. The analysis of the data

In my analysis of the responses to the compliments I assigned them to 3 main categories: (A) the speaker agrees with the compliment, e.g. in (1); (B) when the speaker disagrees with the compliment, as in (2); and (C) when the speaker tries to avoid self-praise, as in (3).

(1)	*Мо н*	*туж*	*шумпот-йсько-Ø*	*тыныд*	*кельш-е*	*шуыса.*
	I	very	be.happy-PRES.1SG	you.DAT	appeal-PRES.3SG	that.
	'I'm very happy that you like it.'					

(AAN11012124)

(2)	*Ой,*	*чик*	*озьы*	*öвöл*	*со.*
	Oh,	totally	that.way	to.be.NEG.PRES	that.
	'Oh, that is not true at all.'				

(JD11012124)

(3)	*Асьтэ-лы*	*тау*	*лыктэм-ды*	*понна.*	
	yourselves-DAT.	thanks	arrival-Px.2PL	because.of	
	'Thank you to you for coming.'				

(AAN11012124)

Of course, these categories have many sub-strategies, because there are various ways to express agreement or disagreement with a compliment, and there are various strategies in which one can avoid self-praising. Here is the complete list of all the strategies (based on Szili 2013: 159–162):

(A) Expressing agreement with the compliment
    (A1) Expressing acceptance
    (A2) Thanking
    (A3) Expressing appeal
    (A4) Counter-compliment/offering the target of the compliment for the other person
    (A5) Joking about the compliment

(B) Expressing disagreement with the compliment
    (B1) Devaluating the target of the compliment
    (B2) Expressing uneasiness/discomfort about the compliment

(B3) Devaluating the compliment itself

(B4) Describing the basis of the disagreement with the compliment

(C) Avoiding self-praise

(C1) Name another person who should be complimented

(C2) changing topics

(C3) expressing how much effort was needed to achieve the result

(C4) not taking the compliment into consideration

(C5) skepticism, seeking for certainty

These strategies can be used on their own (4) or be combined with each other (5) as well.

(4) *Бен,*   *мыным*   *но*   *яра*
     yes   I.DAT   also   like
     'Yeah, I also like it.'

(AP21003115)

(5) *Тау*   *бадӟым,*   *ачид*   *но*   *туж*   *чебер*
     thanks   big,   yourself   also   very   nice
     'Thank you very much, you are pretty as well.'

(AGV21003115)

In this study I am looking for answers to the following questions:

- What are the most commonly used strategies and strategy combinations?
- Are there politeness related expressions borrowed from Russian despite of having equivalents in Udmurt?

My expectations based on my previous knowledge of spoken Udmurt are the following: in the case of the first question I expect them to use mainly short answers using only one strategy on its own, or combinations of two strategies. In my opinion the most common choice is the option to just say *thank you* without any other strategies. But it is not likely to be the most commonly used strategy because of the pressure of society to always give a positive answer, like in the case of some languages like English, because, like in the case of Hungarian, the speaker has the possibility to give negative answers to the questions. This is much more connected to the fact that one of the most well-known stereotypes that happens to be one of the most important values among the Udmurts is modesty. So, in their case the reason behind saying only *thank you* is much more connected to how this way they give a polite answer, and at the same time, they close the given situation (as they do not say anything that could

make the conversation go on, so there is a greater chance that the other person changes the topic).

In connection with the second question, my expectation is that Udmurt speakers do not use any politeness related expression that is borrowed from Russian. This expectation is also based on my personal experiences. There are many politeness related expressions in other situations that are borrowed from Russian, like *пожалуйста*, *извини*(*те*), *прости*(*те*) etc. but they are not expected in responses to compliments. Also, as I mentioned above, I expect speakers to use mainly thanking, and it hardly occurs that they would use the Russian *спасибо* 'thank you' or *спасибо большое* 'thank you very much' instead of the Udmurt equivalents *тау* and *тау бадӟым,* respectively. So, although I expect that speakers would use Russian words or expressions, in my opinion, they use only neutral expressions in connection with politeness.

6. Results

In the evaluation of my results first I discuss them as a whole, considering all the strategies and strategy combinations that were used by my informants, and then I focus on the most used ones. I categorized strategies and strategy combinations in the group of 'most used ones' that appeared at least five times. I chose five as the cutoff mark because if it was used five times, it means that, statistically, it was used by at least one student from each year plus once more.

When we take all the answers into consideration we find 98 different strategies and strategy combinations that were used by these students. Of all the answers, 658 included thanking, at least in some kind of combination. This means that it was used in 75% of the cases. If these cases are narrowed down to those where it was used as the one and only strategy, the result is 193, which is 22% of the cases, so almost in one quarter of the answers students only thanked their interlocutor for the compliment, which made it the leading answer, as was expected. If we analyze responses from the point of view of how many strategies were combined in the answers, we can see that there were 17 variants in which only one strategy was used (17% of the cases), there were 35 variants in which two strategies were combined (36% of the cases), and 46 variants where three or more strategies were combined (47% of the cases). According to these numbers, although the number of combinations including only one or two strategies is greater than the ones combining 3 or more, but there is only a slight difference (53% vs. 47%).

Figure 1. Number of strategies combined

But we should take it into consideration that in the first group there is a strategy that was used almost 200 times, while in the second group, the most used combination was used only six times, there were only three that were used five times, and all the others were used less than five times each. So although the number of combinations is almost as big as the ones with two or less combinations, but the 3+ combinations category was much more rarely used. So soon, when the results are narrowed down to the most used combinations only, their slice in the pie-chart will be much smaller.

The total number of strategies and combinations that belong to the most used ones is 31. This is a bit shy of 32% of all the cases, but actually they were in use in 87% of the cases. So, as mentioned above, of the 46 combinations containing 3+ elements only 4 remained. So although the number of variants was quite high, they were rarely used. Among these narrowed down variants, there were 583 answers that included thanking (89% of the answers) and 193 of them included thanking only (which is 25% of the answers). If we look at the answers from the point of view of strategies combined, we see that there were 12 variants where only one strategy was used (39% of the combinations), 15 variants where 2 strategies were combined (48% of the combinations), and only 4 where more than two strategies were combined (13% of the combinations). Another point that shows the dominance of short answers is that even the first combination that includes 3 strategies appears quite far down the list, it is in the 25th position, and was only used 6 times.

Figure 2. Number of strategies combined in the Top31

The use of the strategy of thanking was so dominant that it was not only the most frequently applied strategy with 193 occurrences, but it is followed by 5 combinations, all including it, and all of them were used more than 40 times each (A2C3: 67, A2A3: 54, A2C2: 51, A2A4: 45, A2A1: 44, A2B1: 40). Also, in the top 31, twelve of the most used strategies and combinations were strategies on their own, so almost all the strategies appeared without being combined with another.

7. Conclusion

As is visible from the results, my expectations have been met by the answers. Thanking was commonly used by the speakers. When we take the combinations and this strategy used alone together, they make up 74% of the cases, which increases to 89% when we narrow it down to the most used combinations. The use of thanking alone appeared in 22% of all the cases and in the case of the most used strategies it appeared in one-quarter of the cases. According to these numbers it is safe to state that although thanking is not an obligatory element of a response to a compliment, it is heavily expected. According to the numbers, it was also true that Udmurt students use mainly one or two strategies, because only 4 combinations (although these make up almost half of the combination variants used by the informants) consisted of 3 strategies, and almost all the strategies appeared on their own.

Getting further from the actual results and looking at them as a whole, it seems that, in the case of the responses to the compliments, the choice of the strategy is much more important than the length of the actual response itself. On the one hand, this is supported by the fact that the responses usually were quite short, and, on the other

hand, by the fact that even if speakers combined different strategies, it was mainly a combination of two. The reason behind this could be either because modesty is a highly appreciated trait among the Udmurts, or it is also plausible that the most frequently used strategies and strategy combinations are so powerful that the speakers do not feel the urge to further support their response, or by the nature of the situation itself, as in these situations the speakers' task is to avoid a possible threat of the face, in contrast with the apologies, where the threat is already present. The latest one is also supported by the fact that, in the case of apologies, the speakers often gave long answers in which they tried to explain the situation, often providing multiple explanations. Of course, these three possibilities can overlap each other as well, and the short answers are the result of combining them.

My expectations in connection with the use of Russian expressions were also met. I have not come across any answer that would have included *спасибо*, *спасибо большое* or any other politeness related words or expressions. Of course, there were other Russian words that were used by the speakers as a result of borrowing or code-switching, but all of them were neutral from the point of view of politeness.

Abbreviations

1	first person
3	third person
DAT	dative case
NEG	negation
PL	plural
PRES	present tense
PX	possessive suffix
SG	singular

References

Blum-Kulka, Shoshana and Elite Olshtain 1984. Requests and apologies: Cross-Cultural Study of Speech Act Realization patterns. *Applied Linguistics*, 5(3): 196–213.

Brown, Penelope, and Stephen C. Levinson 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Edygarova, Svetlana 2014. The varieties of the modern Udmurt Language. *Finnisch-Ugrishe Forschungen* 62. 376–398.

Foley, William 1997. *Anthropological linguistics: An introduction*. Oxford: Blackwell.

Goffman, Erving 1967. *Interaction ritual: Essays on face-to-face behavior*. New York: Doubleday.

Grice, Paul 1975. Logic and conversation. In: Cole, Peter and Morgan, Jerry L. (eds.) *Syntax and semantics 3: Speech acts*. New York: Academic Press. 41–58.

Leech, Geoffrey 1983. *Principles of pragmatics*. London, New York: Longman Group Ltd.

Pischlöger, Christian 2016. Udmurt on Social Network Sites: A comparison with the Welsh Case. In: Saarikivi, Janne and Toivanen, Reetta eds. *Linguistic genocide or superdiversity? New and old language diversities*. New York: Channel View Publications, 108-132.

Searle, John R. 1975. A taxonomy of illocutionary acts. In: Gunderson, Keith (ed.): *Language, mind and knowledge*. Minneapolis: University of Minnesota Press.

Shirobokova, Larisa 2011. *Az udmurt–orosz kétnyelvűség (Udmurt Köztársaság, Sarkan járás, Muvyr község)* [Udmurt–Russian bilingualism (Muvyr village, Sarkan district, Udmurt Republic)]. Budapest: ELTE PhD dissertation.

Szili, Katalin 2013 (2004). *Tetté vált szavak: A beszédaktusok elmélete és gyakorlata* [Words become acts: The theory and practice of speech acts]. Budapest: Tinta Könyvkiadó.

Winkler, Eberhard 2001 *Udmurt*. Languages of the World/Materials 212. München: Lincom Europa.

On some Chuvash–Mari shared lexemes and Agyagási's "Late Gorodets" hypothesis

Christopher Culver

1. Introduction

The Chuvash and Mari languages have been in contact for centuries and during that time loanwords have flowed in both directions, Chuvash–Mari and Mari–Chuvash (see e.g. Räsänen 1920; Fedotov 1990). Some of these shared lexemes can be traced to the Proto-Uralic language ancestral to Mari and were borrowed from Mari into Chuvash. Others were inherited by Chuvash from Proto-Turkic and then transmitted to Mari. Still others can be ultimately derived from Russian or Persian and were mediated through Chuvash to Mari.

There remain, however, a number of shared lexemes in Mari and Chuvash that have lacked a clear etymology. In a series of publications, Agyagási (2000, 2001, 2002) has posited that Mari and Chuvash were both in contact with an unknown third language, spoken by a people that was assimilated by the Mari and Chuvash after the large-scale population movements following the Mongol invasion in the 13th century. Agyagási initially identified this unknown language with the Late Gorodets archaeological population. Furthermore, she suggests that the exonym "Cheremis" – never used by the Mari themselves – originally referred to the people of the Late Gorodets archaeological population, and the ethnonym came to refer to the Mari only after the Late Gorodets population disappeared as a distinct people through assimilation. In a more recent publication on historical interaction between Chuvash and Mari, Agyagási (2019) no longer points to the Late Gorodets archaeological population specifically and, therefore, ceases to employ this nomenclature, but she continues to believe that there was a mysterious third population in contact with both Mari and Chuvash, and now she identifies this language as mediating lexical material from the West Baltic branch of Indo-European to Mari.

Agyagási bases her hypothesis of a third language in the region on a corpus of sixty-four shared Mari–Chuvash words that lack an etymology (Agyagási 2000, hereinafter referred to as the "Late Gorodets" material as a convenient label for the collection in spite of Agyagási's more recent turn away from ascribing it to the Late Gorodets archaeological population). This corpus includes such etymological cruxes

as Mari *meraŋ* ~ Cv. *mărin* 'hare'. On one hand, Chuvash *mărin* cannot be an inherited Turkic word due to the initial *m-* here (on the impermissibility of initial **m-* in Proto-Turkic, see Tenišev 1984: 310ff.). At the same time, Mari *meraŋ* is difficult to explain as an inherited lexeme in Mari due to the irregular vowel and consonant correspondences among the Mari dialects. Thus, as an example, in Agyagási's 2000 publication compiling the "Late Gorodets" material, she brings together the data on Mari *meraŋ* ~ Cv. *mărin* as follows:

> **23.** V(Morg.) *mărin*, *mărik* 'заяц' (Ašm. 8. 317); (Kozm.)
> *moŕan* 'Hase' (Munk.); (ohne Ortsangabe) *mărem* 'заяц' (Ašm.uo.);
>
> vgl. mar. mund. 1775: *меря́ндь* [*merang*] 'заец' (Sebeok and Raun
> 30); P *mer'aŋ*, *meraŋ* 'Hase' (Genetz); KB *moren* (Ramstedt); P B UP
> USj C Č *meraŋ*, M *meraŋ*, MK *mer'äŋ*, UJ *meraŋe*, CČ *mor'an*,
> JT *möran*, JO *möraŋ∂*, V *möräŋg∂* K *moren* 1.'nyúl; 2. fattyú,
> zabigyerek' (Beke). ⊗ Räsänen 1920. 255; Fedotov 1968. 204;
> Arslanov and Isanbaev 1984. 110.

Then, in a subsequent publication Agyagási (2001: 36) attempts to explain the unusual vowel correspondences between the various dialectal forms by positing borrowing into the already distinct Mari dialects from two different dialects of this mysterious neighbor: "Western Late Gorodets" contributed a form, *mor'*, to Hill Mari, while "Eastern Late Gorodets" contributed a different form, *mer'*, to Meadow Mari.

In spite of its convenient compilation of Mari and Chuvash words demanding etymological clarification, Agyagási's corpus of "Late Gorodets" lexical material has drawn almost no commentary from other scholars as of yet. The "Late Gorodets" corpus merits special attention now, when research into the Mari lexicon and Mari historical phonology has made great advances in the intervening years. First, in compiling her wordlist Agyagási relied overwhelmingly on Beke's dictionary, which was published in 1997–2001, and at the time of Agyagási's initial research it was the only dictionary of Mari that drew together material from across the bulk of the Mari dialects. However, scholars today have a fuller picture of the Mari lexicon due to the publication of *Tscheremissches Wörterbuch* (TschWb) in 2008, a compilation of material collected by a number of late-19th century and early-20th century fieldworkers, which represents a resource as ample and valuable as Beke's, and also the dictionary of Mari dialects of Udmurtia and Tatarstan edited by Veršinin (2011). One can also mention the 18th and 19th century manuscript wordlists of Mari dialects described by Sergeev (2002), which occasionally provide dialectal forms or meanings that are absent from later sources. Finally, though it was published three decades before Agyagási's work, the Northwestern Mari dictionary by Ivanov and Tužarov is

not cited in Agyagási's compilation of the "Late Gorodets" material, and yet in some cases it provides unique data unavailable elsewhere.

Second, Agyagási's view of Mari historical phonology hews to the reconstructions proposed by Bereczki. More recently, however, Aikio (2014a) has proposed a new reconstruction of Proto-Mari vocalism that adheres to strict sound laws instead of so readily admitting "sporadic" sound changes. Aikio's reconstruction of Proto-Mari also grounds Mari in the latest innovations in the historical phonology of its parent Proto-Uralic protolanguage, as found in recent publications by Aikio and Zhivlov (see Aikio forthcoming for a survey and the relevant literature).

By drawing on the modern state of the art in Mari historical phonology, dialectology, and lexicology, the present contribution challenges Agyagási's ascription of several shared Mari and Chuvash words to a late shared substrate. I aim to show that a number of these words cannot serve as proof of contact with an unknown third language in the early second millennium CE.[1]

2. Mari and Chuvash historical phonology and the "Late Gorodets" material

It goes without saying that a shared substrate is not the only possible source of unetymologized lexical material found in both Mari and Chuvash. That is, the lexicons of each of these languages stem from a number of sources. Mari is a descendant of Proto-Uralic, and, therefore, it has retained lexical material dating from as far back as the era when Proto-Uralic was spoken; work has continually progressed in identifying new Uralic etymologies for Mari words that were formerly considered to be of uncertain origin (see e.g. Aikio 2014b, Metsäranta 2020). Additionally, Mari has undergone contact with other languages since the Proto-Uralic era. Besides possible contacts between pre-Proto-Mari speakers and prehistoric non-Uralic languages of European Russia that we can only speculate about, scholars have identified in the Mari lexicon early borrowings from Baltic (Gordeev 1967), various stages of Iranian (Joki 1973, Gordeev 1967) and Permian (Bereczki 1977; 1987; 1992–1994: II 97–129; 2005, see also Bereczki et al. 2013). In more recent centuries, Mari has been in contact with Chuvash, Tatar (Isabaev 1989–1994), and Russian (Savatkova 1969).

By the same token, Chuvash is a descendant of Proto-Turkic, and among those unetymologized items in its lexicon there may be words hitherto unidentified for their Proto-Turkic origin. Additionally, Chuvash may have undergone contact with

unknown languages after its split from Proto-Turkic and during West Old Turkic speakers' migration from Asia to the southern Russian steppes and up to the Middle Volga, and scholars have identified borrowings from Persian and Arabic (Scherner 1977), Alanic (Dobrodomov 1980) and Permian (Rédei and Róna-Tas 1982). More recently, Chuvash too has been in contact with Russian and Tatar (see Fedotov 1996).

The first criticism of the "Late Gorodets" material as proof of a substrate shared by Mari and Chuvash is that, on phonological grounds, certain words can be shown to represent lexical material that was already present at the Proto-Mari stage and was then borrowed by Chuvash, or vice versa. Even if a given word still lacks an ultimate etymology, then that word may represent a borrowing in Mari or Chuvash, but it could have been borrowed into either of these languages before the population movements of the 13th century. Such words cannot serve as evidence for an unknown language of the Middle Volga in contact with both Mari and Chuvash, but rather they could be the result of other language contact situations that we know Mari and Chuvash participated in.

To illustrate this point, in this paper Mari historical phonology is viewed according to the recent reconstruction by Aikio (2014a). Briefly, Aikio's reconstruction differs from the reconstruction by Bereczki on which Agyagási generally relied in the following ways.

Bereczki (1992–1994: I 115) reconstructed the following first-syllable vowel system for the parent language of the Mari dialects:

i	ü	u
e	ö	o
		a

Bereczki explains the reduced labial vowels of the Mari dialects as the result of post-Proto-Mari reduction of what were full labial vowels in Proto-Mari; the correspondence Meadow Mari *o* ~ Hill Mari *a* is seen as the result of lowering of original *o in Hill Mari; and Hill Mari *ä* is seen as the result of fronting of original *a.

Aikio proposes, instead, the following vowel inventory for Proto-Mari:

i	ü		u
ĭ	ǔ		ŭ
e			o
ä		a	å

Thus, key differences in Aikio's reconstruction are that both full and reduced labial vowels must be reconstructed for Proto-Mari, and there was an opposition between *å and *o and between *ä and *a. The subsequent developments in the Mari dialects are broadly as follows:

- PMari *u > MariE NW W *u*
- PMari *ü > MariE NW W *ü*
- PMari *ŭ > MariE *u* NW *ŭ* W *ə̂*
- PMari *ǔ > MariE *ü* NW *ǔ* W *ə*
- PMari *å > MariE *o* NW W *a*
- PMari *o > MariE NW W *o*
- PMari *ä > MariE *a* NW W *ä*
- PMari *a > MariE *a* NW W *a*

In the appendix to her recent monograph on Chuvash historical phonology, Agyagási (2019: 289ff) rejects Aikio's reconstruction. Though Agyagási accepts *ə as a member of the Proto-Mari vowel system, she attempts to revive the Bereczki-era reconstruction in terms of the high rounded vowels. She claims that Proto-Mari had solely full high labial vowels, and reduced labial vowels only arose later in Northwestern Mari, Hill Mari, and in the Upša subdialect of Meadow Mari. She goes on to propose various conditioning environments for this vowel reduction.

Nevertheless, Agyagási's arguments against Aikio's reconstruction are unsound. First, her survey of reduced labial vowels in the Mari dialects fails to take into account data from the Bol'šaja Šija and Menzelinsk dialects in the Eastern Mari area as documented by Veršinin (2011); these dialects possess both full and reduced labial vowels (and broadly in the places we would expect from Aikio's reconstruction) but they differ in other aspects from the Northwestern Mari–Hill Mari–Upša complex. Second, it is easy to find exceptions to the conditioning environments which Agyagási proposes. For example, although Agyagási (2019: 297) claims that the first-syllable vowel in Proto-Mari *ü–ə̂ structures undergoes reduction, this is never the case with

MariE W NW *šüðö* '100', which shows a full vowel across the Mari dialects. This in fact relates to one of the key insights afforded by Aikio's 2014 article: among the Mari lexical material inherited from Proto-Uralic, reduced and non-reduced labial vowels go back to different PU vowels. The high front labial reflex of PU **i̮* does not appear as a reduced vowel anywhere across the Mari dialects, cf. PU **śi̮ta* '100' > MariE W NW *šüðö*. Additionally, in the Mari inherited lexicon we find a different treatment of high front labial vowels before *r*, i.e. in one set of words the reflex among the Mari dialects is *ö*, e.g. MariE W *mör* 'strawberry' from PU **mi̮rja*, while other words do not show this lowering, e.g. MariE *βür* NW *βŭr* W *βər* 'blood' < PU **wire* (UEW 264–265, 576). Thus, the distinction between the two types of labial vowels represents an inherited feature in Mari that was already present when Mari and Chuvash came into contact.

There is wider consensus around the Chuvash historical vowel inventory than the Mari one (see Agyagási 2019), though as Savelyev (2018) has recently noted, the Northwestern dialect of Chuvash has been given insufficient attention and some of its distinctive sound shifts are relevant for detecting loans between Mari and Chuvash.

With a firm grounding in Chuvash and Mari historical phonology, evidence that the "Late Gorodets" words must represent material extraneous to Mari and Chuvash evaporates. Many of these words do not show irregularities between the Mari dialects at the post-Proto-Mari period, rather they can be traced back to Proto-Mari. One can then account for the data listed under many of the "Late Gorodets" items through simple borrowing of inherited Mari material into Chuvash.

As an example, consider No. 52 in the "Late Gorodets" material: MariW *βažar* 'cross-grained (of wood)' versus Cv. A *vušar* id. These words both regularly go back to Proto-Mari **βåžar* and Middle Chuvash **våšar*, respectively. As their vocalism is identical, one word can simply represent a borrowing from the other, presumably from Mari into Chuvash as there is no regular source for Chuvash *š* in this environment.

A similar situation holds for MariE *mužôr* NW *mŭžŭr* W *môžôr* 'pair; spouse' versus Cv. V *mŏšŏr* A *mǎšǎr* id. The Mari forms regularly go back to PMari **mŭžər*,[2] while for the Chuvash forms we can reconstruct Middle Chuvash **mŏšŏr*. The

---

[2] Agyagási cites a Malmyž dialect form "mužŭn" from Beke's dictionary, but this is in fact a typo on her part: we find expected *mužur* in Beke. She also cites a supposed Hill Mari form "mužor" from Budenz's dictionary, but Budenz does not actually ascribe the word to Hill Mari, and, on the basis of every other Mari lexical reference, we should assume that Budenz documents a Meadow Mari form. With regard to the second-syllable vowel of Budenz's "mužor", Budenz uses a pre-modern transcription that often fails to correctly reflect Mari's reduced vowels, and so the word should be read as the same form *mužôr/mužŭr* known from other Meadow Mari lexical resources.

protoforms in both languages are essentially identical, and therefore borrowing from Mari into Chuvash or vice versa is the most straightforward explanation. Again, this is probably a borrowing from Mari into Chuvash on the basis of Cv. *š* in this environment.

Let us next turn to the case of the PMari *a and *ä. Proto-Uralic *ä and *a were both raised in pre-Proto-Mari, and consequently any Mari word with a first-syllable vowel going back to Proto-Mari *ä or *a must be a loanword instead of inherited Uralic material. The majority of these words are known to be relatively recent borrowings from Chuvash. However, as noted by Bereczki (2005: 193), we find *a and *ä even in Iranian borrowings in Mari that are presumably much older than the 13th century – an example is the Mari ethnonym *mari itself (Joki 1973). PMari *ä is also found in the earliest layer of loanwords from Permian, cf. MariE *jakte* NW W *jäkte* 'pine tree' < Komi *jag* 'Kiefernwald (auf Sandboden)', Udm. *jag*, *Íag*, *d'ag*. 'Fichtenwald auf Sandboden' (Bereczki 1992–1994: II 98).

Consequently, a word common to Mari and Chuvash that shows PMari *ä or *a and has no Chuvash etymology need not be a borrowing into both Mari and Chuvash from a third language in which they were both in contact beginning from the 13th century. Consider the "Late Gorodets" item consisting of MariE *kaće* NW *käcə* 'young man' versus Cv. A *kaččă* id. (No. 5). The geminate consonant in the Chuvash word is a feature generally found in loanwords, including those from Mari, cf. Cv. V *merčče-* 'чахнуть' < MariE *merćem* W *mercem* id. (Fedotov 1990: 309). The dialectal forms underlying the Mari counterpart to Cv. *kaččă* regularly go back to PMari *käćə. Therefore, it cannot be excluded that the Mari word is a pre-13th-century borrowing from some unknown source,[3] which was then borrowed by Chuvash.

For words where the Chuvash material is known only from the northwestern Chuvash area, where contacts with Mari had been maintained until recent centuries, we must also take into account distinctive Hill Mari sound shifts. Thus, in the case of MariE *nol'*, *nol'o* 'sediment' and W *nal'ə* 'red clay' versus Cv. V *nal'ə/nay* 'тина' (No. 25), we find a word that can be reconstructed as PMari *nål'ə, but in Chuvash the word is attested only in dialects in the northwest of the Chuvash-speaking area, in close proximity to Hill Mari. Consequently, the most parsimonious explanation is that Viryal Chuvash borrowed the word subsequent to the shift of PMari *å > Hill Mari *a*.

A similar objection can be made to the ascription of MariE *tuškem* NW *tŭškem* W *tⱥškem* 'смешивать, перемешивать' ~ Cv. *tăška* id. (No. 44) to a shared substrate. The Mari words regularly go back to PMari *tŭškem. While the dialectal source of

_____

[3] Gordeev (1967: 196) in fact adduces Iranian parallels.

Ašmarin's word is unclear,[4] the most parsimonious explanation is that this represents material from the Viryal dialect, and the back unrounded reduced vowel is simply due to Viryal Chuvash borrowing the word from Hill Mari subsequent to the delabialization of PMari *ŭ > MariW *ə̂. (Again, *š in this position would be normal for inherited Proto-Mari vocabulary, but irregular in Chuvash.)

Finally, let us consider the case of Cv. *păčă, păčkata* ~ MariE *pič́* NW W *pəc* 'stuffy; dark' (No. 29–30). These items were first included in the "Late Gorodets" corpus. Then, in a more recent work Agyagási (2019: 280–281) has argued that the Mari and Chuvash words are not related as such, but nevertheless they are borrowings from the "Cheremis substrate" which supposedly transmitted West Baltic lexical material into both Chuvash and Mari. Thus, she traces Cv. *păčă* to a West Baltic dial. *pũčio* 'suffocate, stifle; blow up', while Mari *pič́* is ultimately a derivation of a West Baltic verb *pū* 'blow up' with a nominal derivational suffix *-č*.[5]

Agyagási's argumentation that these words must have been borrowed independently into Mari and Chuvash from a third, substrate language is flawed on phonetic grounds. First, she reconstructs the Proto-Mari form for the Mari material as *pič. In fact, on the basis of the correspondence MariE *č́* NW W *c*, the Mari word originally featured the consonant *ć and not *č, and therefore we must reconstruct PMari *pəć.

PMari *ć has a regular historical source. On the one hand, it represents PU *ć in word-final or pre-consonantal position (intervocalically pre-PMari *-ć- > PMari *-ź-). On the other hand, it can appear as the result of a shift PU *t > PMari *ć after PU *i, cf. Mari *βič́* '5' < PU *witti (UEW 557). Indeed, the vowel and consonant correspondences of Mari *pič́* are completely identical across the Mari dialects to those of Mari *βič́* '5', and therefore Mari *pič́* need not be any kind of late loan from an unknown substrate language, rather it may well go all the way back to Proto-Uralic.

On the Chuvash side, by contrast, the *č* in Cv. *păčkata* has no regular source; Cv. *č* results regularly only from the affricatization of *t before *i. Cv. *č* would, however, be expected as a reflex of PMari *ć in borrowings from Mari (e.g. Cv. *merčče* 'чахнуть' mentioned above). Consequently, here too the most parsimonious explanation is that Chuvash borrowed material from Mari that had already existed in the latter language since some unknown era.

The adduced examples above are likely loans from Mari into Chuvash. However, among the "Late Gorodets" material there are also words where borrowing from one

---

[4] Ašmarin gives the source as "СПВВ. ВА", but this abbreviation is not found in the list of abbreviations at the end of the dictionary and has apparently never been deciphered by researchers of Chuvash.

[5] This solution is extremely *ad hoc*, as no derivational suffix *-č* is known.

of these languages into the other can be supposed, but the direction of borrowing is unclear. In the comparison in No. 27, Chuvash V *ora* A *ura* 'haystack' regularly goes back to Middle Chuvash *åra, while its Mari counterpart E *ora* NW W *ara* id. regularly goes back to PMari *åra. The words, therefore, show no irregularities for which we must posit a third language in contact with both Mari and Chuvash, but the task is left to future etymological research to determine if this represents inherited Chuvash material transmitted into Mari or inherited Mari material transmitted into Chuvash.

3. Etymologizable words

Besides the general objection that the "Late Gorodets" words may represent inherited, albeit still unetymologized material in Mari borrowed into Chuvash or vice versa, there are cases where more concrete etymologies can be proposed.

One such etymology has in fact already been published. The inclusion of MariW *pȃjȃrka*, *pȃjȃrt* versus Cv. V *păyărt*, *păt'ărt* 'a little bit' in the "Late Gorodets" material (No. 31) overlooks a convincing Russian loan etymology that had already been established decades earlier by Rédei and Róna-Tas (1983: 37). As an aside in their discussion of the mistaken comparison of Komi *parga* 'in der Flachshechel zurückgebliebener flockenförmiger, reiner Abfall vom gehechelten Flachs' with Chuvash *parga* 'Büschel', Rédei and Róna-Tas identify this Mari and Chuvash material – as well as Tat. dial. *payarka* – as adoptions of the Russian *поярок* 'шерсть ягнят (первой стрижки)' with the semantic development 'small heap of wool' → 'small heap, bundle'.[6]

Another instance where a clear acceptable etymology can be proposed is the equation Cv. *kănăš* 'rubbish' ~ Mari *kunȃž* id. (No. 7). The Chuvash forms here V *kŏnŏš* A *kănăš* regularly go back to earlier *kŏnŏš. On the Mari side, the correspondence MariE *u* ~ NW *ŭ* ~ W *ȃ* is regular and goes back to PMari *ŭ according to Aikio (2014a), and therefore Aikio traces the Mari word back to PMari *kŭnəž.[7]

---

[6] See the entry for *поярок* in Dal's dialectal dictionary for the full range of semantics in Russian. When lambs are shorn for the first time, they produce a quite small amount of wool, and the example sentences that Beke's dictionary gives for Mari *pajȃrka* suggest the word was mainly applied to small amounts of material (wool/straw/bast).

[7] Agyagási cites a Mari dialectal form *kunəž* from the *Черемисско-русский словарь* of V. M. Vasil'ev, published in Kazan in 1911. This form is also found in Szilasi's dictionary (drawn in fact from Troitskij's earlier dictionary) and ascribed to the Hill Mari variety. The most parsimonious explanation for this form is that Hill Mari reborrowed the word from Viryal

The vocalism of the reconstructed Proto-Mari and Middle Chuvash forms matches exactly, and borrowing in either direction could be posited on that basis without the need to involve a third, unknown language. Furthermore, as Alexander Savelyev (p.c.) points out, there is another definition of this word besides the ones that Agyagási cites. Namely, the dictionary of Sergeev glosses Cv. *kănăš* as 'всякая всячина в избе, лишние предметы', and we can also compare this word to Cv. *kăn-kan* 'там и сям торчащие в разные стороны предметы'. In Savelyev's view, this appears to be none other than a derivation in *-ăš* of the same root found in Cv. *kăn-* 'вытягиваться, разгибаться', *kăntăran-* 'торчать, вытягиваться'. The root *kăn* in turn goes back to Proto-Turkic *kön (Fedotov 1996: I 254). Thus, an original root meaning 'stick out' was extended to refer to assorted junk around one's home that got in one's way, and then its meaning broadened to 'rubbish; filth' in general.

Savelyev's proposed etymology is attractive, because an earlier sense 'encumbrance' is attested for Mari *kunə̂ž* as well: Beke's dictionary attests MariNW *kŭnŭžlanem* 'im Wege sein', with the Mari informant defining the word as *ere aβašt βelen pižə̂t, okoltep pašaš* 'sie hängen fortwährend an der Mutter, lassen sie nicht arbeiten'.

Also, Savelyev's explanation has predictive power for other Mari material that has hitherto been unetymologized. In his dictionary of Mari dialects of Udmurtia and Tatarstan, Veršinin attests MariE *kunčó* 'надоедливый компаньон (тж. нежеланный и неуходящий, несмотря на намёки, гость); ребёнок, ни на шаг не желающий отходить от своей матери'. Sergeev (2002: 177, 32–33) documents *кýнзĭ̈о* 'воз' from an 18th century manuscript prepared for the *Vocabularia comparativa* of Peter Simon Pallas that reflects mainly the Malmyž dialect, and here we are probably dealing with the same word. This Mari word can be derived from the same Cv. *kăn-* root with a different derivational suffix *-čŏ*,[8] i.e. Middle Chuvash *kŏnčŏ was borrowed into Mari as *kŭnćə̂, and this ultimately became expected *kunčó* in Eastern Mari. All of the above-mentioned words would, thus, be examples of the same polysemy found in Russian *торчать* 'stick out; hang around, be in the way, be bothersome'.

With regard to Cv. *yapala* ~ Mari *jaβala* 'thing' (No. 55), an explanation for the Chuvash word was given by Ašmarin in the early 20th century in his massive Chuvash dictionary. In the entry for the Chuvash interrogative word *уере* 'как?', Ašmarin

---

Chuvash – which preserved a labial vowel – at a time subsequent to the Hill Mari unrounding of Proto-Mari *ŭ.

[8] For a parallel, cf. Mari *olð̂ə̂rčo* 'скальница' (part of a weaving loom) < Chuv. *xultărčă* id., which in turn is considered to be a derivation from a lost verb *xultăr- 'дрожать, трястись' reconstructed on the basis of Common Turkic *qaltïr (Fedotov 1996: II 356).

(IV 278) writes "отсюда происх[одит] слово йапала". Viewing Cv. *yapala* 'thing' as a derivation from *yepe* 'how?' is reasonable inasmuch as the sheer bulk of usages of Cv. *yapala* documented by Ašmarin (IV 204–206) suggest that the word originated as a metasyntactic expression which was ultimately lexicalized, cf. English *whatsit* 'thing' < *what is it* and *whatchamacallit* 'thing' < *what you may call it* for parallels.

Ašmarin's explanation for Cv. *yapala* was overlooked by the Chuvash etymological dictionaries of Egorov (1964) and Fedotov (1996), and therefore the word may have seemed more mysterious than it in fact is. Accepting the Chuvash etymology from *yepe* 'how?' allows viewing the Mari word as a borrowing from Chuvash; the Mari word is, after all, found only in the Volga dialect, which maintained contact with Chuvash as late as the 20th century. There are no phonological obstacles to borrowing from Chuvash into Mari in this case, though the chronology of the borrowing is ambiguous. The existence of forms *yepele* and *yapala* in Chuvash points to original front vocalism. Thus, the word must either have been borrowed prior to the Meadow Mari backing of PMari *ä, i.e. Middle Chuvash *yäpälä > Mari *jäβälä (> modern Volga Mari *jaβala*), or instead the Chuvash word was borrowed into Mari subsequent to the backing of Chuvash *ä > a, i.e. Cv. *yapala* > Volga Mari *jaβala.*

While the etymology of the Chuvash interrogative adverb *yepe* 'how?' (along with variants such as *leple* and *neple* with differing initial consonantism) remains unclear, again this Mari–Chuvash lexical parallel does not serve as proof of contacts between both Mari and Chuvash and a third language of the region, and an explanation should be sought within the separate history of Chuvash.

Finally, with regard to Mari *jokrok* ~ Cv. *yăkraka* 'boring' (No. 57), it should be noted that, across the Mari dialects, the meanings 'empty; creepy; lonely' are also attested for the Mari word. In his recent dialectal dictionary Veršinin (2011: 117) has documented the word *joŋrok* 'один-одинёшенек' from the Verkhnjaja Iž-Bob'ja dialect of Eastern Mari, and we can assume that this represents the same word due to the matching semantics. Consequently, we can reconstruct PMari *joŋrok 'boring; lonely', with the bulk of Mari dialects undergoing cluster simplification to *jokrok*.[9] This form, *jokrok,* with *k* instead of earlier *ŋ would then have been borrowed from Mari into Chuvash; the opposite direction of borrowing is not possible, as Chuvash shows a different resolution of the -*ŋr*- cluster, cf. Cv. *tură* 'god' < Proto-Turkic *teŋri (Fedotov 1996: II 252–253). The variant Hill Mari forms *jə̂k(ə̂)rikä* and *jə̂k(ə̂)raka* can be explained through reborrowing from Chuvash.

---

[9] Clusters involving *ŋ* followed by a consonant have been unstable in the history of Mari, cf. for example MariE *umdo* W *ə̂ŋgə̂δə̂* 'spear' <PMari *ŭŋδə̂.

4. Onomatopoeia/sound symbolism

As a third objection, some of the items in the "Late Gorodets" corpus can be shown to be of an onomatopoetic nature or cases of sound symbolism, and in such words irregular phonetic correspondences are commonplace and expected. On the basis of solely such irregular phonetic correspondences, these words do not serve as proof of borrowing from a third language.

An example of this is Agyagási's comparison (as No. 17) of Cv. V *lŏnčŏ*, *lĕnčĕ* 'слабый, нетуго натянутый' with MariE *lânč̂ᵊra* NW *lânẑᵊra* W *lənzəra* 'schwach, weich (Mensch, Pferd); dünngewebt (Stoff); zerfetzt (Kleid)'. If we draw on the full breadth of Mari lexical resources cited herein, we find the same root 'soft' in a highly varied number of forms: MariE *lanč̂ᵊra, luńč̂ᵊra*, *lazᵊra, lânč̌e-lańče* NW *länzərä, lŭnzᵊra* W *l'änzərä, l'äzərä, lənzəra, lânẑᵊra*. Confronted by this wild difference in vocalism and the presence or absence of the nasal, it would be far more straightforward to consider this a sound-symbolism root where either *ŭ, *ä or *ə could be plugged into the consonantal skeleton *lV(n)ć-, than to consider this a post-1200 CE borrowing into both Mari and Chuvash. Agyagási's attempt to argue for contact with dialects of an unknown third language by divergent forms from one Mari dialect to another (see her treatment of Mari *meraŋ* above) breaks down when we consider that even within the same dialect we find a variety. For example, speakers from the Bol'šoj Kil'mez dialect provided both *lańč̂ᵊra* and *lazᵊra*. The word in Chuvash must be a loanword owing to the initial *l*-, but the word can be explained as a borrowing from Mari, and the above-cited Mari forms with a labial vowel serve to explain the back labial reduced vowel in Viryal Chuvash *lŏnčŏ*.

Similarly, the comparison Cv. *kăsăya* ~ Mari *kisa* 'синица' (No. 10) is found in the "Late Gorodets" corpus. While the forms attested in the Mari and Chuvash dialects show irregular correspondences, the tit bird has a distinctive call that lends itself well to onomatopoeia.

5. Conclusions

The objections raised in the course of this work ultimately affect approximately 45 items of the 64 in the "Late Gorodets" wordlist. Some of the Mari words may well go back to Proto-Uralic, inasmuch as phonologically nothing speaks against tracing them to a Proto-Uralic ancestor. In other cases, we might seek an origin in post-Proto-Uralic contacts with other languages of the region. The Chuvash material in most cases can be explained as loans from Mari. This in fact represents a return to the *status quo ante*, as Lukojanov (1974) and Fedotov (1990) prior to Agyagási's "Late Gorodets" hypothesis were inclined to treat many of these words as just like any other Mari loans into Chuvash.

Consequently, for those remaining items where irregularities do appear among the Mari dialectal forms, other explanations should be sought. With regard to Mari *meraŋ* 'hare', we might revive the suggestion by Räsänen (1933: 360) that this word is a compound word consisting of Mari *mü 'earth' (reconstructed as the root of attested *mlande* id.) and perhaps a lost Viryal Chuvash noun *xoraŋ that can be reconstructed on the basis of cognates in other Turkic languages, cf. Tatar *kujan*, Altai *kojan*, etc., and consequently the Mari dialectal forms are the result of differing processes of vowel contraction in the post-Proto-Mari period.

While the exact identity of the Cheremis people attested in historical sources remains an outstanding problem of the history of the Volga–Kama region, evidence is lacking that the material collected in the "Late Gorodets" corpus must necessarily be traced to the language spoken by this people instead of to other possible sources for Mari and Chuvash vocabulary.

Abbreviations

MariE Eastern and Meadow Mari
    W Western (Hill) Mari
    NW Northwestern Mari
Cv. Chuvash
    A Anatri dialect of Chuvash
    V Viryal dialect of Chuvash
Tat. Tatar

References

Aikio, Ante 2014a. On the reconstruction of Proto-Mari vocalism. *Journal of Language Relationship,* 11: 125–157.

Aikio, Ante 2014b. Studies in Uralic etymology III: Mari etymologies. *Linguistica Uralica,* 2014(2): 81–93.

Aikio, Ante. forthcoming. Proto-Uralic. In: Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik (eds.), *The Oxford guide to the Uralic languages*. Oxford: Oxford University Press.

Agyagási, Klára 2000. Der sprachliche Nachlaß der Spät-Gorodec Bevölkerung in den tschuwaschischen und mariischen Mundarten. *Folia Uralica Debreceniensia*, 7: 3–26.

Agyagási, Klára 2001a. Die Spuren der Sprache der Spät-Gorodec Bevölkerung in den tschuwaschischen und mariischen Mundarten. Congressus nonus internationalis Fenno-ugristarum Pars IV. 35-39.

Agyagási, Klára 2001b. Nyelvi kapcsolatok történeti típusainak vizsgálata a csuvas–cseremisz nyelvviszonyban [An examination of the types of language contact in relation to Chuvash and Cheremis]. *Folia Uralica Debreceniensia* 8: 27–33.

Agyagási, Klára 2002. Название черемис в средневековых источниках. In: И. В. Тараканов (ed.), *Пермистика-9. Вопросы пермской и финно-угорской филологии.* Ижевск: Изд. дом "Удм. ун-т", 102–109.

Agyagási, Klára 2019. *Chuvash historical phonetics: An areal linguistic study. With an Appendix on the role of Proto-Mari in the history of Chuvash vocalism.* Wiesbaden: Harrassowitz.

Ašmarin = Ашмарин, Н. И. 1928–1950. *Словарь чувашского языка 1–17.* [Dictionary of the Chuvash language] Казань & Чебоксары: Издание Народного комиссариата по Просвещению Чувашской Автономной Советской Социалистической Республики и Чувашского Государственного Университета.

Beke, Ödön 1997–2001: *Mari nyelvjárási szótár (Tscheremissisches Dialektwörterbuch) 1–9.* Bibliotheca ceremissica 4. Szombathely: Berzsenyi Dániel Tanárképző Főiskola.

Bereczki, Gábor 1977. Permi-cseremisz lexikális kölcsönzések [Permian–Cheremis lexical borrowings]. *Nyelvtudományi Közlemények*, 79: 57–77.

Bereczki, Gábor 1987. Пермско-марийские лексические связи: Сущность, развитие и функции языка. Москва: Наука, 112–115.

Bereczki, Gábor 1992–1994: Grundzüge der tscheremissischen Sprachgeschichte I–II. Debrecen: Universitas Szegediensis de Attila József nominata.

Bereczki, Gábor 2003. Adalékok a permi–cseremisz nyelvi kapcsolatokhoz. In: Zoltán Molnár and Gábor Zaicz (eds.), *Permistica et Uralica: Köszöntő könyv Csúcs Sándor tiszteletére* [Permistica et Uralica: Festschrift for Sándor Csúcs]. Fenno-Ugrica Pázmániensia 1. Piliscsaba: Pázmány Péter Katolikus Egyetem, 39–41.

Bereczki, Gábor 2004. Újabb adalékok a permi-cseremisz lexikális kölcsönhatáshoz [New data on Permian–Cheremis lexical contact effects]. In: Eszter Várady and Márta Csepregi (eds.), *Permiek, finnek, magyarok: Írások Szíj Enikő 60. születésnapjára* [Permians, Finns, and Hungarians: Festschrift for Enikő Szíj on the occasion of her 60th birthday]. Budapest: ELTE BTK Finnugor Tanszék, 14–17.

Bereczki 2005 = Берецки, Габор 2005. Пермско-марийские лексические совпадения – заимствования или общие субстратные элементы? *Linguistica Uralica*, 41(3): 187–200.

Bereczki, Gábor et al. 2013. *Etymologisches Wörterbuch des Tscheremissischen (Mari). Der einheimische Wortschatz. Nach dem Tode des Verfassers herausgegeben von Klára Agyagási und Eberhard Winkler*. Veröffentlichungen der Societas Uralo-Altaica 86. Wiesbaden: Harrassowitz.

Dobrodomov = Добродомов, И. Г. 1980. *О некоторых гиппологизмах и созвучных словах: Из аланского пласта иранских заимствований чувашского языка.* In: *Проблемы исторической лексикологии чувашского языка*. Труды ЧНИИ 97. Чебоксары: Научно-исследовательский институт языка, литературы, истории и экономики при совете министров чувашской АССР, 21–29.

Fedotov 1990 = Федотов, М. Р. 1990. *Чувашско-марийские языковые взаимосвязи*. Саранск: Издательство саратовского университета, саранский филиал.

Fedotov 1996 = Федотов М. Р. 1996. *Этимологический словарь чувашского языка I–II*. Чебоксары: Чувашский государственный институт гуманитарных наук.

Gordeev 1967 = Гордеев Ф. И. 1967. *Балтийские и иранские заимствования в марийском языке*. In: *Происхождение марийского народа*. Йошкар-Ола: Марийское книжное издательство, 180–202.

Isanbaev = Исанбаев, Н. И. 1989–1994. *Марийско-тюркские языковые контакты I–II*. Йошкар-Ола: Научно-исследовательский институт языка, литературы и истории им. В. М. Васильева.

Ivanov and Tužarov = Иванов, И. Г. and Тужаров, Г. М. 1971. *Словарь северо-западное наречия марийсково языка*. Йошкар-Ола: Марийский научно-исследовательский институт.

Joki, Aulis 1973. Uralier und Indogermanen. Die älteren Berührungen zwischen den uralischen und indogermanischen Sprachen. Mémoires de la Société Finno-ougrienne 151. Helsinki: Suomalais-Ugrilainen Seura.

Lukojanov = Лукоянов, Г. В. 1974. *Марийские заимствования в чувашском языке*. Чебоксары: Чувашское книжное издательство.

Metsäranta, Niklas 2020: Läpiä päähän – huomioita itämerensuomalaisesta läpi-sanueesta. In: Sampsa Holopainen et al. (eds.), *Scripta miscellanea in honorem Ulla-Maija Forsberg*. Mémoires de la Société Finno-ougrienne 275. Helsinki: Suomalais-Ugrilainen Seura, 228–232.

Räsänen, Martti 1920. *Die tschuwassischen Lehnwörter im Tscheremissischen*. Mémoires de la Société Finno-ougrienne 48. Helsinki: Société Finno-ougrienne.

Räsänen, Martti 1933. *Tscheremissische u.a. Etymologien. – Liber semisaecularis Societatis Fenno-Ugricae*. Mémoires de la Société Finno-ougrienne 67. Helsinki: Suomalais-Ugrilainen Seura.

Rédei, K. & Róna-Tas, A. 1982. Proto-Permian and Votyak loan-words in Chuvash. In: András Róna-Tas (ed.), *Studies in Chuvash etymology I*. Studia Uralo-Altaica 17. Szeged: Universitas szegediensis de József Attila nominata, 158–179.

Rédei, Károly and Róna-Tas, András 1983. Early Bulgarian loanwords in the Permian languages. *Acta Orientalia Academiae Scientarum Hung*., 37(1–3): 3–41

Savatkova = Саваткова А. А. 1969. *Русские заимствования в марийском языке*. Йошкар-Ола: Марийское книжное издательство.

Savelyev 2018 = Савельев, А. В. 2018. К уточнению сценария чувашско-марийских контактов. In: А. М. Иванова & Э. В. Фомин (eds.) *Языковые контакты народов Поволжья и Урала. Сборник статей XI Международного симпозиума.* Чебоксары: Чувашский государственный университет имени И. Н. Ульянова, 95–104.

Scherner, Bernd 1977. *Arabische und neupersische Lehnwörter im Tschuwaschischen*. Wiesbaden: Franz Steiner Verlag.

Sergeev = Сергеев, О. А. 2002. *Истоки марийской письменности*. Йошкар-Ола: Марийское книжное издательство.

Tenišev = Тенишев Э. Р. (ed.) 1984. *Сравительно-историческая грамматика тюркских языков: Фонетика.* Москва: Наука.

TschWb = Arto Moisio and Sirkka Saarinen (eds.) 2008. *Tscheremissisches Wörterbuch*. Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen. Lexica Societatis Fenno-Ugricae 32. Helsinki: Suomalais-ugrilainen Seura – Kotimaisten kielten tutkimuskeskus.

UEW = Rédei, Károly 1988–1991. *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Veršinin 2011 = Вершинин, В. И. 2011. *Словарь марийских говоров Татарстана и Удмуртии.* Йошкар-Ола: Научно-исследовательский институт языка, литературы и истории им. В. М. Васильева.

"Сувениры Севера"
Minority identity and discourse:
Representation of indigenous minorities of
Northern Russia in the digital media
The case of Dudinka[1]

Zsuzsa Várnai and Ágnes Hámori

Hungarian Research Centre for Linguistics, Budapest

1. Introduction

This paper examines the representation of indigenous minorities of Northern Russia in the local online media, focusing on the relationship between language, culture, and ethnic identity. Using a critical sociolinguistic and discourse analytic approach, we analyze the discursive representation of five indigenous minorities on three digital media platforms: in the news of a state-run online newspaper, and the social media platforms of two indigenous cultural centers.

Based on the results, we show how indigenous languages and identities are represented in the online discourse, which elements of minority identity are referred to and in what ways, and how they are also mixed and intertwined with other local and global identities. The media representation of the indigenous languages and communities reveals a lot about how people see identity categories and relations, which attitudes, ideologies and values are tied to them, and, therefore, it helps to identify locally relevant status and factors concerning minority language and identity. Furthermore, public representations are important because they not only reflect but also construct language and identity categories and, shaping how people see themselves and others or indigenous languages in general, they may influence the status and processes related to minority identity and indigenous languages in the examined communities.

1.1. Aims and research questions

The research reported on here is part of a larger project aimed at examining linguistic and cultural identity and the integration of minorities into a multilingual urban environment among representatives of northern indigenous multilingual communities living in the urban settlements of Far North Russia and Finland. In this study, we focus on five indigenous minorities (the Nenets, the Dolgans, the Nganasans, the Evenkis, and the Enets) living in Dudinka, a city in northern Russia, and investigate their representations on the online media platforms in terms of the connection between language, culture, and minority identity.

The main research questions of the study are as follows: How are the indigenous language and minorities represented in the discourses of local digital media? What are the main social identity categories represented in these texts? What relationships are shown between the indigenous language, culture, and minority community? How is the minority language used and/or mentioned in these discourses, and to what extent is language presented as an essential part of ethnic identity? Which other markers of ethnic minority are also used and maintained as elements of identity after indigenous people leave their traditional way of life and move to the urban environment of a modern city? How is the minority identity connected or intertwined with other – local and global – identities?

The theoretical and methodological resources of the study – beyond common sociolinguistic knowledge concerning minority languages and communities – are from critical sociolinguistics (Blommaert 2016, Pietikäinen 2008) and critical discourse analysis (Fairchlough 1992), combined with a digital ethnographic perspective (Varis 2015).

The aim of the analysis is to gain insight into the recent concepts and categories connected to language and minority identity of the indigenous communities in this northern Russian city. To do that, this article seeks to involve new methods of sociolinguistic investigation – especially discourse analysis – and to expand research to the discourse of digital media, which has become a central space of public communication of our time. It plays a pivotal role also in the representation and construction of social relations and identities, typically connected to various communication practices and techniques.

Very few sociolinguistic studies have been conducted on these small indigenous languages and minorities of Northern Russia such as the Nenets, Dolgan, Nganasan, Enets, and Evenki communities. Although many linguistic works investigate them, but these focus on the grammatical description of languages, while sociolinguistic aspects are much less researched.

Most sociolinguistic studies use traditional methods, e.g. questionnaires, interviews, or field observation. Digital discourses, online media, and social media

have not so far been included in these analyses – this also increases the novelty and relevance of the present study. An important exception is the outstanding work by Pietikäinen, primarily on the Sami communities and languages, in which she uses critical discourse analysis and a critical sociolinguistic and ethnographic approaches and, in analyzing media discourse (Pietikäinen 2003, 2008, 2018), provides significant information about Sami people and language use.

These minorities in Northern Russia are in a similar sociolinguistic situation as the Sami, but there are some differences as well, e.g. a smaller number of speakers or the even more severely endangered situation of the languages (Mosely 2010) with minimal or no revitalization tendencies. Our analysis will hopefully help explore the role of language and its relationship with minority identity among these communities, and other sociolinguistic phenomena as well, which are locally relevant for them. The results contribute to a better understanding of the recent sociolinguistic situation of these languages, but from a broader perspective, they may be related to other endangered languages or minorities as well.

1.2. Theoretic background and methods

The theoretical and methodological background of this study drew mainly on theories and concepts of critical sociolinguistics, combined with critical discourse analysis (Fairchlough 1995, Wodak and Meyer 2009) and a digital ethnographic perspective (Varis 2015). We relied especially on the work of Blommaert on globalization, superdiversity, mobility and complexity (Blommaert 2016, Blommaert and Rampton 2011), and the research and concepts of Pietikäinen concerning cultural hybridization, news representation, identity, and the language use of the Sami people (Pietikäinen 2003, 2008, 2010, and 2018).

Research on minority languages and identity has had a long history in traditional sociolinguistics and, recently, in the new sociolinguistic work, often referred to under the umbrella term "critical sociolinguistics" or "sociolinguistics of globalization". This latter approach highlights the pivotal role of globalization, digitalization, mobility and diversity in sociolinguistic processes:

> "Over the past two decades, globalization has altered the face of social, cultural and linguistic diversity in societies all over the world. […] The demographic and social changes are complicated by the emergence of new media and technologies of communication and information circulation", […] with people getting involved in transnational networks that offer potentially altered forms of identity, community formation and cooperation" (Blommaert and Rampton 2011: 1.).

To describe this ongoing, multiple, and complex sociolinguistic processes, recent sociolinguistic studies have provided a number of new concepts and terms such as, for instance, superdiversity, translanguaging, hybridity, transnationalism, multilingualism, plurilingualism etc., emphasizing the need for a more complex, dynamic view of sociolinguistic perspective", in which, mobility, mixing, political dynamics central concerns in the study of languages, and communication" (Blommaert 2016, Pennycook 2016, Bartha ed. 2016, Bodó and Heltai 2018 for an overview).

In this approach, based on social constructionists theories, identity is viewed as a socially constructed, multiple and complex category, as "multiple, dynamic sites of struggle and investment, as the object of multiple discourses" (Pietikäinen 2018: 185), which has dynamically changing boundaries and is "unfixed, fluid, multiple, as process continuing across times, places, positions, practices and discourses […], a matter of becoming as well as of being" (Hall 1996, Pietikäinen 2003). Our study draws upon this approach of social identity, implying that discourse, communication practice and, especially, public discourse and media representations have a pivotal role with regard to identity relations: they reflect and reveal the boundaries of identity categories and various attitudes, ideologies and evaluations connected to them, but they also construct these categories, shaping how people see themselves and others. This is particularly important in relation to language and connections between minority language and identity; there is a "whole web of intertwined discourses, practices, and emotions related to the production of identity and language categories, and the relationship between the two" (Pietikäinen 2018).

This approach and concepts are especially useful for investigating indigenous languages and communities in northern Russia, since they are particularly affected by political, economic and social changes, mobility and the multiple effects of globalization and a modern urban environment. The term mobility in this case does not only imply geographic movement (from rural to urban areas) but also profound social and economic changes, which all have severe effects on minority languages and identity, causing various – often adversarial – processes and changes. As Pietikäinen describes, based on her longitudinal research on language and identity of the Sami people:

> "amid the turbulent currents of global changes, the once-peripheral spaces of the Arctic now lie at the epicenter of an ambivalent conjunction of at least three major forces: climate change, expanding economic interest and cultural transformation. Under these changing circumstances, language and identity boundaries that have so far been used primarily for social structuration and political projects have become also resources for economic development in

the context of the new economy of experience tourism and markets of authenticity" (Pietikäinen 2018: 184).

Also, "linguistic and other semiotic resources are used not only for communication and building an indigenous community, but they function also as a commodity in global music markets and tourism, or as a resource for creative identity performance" (Pietikäinen 2003: 79).

By analyzing minority representation in media news and posts, we rely on a critical discourse analytic approach (Fairclough 1995, Wodak and Meyer 2009). Both from the field of sociolinguistics and critical discourse analysis, numerous studies have proven that public discourses and, especially, digital media and news play an important role concerning minorities, their languages, and minority identity:

> "identity consists partly of the ways in which we are represented. The media are an important and influential public space of contemporary society […]; among various other public portrayals, news representations are crucial in representing culture, people, politics and social life: news representations contribute to ways in which people see themselves, their own identity and that of others, and the relations between 'us' and 'them' . […] News, therefore, reveals a great deal about a society and its power relation" (Pietikäinen 2003: 589).

Public discourse reflects a lot about boundaries and elements of identities and reveals locally relevant categories, attitudes, relations and evaluations. In this paper, we analyze the news mainly from this perspective, to explore general characteristics of locally relevant categories and relations. However, also the constructive potential of these representations must be strongly emphasized: discourse is not only shaped by, but also shapes social structures and relations, it "contributes to the creation and constant recreation of the relations, subjects […] and objects which populate the social world" (Fairclough 1995).

It is important to notice that not only the representations found in the texts are important: something that is *not* represented plays a role as well, i.e. what is absent or omitted from texts, since the lack of representation is also a way of construction (Wodak and Meyer 2009: 114, Fairclough 1995: 210). In addition to discourse analysis, we also use a digital ethnographic perspective to reveal locally specific digital practices and semiotizations related to social identity categories and the "lived local realities" (Varis 2015: 55).

For investigating the representation of minorities, we collected and analyzed news articles and posts from three local digital media sources: (1) the Krasnoyarsk regional online newsletter *Taimyr Telegraf* (TT), (2) the official website of the City Folk Art Center of Dudinka (FAC); and, finally, (3) the Facebook page of the House of Folk

Arts of Taimyr – Chum (Ch). The three sources have slightly diverse genres (newspaper vs. social media posts) and different perspectives (from the position of the state and the members of community itself), thus, the database created from this collection of news covers a wider range of local discourses and representations. For this paper, we analyze this database as a whole, focusing on a general description of the characteristics, main aspects and elements of minority representation; we do not address further questions such as, for instance, differences between the state-run media and the community platforms, or longitudinal changes in the media representations of minorities etc., as these aspects should be explored in future studies.

In line with critical sociolinguistic and digital ethnographic principles, we completed the textual investigation with multimodal analysis involving pictures and other visual elements from the news, "moving from 'language' in the strict sense towards semiosis as our focus of inquiry, and from 'linguistics' towards a new sociolinguistically informed semiotics as our disciplinary space" (Blommaert and Rampton 2011).

In 2019, the relationship between languages and minorities was a particularly significant issue, as the UN has declared 2019 the International Year of Indigenous Languages. Therefore, we placed this period at the center of our research and built the database form the news articles and posts published during this year, from January 1 to December 31, 2019. We analyzed the news from various perspectives, for instance: How many posts and articles were published with regard to indigenous issues? Which topics were frequently involved? Were indigenous languages used or mentioned and if so, in what form or by which terms? Which ethnic groups and other social identity categories appeared in these discourses? What kind of representations were put forward, and how do they contribute to the construction of indigenous identity, and/or to the position of the ethnicity? What were the main textual manifestations of ethnic representation? What kind of quotation patterns, key concepts and collocations, metaphors were typically used in news reports? Did also non-linguistic indices of indigenous identity – e.g. clothes, food, objects on pictures – appear in the observed media, and what role did they play as part of the visual and multimodal semiotic repertoire?

Although we conducted fieldwork in the communities in question to gain insight into local practices and situation, the analysis was not based on a priori categories but focused on the emergent categories and meanings in the discourse. Built on the findings, we present some quantitative data, for instance, on frequency, however, exact numbers do not have a central role, they serve mainly to show proportions and tendencies in the discourse.

In the next subsection, we give a brief overview of the role and specific features of digital media and communication on the Internet. Section 2 provides a brief

summary on the sociolinguistic background of the indigenous languages and minorities investigated in this paper. In section 3, a detailed analysis of various aspects of the news is presented, and, finally, section 4 contains a short discussion and conclusion.

1.3. The discourse in digital media: the role of the internet as a multimodal semiotic space of communication

In our age, the Internet is one of the most important spaces of social publicity. Information and communication technology has created new types of space-time relationships and is changing our way of life, our practical guide to communication, and identity formation. Cyberspace has the greatest impact not only on the delivery and processing of information but also on the development of social relations. It affects self-awareness and communities. The interactive social contact that takes place in it has a significant impact on individual people, changing their worldview and values (McRae: 1997). The Internet is both a public and a private space, providing a new terrain for experiencing the difference between individual and collective identity at the same time.

Social networks play an important role not only as a communication channel but also as a means of building organizations and mobilizing people. The fast and cheap access to information provided by the Internet encourages civic activism, which brings people together and helps them cross geographical and social boundaries, increasing public participation in social and political arenas. This is a public space that also provides an opportunity for social and political debates (Papacharissi 2003; Poster 1995).

Regarding the communication effects of cyberspace, two main factors can be highlighted. One is interactivity, which is primarily dominant in this medium, and whose wide range of possibilities has made social media sites, especially Facebook, unique. There is a need for interaction, an active collection of opinions, and "personal" contact with the audience. This was linked to various commenting opportunities; the social network thus created a specific interactivity that differs from both face-to-face and online forms of correspondence. For these, Facebook in particular provides many opportunities: liking, sharing, and commenting, through which the audience is mobilized, and not only activity, but also sympathy towards and commitment to a particular cause can grow. Manifestations expressing and/or creating identity are of paramount importance in this interactive social and societal space. Another defining feature of cyberspace is fragmentation. The information obtained online is mostly fragmented in nature, into fragments that show only one side of a question or are made up of randomly compiled opinions and unverified facts. At the same time, it also

shapes the mental environment, the context in which we create meanings and values, and in which reality is represented and constructed (Schement and Curtis 1997: 120).

Given the signifying power of the news media, the ways in which news represents various ethnic communities, issues related to multiculturalization and a changing society have significance in relation to the positions of various groups and to the relations between them. News is one of the most influential public spaces of contemporary society, perceived as offering stories about the world, its events and people. Minority media are often considered as a resource for promoting a sense of belonging and shared identity of the community, as well as offering a new domain for cultural and political participation. Indigenous people seek ways to take part in producing media publicity in their own terms because media space is seen as an important tool for identity politics, construction of community/nation, and revitalization of endangered languages. Minority language media are often considered to be an important element in revitalization of endangered languages. The terrain of minority media production and consumption is increasingly multilingual. The news can constitute an arena for constructing identity where the interests of different groups and regions will be heard and discussed (Allan 1999).

Many people believe that information technology and the cyberspace thus created destroy traditional forms of cultural and social relations, promote the expansion of global culture, and thus weaken local customs, cultures and traditions. At the same time, as the present study also points out, the interactions between the local and global spheres and the emergence of new forms of local character can also be observed. Many communities around the world have recognized the opportunities offered by cyberspace and they are working to strike a balance between the global and the local. They try to forge a capital from the uniqueness of their culture and from local nature (Mészáros 2001).

Cyberspace is a specific medium with significant economic implications. More and more people are realizing that a unique culture is a desirable product that attracts consumers, investment, and tourists. Therefore, in many cases, cities sell their cultural resources to gain capital (Lash and Urry, 1994). More and more city governing bodies and cultural institutions are establishing a virtual space for their city, with the intention that their presence on the Web will act as an economic factor for the benefit of the city. The cities of the world seem to be developing in two directions at once. They are less and less distinguishable from each other, becoming more global in nature and more homogeneous than before; at the same time, there is an emphasis on uniqueness and peculiarities, by which they try to present themselves as unique, different from all other places, in order to attract consumers. From these processes, a complex interaction of local and global traits emerges (Mészáros 2001).

2. Indigenous communities in Dudinka: geographic, demographic, and sociolinguistic background

Dudinka is a port city on the banks of the Yenisei, in the northernmost part of Siberia, beyond the Arctic Circle in Russia. Administratively, it also includes five villages: Volochanka, Levinskie Peski, Potapovo, Ust-Avam, and Khantaiskoe Ozero, which are located between 90 km to 340 km from Dudinka. It is the seat of the Taimyr district (Dolgan-Nenets) in the Krasnoyarsk Krai.

There are 21,978 people living in Dudinka (according to the city's official website on the basis of data from January 1, 2018) and 22,175 people according to the 2010 Russian census (the corresponding numbers were 25,132 in 2002, and 32,325 in 1989). The city is inhabited by five indigenous minorities in addition to the Russian majority: the Nenets and Dolgans in the largest numbers, as well as Nganasans, Evenkis, and Enets. Accurate data on the ethnic distribution of the urban population are rarely published: according to 2005 data, 1,328 of the 25,000 inhabitants were indigenous, representing only 5.3% of the urban population (Siegl 2013). According to the latest, 2019 data from the local administration, the city is home to 786 indigenous people, 390 of them Nenets, 111 Nganasans, 71 Evenkis and 24 Enets. Thus, indigenous peoples make up 9.9% of the town's population.

The official language of the city is Russian. The members of the indigenous population are very different and diverse in terms of language skills depending on the individual and the ethnic community, and their language use habits are also very variable. Whereas for some a language is their mother tongue, for others it is learnt in school or later in life. There are also people who hardly speak their heritage language.

The older generations are mainly those who speak and use the indigenous language. They work mostly in the educational and cultural centers, in the media, where the indigenous minority culture, language and identity have been preserved. However, the larger proportion of communities have little or no knowledge of their mother tongue.

On the list of endangered languages by UNESCO, each of these languages is listed as severely endangered (see Moseley 2010). Although to a very small extent, in recent decades revitalization efforts can be observed in the city: family clubs have been established (in the Nenets, Dolgan, and Nganasan communities), events, classes, language courses have started, "language nests" and summer schools have been established. However, these are mainly ad hoc, project-based initiatives of the state. The teaching of minority languages is the prerogative of some schools only, and they teach only one indigenous language (Nenets, Nganasan, or Dolgan) even in the absence of mother tongue teachers and teaching materials. Students can learn them as a foreign language in 2 to 4 hours a week.

3. Analysis

3.1. The corpus

For the present study, the preparation and data collection lasted one year, from January to December 2019. In the course of this, we processed three different online media sources of different genres, themes, scope, and operational backgrounds: (1) the news published on the official website of the City Folk Art Center of Dudinka (FAC); (2) the news on the Facebook page of the House of Folk Arts of Taimyr – Chum (Ch), and finally (3) the news of the Krasnoyarsk regional online newsletter *Taimyr Telegraf* (TT).

The first two media sources are connected to the folk art centers: they are maintained, edited, and operated by their staff, their primary purpose and aim is to report on the events of the centers. The *Taimyr Telegraf* is a broader, general purpose news portal, a state run and centrally edited online newspaper that provides news about the Krasnoyarsk region, with a more comprehensive view, not limited to indigenous minorities or cultural events.

	Taimyr Telegraf (TT)		Facebook page of the House of Folk Arts of Taimyr – Chum (Ch)		official website of the City Folk Art Center of Dudinka (FAC)	
2019	total	indigenous	total	indige-nous	total	indigenous
January	177	5	3	3	1	0
February	221	6	5	4	6	3
March	214	3	4	4	2	0
April	224	2	7	6	4	2
May	208	4	8	8	4	2
June	208	2	0	0	1	0
July	242	4	1	1	1	1
August	196	6	5	5	2	2
September	181	8	4	4	4	4
October	244	7	10	9	7	7
November	215	27	4	3	2	2
December	220	9	10	8	3	2
Σ	2,550	47	63	57	37	25

Table 1. Distribution of news articles by month, highlighting those dealing with indigenous peoples

Examining the year 2019, we can see from the data that, due to its nature and genre, the Taimyr newsletter is much more extensive, therefore we find a significant difference in the occurrence of indigenous-themed news compared to the other two media sources (only 1.8% of the news deals with indigenous peoples). On the other two examined sites, indigenous-themed statements dominate (Ch: 90%, FAC: 67%). Due to its genre, the Taimyr newsletter is thematically broader, not only the topics of folk art, culture, language, and customs appear in the news, but also the issues of economy and politics are given a strong emphasis. It is a state medium, where news is edited and texts are written by employees, while the news and entries of the two folk art centers are written by the members of the indigenous minorities themselves. All of this results in different perspectives and modes of construction; it is precisely these differences that justify the joint examination of the three media sources.

3.2. The representation of indigenous languages in texts: use and naming. The construction of identity categories

How do indigenous languages appear in media discourse? Indigenous language statements are seldom found in the examined texts only in very small numbers (Ch: 17 / FAC: 14 / TT: 13).

It can be ascertained that the use of indigenous languages in communication functions is not typical and frequent in these discourses, not even on the social media platforms, i.e. in the discourses of the indigenous community. These minority languages are used primarily in indexical functions, typically in the use of a single term, usually a name, such as that of an art group, person, place, or object related to indigenous culture:

Клуб «Май"ма»	Club '*May'ma*'
участники семейного долганского клуба «Биирге»	members of the family Dolgan club '*Biirge*'
по сказке «Нохо-богатырь»	according to the tale '*Noho*-hero'
Счастливый из народа «Ня».	Happy from the people of '*Nya*'

3.3. Labelling the members of the ethnic groups

Next we want to focus on a set of textual features of the news that are assumed to be central in representing indigenous minorities of Taimyr. We analyzed the lexicalizations referring to people living in Taimyr to find out how the members of the ethnic groups are referred in news texts. For the analysis we counted all words referring to indigenous peoples living in this area. Figure 1 shows the most common

labelling with numbers of occurrence and the list below it contains the names used in public groups living in Taimyr.
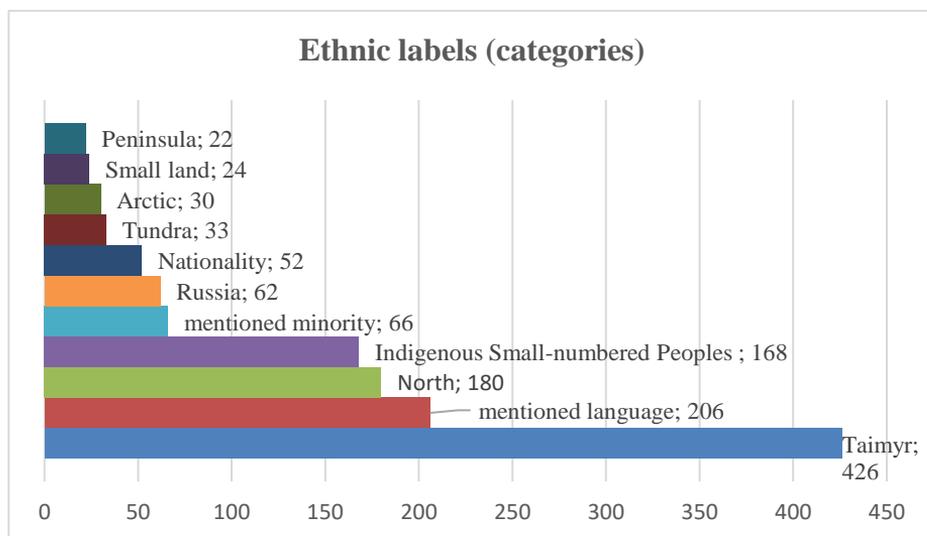
Figure 1. Ethnic labels in the articles

The lexicalizations referring to people living in Taimyr in the articles are the following:

Official names

Indigenous Small-numbered Peoples of the North, Siberia and the Far East of the Russian Federation
Indigenous Small-numbered Peoples of the North, Siberia and the Far East
Indigenous Small-numbered Peoples of the Far North
Indigenous Small-numbered Peoples of the North of the Krasnoyarsk Territory

Indigenous

indigenous dweller
indigenous peoples
indigenous population
indigenous ethnic groups: Nenets, Dolgans, Nganasans, Evenkis and Enets

Taimyr

indigenous dwellers of Taimyr
indigenous people of Taimyr
indigenous peoples of Taimyr

indigenous small-numbered peoples of Taimyr
indigenous dwellers of our peninsula
indigenous dwellers of the area
indigenous peoples of Russia
Taimyr
Taimyr treasures
Taimyrer / SUPER Taimyrer

North/Northerer

indigenous peoples of the North
indigenous small-numbered peoples living in the northern region
small-numbered peoples of the North
peoples of the Far North
peoples of the North
treasures of the North
northern peoples
northerners
nomad North
indigenous northener-nomads
indigenous northern peoples
indigenous northerners

Arctic

Arctic Talents
Guardians of the Arctic
the polar capital
Arctic Circle Artists

Tundra

tundra people
people of the tundra
indigenous tundra population

Homeland

owners of our land
compatriot
Taimyrer compatriot
compatriot from the indigenous peoples of our peninsula
great patriots of the northern land
land of our ancestors

Examining the different group names and categories in the news, we can state the following: several larger categories appear and are equally common, such as: *Russia, Small-numbered Indigenous Peoples of Russia, Small-numbered Indigenous peoples of the North and the Far East,* etc. It is important to mention that the minority vs. majority opposition is not observed at all. The concept of "minority existence" does not appear at all in these texts, these peoples are referred to by other names as indigenous peoples (according to the general use in Russia).

The ethnic categories of each nation are given an important role; in addition, local categories referring to the geographical location, such as *Taimyrian*, *Northerners*, *people of the tundra*, also appear prominently. We can make observations similar to Pietikainen's 2003 observations in the Sami community. Taken together, these outline a distinctive northern / Taimyrian identity that unites these five indigenous minorities and is based on the experience of a common destiny, habitat, and way of life. The indigenous identity of the peoples in the Taimyr has been articulated through times and territories, and across changing political, cultural, and legal circumstances This identity work has been a dialogue between the past and the present, between traditional and new conditions. This identity is fluid and multiple, it is a process across the times and places. It does not only involve language, livelihood or living in a specific area, but it also represents a deeper and more complex sense of belonging to a culture, family, and heritage that stretches over the North and across centuries. This Northern identity is also profoundly tied in with the region and the shared experience. Regardless of their ethnicity, people living in the same region share similar problems as inhabitants in their local communities. This is the familiar story of mixed marriages, of the gradual disappearance of minority culture, and of the oppression of the minority, resulting in, for example, their language shift; a high unemployment rate, the countryside emptying of people, and diminishing welfare services; the fragmentation of society and the hardening of walls between various ethnic groups.

3.4. Correlation with identity elements

Identity construction and the link between identity and language are the central topics of this investigation, thus, we seek a deeper understanding of what exactly the linguistic and non-linguistic markers of this identity mean for the indigenous peoples. To investigate this, we previously developed a special method for identity interviews called "Identity game" based on minority identity elements determined with the help of fieldwork observations and the results of our previous research (Duray 2019, Várnai 2018 and 2020). The following concepts were defined as the main components of the identity of the examined minorities: kinship, nationality of parents; traditions and culture; language; religion; knowledge of the past; and genetic similarity. Based on these, we constructed what we called the 'Identity game'. During the 'Identity

game' conducted during our fieldwork in Enontekiö in 2016 and in Dudinka in 2019, we requested the participants to arrange constituents of the indigenous culture in order of importance with regard to their own ethnic identity. The items related were printed on pieces of paper, and participants were asked to arrange them in order of importance, i.e. to put the item that describes their ethnic identity the best in the first place and the one that describes it the least in the last place.

Comparing the data during the analysis we examine the individual elements in their relation to each other, so we get an overview of which elements of the minority identity of the communities undergoing language shift are the ones that come to the fore during assimilation and language shift. Each element was identified separately for the different urban communities studied. For the sake of comparability, we sought uniformity, but this principle could not be fully enforced. There were elements that all indigenous communities could claim, but there were some that were unique to one or another: for example, while being listed in the electoral register is an important element of the minority identity of the Sámi community in Finland, it is completely negligible for those living in Russia due to their different political status. Or, while traditional food is an important part of identity of communities in Russia, it should not be included among the identity elements in the case of Sami in Finland, as Sami traditional food is not different from the food of the majority Finnish cuisine. Table 2 illustrates the relative importance of each item as participants perceived them.

Results for Enontekiö (Zsuzsa Duray)	Results for Dudinka (Zsuzsa Várnai)
Close Sámi relatives	Close relatives
The Sámi language	Language
Reindeer herding and related activities	Religion, traditional holidays
Traditional Sámi costumes	Folk art, handicrafts
Keeping contacts with the Sámi community	Traditional way of life (reindeer hunting/herding, fishing, living in the tundra)
Sámi handicrafts	Folklore, fairy tales, stories, and songs
Fishing	Living in Taimyr
Living in Lapland	Active participation in the activities of local cultural organizations
Sámi indigenous religion	Contact with the community, participation in cultural events

Results for Enontekiö (Zsuzsa Duray)	Results for Dudinka (Zsuzsa Várnai)
Taking part in the activities of Sámi organizations	Eating habits, cooking a traditional dish
Being listed in the electoral register of the Sámi Parliament	External / internal properties
The Sámi joiks/Joiking	
Staying in a Sámi tent	

Table 2. Aspects and components of minority identity in Enontekiö and Dudinka

For a deeper understanding and a comparison of the results provided by different methods, we in*vestigated how the aspects of minority identity examined in the 'Identity game' were represented – and if they were represented – in the self-initiated context of news discourses. The results are shown in Figure 2:
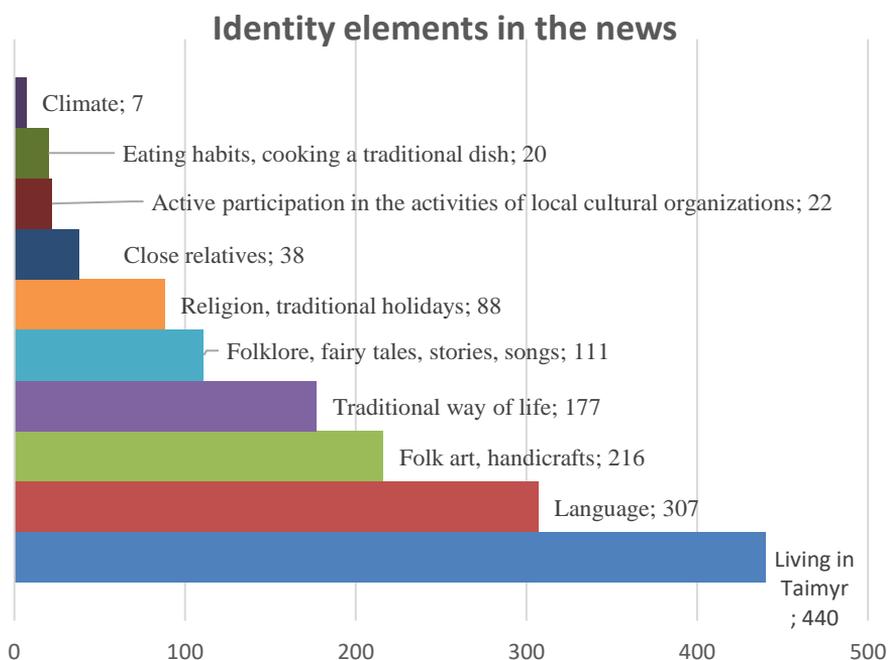

Figure 2. Representation of identity elements in the news

We can state that there are strong similarities between the lists of core components of ethnic minorities, both on the basis of data from a face-to-face interview-like situation and from public discourses.

Regarding the order of components, it should be noted that the genre characteristics of the media can influence the order of the elements to some extent: for example, close relatives as a topic does not play such a prominent role in the public media discourse as in the life of the individuals. Surprisingly, the category of local identity is especially prominent: 'Living in Taimyr' took first place. We observed in the articles a new element of identity as well: tolerance of the harsh climate and intemperate weather are mentioned with a positive emphasis, as a community-forming force:

"Как мужество, терпение и мудрость помогли выжить им в этих вечных снегах в течение долгих тысячелетий?"
"How could courage, patience, and wisdom help them to survive in eternal snow for millennia?"

"Каким образом в нечеловеческих условиях арктического климата нашли они силы и место в душе для любви к этой суровой, но родной для них земле?"
"How did they find enough strength and a place in their soul to love their harsh but native land in extreme conditions of the Arctic climate,"

"красота северянок в контрасте с суровыми условиями жизни их сила воли, умение сохранять и передавать традиции предков новому поколению"
"the beauty of northerners is their willpower and ability to preserve and transmit the traditions of their ancestors to the new generations in contrast to the harsh living conditions."

"В описуемых условиах могут выжить только сильные телом и духом люди! Не представляю, чтобы коренные северные народы согласились-бы жить в Африке."
"In those conditions described, only people strong in body and spirit can survive! I can't imagine the indigenous northern peoples agreeing to live in Africa."

A significant finding is that the minority language takes a prominent position, too, occupying second place among the most frequently mentioned identity elements, which signals an outstanding role of language in minority identity. It is even more impressive if we consider that these minority languages were not used in communicative function in these discourses at all – not even in the news of the minority cultural centers – and appeared only in symbolic (indexical) function. In spite of their endangered status, the indigenous languages still seem to be considered as something particularly important and valuable for minority identity. These representations, however, may have been influenced by the "Year of Indigenous Languages" e.g. because of more frequent mentions – future research should explore the longitudinal tendencies in this respect.

3.5. Key concepts and collocations related to language

For further investigation of the role of language and the values connected to it, in this section we review the phrases in which language – or *a* language – is explicitly mentioned in the news. We have already seen that it is mentioned in a large number of articles.

нганасанский/энецкий/долганский/н енецкий язык	Nenets / Dolgan / Nganasan / Evenki / Enets languages
уральскими тунгусскими языками / самодийских языков/ Языки Евразии	Uralic and Tungusic languages / Samoyed languages / Eurasian languages
языках коренных народов / Языки народов Таймыра / Языки и культура народов Севера	Languages of Indigenous Peoples / Languages of the Peoples of Taimyr / Language and Culture of the Northern Peoples
Международном Симпозиуме по языкам *лингвистической типологии* *изучение синтаксических структур языков* *отработать языковые грамматику искусственный язык*	International Language Symposium of language typology studying the syntactic structures of languages develop grammar artificial language
Международный год родных языков/ *Международному году языков коренных народов*	International Year of Mother Tongues International Year of Indigenous Languages
Дню языков коренных народов	Day of Indigenous Languages
приняли участие в языковых играх	they participated in language games
сохранения языков коренных народов Севера	preserve the indigenous languages of the North
развития языкового многообразия	developing linguistic diversity

Table 3. Collocations in which "language" occurs

We can observe phrases where the languages of language families are mentioned, by linguistic terms, or phrases with the year of indigenous languages, most of them are related to each language. There is also a significant group of expressions where language is mentioned in connection with positive concepts, e.g. as a valuable phenomenon to be preserved; something that ought to be learned and loved; or as the language of the ancestors; expressions referring to the charm of the language, and

language as a great gift. In this group, language is presented in significantly positive collocations, conveying a positive attitude towards it, which is also an important message for the whole community.

желание обучаться языкам	desire to learn languages
язык моих предков	language of my ancestors
изучать и любить свой язык	learn and love your own language
сохранение и развитие языков	preservation and development of languages
язык великий дар	language is a great gift
„не ослабевает желание обучаться языкам, на которых говорят коренные народы нашего полуострова"	"The desire to learn the languages spoken by the Indigenous Peoples of our Peninsula does not diminish"
«Волшебство родной речи»	"The magic of the mother tongue"

Table 4. Collocations in which "language" occurs with positive concepts

3.6. Multimodal components of identity and visual representation

Examining the elements of indigenous identity in visual signs and pictures, they were found so numerous and connected with so many meanings that a broader analysis would require an independent, more extensive study. Hence, here we highlight only a few features that are particularly important for our present topic and represent the emergence of indigenous identity in an urban setting. Regarding visual indices, it can be stated that one of the most significant phenomena is the mixing of different contexts and (identity) domains: e.g. the mixing and co-occurrence of times, past and present. Similarly, the mixing of indices belonging to different cultural contexts and cultural identities – resulting in a "cultural hybridization" (Pietikäinen 2008) – is also a central phenomenon.

Figure 3. Indigenous woman celebrating the Day of the Indigenous Peoples of
Taimyr (Source: Chum/facebook.com, 10. 08. 2019)

In the Figure 3 we can see a block of flats and a house with legs next to each other,
an indigenous tent made of modern materials, decorated with Nenets ornaments. In
this setting an indigenous woman takes part in the celebration of Argis, standing in
front of them. She is adorned with jewelry sewn from traditional pearls, while at the
same time wearing a modern T-shirt – connected to modernity, maybe even sportiness,
and to the urban environment – with plastic-color-printed minority figures dressed in
indigenous folk costumes.

Figure 4. Dolgans dressed in folk costumes with portraits of heroes of the Great
Patriotic War (Source: Chum/facebook.com, 9. 05. 2019)

Figure 5. Children taking part in a celebration of the Great Patriotic War (Source: Chum/facebook.com, 8. 05. 2019)

In both Figures above, 4 and 5, we see the actors of the state holiday (the Great Patriotic War) show. On Figure 2 the members of the Dolgan dance group dressed in folk costumes are holding the portraits of heroes killed in the war. On Figure 3 we see children dressed in military uniforms playing on traditional instruments. The elements of different cultures, communities and historical events represented together on these pictures (and at these events in the real life) results also here in a shared – complex and hybrid – cultural category, which is, at the same time, a mixing and a kind of integrating frame. This hybrid identity category eliminates traditional boundaries and makes it possible, that people with different sociocultural backgrounds share a common identity and feel themselves to belong to a same community.

Figure 6. Greeting with flowers for State Women's Day, an indigenous shaman woman drumming in the background (Source: Chum/facebook.com, 7. 03. 2019)
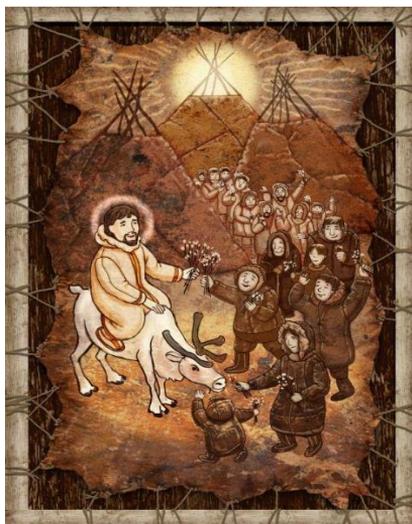
Figure 7. The cover of the Bible translated to Dolgan (Source: Chum/facebook.com, April 30, 2019)

In Figure 6, the elements of traditional culture (shaman woman with drum) are mixed with the motifs of a modern public holiday, belonging originally to the culture of the modernity and globalization (Women's Day). Figure 7 represents a leather picture stretched on a stylized wooden frame fashioned in traditional culture. In this traditional indigenous context, we see a scene of the arrival of Jesus – a central person of the Christian religion – who is sitting on a white reindeer, in traditional clothes among the natives with tents in the background.

On all this blending, one can observe the complex interaction of present and past, uniquely local and global, traditional and modern.

4. Conclusion

Summarizing the results of the analyses, the following can be stated:

1. The indigenous communities investigated here were represented frequently and salient in the local digital media discourse.
2. Their appearance on social media were typically accompanied by positive collocations and a very positive attitude (likes, comments).
3. In the content of the news, in the pictures and in the accompanied discourse – and by discourse – a peculiar world has been constructed, which, however, differs from the objective reality. It can be observed, for example, that problems and controversial issues affecting the indigenous people are not

represented in the news, they only can be inferred implicitly from other expressions:

"Компания поддерживает программы, направленные на улучшение качества жизни коренного населения Таймыра, уделяет внимание проблеме сохранения языков коренных народов Севера"	"The company supports programs aimed at improving the quality of life of the indigenous people of Taimyr, pays attention to the problem of preserving the languages of the indigenous peoples of the North"
"Среди участников не раз была делегация Таймыра. Но в этом году в Якутск из-за отсутствия финансирования она не поехала."	"Among the participants there was more than once a delegation from the Taimyr. But this year, due to lack of funding, they did not go to Yakutsk.

These findings are remarkable if one considers that not only representations but also non-representations in discourse – absence of some groups or names, omitted information or implicitly-left contents –, which make things invisible or backgrounded, are important communication techniques that can shape and construct social categories, relations and, finally, even reality.

Several different group and identity categories appeared in news texts. Ethnic identity was a frequently mentioned category, represented in many ways and with many various components. However, interestingly, there was an important identity category, a kind of a local identity which occupied an outstanding position – first place – among identity aspects, and seems to be a central and particular prominent category, as a kind of broader "tundra", "northern", "Taimyr" identity, defined by the central concepts of the broader local (geographic) area and patriotism:

"людей, которые проживают на Таймыре, – самое лучшее, яркое, достойное"	"The people who live in Taimyr are the best, the greatest, the most honorable"
"Самые самые СУПЕР Таймырцы"	"The most, most SUPER Taimyrians"
"великие патриоты северной земли"	"Great patriots of the northern land"

Concerning the role of langauge, a significant finding is that minority language was a very freqently mentioned identity element which has tied to an outstanding position and value in minority identity. This is even more impressive if we consider that this minority languages were not used in communicative function in these

discourses, only in symbolic (indexical) function, and, according to sociolinguistic research, they have a weak, severely endangered position or even near to extinct. Still is the concept of minority language important, and closely connected to the minority identity, positive attitude and values. However, these positive representations can also be – at least partly – only a result of influence of the "Year of Indigenous Languages".

Taking advantage of the opportunities offered by cyberspace, it can also be observed in the online discourse that the city seeks to attract consumption by "selling" itself on the Web (similar commodification is described by Pietikäinen within Sami communities, Pietikäinen 2010: 94). Emphasizing the unique local character, for instance, even a "Taimyr brand" is formulated:

"Марка "Сделано в Арктике"	"Brand "Made in the Arctic""
"высокий туристический потенциал полуострова"	"High tourist potential of the peninsula"
"о фестивале «Фольклорная классика Таймыра», который значится культурным брендом северной территории"	"About the festival" Folk Classics of Taimyr ", which is a cultural brand of the northern territory"

Indigenous languages appeared only in indexical use in articles, but they were particularly salient (often mentioned). Thus, they convey a strong positive attitude towards the community. In all of this, it can also be observed that the positive display of indigenous identity and language, and the idealized, problem-free-constructed world represent even the indigenous people as a kind of "jewel", as souvenirs of the North: "Сувениры Севера". This representation is the same – only constructed by linguistic devices, in and through discourse – as representations of minorities in the typical tourist brochures, where indigenous people are shown dressed in traditional folk costumes standing in the front of their tents and surrounded by reindeers in vast Nord scenery, similar to the other stereotypical representations of the nature-dominated aboriginal people (Bird 1999; Tan et al. 1997; Wilson 1998). However, the concept of "souvenir" – as the meaning of this representations as well – is twofold: it means not only a valuable treasure, but also an additional "decoration", a constructed scenery. Knowing the real linguistic and cultural situation of minorities, the latter meaning is rather relevant. However, even this idealized presentation, despite its false overtones, may have a constructing power and a positive, retaining effect on minorities: value-adding, appreciative, idealizing representation can have a positive impact on the attitudes of the communities. Thus, a possible effect is that this indigenous identity – which also includes the language that has hardly been spoken, but is considered valuable – will be considered more and more prominent in the city

and has chance to maintain, even if in some hybridized categories, a relevant part of his components and values.

References

Allan, Stuart. 1999. *News culture*. Buckingham: Open University Press.

Barker, Chris. 1999. *Television, globalization and cultural identities*. Buckingham: University Press.

Bartha, Csilla (ed.) 2016. *A többnyelvűség dimenziói: Terek, kontextusok, kutatási távlatok*. [Dimensions of multilingualism: Spaces, contexts, research perspectives] Általános Nyelvészeti Tanulmányok 28. Budapest: Akadémiai Kiadó.

Bell, Allan. 1991. The *language of media*. Oxford: Blackwell.

Bell, Daniel 1981. The social framework of the information society. In: Forester, Tom (ed.), *The microelectronics revolution*. Cambridge: MIT Press, 500–549.

Bellah, Robert N., Madsen, Richard, Sullivan, William M., Swidler, Ann and Tipton Steven M. 1985. *Habits of the heart*. Berkeley: University of California Press.

Bird, S.Elisabeth. 1999. Gendered construction of the American Indian in popular media. *Journ al of Communication*, 49 (3): 61–83.

Bodó, Csanád and Heltai, János Imre 2018. Mi a kritikai szociolingvisztika? [What is critical sociolinguistics?] *Magyar Nyelvőr,* 142 (4): 505–523.

Duray, Zsuzsa 2018. "If you don't speak the language, you're excluded. You're a tough one, so to say": Conceptions of being a Sámi today as reflected in interviews on language and identity with Sámi people in Enontekiö, Finland. *Finnisch-Ugrische Mitteilungen,* 42: 1–51.

Fairclough, Norman. 1995. *Critical discourse analysis: the critical study of language*. London: Longman

Grossberg, Lawrence. 1996. Identity and cultural studies: Is that all there is? In: S. Hall and P. Du Gay (Eds.), *Questions of cultural identity*. London: Sage, 87–107.

Hall, Stuart 1996. 'Who needs 'identity'?'. In: S. Hall and P. du Gay (eds.), *Questions of cultural identity*. London: Sage, 1–17.

Lash, Scott and Urry, John 1994. *Economies of the sign and spaces*. London: Sage.

Kling, Rob 1996. Hopes and horrors: Technological utopianism and anti-utopianism in narratives of computerization. In: Kling, Rob (ed.), *Computerization and controversy: Value conflicts and social choices.* Boston: Academic Press, 40–58.

McRae, Shannon 1997. Flesh made word: Sex, text and the virtual body. In: Porter, David (ed.), *Internet culture*. London: Routledge, 73–86.

Mészáros, Rezső 2001. A kibertér társadalomföldrajzi megközelítése. [Socio-geographical approach to cyberspace] *Magyar Tudomány,* 46: 769–779.

Moseley, Christopher (ed.) 2010. *Atlas of the world's languages in danger*. 3rd ed. Paris: UNESCO Publishing.
Online version: http://www.unesco.org/culture/en/endangeredlanguages/atlas.
Date of access: December 4, 2020.

Negroponte, Nicholas 1998. Beyond digital. *Wired*, 6 (12). https://web.media. mit.edu/~nicholas/Wired/WIRED6-12.html. Date of access: December 4, 2020.

Papacharissi, Zizi 2003. A virtuális szféra. [Virtual sphere] *Médiakutató,* 4 (1): 121–137.

Pennycook, Alistair 2016. Mobile times, mobile terms: The trans-super-poly-metro movement. In: Nicholas Coupland (ed.), *Sociolinguistics: Theoretical debates*. Cambridge: Cambridge University Press, 201–216.
doi:10.1017/CBO9781107449787.010

Pietikäinen, Sari 2000. *Discourses of differentiation: Ethnic representations in newspaper texts*. Jyväskylä University, Studies in Communication 12. Jyväskylä: University of Jyväskylä.

Pietikäinen, Sari 2003. Indigenous identity in print: representations of the Sami in news discourse. *Discourse & Society*, 14 (5): 581–609

Pietikäinen, Sari 2008. Sami in the media: Questions of language vitality and cultural hybridisation. *Journal of Multicultural Discourses,* 3 (1): 22–35.

Pietikäinen, Sari 2010. Sami language mobility: Scales and discourses of multilingualism in a polycentric environment. *International Journal of Sociology of Language*, 202: 79–101.

Pietikäinen, Sari 2018. *Investing in indigenous multilingualism in the Arctic*. Language and Communication 62.
https://doi.org/10.1016/j.langcom.2018.05.001. Date of access: December 4, 2020.

Poster, Mark 1995. The Internet as a public sphere? *Wired,* 3 (1). https://www.wired. com/1995/11/poster-if/. Date of access: March 28, 2020.

Schement, Jorge and Curtis, Terry 1997. *Tendencies and tensions of the information age: The production and distribution of information in the United States.* New Brunswick, NJ: Transaction.

Siegel, Florian 2013. The sociolinguistic status quo on the Taimyr Peninsula. *Études finno-ougriennes,* 45: 239–280.

Tan, A., Fujioka, Y. and Lucht, N. 1997. Native American stereotypes, TV portrayals, and personal contact. *Journalism and Mass Communication Quarterly,* 74(2): 265–284.

Varis, Piia 2015. Digital ethnography. In: Alexandra Georgakopoulou and Teresa Spilioti, eds. *The Routledge handbook of language and digital communication*. London: Routledge.
https://www.routledgehandbooks.com/doi/10.4324/9781315694344.ch3.
Date of access: March 28, 2017.

Várnai, Zsuzsa 2018. Kisebbségi nyelvek az urbanizáció folyamatában: A városi többnyelvűség összehasonlító vizsgálata sarkköri őshonos közösségekben – Dugyinka. [Minority languages in the process of urbanization: A comparative study of urban multilingualism in Arctic indigenous communities] In: Mária Pásztor Kicsi (ed.). Kétnyelvűség – hátrány vagy esély? Újvidéki Egyetem, Magyar Nyelv és Irodalom Tanszék, Újvidék. 99–108.

Várnai, Zsuzsa 2020. *„Identitásjáték”: Az etnikus identitás összetevőinek vizsgálata arktikus népeknél.* ["Identity Game ": An examination of the components of ethnic identity in Arctic peoples] Unpublished manuscript.

Várnai, Zsuzsa. "*Who am I?" Indigenous identity in arctic urban communities: Sociolinguistic interviews in Dudinka.* Unpublished manuscript.

Wodak, Ruth and Meyer, Michael 2009. *Methods of critical discourse analysis*. London: Sage.

Wilson, Pamela. 1998. Confronting "The Indian Problem". In: S. Torres (ed.), *Living colour: Race and television in the United States*. Durham, NC: Duke University Press, 35–61.

The official website of the City Folk Art Center of Dudinka: http://www.gorod-dudinka.ru/

The the Facebook page of the House of Folk Arts of Taimyr – Chum: https://www.facebook.com/kgbuktdnt/

Taimyr Telegraf (TT): http://ttelegraf.ru/

Reconsidering the Nganasan vowel system

László Fejes

Hungarian Research Centre for Linguistics, Budapest

The description of the Nganasan vowel system has not radically changed after Helimski (1998: 485–486), and his description is quite close to Tereščenko's (1979). The most considerable difference between the two is that Helimski introduced a new category "between" simple vowels and diphthongs: diphthongoid $^i a$ (instead of Tereščenko's simple vowel *æ*) and diphthongoid $^u a$ (in addition to the diphthong *ua*: Tereščenko did not distinguish these two). This distinction is based on (or at least parallel with) Tibor Mikola's fieldwork results, although Mikola suggested that in these cases the articulation gestures before *a* belong to the preceding consonant, that is, the consonants are palatalized or labialized (Mikola 1970: 60–61).

In this paper, I give an overview of the presentations of the Nganasan vowel system and attempt to ascertain which of the competing views are more adequate for the description of the language.[1] Moreover, I suggest some modifications which are are without precedent in the history of the research. In Section 1, I present the descriptions and define their differences. In Section 2, I examine whether the distinction of central vowels is phonologically necessary. In Section 3, I explore whether complex vowel constructions should be analyzed as diphthongs and long vowels or as vowel sequences. In Section 4, I reconsider the position of *a* in the vowel system. Finally, in Section 5, I argue that $^u a$ is probably not a phoneme in Nganasan.

1. Introduction: differences in the presentation and the description of the vowel system (1998–2018)

Although Helimski introduced the new category of diphthongoids into the vowel system of Nganasan, he placed these phonemes into the table presenting the simple vowels (1998: 483):[2]

	Front		**Non-front**	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	*i*	*ü*	*ɪ*[3]	*u*
Mid	*e*		*ə*	*o*
Low	*ⁱa*		*a*	*ᵘa*

Table 1. Nganasan vowels, Helimski (1998)

Diphthongoid is a rarely used term, and (e.g. for English) it usually denotes a long vowel pronounced in a diphthong-like way (where some phonetic quality of the beginning and the end of the vowel differs, but not to such a great extent as usually the quality of the simple vowels of the given language usually does). Although Helimski did not state it explicitly, it seems that his diphthongoids are short vowels which are realized as diphthongs (e.g. *ä* and *å* – therefore, he listed them among short vowels). Moreover, Helimski suggests about long vowels and his diphthongs that "it is reasonable to treat them as vocalic sequences" (Helimski 1998: 485). Let us notice that if "diphthongoids" are short vowels and "diphthongs" are vowel sequences, it means that Nganasan syllables have no branching nuclei.

Várnai (2000: 33) presents the simple vowels of Nganasan slightly differently:

	Front		**Central**	**Back**	
	Illabial	**Labial**	**Illabial**	**Illabial**	**Labial**
High	*i*	*ü*	*ɪ*		*u*
Mid	*e*		*ə*		*o*
Low	*ⁱa*		*a*		*ᵘa*

Table 2. Nganasan vowels, Várnai (2000)

---

[2] Katzschmann (2008: 318) provides the same system with one irrelevant difference: he changed non-front to back (*hintere*).

[3] Although different sources mark this vowel with different modified forms of *i* (I am aware of at least *ĭ*, *ɨ* and *i̭*), in this article, following Turkish orthography, I will consistently use the simplest possibility, *ɪ*.

As it is obvious, Várnai introduces a new tongue position into the Nganasan vowel system; however, she does not explain why it is necessary. In addition, she treats Helimski's diphthongs as vowel sequences and calls Helimski's diphthongoids diphthongs (despite this, she presents them among simple vowels). However, she also speaks about long vowels (33–34), which are not treated as vowel sequences, as Helimski suggests, although Várnai's arguments (see below) are applicable to them, too. Nonetheless, later, on page 36, Várnai states that the elements of long vowels, similarly to the elements of vowel sequences, belong to different syllables.

Wagner-Nagy (2018: 48) treats Nganasan simple vowels in a slightly different way again:

	Front		Central	Back	
	unrounded	rounded	unrounded	unrounded	rounded
High	*i*	*ü*	*ι*		*u*
Mid	*e*		*ə*		*o*
Low				*a*	

Table 3. Nganasan vowels, Wagner-Nagy (2018)

Wagner-Nagy, agreeing with Várnai, calls *ⁱa* and *ᵘa* diphthongs and all the other non-simple vowels vowel sequences. Moreover, she also considers Helimski's and Várnai's long vowels to be vowel sequences. Consistently with this analysis, her table of simple vowels does not contain *ⁱa* and *ᵘa*. Similarly to Várnai, Wagner-Nagy also distinguishes central and back vowels. But while Várnai has no back unrounded vowels, Wagner-Nagy places *a* into the back unrounded column, and *ι* and *ə* into the central unrounded column.

The main differences of the above analyses can be summarized as follows:

Difference	Helimski 1998	Várnai 2000	W-N 2018
existence of central vowel phonemes	✗	✓	✓
position of *a* (= *ι* and *ə*?)	✓[4]	✓	✗
ⁱa and *ᵘa* among simple vowels[5]	✓	✓	✗
ⁱa and *ᵘa* are called diphthongs	✗	✓	✓
"long vowels" are vowel sequences	✗	✗✓	✓

Table 4. Nganasan vowels, systems compared

As we can see, the Wager-Nagy's and Helimski's analyses are always different according to these parameters, while Várnai represents a transition between the two: her description agrees with both Helimski's and Wagner-Nagy's in two parameters, while her description is somewhat inconsistent in the question of "long vowels".

2. The problem of central vowels

This problem is relatively simple, since we find no case where the distinction of central vowels were necessary: there are no two otherwise similar vowels between which the only difference is that one is back but the other is central. Várnai's central vowels can be analyzed as unrounded back vowels which are centralized by a redundancy rule. Wagner-Nagy's opinion, namely, that *ι* and *ə* are centralized but *a* is not, can be reconciled with a redundancy rule, according to which non-low unrounded back vowels are centralized.

For the pure satisfaction of being exact, I checked whether Várnai or Wagner-Nagy refers to the centrality of vowels. I found only one reference: "The stress,

---

[4] The reviewer argues that Helimski (1998) is incomparable from this point of view to the other two grammars since he uses different categories. Nonetheless, the difference between the categorization of unrounded vowels other than *i* and *e* is in labeling them similarly, contrary to Wagner-Nagy, who treats them differently.

[5] The reviewer suggests that I misinterpret Várnai's description. According to them, Várnai includes the diphthongs in her table of vowels just because they behave as simple vowels do. However, as I see it, in phonology, the best reason to categorize two surfacially different entities into the same category is that they behave in the same way. Despite that, I can imagine that there are no crucial differences in the analysis of the vowel system behind these differences of presentation.

however, never moves from a high vowel to a central vowel or from a central vowel to a high one" (Wagner-Nagy 2018: 73). Since here central vowels are opposed to high vowels, we must suggest that *central* is a misprint instead of *mid*.[6] Even if it was not the case, *central* could be replaced by *non-low unrounded back* without any consequences.

As a conclusion, I argue that it is not reasonable to distinguish central vowels in the Nganasan vowel system.

3. The problem of diphthongs

As has been shown before, although Helimski (1998: 485) calls complex vowel structures (except for ⁱ*a* and ᵘ*a*) diphthongs, he also mentions that "it is reasonable to treat [...] them as vocalic sequences". He also sketches some arguments, and many other arguments were added by Várnai (2000, 2005). In what follows, these arguments are discussed in detail.

3.1. Length and morpheme boundary

One argument for the vowel sequence analysis is that Nganasan long vowels and diphthongs are twice as long as a single vowel (Helimski 1998: 485). This must sound strange for a Uralist, since Finnish long vowels and diphthongs are also twice as long or even longer than short vowels (Hakulinen 1941: 22[7]); however, they are not analyzed as vowel sequences in grammars.[8]

---

[6] Beáta Wagner-Nagy confirmed my supposition (personal communication). Nonetheless, the reviewer mentioned that in Wagner-Nagy's two examples, *čühənu* 'during' and *tənini* 'there', both contain *ə*, which is both mid and central. The important point here is that the term is opposed to high (and not front or back). They also lack a citation to Vaysman (2009). There are two main reasons for that. First of all, here I am not interested in stress: I refer to the given locus by Wagner-Nagy only because she mentions central vowels there. The other important reason is that Vaysman's approach is so different from Wagner-Nagy's that it is difficult to compare the two analyses (not to mention the differences of the linguistic data). Although Vaysman (2009: 25–40) writes about the peculiarities of *ɪ* and *ə*, these can be referred to as non-low unrounded back vowels even if *a* is analyzed as a low unrounded back vowel.

[7] "Lyhyen ja pitkän vokaalin kestoero suomessa on yhtä jyrkkä kuin yksinäis- ja geminaattakonsonantin: suhde on tavallisesti noin 1 : 2 tai 1 : 2½, mutta saattaa usein olla jopa 1 : 3." – "The difference between the length of short and long vowels in Finnish is as radical as between simple and geminate consonants: their proportion is usually 1 : 2 or 2½, but it can be frequently even 1 : 3."

[8] The reviewer interprets my arguments here and below in a way that I compare Nganasan to Finnish, and I suggest that the former should follow the latter. Nonetheless, my approach is completely different. Theoretically, I can imagine that even two different analyses work

Another argument by Helimski (1998: 485, without examples) is that these vowel sequences are often divided by morpheme boundaries;[9] nonetheless, this is also the case in Finnish, including nominal cases (*kala* 'fish' : *kala-a* 'fish-PART' : *kala-an* 'fish-ILL' – although see also *talo-a* 'house-PART', in which *oa* is considered to be a vowel sequence), possessive marking (*talo-ssa-an* 'house-INE-3'), plural marking (*talo-i-ssa* 'house-PL-INE'), and past tense in verbal inflection (*sano-i* 'say-PST.3SG)' etc.

These arguments are not repeated later by other researchers.

3.2. Number of possible combinations

Várnai (2005: 116) argues that there would be too many diphthongs (this time, including long vowels) if we did not consider them as vowel sequences. She refers to typological works (Maddieson 1984; Ladefoged and Maddieson 1996) according to which the typical number of diphthongs is under 10: she finds only one language with 22 diphthongs, but she considers the case an error. According to her, there are over 20 vowel sequences in Nganasan (Várnai 2005: 115–116; Várnai 2000: 34; Wagner-Nagy 2018: 51):

---

satisfyingly for the same language. However, there are arguments for one kind of analysis (and against another) which refer to some facts of a language (namely, Nganasan) which are also true of another (i.e. Finnish). The only thing I state is that if these were conclusive arguments in the case of Nganasan, they should be the same for Finnish. Even so, not accepting these arguments does not mean that I cannot imagine the analysis of these constructions as vowel sequences, it means only that I cannot exclude the analysis of them as long vowles and diphthongs diphthongs. Additionally, the reviewer states that the length proportion of short and long vowels (diphthongs) is not known – anyway, I just refer to the statement by Helimski.

[9] It remains unclear whether Helimski means that these vowel sequences are usually divided by morpheme boundaries in texts (no statistics included) or that most of them occur only at morpheme boundaries. It seems that the latter is not true, just two of them (*ui* and *ⁱai* – Várnai 2000: 34; 2005: 115–116; Wagner-Nagy 2018: 50–51) are not attested morpheme internally (in stems).

V₁/V₂	i	$ü$	$ɩ$	u	e	$ə$	o	a	$^i a$	$^u a$
i	+	×	×	×	*	+	*	+	−	÷
$ü$	×	+	×	×	*	+	*	V	−	÷
$ɩ$	×	×	+	×	*	+	*	+	−	÷
u	(+)	×	×	+	*	+	*	+	−	÷
e	+	+	−	−	*	¤	*	−	−	÷
$ə$	+	WN	−	+	*	+	*	¤	−	÷
o	+	−	−	+	*	¤	*	+	−	÷
a	+	+	−	+	*	¤	*	+	−	÷
$^i a$	(+)	WN	−	−	*	¤	*	WN	−	÷
$^u a$	+	−	−	−	*	¤	*	+	−	÷

Table 5. Nganasan vowel sequences

Combinations against a darker background are attested. Combinations marked by + occur word internally as well, while sound combinations marked by (+) occur only across morpheme boundaries. The combination $^u a$ marked by V occurs only according to Várnai (2000: 34; 2005: 115), the combinations $əü$, $^i aü$ and $^i aa$ marked by WN only appear in Wagner-Nagy (2018: 51). The lack of combinations marked by *, ÷, ¤ and × can be explained (see below). Combinations marked by – are absent without evident reason.

Vowels e and o do not occur in non-initial syllables (*); therefore, it is self-evident that they cannot occur as a second vowel of a vowel sequence either. The fact that they do not occur there could be an argument for the stance that these are vowel sequences – however, it is not clear whether vowels restricted to initial syllables can occur as second elements of diphthongs cross-linguistically.

The lack of $^u a$ as a second element (÷) is not very surprising, since this element is very rare (however, another explanation will be suggested in section 5). This can also be a reason for it to appear only in some combinations. The reason for the lack of $^i a$

as a second element and its restricted occurrence as a first element is not so self-evident.

The most frequent vowel, ə, occurs only in combination with high vowels and itself (other possible cases: ¤). The explanation can be that ə is fully assimilated in these combinations. Interestingly, Várnai (2000: 43) mentions this, but all of her examples involve high vowels. The explanation can be that the assimilation with high vowels is optional, but it is obligatory in other cases.

Different high vowels are not combined (×). This is probably a result of a strong harmonic tendency, according to which high vowels assimilate to each other in both roundedness and frontness (cf. Fejes 2019: 110–114).

If we exclude *e*, *o* and *ᵘa* as second elements, 70 possible combinations remain, of which 31 are attested, that means 44%. For the sake of comparison, in Finnish, according to ISKo §21, the following combinations are possible:

V₁/V₂	*i*	*y*	*u*	*e*	*ö*	*o*	*ä*	*a*
i	+	+	+	+	–	–	–	–
y	+	+	×	–	+	×	–	×
u	+	×	+	–	×	+	×	–
e	+	+	+	+	*	*	–	–
ö	+	+	×	*	+	×	–	×
o	+	×	+	*	×	+	×	–
ä	+	+	×	÷	÷	×	+	×
a	+	×	+	–	×	–	×	+

Table 6. Finnish vowel combinations

In Finnish, we have 64 possible combinations, 26 of which are attested in diphthongs: that is about 41%, not crucially different from what we find in Nganasan. However, the Finnish system is much more systematic: the reason for lacking a combination is vowel harmony (×), a restriction that closing diphthongs cannot end in a mid vowel (÷), non-high (mid) vowels do not combine with each other (*), and

opening diphthongs must end in a mid vowel of identical roundedness and frontness (otherwise –). (Of course, in some cases these restrictions overlap.) Similarly to Nganasan, the combination of identical vowels is always allowed.

In Finnish, some vowel combinations which do not occur in diphthongs can occur in vowel sequences, that is, when speakers consider them to belong to different syllables. Sometimes suffixation is also an argument to treat these as diphthongs or vowel sequences (see below; sometimes there can be also vacillation, cf. ISKo §22). In Nganasan, it seems that nobody has suggested treating these combinations differently or mentioned vacillation.

The reason for the absence of certain combinations in Nganasan remains unclear. While high vowel *i* can be a second element after any non-high vowel, *ɩ* does not occur after any of them. The most problematic issue is that diphthong(oid)s cannot follow any vowel. It can be easily explained if these combinations are treated as diphthongs: triphthongs are prohibited. However, this fact is also accompanied by another phenomenon, i.e. these entities do not occur word initially (cf. Wagner-Nagy 2018: 49–50): diphthong(oid)s can occur only in syllables with a filled onset.

3.3. Radical consonant gradation

In radical (or syllabic) consonant gradation some consonants or consonant clusters alternate due to the open or closed nature of the syllable. When the syllable is open, the consonant (cluster) before the vowel is in the so-called strong grade; when it is closed, it is in the so-called weak grade. E. g. *basa* 'iron, money' : *baďaʔ* 'iron, money-PL' (Várnai 2000: 63–64). According to Várnai (2005: 117), we always find the strong grade before these complex vowel structures. This means that in these cases the expectedly alternating consonant (cluster) is always followed by an open syllable (i.e. by a vowel sequence): *kasuə* 'cortex' : *kasuəʔ* (*kaďuəʔ*) 'cortex-PL'.

However, we find the same phenomenon in Finnish before long vowels and diphthongs: *vapaa* 'free' : *vapaa-n* (*vavaa-n*) 'free-GEN'; *tupakoi-da* 'smoke-INF' : *tupakoi-n* (*tupaoi-n*) 'smoke-1SG'; *laatikko* 'box' : *laatikko-on* (*laatiko-on*) 'box-ILL' etc. However, there are also cases when weak grade is used before a diphthong: *anto-i(-Ø)* 'give-PST(-3SG)' : *anno-i-n* 'give-PST-1SG'. This kind of behavior can be explained by the fact that some of the long vowels and diphthongs of Finnish have developed from simple vowels of different syllables. Radical consonant gradation is not an active phonological process anymore, it is rather morphologized, sometimes reflecting the phonology of an earlier stage.

This is also true of Nganasan. For instance, Várnai (2000: 64) mentions the phenomenon of reversed consonant gradation. We find weak grade in a word form with an open syllable, but a strong grade in a word form with a closed syllable: *koðu*

'storm' : *kotuʔ* 'storm-PL'. Therefore, it is strongly questionable whether radical consonant gradation can be an argument in this debate.[10]

3.4. Rhythmical consonant gradation

Rhythmical consonant gradation is typical of suffixes: consonants are in the strong grade in even syllables (i.e. after stems with an odd number of syllables) and in the weak grade in odd syllables (i.e. after stems with an even number of syllables): *nı-tı* 'wife-3SG', *mənu-ðu* 'egg-3SG', *bakunu-tu* 'sturgeon-3SG', *kaanaʔa-ðu* 'creek-3SG' etc. (Várnai 2000: 62). As the last example shows, *aa* has to be counted as two syllables to get four syllables in the stem.

Earlier Hajdú (1962, 1964) suggested that this kind of phenomenon can be explained by mora counting: short vowels count as one mora, while long vowels and diphthongs count as two moras. Várnai (2005: 125) suggests that instead of counting moras it is enough to count syllables, if we speak about vowel sequences instead of long vowels and diphthongs. However, it seems that neither mora counting nor syllable counting is a superior solution to the other.[11]

Moreover, there are some cases which show that the number of syllables is not the only factor which can affect rhythmical consonant gradation. For us, the most important fact is that vowel sequences are always followed by the weak grade, even if the stem has an odd number of syllables: *latəə-ðu* 'bone-3SG', *kümaa-ðu* 'knife-3SG', *kuhuə-ðu* 'skin-3SG' etc. (Várnai 2000: 63). This phenomenon cannot be explained by interpreting these complex vowel structures either (i) as long vowels and

---

[10] The reviewer does not agree with my conclusions and argues that the system of gradation did not collapse in Nganasan as it did in Finnish. As an argument, they refer to the fact that some speakers apply the rules of (radical) consonant gradation even in Russian loanwords. However, this is also true of Finnish: some loanwords obey the rules of consonant gradation (*netti* 'net-SG.NOM' : *neti-n* 'net-SG.GEN') and even introduce new types of gradation (of consonants occurring only in loanwords, cf. *bloga-ta* 'blog-INF' : *bloggaa-n* 'blog-1SG'). This phenomenon does not inevitably confirm that gradation is active phonologically. Instead, it can be analyzed as the loanword's adaptation to the existing morpho(phono)logical patterns of the language.

[11] After my presentation Tapani Salminen referred to the Occam's razor principle, saying that syllable counting is better than mora counting, because we do not have to adduce moras: the simpler analysis is always better. Then I had to agree with this argument, but now I think Occam's razor can also cut in the opposite direction. Since phonetically we have long vowels and diphthongs, it is an obvious and straightforward solution to treat them as long vowels and diphthongs in phonology as well, and to refer to their length (moras) when it is needed. Supposing a different phonological structure when it is not really needed is superfluous – or, at least, neither of the two solutions can be considered more complicated than the other one.

diphthongs or (ii) as vowel sequences. Another fact is that we can always find the strong grade after consonant-final stems, including those which are vowel-final on the surface but a stem final consonant-like element (called quasi-nasal by Várnai 2000: 63, quasi-consonant by Várnai 2010, and empty slot by Wagner-Nagy 2018: 67, etc.) can be found: *koðu-tu* 'storm-PL'. Anyway, this kind of element has to be assumed because rhythmical consonant gradation is not a purely phonological phenomenon in contemporary Nganasan, and we have to introduce historical elements to account for some morphologized phenomena as a part of contemporary phonology.

Although Finnish used to have rhythmical consonant gradation as well, it has just traces of it now. Thus, in contemporary Finnish there is no phenomenon comparable to Nganasan rhythmical consonant gradation.

3.5. Transparency in vowel harmony

In Nganasan, high vowels following each other strongly tend to agree in frontness/backness (front/back vowel harmony). However, in the following cases, high vowels do not agree with each other: *hüə-ðu* (**hüə-ðü*) 'year-3SG', *küə-ðu* (**küə-ðü*) 'pine-3SG', *biə-ðı* (**biə-ðı*) 'year-3SG' (Várnai 2005: 117–118). Várnai cites these examples as arguments in support of treating complex vowel structures like *üə* or *iə* as vowel sequences. According to her, if *üə* and *iə* were diphthongs, vowel harmony should work here. The preconception in this argumentation is that even if a vowel (in this case *ə*) is opaque as the nucleus of a syllable, it must be transparent as a second element of a diphthong.[12]

However, this preconception is not supported by any fact. Moreover, we can cite examples from Finnish which contradict this assumption. Although Finnish has no opaque neutral vowels in vowel harmony, there is another kind of phonological phenomenon which indicates something else. One group of the allomorphs of the illative case is *-hVn*, in which V is identical to the vowel before *h*: *tä-hän* 'this-ILL', dialectal/archaic *koulu-hun* 'school-ILL'. In Modern Standard Finnish, this ending usually occurs after long vowels and diphthongs, and in the latter case V is identical to the second element of the diphthong: *työ-hön* (**työ-hyn*) 'work-ILL', *töi-hin* (**töi-hön*, **töi-hyn*) 'work.PL-ILL'. As these cases show, the height of *y* in the first example

---

[12] The reviewer wonders why this preconception is necessary in this argumentation. Actually, it seems to be clear for me that if someone uses a phenomenon as an argument in favor of analyzing an element as something, they inevitably must think that the given phenomenon could not work in the same way if the element was analyzed differently. Additionally, Várnai (2005: 117) herself writes: "if the two adjacent vowels belonged to the same syllable (if they constituted a diphthong), vowel harmony would be expected to apply". As I demonstrate below, this is not the case.

or the labiality (and again, height) of *ö* in the second example do not affect the vowel of the suffix. The behavior of the second element of a diphthong and a vowel in a syllable nucleus is not different when locality plays a role. This can be the case in Nganasan as well; therefore, the above mentioned forms cannot be used as arguments in the debate about long vowels and diphthongs vs. vowel sequences.

3.6. Syllable structure

In many languages, the complexity of syllable structure is restricted. This can mean that when the nucleus is branching, the code cannot be filled or cannot be branching. This provides a promising method[13] to decide which structures are diphthongs (long vowels) in Nganasan:

If the distribution of short vowels is similar to that of *ⁱa* and *ᵘa* but differs from that of others, we have to speak about "diphthongoids" and diphthongs.

If the distribution of short vowels differs from that of *ⁱa* and *ᵘa* but is similar to that of others, then we have to speak about diphthongs and vowel sequences.

A statistical evaluation of the material in Brykina et al. (2016)[14] gave the following results:

---

[13] To my knowledge, nobody has ever tried to use this method to inspect this problematic field of Nganasan phonology.

[14] The reviewer criticizes me for using the corpus with such methods. They argue that the material of the corpus (and any earlier materials on Nganasan!) cannot be used for phonetic or phonological investigations, because the transcription is neither phonetic or phonological. This kind of argumentation raises several issues. First of all, any current or earlier study of Nganasan phonology can be criticized similarly. Additionally, phonological analysis must precede a phonological (phonemic) transcription: there is no phonemic transcription until the phoneme inventory is specified (including allomorphy). Analysis and transcription are always in interaction: they overwrite each other. According to the reviewer, I should have checked the sound material. However, not being an expert of Nganasan, I do not consider myself competent to correct the transcription of the fieldworkers, and it would hardly be convincing if I did that. Moreover, I could be criticized for not doing phonetic measurements – and if I did that, it would be another paper. To sum up: I am aware of the limits of my investigation and I do not use the corpus because I believe that it is completely reliable, but because that is the most comprehensive and best researchable source available.

nucleus/coda	empty (∅)	simple (C)	branching (CC)	∅ : C
V	18,335	5,906	24	3.1
^{i}a, ^{u}a	715	368	0	1.95
VV	3,237	1,113	0	2.9
V: ^{i}a, ^{u}a	25.6	16	–	
V : VV	5.7	5.3	–	

Table 7. The distribution of vowels (diphthongs, vowel sequencies etc.) in different syllable structures

First of all, we have to conclude that branching codas are practically unknown in Nganasan (cf. Várnai 2000: 39, Wagner-Nagy 2018: 66). All the words with branching codas attested in the corpus are foreign words (most frequently *maskvaa ~ məskvaa* 'Moscow') or typos. (Nonetheless, there is neither branching onset nor branching coda at the word periphery in native Nganasan words. Therefore, it is impossible to decide whether it is the onset or the coda branching inside this kind of loanwords.)

It is salient that the distribution of V and VV is very similar: in both cases, every fourth example is followed by a consonant. (Or, every fifth or sixth coda is preceded by VV, independently of whether they are filled or not.) This suggests that VVs are vowel sequences, since their distribution is similar to that of simple vowels. In this case, we have to suppose that the second element is in a different syllable, and this is the reason why it is followed by a consonant in the same proportion as any other vowel. In this case, we expect that *^{i}a* and *^{u}a* are diphthongs, and therefore they will be followed by a coda more rarely than simple vowels. However, the case is the opposite: these sounds are followed by a coda in every second case. Therefore, these statistics do not prove anything.

We have to conclude that the complexity of the nucleus does not restrict the complexity of coda in Nganasan. Therefore, this method cannot be used to decide whether Nganasan has a rich system of long vowels and diphthongs or vowel sequences.

3.7. Conclusion

To sum up, despite the fact that current literature on Nganasan phonology states that there are vowel sequences in Nganasan, there is no conclusive argument which would decide whether these are really vowel sequences or diphthongs and long vowels. It is also questionable whether this is a crucial problem of Nganasan phonology. It plays the most important role in the description of rhythmical consonant gradation, and even there both can work to the same extent. However, rhythmical consonant gradation is not an active phonological process but rather a morhologized regularity, which more or less reflects the phonology of an earlier stage in the history of the language.

The currently accepted analysis with vowel sequences causes two different kinds of problems. On the one hand, according to the analysis with vowel sequences, Nganasan has a lot of hiatus, although vowel initial words are rare (Wagner-Nagy 2018: 49). Cross-linguistically, even languages in which vowel initial words are common tend to avoid hiatus and use different kinds of strategies to resolve them (mainly throught the deletion of one of the vowels or consonant epenthesis). On the other hand, if we analyze Nganasan complex vowel structures as vowel sequences, it would be consistent to do so with the similar Finnish ones. However, the necessity of this kind of analysis has never emerged in Finnish phonology. Anyway, I cannot see such crucial differences between the phonetic and phonological phenomena of Finnish and Nganasan that they should have to be analyzed in such a radically different way.

4. The problem of *a*

Nganasan has two kinds of vowel harmony. One is the already mentioned front/back harmony, which is quite regular but restricted to high vowels. In suffixation it is active only when the other kind of harmony is active as well. This other kind of harmony can be called quasi-labial, and it works throughout Nganasan morphophonology, however, with many exceptions and irregularities (see Fejes 2018a; Fejes 2018b). It can be called quasi-labial because one of the harmonic classes contains rounded vowels and *a* (bolded in Table 8), the other contains unrounded vowels (except for *ə*, which does not belong to any of the classes; underlined in Table 8):

	Front		Back	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	*i̱*	*ü*	*ı̱*	*u*
Mid	*e̱*		*ə̱*	*o*
Low			*a*	
Diphthongs	*ⁱa̱*			*ᵘa*

Table 8. Harmonic classes in Nganasan

This suggests that, although *a* is phonetically unrounded, phonologically it belongs to the rounded vowels. There is no known phonological phenomenon in Nganasan which would support that it is a phonologically unrounded vowel. Since in the case of low vowels roundedness does not have such a strong distinctive power as in the case of higher vowels, a redundancy rule causing a low vowel to become unrounded is not very strange. Therefore, the vowel system should look like this:

	Front		Back	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	*i̱*	*ü*	*ɩ̱*	*u*
Mid	*e̱*		*ə*	*o*
Low				*a*
Diphthongs	*ⁱa̱*			*ᵘa*

Table 9. Harmonic classes in Nganasan with a analyzed as rounded

However, if *ⁱa* and *ᵘa* are analyzed as simple vowels realized in a diphthong-like way, we get a system with two low rounded back vowels:

	Front		Back	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	*i̱*	*ü*	*ɩ̱*	*u*
Mid	*e̱*		*ə*	*o*
Low	*ⁱa̱*			*a, ᵘa*

Table 10. Harmonic classes in Nganasan with a analyzed as rounded and diphthons analyzed as monophthongs underlyingly

Does this mean that *ⁱa* and *ᵘa* must inevitably be analyzed as diphthongs? To answer this question, it is worth taking a closer look at the distribution of *ᵘa*.

5. The problem of *ᵘa*

To decide whether *ᵘa* must be considered to have phonemic status in Nganasan, we have to take a closer look at its distribution. Since it occurs most of the time after *h* where it also alternates with *a*, which almost never occurs after *h*, in Section 5.1 I concentrate on the distribution of *a* and *ᵘa* after *h*. In Section 5.2, I overview the cases when *ᵘa* occurs after consonants other than *h*. In Section 5.3, I sum up the facts on the distribution of *ᵘa* and argue that it is not a phoneme in Nganasan.

5.1. The distribution of *a* and *ᵘa* after *h*

The *ᵘa* element seems to be in an almost perfect complementary distribution with *a* after *h*. Wagner-Nagy (2018: 85) writes: "[t]he consonant and vowel combination /ha/ exists only in one word, *tahar͡ia* 'now' which is an old loanword from Russian (*taperja* 'now'). In other words, /u͡a/ appears in this position, and the sound combination is pronounced as [hʷa] [...]". The loanword must be considered old, since Russian *p* is reflected by Nganasan *h* due to the *p* (> *f*) > *h* change. (The *f* state is attested in Castrén's materials, cf. e.g. Helimski 1998: 484–485.) The vowel *e*, following the original *p*, is atypical in non-initial syllables in Nganasan. We could expect that it is substituted by the only mid vowel possible in a non-initial syllable, *ə*, but the *hə* sequence seems to be prohibited as well: "/h/: There are no data for a /hə/ sequence. /ha/ only appears in one word, *tahar͡ia*, which is an old loanword from Russian. The sequence /he/ is attested in the word-initial position" (Wagner-Nagy 2018: 59). Nonetheless, the statement about the absence of *hə* seems to be a mistake, since both Brykina et al. (2016) and even Wagner-Nagy (2018) give several words containing this combination, including *čühə* 'date, time, period', *səŋhəlʲaŋkə* '5', *nəŋhə* 'wrong, bad', *həkəgəə* 'warm' etc. However, although *taharʲa* occurs several times and in different forms in Brykina et al. (2016), it has no form with a vowel other than *a* after *h*. (The most frequent form of it is *tahar͡iaa*, which is the second most frequent word form of the whole corpus: only *tə* 'well' is more frequent (by ~30%), and it is more frequent (by ~60%), than the third most frequent *təti* 'that'.)

Another remarkable fact is that Helimski (1998: 485) mentions that even his elderly female informants pronounced labial fricatives ([f] or [ɸ]) or labiovelar [hw] instead of [h]. This suggests that originally *h* was labialized, and delabialization happened everywhere but before *a*. In this case, we must suppose that either the emergence of the second *a* in *taharʲa* is relatively late and earlier another vowel stood after *h*, or we have a case of irregular delabialization. In any case, it remains unclear why such a sequence, atypical for Nganasan, has developed.

In Brykina et al. (2016), we find some other cases of *ha* combinations. However, these are all in words of Russian origin, predominantly proper names (and even the possibility of code switching cannot be excluded). In all these cases, the source of *h* in the original Russian word is [x].

Despite these facts, we can consider *ᵘa* an allophone of *a* after *h*, mainly because of its alternations. In consonant gradation, *b* alternates with *h*, and when *b* is followed by *a* in the weak grade allomorph, *h* is always followed by *ᵘa* in the strong grade (see the allomorphs of the so-called renarrative mood suffix in Helimski 1998: 489). However, it must be pointed out that sometimes we find *a* after *h* in some suffixes in Brykina et al. (2016):

     bəð͡ua-hatu grow(intr)|вырасти-INFER.[3SG.S]

     basa-han-sa hunt|охотиться-INTENT-INF

     d'üəra-lu-tə-baha-ndəʔ paint|краска.[VBLZ]-PASS-IPFV-NAR-3PL.R

However, we can find identical or very similar word forms in which ᵘ*a* is an allophone of *a* after *h*:

     tuj-hu͡atu come|прийти-INFER.[3SG.S]

     basa-hu͡an-sa hunt|охотиться-INTENT-INF

     d'üəra-lu-tə-bahu͡a-ndəʔ paint|краска.[VBLZ]-PASS-IPFV-NAR-3PL.R

Therefore we must suppose that these forms with ha are typos, slips of the language consultant's tongue, or alternative but rare realizations of /ha/.

5.2. The occurrence of ᵘ*a* after consonants other than *h*

In some cases, ᵘ*a* is also attested after consonants other than *h*. However, the number of such consonants is relatively small. In Table 11 below, I show some data on the distribution of some vowels (including diphthong(oid)s) after different consonants from Brykina et al. (2016). Items are ordered according to frequency. The first column contains the vowel, the second shows the number of occurrences, the third displays after how many different consonants it is attested, while the fourth and fifth show which the most frequent consonant preceding the vowel is, and how many times it occurs before the vowel. The last column shows the proportion of the cases in which the most frequent consonant (column 4) precedes the vowel out of the cases when the given vowel occurs.

Vowel	∑	Cs	most freq.	times	%
ə	9,454	23	*t*	1,943	21%
u	5,257	20	*t*	1,263	24%
a	4,760	32	*m*	613	13%
i	4,667	25	*n*	467	10%
ı	3,390	17	*t*	825	24%
ü	3,294	20	*t'*	423	13%
o	1,338	20	*k*	466	35%
e	899	19	*d', ń*	199	22-22%
^ia	788	19	*ʔ*	199	25%
^ua	389	8	*h*	265	68%
^ıa	63	2	*ʔ*	61	98%
^üa	34	8	*ð*	12	35%
$^uə$	8	2	*h*	5	63%
$^üə$	3	1	*ʔ*	3	100%
$^iə$	2	1	*ʔ*	2	100%
a^i	1	1	*ʔ*	1	100%
$ə^i$	1	1	*ʔ*	1	100%

Table 11. The number and distribution of vowels (including diphthongs) in the Nganasan corpus

The number of consonants in the third column can be over 21, the number of consonant phonemes in Nganasan (e.g. Wagner-Nagy 2018: 34), because the corpus also contains foreign words not adapted to Nganasan phonology. The vowels under the row of ^ua are considered to be allophones of *a*, ^ia or ^ua. Wagner-Nagy (2018: 48) states that ^üa occurs after palatal consonants. This, however, is certainly not justified by the data in the corpus: ^üa occurs twice after *d'* and once after *t'*, but after non-palatal consonants in all other 31 cases. Wagner-Nagy (2018: 49) also states that ^ua can be palatalized ($^üæ$, IPA [y͡æ]) in a syllable with a syllable following a palatal vowel. If we accept that ^üa in the corpus indicates these cases, this seems to be true for all of the cases, except for the stem *bən^{ü}a-* 'stretch oneself'.

In stems, ^ua occurs after stops and sonorants. However, in most cases, we can find the same stems with *u*, or more rarely, *a* or *ua* instead of ^ua.

d:

kund^{u}a- 'sleep, fall asleep' occurs 27 times, but we also find *kundu-kə-ta-ðə* '~-ITER-PRS-3SG.R' beside *kund^{u}a-kə-tu* '~-ITER-PRS.[3SG.S]';

ð:

bəðᵘa- 'grow (intr–tr)' 5 times, but also *bətu-baðu-ŋ* 'grow(tr)-INFER-2SG.S' twice;

t:

ďomtᵘa- 'fight' 10 times, but also *ďomta-*, *ďomtu-* (7 and 10 times, respectively),

kobtᵘa- 'girl' (over 150 times), also *kobtu-ʔᵘa*, 'girl-AUG' (7 times),

ťetᵘa 'very' (34 times), also *ťetua* 'very' (24 times);

k:

təjkᵘa- 'be in difficulty' (once) and *təjkᵘatu-* 'hindrance' (once), but also *təjku-* 'be in difficulty' (once)

kᵘańďa- 'fresh' (once), but also *kuańďa-* 'fresh' (3 times);

l:

bəlᵘaťa- ~ *bəluaťu-* 'become angry' (once both), *bəlᵘaðatəgibtü-* 'get angry' (twice);

n:

tanᵘa- 'few' (once), also *tanə-* (4 times);

r :

horᵘa- 'take up, deal with' eight times, also *hora-* for times (once even *horu-*, but without meaning): *horua-tuə* 'take up-PTCP.PRS[NOM.SG]' but *horu-tuə-gümü-ʔ* '~-PTCP.PRS-EMPH-NOM.PL', cf. also *horuabutə-* 'sort out' but *horuəbtuʔtə-* 'have floors',

mərᵘa- 'break' (tr: ломать) once: once also *məru-* 'break' (intr: сломаться).

It seems that there is only a very restricted set of consonants after which we find ᵘ*a*; even in those forms, ᵘ*a* occurs just in a very few (one to three) stems after the given consonant. Even in these stems we can observe more or less vacillation between ᵘ*a* and some other vowels or vowel sequences.

I have to mention here that the reviewer strongly doubts that some of the data above really prove vacillation. According to them, *kundᵘa-* ('sleep'; Kosterkina et al. 2001, 72: *кундуаса* 'спать') and *kuntu-* (: *kundu-*) ('fall asleep'; Kosterkina et al. 2001, 72: *кундудя* 'заснуть') are two different verbs. Nonetheless, while *kundu-* and *kuntu-* (and also *kunda-*) are consistently tagged as 'fall.asleep' in the corpus, *kundᵘa-* is tagged 16 times as 'sleep', and 10 times as 'fall.asleep'. This suggests that vacillation is attested at least in the meaning 'fall asleep'. Wagner-Nagy (2018: 427, 510) also cites examples of the *kundᵘa-* stem with the glossing 'fall asleep', although I see also the glossing 'sleep' as acceptable in the given sentences.

The reviewer also states that the corpus evidently handles *bəðᵘa-* 'grow (intr–tr)' and *bətu-* 'grow (tr)' as two different verbs. However, the latter occurs twice, and the former twice as intransitive and three times as transitive. Therefore, it does not seem to be evident that these are two different verbs. According to Kosterkina et al. (2001),

bəðᵘa- (29: *бэзӱаса*) is intransitive and *bətu-* (34: *бэтудя*) is transitive. I also have to mention that Wagner-Nagy in some cases (2018: 96, 515 three times) writes *bəðua-* speaking about the action noun *bəðuamu* 'growing' (which I could not attest in the corpus).

According to the reviewer, the forms *ďomtᵘa-*, *ďomtu-* and *ďomta-* belong to two different verbs again. Nonetheless, this does not seem to be evident. The form *ďomtu-* is glossed as 'scold' in five cases and as 'fight' in two cases, while the forms *ďomta-* and *ďomtᵘa-* as 'fight' in all the cases. Kosterkina et al. (2001: 45) contains only *ďomtᵘa-* (*дэмтуаса*) and gives both meanings.

The reviewer also doubts, based on the sound records, that the augmentative suffix is present in the forms written as *kobtu-ʔᵘa*. I leave this question open: I hesitate to believe that a non-existent suffix is written seven times. They also remark that all the seven forms occur in one and the same text, so it can be a personal variant. However, I believe that even individual language use can reflect the general characteristics of language, cf. how a feature of phonology can be attested even in a form with specific morphology. They also add that other augmentative forms of the word are more frequent: it is true that *kobtᵘa-rbaʔa* occurs 16 times unsuffixed and 4 times with other suffixes, and *kobtᵘa-naʔa* also exists, but I cannot see its relevance to the case (and to the existence) of *kobtu-ʔᵘa*.

The reviewer states that the forms *ťetᵘa* 'very' and *ťetua* are just traditional forms of recording, and most likely just one of them is correct, although it is difficult to decide which one. However, according to the descriptions, these two should be distinguishable based on the phonetic facts. Mikola (1970: 60) states that *ᵘa* sounds (and, according to him, should be analyzed) as a sequence of a labialized consonant and a vowel *a*. Várnai (2000: 40–41) and Wagner-Nagy (2008: 49) also claim that the first element of the allomorphs of *ᵘa* is very short and even likely to remain unpronounced, or its liability can surface on the preceding consonant or on the second half of the diphthong (differently in different environments). In addition, Helimski (1998: 485) states that vowel sequences are twice as long as simple vowels. Várnai (2005: 119) writes that "in casual pronunciation, diphthongs alternate with short vowels, whereas vowel clusters alternate with long vowels" (on the next page arguing that there are no long vowels in Nganasan, just clusters of two identical vowels, so this must be understood as different vowels immediately following each other being able to assimilate to each other). If these statements are all correct, *ᵘa* and *ua* should be phonetically well distinguishable. If they are not, then we must suppose that the reason for the inconsistency is that 'very' is a word that cannot be suffixed. Thus, it cannot be attested whether it behaves like a simple vowel or not. In that case, the question is whether there is a real difference in the pronunciation of the suffixable stems, or the decision between recording *ᵘa* or *ua* is simply based on morphophonological analysis.

According to the reviewer, one of *kᵘańd'a-* and *kuańd'a-* is simply an inconsistency in transcription. However, they do not say which of these should be wrong. The form *ku͡an'd'a-* occurs only once, but it is much more possible that the diphthong marker diacritic sign was left out by mistake. They also accuse me that I do not take into consideration the possibility of typos. I definitely do – however, when I found that in every instance when *ᵘa* follows a consonant other than *h*, there is also a similar form with *u*, *a* or *ua* instead of *ᵘa*, it does not seem to be the most probable possibility that these are all typos.

We must suppose that at least in a part of the cases, when the corpus contains a form with *ᵘa*, the informant pronounced a vowel without labialization. Wagner-Nagy (2018: 49) writes: "[t]he pronunciation of [u͡a][15] is rather varied, it may be realized as [ɐ], e.g. *kəbtuʔu͡a* [koptuʔɐ] 'went out [fire]' or even as [ʌ], e.g. *tanu͡amsa* [tanʌmsa] or [tanamsa] 'get used to'." The question is whether the diphthong is simplified in these cases, or (a) simple vowel(s) has/have a diphthong-like allophone. The latter seems to be less favorable if we take it into account that these supposed allophones occur just in a small set of stems, and there is no clearly identifiable environment in which they arise. The former seems to be unlikely since the supposed phoneme is extremely rare, its primary realization is more frequent as a secondary realization of another phoneme.

As for suffixes, there is only one form, *-ʔᵘa*, with the vowel *ᵘa*. The suffix appears in two positions after nominal stems and after postpositional stems, being an augmentative suffix (AUG in the corpus) after the former and both an adverbalizer (ADVZ.STEM) or a lative pronominalizer (LATPRON). The distribution of the different forms of these suffixes exhibits some difference. For the augmentative suffix, the allomorph *-ʔa* is the most usual form, but sometimes *a* is changed into a diphthong: in most cases, the vowel preceding the glottal stop seems to be copied after it, before *a*. The form *-ʔᵘa* occurs exclusively after stem final *u*, cf. *kobtu-ʔᵘa*, 'girl-AUG' above. However, this kind of assimilation does not seem to be obligatory, since there are at least two cases in the corpus with a: *d'üntuu-ʔa* 'middle-AUG' (twice) and *muŋku-ʔa* 'tree-AUG'.

In the case of the adverbalizer (19 times) and the lative pronominalizer (15 times), the suffix seems to be even more regular:

> *-ʔᵘa* after *u,*
> *-ʔⁱa* a after *i* and *ı,*
> *-ʔᵊa* a after *ı,*
> *-ʔa* after *a.*

---

[15] Supposedly a typo, instead of /u͡a/.

5.3. Conclusion

It has been demonstrated that there are only some restricted cases when *ᵘa* occurs in Nganasan. Most frequently, both in stems and suffixes, it is attested after *h*, where it is in an almost perfect complementary distribution with *a*. The emergence of *ᵘa* after *h* is historically clear, basically, the labiality of the consonant has been moved to the following vowel. The only problematic issue is that of *taharⁱa*, but it is not enough to reject the analysis of *ᵘa* after *h* as an allophone of *a*.

In suffixes, *ᵘa* also occurs after *ʔ*. In these cases it is clearly an allomorph of *a*, which has even an allomorph *ⁱa* in such position (cf. Wagner-Nagy 2018: 48).

In stems, *ᵘa* also occurs after some stops and sonorants, but the number of stems in which it occurs is very low, and even then we can find vacillation between *ᵘa* and other vowels or the vowel sequence *ua*. These cases are far from unproblematic, since we find no explanation for this kind of allophony. Nonetheless, if we analyze *ᵘa* in these stems as a phoneme, it will be an extremely rare phoneme, the realization of which is much more frequent as a secondary, allophonic realization of another phoneme. In addition, the emergence of *ᵘa* in these environments lacks historical explanation.

Moreover, we are not aware of any minimal pairs in which one word contains an *ᵘa* and a similar word contains another simple vowel, diphthong or vowel sequence in the same position, and their meanings differ.

To sum up, we do not have enough supporting evidence for *ᵘa* being a phoneme, and, at the moment, it seems to be much more likely that we can find an explanation for allophony in the future than convincing arguments to treat it as a phoneme.

6. A modest proposal

Based on the above, the case of central vowels (2) is the most evident: we need no distinction of central vowels in Nganasan, centrality is just a phonetic feature of (non-low?) unrounded back vowels.

The phonemic labiality of *a* (4) also seems to be unambiguous, at least until we discover some phonemic regularities which show that *a* sometimes behaves like unrounded vowels. However, two problems arise in connection with this. On the one hand, if we want to treat *a* as a phonetically central vowel, we have to order redundancy rules in a way that a first delabializes and then centralizes. If we want to see *a* as a phonetically back vowel, we have to apply the rules in the reverse order. On the other hand, if seeing the high number of irregularities and unusual behavior in quasi-labial harmony, we decide to treat this phenomenon as reflecting earlier but not any more active phonological processes (cf. Fejes 2018: 68), the analysis of *a* as a labial vowel is not justified anymore.

The deletion of ^{u}a from the phonemic inventory (5) – independently of the fact whether we analyze it as a diphthong, a simple vowel realized in a diphthong-like way or anything else – is much more problematic, but in general well-founded. In most of the cases ^{u}a shows a clearly allophonic distribution, and there are only a handful of stems where we have no clear explanation for allophony. However, it is not enough to analyze it as a phoneme.

Finally, the problem of diphthongs-or-vowel-sequences (Section 3) is far from evident. The only thing which is worth emphasizing is that present-day mainstream literature on Nganasan phonology unequivocally supports the vowel sequence analysis, although this analysis is not unquestionably better than the other one involving diphthongs. The main issue is the inconsistency with mainstream Finnish phonology: to accept the vowel sequence analysis entirely, either we have to show the crucial differences between Nganasan and Finnish phonology which support the different analysis, or we have to rewrite Finnish phonology.

To sum up, I recommend the following table to present the Nganasan vowel system:

	Front		Back	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	*i*	*ü*	*ɩ*	*u*
Mid	*e*		*ə*	*o*
Low	(*ⁱa*)			*a*

Table 12. Nganasan vowels – a proposal

Abbreviations

3	3rd person possessive suffix
3SG	3rd person singular possessive/verbal suffix
ADVZ.STEM	adverbializer
AUG	augmentative
EMPH	emphatic
ILL	illative
INE	inessive
INF	infinitive
INFER	inferential
INTENT	intentional
IPFV	imperfective
ITER	iterative

LATPRON	lative pronominalizer
NAR	narrative
NOM	nominative
PART	partitive
PASS	passive
PL	plural
PTCP	participle
PRS	present
R	reflexive
S	subjective
SG	singular
VBLZ	verbalizer

References

Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor, and Wagner-Nagy, Beáta. 2016. *Nganasan Spoken Language Corpus (NSLC): Archived in Hamburger Zentrum für Sprachkorpora. Version 0.1*. Publication date December 23, 2016. http://hdl.handle.net/11022/0000-0001-B36C-C

Fejes, László 2018a. Utilization of Nganasan digital resources: a statistical approach to vowel harmony. In: Pirinen, Tommi A., Rießler, Michael, Rueter, Jack, Trosterud, Trond, and Tyers, Francis M., eds. Proceedings of the 4th International Workshop for Computational Linguistics for Uralic Languages (IWCLUL 2018). Helsinki: Association for Computational Linguistics, 121–140.

Fejes, László 2018b. A rejtőzködő harmónia [Harmony in hiding]. *Általános Nyelvészeti Tanulmányok* 30: 37–70.

Fejes, László 2019. Mennyire új a nganaszan palatoveláris harmónia? [How new is the Nganasan palato-velar harmony?] In: Forgács, Tamás, Németh, Miklós, and Sinkovics, Balázs (eds.) *A nyelvtörténeti kutatások újabb eredményei. 10. kötet* [Recent findings in historical linguistics, vol. 10]. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék, 105–117.

Hajdú, Péter 1962. Die Frage des Stufenwechsels in den samojedischen Sprachen. *Ural-altaische Jahrbücher,* 34: 41–54.

Hajdú Péter 1964. Morás nyelv-e a nganaszan? [Is Nganasan a moraic language?] *Nyelvtudományi Közlemények*, 66: 397–400.

Hakulinen, Lauri 1941. *Suomen kielen rakenne ja kehitys. Ensimmäinen osa*. Ääne- ja muoto-oppia. [The structure and development of the Finnish language. Vol. 1]. Helsinki: Otava.

Helimski 1993 = Хелимский, Евгений 1993. Прасамодийские ə и ӭ: прауральские источники и нганасанские рефлексы. [The Proto-Samoyed ə and ӭ: their Proto-Uralic sources and Nganasan reflexes]: In: Sz. Bakró-Nagy, Marianne and Szíj, Enikő (eds.): *Hajdú Péter 70 éves* [Festschrift for Péter Hajdú, 70]. Budapest: MTA Nyelvtudományi Intézet, 125–133.

Helimski, Eugene 1998. Nganasan. In: Daniel Abondolo (ed.): *The Uralic languages*. London – New York: Routledge, 480–515.

ISKo: Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta and Alho, Irja 2004. *Iso suomen kielioppi* [A big Finnish grammar]. *Suomalaisen Kirjallisuuden Seuran Toimituksia 950*. Hämeenlinna: Suomalaisen Kirjallisuuden Seura.

Kosterkina et al. 2001 = Костеркина, Н. Т., Момде, А. Ч. and Жданова, Т. Ю. *Словарь нганасанско-русский и русско-нганасанский* [Nganasan–Russian and Russian–Nganasan dictionary]. Санкт-Петербург: Филиал издательства «Просвещение».

Katzschmann, Michael 2008. *Chrestomathia Nganasanica. Texte – Übersetzung – Glossar – Grammatik. Bearbeitung der Нганасанская фольклорная хрестоматия zusammengestellt von Kazis I. Labanauskas unter Berücksichtigung des Словарь нганасанско-русский и русско-нганасанский.* Norderstedt: Books on Demand.

Ladefoged, Peter and Maddieson, Ian (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Mikola, Tibor 1970: Adalékok a nganaszan nyelv ismeretéhez [Additional information on Nganasan]. *Nyelvtudomány Közlemének*, 72: 59–92.

Tereščenko 1979 = Терещенко, Н. М. *Нганасанский язык* [Nganasan language]. Ленинград: Наука.

Várnai, Zsuzsa 2002. Hangtan. In: Wagner-Nagy, Beáta (ed.) *Chrestomathia Nganasanica*. Szeged – Budapest: SZTE Finnugor Tanszék – MTA Nyelvtudományi Intézet.

Várnai, Zsuzsa 2005. Some problems of Nganasan phonology: Mora or syllable? In: Wagner-Nagy, Beáta (ed.). *Mikola-konferencia 2004* [A Mikola conference]. Budapest—Szeged: Finnugor Tanszék, 113–126.

Várnai, Zsuzsa 2010. Quasi-consonantal stems in Nganasan: A possible analysis. *Finnisch-Ugrische Mitteilungen*, 32/33: 607–619.

Vaysman, Olga 2009. *Segmental alternations and metrical theory*. Cambridge, MA: MIT PhD dissertation.

Wagner-Nagy, Beáta 2018. *A grammar of Nganasan.* Leiden – Boston: Brill

New aspects in the study of Mari, Udmurt, and Komi-Permyak: The *Typological Database of the Volga Area Finno-Ugric Languages*[1]

Erika Asztalos,[1,2] Nikolett F. Gulyás,[2]
Laura Horváth,[2] and Bogáta Timár[2]

[1]Hungarian Research Centre for Linguistics, Budapest
[2] Eötvös Loránd University, Budapest

The present paper provides a brief overview of the state of the art of an ongoing research project, the *Typological Database of the Volga Area Finno-Ugric Languages* focusing on three case studies. The database is a collection of updated research material from three Finno-Ugric languages of the Volga-Kama linguistic area, namely, Meadow Mari (Finno-Ugric, Uralic), Udmurt, and Komi-Permyak (Permic, Uralic). The case studies cover both morphological and syntactic properties of these languages. The paper is structured as follows: in section 1, we introduce the aims of the project and give an overview of the key notions of the database. Then we examine three research topics that are discussed in the database. Section 2 gives an overview of the encoding and usage of predicative possessive constructions in Udmurt. Section 3 addresses the topic of person marking on nominal adpositions in Meadow Mari, Udmurt, and Komi-Permyak. Finally, we examine the feature of reduplication in all three target languages in section 4, then in section 5, we summarize our findings.

1. Introduction

Linguistic databases differ both in their purpose and structure. Our project aims to create an online typological database of some lesser-described Finno-Ugric languages relying on the concepts of previous typological databases including Dryer (2001) and

---

Dryer and Haspelmath (2013). Our goal is to provide useful material for general linguists and typologists, additionally, the typological framework used in the database can broaden the horizon of recent research within Finno-Ugric studies. In other words, we aim to answer old questions of Uralistics with new methods. Among other questions, we focus on the following general issues:

(i)   What are the typologically defined morphological and syntactic features of the analyzed languages?

(ii) How do these languages fit into the typology of the world's languages?

(iii) How can the database be suitable for analyzing relevant relationships between languages?

The current project is a follow-up of a previous project called the *Typological Database of the Ugric languages* (Havas et al. 2015), thus the key notions we use have remained the same. The database consists of parameters, i.e. cross-linguistically comparable grammatical properties (e.g. Optative mood) and values that are the set of logically possible variants of a certain parameter (e.g. NoOptInfl: There is no distinct verbal paradigm to express the optative mood; OptInfl: There is a distinct verbal paradigm to express the optative mood). The majority of the parameters have more possible values, so there are non-binary distinctions among them.

As we have already mentioned, the *World Atlas of Language Structures Online* (henceforth WALS, Dryer and Haspelmath 2013) served as a point of departure for our project, but there are several differences between the two databases. While there are 192 parameters (or features) in WALS covering all levels of language from phonology to syntax, our database contains 213 parameters focusing on morphosyntactic phenomena only. WALS includes data from more than 2,600 languages, while our database, which is an intragenetic study, contains data on three related Uralic languages. In the previous project, four additional Uralic languages were studied, see Havas et al. (2015). However, our long-term goal is to broaden the database with data from other Uralic languages. Since the main goal of WALS is to illustrate the structural variation and the areal distribution of this variation cross-linguistically, it is unsurprising that it has fewer up-to-date and well-balanced from Finno-Ugric languages. Therefore, we aim to include some new data elicited from native speakers and language experts, as well as from the literature. Additionally, we provide a short description of parameters and the relevant value(s) of the target languages with an addition of linguistic examples with interlinear glosses. An additional difference is that our database is available not only in English but in Hungarian and Russian also.

2. Case study 1: Predicative possessive sentences in Udmurt

Two parameters in the database (Habitive constructions and Possessive marking in habitive constructions) are concerned with the morphosyntactic encoding of predicative possessive sentences, including case marking and the syntactic function of the possessor and the possessee, transitivity/intransitivity and the semantics of the predicate, and whether the possessee is marked with a morpheme expressing possession and/or agreement with the possessor. The parameters aim to define the canonical variant of predicative possessive sentences.

According to the literature, in Udmurt, the possessor (if overtly present) is in the genitive case, whereas the possessee (which bears the role of subject) is in the nominative, and it is marked by a possessive suffix agreeing in person and number with the possessor (1). The predicate of the sentence is the existential verb,[2] which occupies clause-final position:

(1)   Udmurt (Csúcs 1990: 73)
       (tynad)   *pinal-jos-yd*   *vań.*
       you.GEN   child-PL-2SG   be.PRS
       'You have children.'

However, defining the values for the typological parameters of the database also sheds light on the non-canonical variants of the constructions under discussion, and raises the question of how we can account for those variants. In other words, when identifying the values for the parameters one may also have to identify under what conditions the non-canonical variants of the construction occur, and whether they really are non-canonical instances.

Returning to predicative possessive sentences in Udmurt, a more careful examination reveals that their encoding shows some variation in addition to the canonical variant illustrated in (1). First, the existential verb can precede the possessee-subject, as in (2). Second, when the possessor is overtly present in the sentence, the possessive suffix may in some cases be absent (2). Third, there are instances of nominal predicative possessive sentences, that is, the existential verb under certain circumstances may also be absent from the construction (3).

---

[2] Udmurt has two existential verbs: one of them, *luyny* 'to be' has a full paradigm, while the other, *vań* '(there) is' is strongly defective. Thus, there is no general consensus in Udmurt linguistics as to whether the latter actually counts as a verb. Winkler (2001: 65), for example, takes it for a particle. However, for the sake of simplicity, we will refer in this paper to both elements as existential verbs.

(2)   Udmurt (Asztalos 2018a: 127)
  15 ar ortćysa mynam lu-o-z   badźym korka.
  15 year after I.GEN be-FUT-3SG big   house
  'In 15 years I will have a big house.'

(3)   Udmurt (Asztalos 2018a: 125)
  saša-len   kyk   nylpi-jez.
  Sasha-GEN   two   child-3SG
  'Sasha has two children.'

As for the order of the possessee and the verb, Asztalos's investigations (2018a: 123–130) indicate that in contemporary Udmurt both orders, possessee–verb and verb–possessee, can occur in discourse-neutral contexts. This suggests that the latter variant may also be taken for a canonical manifestation of the construction.

Concerning the omission of the possessive suffix (2), the question arises whether the conditions of omission are similar to those that apply in the case of adnominal possessive constructions. Thus, the parameters under discussion relate to other parameters, which examine the encoding of adnominal possessive constructions. In fact, adnominal possessive constructions are similar to predicative possessive ones in that the possessor, when present, bears the genitive case[3] and the possessee, in the vast majority of cases, is marked by a possessive suffix (cf. 4) but may occasionally be absent[4] (see 5).

---

[3] More precisely, the possessor is in the genitive case if the possessee is not the direct object of the clause, whereas it is in the ablative if the possessee fulfils the role of direct object, as in (i):

(i)   Udmurt (Winkler 2001: 71)
  so kolhoz-leś busy-z-e   voźmat-i-z.
  s/he kolkhoz-ABL field-3SG-ACC show-PST-3SG
  'S/he showed the field of the kolkhoz.'

[4] Pleshak (2018: 144) claims that dependent-marking (i.e. morphological marking only of the possessor, but not of the possessee) does not occur in Udmurt with internal possessors. However, she also mentions the difficulties one may encounter when trying to distinguish between external and internal possessive constructions, arising from the fact that the two construction types may be marked the same way (Pleshak 2018: 142), and she does not explain on what grounds she excluded constructions similar to (5) from her analysis. At first glance, sentences like (5) could be analysed both as internal (adnominal) and as external possessive constructions, and deciding about their exact syntactic status is left for future research.

(4) Udmurt (Winkler 2001: 21)
     nyl-len   *derem-ez*
     girl-GEN   dress-3SG
     'a/the girl's dress'

(5) Udmurt (Perevoshchikov et al. 1962: 77)
     śergej-len   *mylkyd*   *lobʒ́-ono*   *kaď.*
     Sergey-GEN   mood   fly_away-PTCP.NEC   like
     'Sergey is in a good mood.' (lit. 'Sergey's mood is about to fly away.').

Preliminary observations suggest that the conditions for possessive suffix omission are not the same in the case of adnominal and of predicative possessive constructions. In adnominal constructions, omission is mostly possible if the possessee is inalienable (5) (F. Gulyás et al. 2018). The example in (2) suggests that this is not necessarily the case for predicative possessive constructions. Our examples of predicative possessive constructions with the possessive suffix absent mostly come from young speakers of Udmurt, whereas adnominal constructions without a possessive marker have also been cited, e.g. from old folklore texts in the literature. It is thus possible that possessive suffix omission in predicative possessive structures is not conditioned by semantic or grammatical factors but is a relatively recent phenomenon, which is simply due to the influence of Russian.[5]

Verbless predicative possessive sentences are examined in Asztalos (2018b) as a subclass of verbless existential sentences. The results suggest that such sentences are associated with the pragmatic-semantic features of presupposition and exhaustivity, and involve a focused subject, or a subject with a focused modifier. Namely, verbless predicative possessive and existential sentences occur in contexts that presuppose the existence of a set of possible referents within the frame established by the (locative or temporal) adverbial or the possessor present in the sentence. The subject constituent of the existential/predicative possessive sentence identifies exhaustively, within a set of possible referents, that referent (or those referents) to which the proposition holds true. Exhaustive identification means that the proposition only holds true, within the set of possible referents, for this referent. For instance, the sentence in (6) is grammatical in contexts which presuppose that there is someone in the room and asserts that the person in the room is a woman.

---

[5] Russian lacks possessive suffixes but otherwise encodes predicative possessive sentences in a similar way to Udmurt (and many other Finno-Ugric languages), the predicate of these sentences being the existential verb and the possessee fulfilling the subject role. On the other hand, the question arises whether possessive suffix omission in adnominal constructions (even if it seems to be an older pattern) may also be at least partly triggered by the influence of Russian.

(6)   Udmurt (Winkler 2011: 141)
        komnata-yn   kyšnomurt.
        room-INE   woman
        'There is a woman in the room.'/'It is a woman who is in the room.'

The sentence in (3) presupposes that Sasha has children, and the numeral modifier *kyk* 'two' of the subject states that Sasha has exactly two of them.

However, as verbless predicative sentences contain a focused subject or a subject with a focused modifier, they cannot count as canonical manifestations of predicative possessive sentences in Udmurt. Therefore, such constructions have to be left out when identifying the values for the related typological parameters.

3. Case study 2: Person marking on nominal adpositions in Meadow Mari

The person marking on adpositions is discussed under the parameter 48A in WALS (Bakker 2013), where it examines occurrences both with a pronominal and a nominal complement. It distinguishes four possible values:

   (a) there are no adpositions in the languages
   (b) there is no person marking on adpositions
   (c) person marking only occurs with a pronoun
   (d) person marking occurs with both a pronoun and a noun.

Our database discusses person marking on adpositions with (i) a personal pronominal complement (7), (ii) a non-personal pronominal complement (8), and (iii) a nominal complement (9)–(10).

(7)   Synja Khanty (F. Gulyás 2015)
        (mā)   šanš-em-n   išńi.
        I   behind-1SG-LOC   window
        'There is a window behind me.'

(8)   Northern Mansi (Németh 2015)
        ta   jot-e
        that   with-3SG
        'with that'

(9)   Hungarian (obsolete) (F. Gulyás 2015)
        a   ház-nak   mellett-e
        ART   house-GEN   next_to-3SG
        'next to the house'

(10)   Hungarian (personal knowledge)
       a   *lány-om*   *mögött*
       ART   daughter-1SG   behind
       'behind my daughter'

This case study only focuses on the latter, with special regards to Meadow Mari. Theoretically, there are two possible ways in which the person marker can occur with a nominal complement. In (9), the person marker on the postposition agrees with the noun, which is therefore always in third person. In (10), the noun displays the function of possessee, and the adposition agrees with the possessor. In this analysis, only the latter is taken into account. We consider the adposition as the head, and the nouns or noun phrases they modify as the complements or objects of the adpositions (Havas 2015). In the description of the parameter, the restriction is made that the person marking must only be realized via affixes and not clitics.

       For the parameter, the following values are ascertained:

NoAdp: The language does not have adpositions.
AdpNonPM: Person marking cannot be expressed on adpositions.
AdpNNonPM: Person marking cannot be expressed on nominal adpositions.
AdpN(PM):   Person marking is optional for adpositions when they appear with nouns.
AdpNPM: Person marking is required for adpositions when they appear with nouns.

Like most Uralic languages, Meadow Mari has adpositions, which are postpositions (Bereczki 1990, Riese et al. 2019, Kangasmaa-Minn 1998). In possessive constructions, person marking is realized by possessive pronouns and possessive suffixes (Alhoniemi 1985: 74–79), both of which can be dropped if the other is present (11)–(12). Co-occurrence of both the possessive suffix and the possessive pronoun is also frequent (13).

(11)   Meadow Mari (Bereczki 1990: 42)
       məj-ən   *šergaš*   *šörtnən.*
       I-GEN   ring   silver
       'My ring is silver.'

(12)   Meadow Mari (Arkhangelskiy 2019)
       marij   *jəlme*   *verč*   *čon-em*   *koršt-a.*
       Mari   language   for   soul-1SG   ache-3SG
       'My soul aches for the Mari language.'

(13) Meadow Mari (Alhoniemi 1985: 77)
məj-ən   *molo*   *šoč-š-em*   *uke.*
I-GEN   other   be_born-PTCP.ACT-1SG   be.NEG
'I have no other child.'

The regular pattern for a possessive construction in Meadow Mari, as well as in Udmurt, is that the noun agrees in person with the possessor, not the adposition. However, instances where the person marker appears on the adposition instead of the noun can be found in all three languages (14)–(16).

(14) Udmurt (elicited)
kńiga-je   *ulyś*   ~   *kńiga*   *ulyśty-m*
book-1SG   from_under   book   from_under-1SG
'from under my book'

(15) Komi-Permyak (elicited)
pyzan   *bok-a-m*
table   next_to-INE-1SG
'next to my table'

(16) Meadow Mari (elicited)
pakča-t   *pokšelne*   ~   *pakča*   *pokšelne-t*
yard-2SG   in_the_middle   yard   in_the_middle-2SG
'in the middle of your yard'

This pattern is not possible on every occasion. Whether the person marker can appear on the adposition first and foremostly depends on the adposition itself. Some adpositions do not show person agreement at all, neither with pronominal nor nominal complements (cf. for Meadow Mari, Riese et al. 2019: 163–185).

(17) Meadow Mari (elicited)
pört-da   *marte*   *kaj-en-na*   ~   **pört*   *marte-da*   *kaj-en-na*
house-2PL   until   go-PST2-1PL   house   until-2PL   go-PST2-1PL
'we went up to your house'

(18) Meadow Mari (elicited)
məj-ən marte ilə-ne-da   *gən*   ~   **məj-ən marte-m ilə-ne-da*   *gən*
I-GEN until live-DES-2PL   if   I-GEN   until-1SG live-DES-2PL if
'if you want to live as long as me'

Some adpositions do allow person agreement with pronominal complements (19), but do not allow the same with a nominal complement (20).

(19)   Meadow Mari (elicited)
       məj-ən nergen voz-en-ət   ~   *məj-ən nergen-em voz-en-ət*
       I-GEN   about   write-PST2-3PL   I-GEN   about-1SG write-PST2-3PL
       'they wrote about me'

(20)   Meadow Mari (elicited)
       pij-et   nergen šuko kalas-et   ~   **pi nergen-et šuko kalas-et*
       dog-2SG about a_lot tell-2SG   dog about-2SG a_lot tell-2SG
       'you talk a lot about your dog'

Whether an adposition can or cannot show person agreement with a noun has been subject of a number of studies (Schlachter 1960, Rédei 1962, Efremov 1955–56, Saarinen 1991, Kubínyi 2015). In her analysis based on Udmurt, Meadow Mari and Komi-Zyrian texts, Kubínyi (2015) argues that in the Permic languages the main defining factor is the so-called "controlledness". In that sense, the possession is more "controlled" if the possessee in the "object"-NP (i.e. the possessor and the noun) is an entity controlled immediately by the possessor, and/or the postposition has a spatial value (such as in 14–16). On the other hand, a possession is "uncontrolled" if the possessee is caused or experienced by the possessor, and/or the semantic characteristics of the postposition are temporal or causal-final (20). The adposition tends to agree with the possessor if the possessive construction is controlled. While this correlation is apparent in Permic languages, the same phenomenon is less clearly visible in Meadow Mari (Kubínyi 2015). So if controlledness is not definitive in terms of irregular person marking in Meadow Mari, then what is?

To answer this question, further research should be carried out. Here are some factors that may influence person marking:

-   animacy hierarchy
-   alienability
-   person and number
-   frequency of usage
-   degree of grammaticalization of the postposition

With such restrictions, the following parameter values have been agreed to be true for Meadow Mari:
Person marking cannot be expressed on adpositions (17)–(18)
Person marking cannot be expressed on nominal adpositions (19)–(20)
Person marking is optional for adpositions when they appear with nouns (16).

However, though these parameter values give a detailed description on the person marking in a given language, by themselves they still do not shed light on the actual problematics of what allows or does not allow the irregular pattern to appear.

Another problem arises with the restriction stated in the parameter description. In order to be taken into consideration, it was required for the person marker to be an affix, not a clitic. In the most widely accepted definition, a clitic is distinguished by its flexibility of attachment, as they attach to the phrase, not the word (Halpern 1998: 101), therefore they can attach to nouns and adpositions alike. As we have seen in (14)–(16), this is exactly the case concerning the languages in question, so it can be argued that the person markers are clitics[6] rather than affixes under certain circumstances, and, therefore should not be taken into account. Naturally, it is a stretch to state that person markers in the Uralic languages of the Volga–Kama region were clitics, especially given how special (and not completely defined) circumstances need to constellate for such a pattern to appear, yet it is worth pointing out, to show how indefinite the borders of various terms and their definitions can be.

4. Case study 3: Reduplication in the Volga area Finno-Ugric languages

Reduplication can be broadly understood as the productively employed repetition of words or parts of words to form a new constituent with a different grammatical or lexical function (Schwaiger 2015: 468). The phenomenon is quite common cross-linguistically and can be discussed, for instance, within the framework of Morphological Doubling Theory (MDT) and Sign-Based Morphology, (see, for instance, Inkelas and Zoll 2005), or Evaluative Morphology (see, for instance, Körtvélyessy 2015).

According to Inkelas and Zoll (2005: 2), two general approaches to duplication are possible: it can be analyzed as phonological copying or as morpho-semantic (MS) feature duplication. There are a few criteria for classifying the two types. First, while typical phonological copying is motivated by phonological well-formedness, morphological reduplication serves a morphological purpose instead. Second, while phonological copying involves a phonological segment (mora, syllable, or foot), morphological doubling involves a morphological constituent (affix, root, stem, or word that can be truncated to a prosodic constituent) (Inkelas and Zoll 2005: 22). Finally, in morphological doubling, the two copies are not necessarily identical phonologically (Inkelas and Zoll 2005: 3) (although phonological identity may occur in these cases as well, cf. Inkelas and Zoll 2005: 22), while phonological reduplication

---

[6] Kubínyi (2015) in fact called the pattern Possessive Clitic Climbing, referring to the person markers as clitics.

necessarily involves phonological identity. What morphological reduplication involves is, however, semantic identity: according to the MDT, the duplicants are required to be identical only semantically (Inkelas and Zoll 2005: 7, 25). Reduplication involving duplicants which are synonymous, antonymous, or semantically closely related is called synonym reduplication in other works (see, for instance, Inkelas 2014: 170). Other subtypes of reduplication are the following: total/full reduplication, partial reduplication, echo reduplication, syntactic doubling (cf. Inkelas 2014: 169–172).

Reduplication can be used to perform a wide range of both derivational and inflectional functions. Reduplication as derivation can change the word class and argument structure, or alter the meaning of a word creating a new lexeme (Inkelas 2014: 174–175). The formation of expressive, evocative or onomatopoetic formations is among the most common functions of reduplication in the languages of our database (see, for instance, Riese 2016, Shlyakhova 2013, Fejes 2016), as well as in the Turkic languages of the area (see, for instance, Iskandarova 2016, Semenova and Ivanova 2016). In the Meadow Mari example (22), reduplication converts an inflected form of a reduplicated verb (< *manaš* 'to say') into a noun. This part of speech-changing process does not seem to be a common productive technique for deriving nouns in Mari, nor is it typical of the other languages of the database:

(21)   Meadow Mari (Riese 2016: 3287)
        man-eš   ~   *man-eš*
        say-3SG   ~   say-3SG
        'gossip'

As an inflectional device, reduplication is frequently employed for encoding plurality in nouns, or less commonly, possession categories; case-marking is an uncommon but attested function of reduplication as well. Reduplication is often used to encode aspectual distinctions: pluractionality of verbs (event pluralization), frequentativity, continuation, progressivity, habituality, imperfectivity, and even perfectivity (cf. Inkelas 2014: 172–176).

Because the derivation-inflection relation also has a cline-like nature itself, there are many examples regarding cases of reduplication that are not so clear-cut: they share the properties of both derivational and inflectional parts of this cline – for instance, diminutives, augmentatives, and other evaluatives (cf. Inkelas 2014: 175–176, Körtvélyessy 2015: 22–31). In this paper, evaluation is treated "as a continuum such that prototypical cases express the meaning of quantity under or above the default value" (Körtvélyessy 2015: 41). The default value is rather a relative concept: these reference points are influenced by extralingual factors: culture, experience and knowledge of the speakers. The default quantity/value is the reference point that is

anchored in categories such as "substance", "action", "quality", and "circumstance". In this system, diminutives and augmentatives are viewed as deviations from the prototypical value in these cognitive categories (Körtvélyessy 2015: 41–47).

In this paper, we aim to concentrate on the cases of morphological doubling expressing evaluative and aspectual functions of reduplication in the languages of the database.

4.1. Evaluatives

Reduplicative evaluative formations are frequently used by all of the languages of our database. It is a common feature of the Turkic languages of the Volga–Kama region (Bashkir, Chuvash, and Tatar) as well (Semenova and Ivanova 2016: 3464, Iskandarova 2016: 3394– 3395, Károly 2016: 3411). Intensification is a very common function of reduplication also cross-linguistically (see, for instance, Štekauer et al. 2012: 123–126).

Reduplication is not regarded as a typical feature of the Uralic proto-language, but it was already used for the expressive intensification of color names and other adjectives in Old Turkic (Erdal 1991: 65). In the languages of our database, reduplication is often used for expressing augmentatives (higher quantity of a quality). Intensification of color names can be employed by total reduplication:

(22)   Udmurt (Keľmakov and Saarinen 1994: 112)
gord   ~   *gord*
red   ~   red
'very red'

(23)   Komi-Permyak (elicited)
görd   ~   *görd*
red   ~   red
'very red'

(24)   Meadow Mari (Riese 2016: 3287)
kande   ~   *kande*
blue   ~   blue
'very blue'

Adverbs and adjectives other than color names can be reduplicated as well:

(25)   Udmurt (Shlyakhova 2013: 1331)
umoj   ~   *umoj*   *vala-ny*
good   ~   good   understand-INF
'to understand very well'

(26)   Komi-Permyak (elicited)
        basök   ~   basök
        beautiful   ~   beautiful
        'very beautiful'

(27)   Meadow Mari (elicited)
        šokšən   ~   šokšən   öndal-eš
        warmly   ~   warmly   hug-3SG
        's/he gives a very warm hug'

In Mari, one special kind of partial reduplication can be used: adjectives beginning with CVC-structure can be reduplicated by doubling the CVC and by the substitution of the second C with *p*. The reduplicant precedes the adjective:

(28)   Meadow Mari (Bradley et al. 2014)
        joltaš-em   tap   ~   taza
        friend-1SG   RED   ~   healthy
        'my friend is fit as a fiddle'

This kind of partial reduplication is considered to be a feature of Old Turkic (Erdal 1991: 65) and occurs in Mari due to Tatar and Chuvash influence (Riese 2016: 3287), cf. the following Tatar example:

(29)   Tatar (Károly 2016: 3411)
        kap   ~   kart
        RED   ~   old
        'very old'

Intensification can also be expressed by echo reduplication in at least Udmurt and Meadow Mari. In these cases, the reduplicant follows the "base" (for the terms "base" and "reduplicant", see Inkelas and Zoll 2005: 7–11), and the beginning of the reduplicant is replaced by a fixed segment (Inkelas 2014: 171):

(30)   Udmurt (Shlyakhova 2013: 1330)
        kyryž   ~   meryž
        crookedly   ~   RED
        'very crookedly'

(31)   Meadow Mari (Riese 2016: 3287)
        kadyr   ~   gudyr
        crooked   ~   RED
        'very crooked'

Syntactic doubling can also encode evaluation in the languages of the database: the duplicants can be separated by affixes (32)–(33), particles (34), affixes and particles (35), postpositions (36), and conjunctions (37)–(38):

(32)   Udmurt (elicited)
   ćeber-leś   *ćeber*
   beautiful-ABL beautiful
   'very beautiful'

(33)   Komi-Permyak (elicited)
   basök-śa   *basök*
   beautiful-PRECL beautiful
   'very beautiful'

(34)   Meadow Mari (elicited)
   motor-at   *motor*
   beautiful-PCL beautiful
   'very beautiful'

(35)   Udmurt (elicited)
   ćeber-leś   *no* *ćeber*
   beautiful-ABL PCL beautiful
   'very beautiful'

(36)   Meadow Mari (elicited)
   motor *deč* *motor*
   beautiful from beautiful
   'very beautiful'

(37)   Northern Udmurt (Karpova 2011: 30)
   kuź *da* *kuź*
   long and long
   'very long'

(38)   Komi- Permyak (Shlyakhova 2013: 1330)
   ʒ́or *i* *ʒ́or*
   unripe and unripe
   'very unripe'

4.2. Reduplication as an aspectual device in the languages of the database

As has already been mentioned, reduplication is often used to encode various aspectual values: cross-linguistically, reduplication is not uncommonly linked with imperfective aspectual categories (Inkelas 2014: 173–174, Štekauer, Valera and Körtvélyessy 2012: 126–128). In the languages of our database, in some cases, reduplicated adverbs have an effect on the aspectual system. They can be employed in habitual utterances: Meadow Mari *ugəč ~ ugəč* 'repeatedly' ← *ugəč* 'anew', Udmurt *vyľyś ~ vyľyś* 'repeatedly' ← *vyľyś* 'anew'.

In some Northern Udmurt dialects, the frequentative suffix *-ly-*, *-ľľa-* can be reduplicated "for intensifying the repetition of the event" (Keľmakov and Saarinen 1994: 125), that is, the pluractionality and habitual value of an action:

(39) Northern Udmurt (Keľmakov and Saarinen 1994: 125)

u̯aľľo	*pios-jos*		*no,*	*nyl-jos*
long_time_ago	man-PL		too	girl-PL
no	*kyrʒa-ľľa-ľľa-zy*		*udmort*	*kreźź-os-ty.*
too	sing-FREQ-FREQ-PST.3PL		Udmurt	song-PL-ACC

'A long time ago, men and girls used to sing Udmurt songs.'

In Komi-Permyak, the frequentative suffix itself can be used in the case of both repeated and non-repeated events:

(40) Komi-Permyak (elicited)

körkö	*me*	*ker-l-i*	*šańga-ez.*
long_time_ago	I	make-FREQ-PST.1SG	shanga-PL

1. 'I once did shangas a long time ago.'
2. 'A long time ago, I used to make shangas (many times).'

However, with two frequentative suffixes attached to the verbal stem, the event can only be interpreted as a pluractional event, that is, reduplication can prevent ambiguousness:

(41) Komi-Permyak (elicited)

körkö	*me*	*ker-l-yvl-i*	*šańga-ez.*
long_time_ago	I	make-FREQ-FREQ-PST.1SG	shanga-PL

'A long time ago, I used to make shangas (many times).'

5. Conclusion

In our paper, we have presented the aims, main research questions and linguistic background of the ongoing project *Typological Database of the Volga Area Finno-Ugric Languages*, and provided analyses of three features of the database.

First, we have discussed the properties of the presumably non-canonical cases of the predicative possessive constructions in Udmurt and attempted to specify the conditions under which these cases occur, taking into consideration the word order and differences between predicative and adnominal possessive constructions. In these kinds of sentences, the existential verb can be absent, and/or the possessive suffix is not attached to the possessee. We have found that while in the case of adnominal constructions, the omission of the possessive suffix is linked with the inalienability of the possessee, this does not seem to be the case when using predicative possessives: in the latter case, omission might be due to the influence of Russian on Udmurt. As for the lack of an existential verb in the predicative possessive constructions, it seems to be useful to take into account notions like focus, presupposition, and exhaustivity as well.

Second, we have provided an analysis concerning person marking on adpositions with a nominal complement, concentrating mainly on Meadow Mari. In Meadow Mari adpositional constructions, in some cases, the person marker can appear either on the noun or on the adposition. However, we have found that some adpositions do not allow person agreement, and some allow it only under certain conditions: they show agreement only with pronominal complements and not with nominal ones. We assume that factors other than controlledness, for instance, animacy hierarchy, alienability, frequency of usage, or the degree of grammaticalization of the postposition should be taken into consideration when analyzing the person marking in these constructions.

In our third case study, we have focused on some formal and functional properties of reduplication in Udmurt, Komi-Permyak and Meadow Mari, concentrating on the evaluative and inflection-like functions of reduplication in these languages. Reduplication is often a tool for expressing augmentatives (a higher quantity of a quality) in both the Turkic and Finno-Ugric languages of the Volga–Kama linguistic area. Reduplication has a role in encoding aspectual values as well: in the languages of the database, different aspectual distinctions linked to imperfective aspect (pluractionality, habituality) are marked with reduplicated suffixes and adverbs.

Abbreviations

1   first person
2   second person
3   third person

ABL	ablative
ACC	accusative
ACT	active
ART	article
DES	desiderative
FREQ	frequentative
FUT	future
GEN	genitive
INE	inessive
INF	infinitive
LOC	locative
NEC	necessive
NEG	negative
PCL	particle
PL	plural
PRS	present
PST	past
PST2	second past
PTCP	participle
PRECL	preclusive
RED	reduplication
SG	singular

References

Alhoniemi, Alho 1985. *Marin kielioppi. Apuneuvoja suomalais-ugrilaisten kielten opintoja varten X.* Helsinki: Suomalais-ugrilainen seura.

Arkhangelskiy, Timofey 2019. Corpora of social media in minority Uralic languages. In: Pirinen, Tommi A., Kaalep, Heiki-Jaan and Tyers, Francis M. (eds.) *Proceedings of the Fifth Workshop on Computational Linguistics for Uralic Languages.* Tartu: Association for Computational Linguistics, 125–140.

Asztalos, Erika 2018a. *Szórendi típusváltás az udmurt nyelvben* [Typological change in word order in the Udmurt language]. Budapest: Eötvös Loránd University PhD dissertation.

Asztalos, Erika 2018b. *Nonverbal existential sentences in Udmurt.* Poster presentation at Syntax of the World's Languages 8. Paris: Institut National des Langues et Civilisations Orientales. 3–5. September 2018. http://ugortip.elte.hu/Volgatip/Nonverbal.pdf. Date of access: March 28, 2020

Bakker, Dik 2013. Person marking on adpositions. In: Dryer, Matthew S. and Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/48. Date of access: March 28, 2020.

Bereczki, Gábor 1990. *Chrestomathia ceremissica*. Budapest: Tankönyvkiadó.

Bradley, Jeremy, Riese, Timothy and Guseva, Elina 2014. *Mari–English dictionary*. Vienna: University of Vienna. http://dict.mari-language.com. Date of access: August 28, 2018.

Csúcs, Sándor 1990. *Chrestomathia Votiacica*. Budapest: Tankönyvkiadó.

Dryer, Matthew S. 2001. *Matthew Dryer's Typological Database*. Buffalo: Department of Linguistics, University at Buffalo. http://www.acsu.buffalo.edu/~dryer/database.htm. Date of access: October 22, 2019.

Dryer, Matthew S. and Haspelmath, Martin (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/. Date of access: October 22, 2019.

Efremov, V. A. 1955. *Poslelogi kak sredstvo vyraženija grammatičeskih otnošenij v marijskom jazyke*. – Istorija, jazyki literatura mari. Trudy Marijskogo Naučno-Issledovatel'skogo Instituta Jazyka, Literatury i Istorii 7. Joškar-Ola: MarNII, 59–85.

Erdal, Marcel 1991. *Old Turkic word formation: A functional approach to the lexicon, Vol. 1*. Wiesbaden: Harrasowitz.

Fejes, László 2016. Permic. In: Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 5*. Berlin and Boston: De Gruyter Mouton, 3260–3274.

F. Gulyás, Nikolett, Asztalos, Erika, Timár, Bogáta, Horváth, Laura and Szabó, Ditta 2018. *Possessive agreement in Meadow Mari, Udmurt, and Komi-Permyak*. Conference presentation at Syntax of the World's Languages 8. Paris: Institut National des Langues et Civilisations Orientales. September 3–5, 2018. http://ugortip.elte.hu/Volgatip/possessive_agr_fu_swl_2018.pdf. Date of access: March 28, 2020.

F. Gulyás, Nikolett 2015. Person marker on nominal adpositions: Hungarian. In: Havas, Ferenc, Csepregi, Márta, F. Gulyás, Nikolett and Németh, Szilvia *Typological Database of the Ugric Languages*. Budapest: ELTE Finnugor Tanszék. http://en.utdb.nullpoint.info/node/1185. Date of access: March 28, 2020.

F. Gulyás, Nikolett 2015. Person marker on nominal adpositions. Synja Khanty. In: Havas, Ferenc, Csepregi, Márta, F. Gulyás, Nikolett and Németh, Szilvia *Typological Database of the Ugric Languages*. Budapest: ELTE Finnugor Tanszék. http://en.utdb.nullpoint.info/node/1184. Date of access: March 28, 2020.

Halpern, Aaron L. 1998. Clitics. In: Spencer, Andrew and Zwicky, Arnold M. (eds.) *The handbook of morphology*. Oxford: Blackwell, 101–122.

Havas, Ferenc (ed.) 2015. Person marking on nominal adpositions. In: Havas, Ferenc; Csepregi, Márta; F. Gulyás, Nikolett and Németh, Szilvia. *Typological Database of the Ugric Languages*. Budapest: ELTE Finnugor Tanszék. http://en.utdb.nullpoint.info/content/person-marking-nominal-adpositions. Date of access: March 28, 2020.

Havas, Ferenc, Csepregi, Márta, F. Gulyás, Nikolett and Németh, Szilvia 2015. *Typological Database of the Ugric Languages*. Budapest: ELTE Finnugor Tanszék. http://utdb.elte.hu. Date of access: October 22, 2019.

Inkelas, Sharon 2014. Non-concatenative derivation: Reduplication. In: Lieber, Rochelle and Štekauer, Pavol (eds.) *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press, 169–189.

Inkelas, Sharon and Zoll, Cheryl 2005. *Reduplication: Doubling in morphology*. Cambridge Studies in Linguistics 106. Cambridge: Cambridge University Press.

Iskandarova, Gulnara 2016. Bashkir. In: Müller, Peter O.; Ohnheiser, Ingeborg; Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 5.* Berlin and Boston: De Gruyter Mouton. 3386–3397.

Kangasmaa-Minn, Eeva 1998. Mari. In: Abondolo, Daniel (ed.) *The Uralic languages*. London and New York: Routledge. 219–248.

Károly, László 2016. Tatar. In: Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 5*. Berlin and Boston: De Gruyter Mouton, 3398–3413.

Karpova, L. L. 2011. O stepenjah kačestva prilagatelnyh v severnoudmurtskih dialektah. *Uralo-altajskije issledovanija,* 1 (4): 24–33.

Keľmakov, V. K. and Saarinen, Sirkka 1994. *Udmurtin murteet*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 47. Turku and Izhevsk: Turun yliopisto.

Körtvélyessy, Lívia 2015. *Evaluative morphology from a cross-linguistic perspective*. Cambridge: Cambridge Scholars Publishing.

Kubínyi, Katalin 2015. *Possessive clitic climbing as a pattern of agreement with the possessor in Permic and Mari postpositional phrases*. XII. International Congress for Finno-Ugric Studies. Poster presentation. Oulu.

Németh, Szilvia 2015. Person marker on nominal adpositions. Northern Mansi. In: Havas, Ferenc, Csepregi, Márta, F. Gulyás, Nikolett and Németh, Szilvia *Typological Database of the Ugric Languages*. Budapest: ELTE Finnugor Tanszék. http://en.utdb.nullpoint.info/node/845. Date of access: March 28, 2020.

Perevoshchikov, P. N., Vakhrushev, V. M., Alatyrev, V. I., Pozdeeva, A. A. and Tarakanov, I. V. (eds.) 1962. *Grammatika sovremennogo udmurtskogo jazyka. Fonetika i morfologija.* Izhevsk: Udmurtskoje kniznoje izdatelstvo.

Pleshak, Polina 2018. Adnominal possessive constructions in Mordvin, Mari and Permic. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri*, 9(1): 139–168.

Rédei, Károly 1962. *Die Postpositionen im Syrjänischen unter Berücksichtigung des Wotjakischen.* Budapest: Akadémiai.

Riese, Timothy 2016. Mari. In: Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 5.* Berlin and Boston: De Gruyter Mouton, 3275–3288.

Riese, Timothy, Bradley, Jeremy, Schötschel, Monika and Yefremova, Tatiana 2019. *Mari (марий йылме): An essential grammar for international learners [draft version].* Vienna: University of Vienna. Published online at http://grammar.mari-language.com. Date of access: August 08, 2019.

Saarinen, Sirkka 1991. *Marilaisen arvoituksen kielioppi*. MSFOu 210. Helsinki: SUS. 81–82.

Schlachter, Wolfgang 1960. *Studien zum Possessivsuffix des Syrjänischen*. Berlin: Akademie-Verlag.

Schwaiger, Thomas 2015. Reduplication. In: Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 1.* Berlin and Boston: De Gruyter Mouton, 467–484.

Semenova, G. N. and Ivanova, A. M. 2016. Chuvash. In: Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan and Rainer, Franz (eds.) *Word-formation, Vol. 5.* Berlin and Boston: De Gruyter Mouton, 3451–3467.

Shlyakhova, S. S. 2013. Study of reduplication in Finno-Ugrian languages of Permic group. *Middle-East Journal of Scientific Research,* 16: 1329–1333.

Štekauer, Pavol, Valera, Salvador, and Körtvélyessy, Lívia 2012. *Word-formation in the world's languagest A Typological survey*. Cambridge: Cambridge University Press.

Winkler, Eberhard 2001. *Udmurt*. Languages of the World 212. München: Lincom Europa.

Winkler, Eberhard 2011. *Udmurtische Grammatik*. Veröffentlichungen der Societas Uralo-Altaica 81. Wiesbaden: Harrassowitz.

Ethnosyntax in Siberian Uralic languages (project[1] overview)

Sándor Szeverényi, Bernadett Bíró, and Katalin Sipőcz

University of Szeged

1. Introduction

It is a well-known fact that language interacts with the environment around it, in another words, "linguistic structure is formed, changed and influenced by different aspects of the human environment" (Busser 2015: 1). Recently, various subfields of linguistics have played a role in studies of extra-linguistic effects on language structures, such as sociolinguistics, ecolinguistics, cultural linguistics, and ethnosyntax.

The root of our intention can be found in the traditions of ethnolinguistics, which is "an interdisciplinary field based on linguistics and social anthropology which is concerned with the relationship between language and culture. [...] Ethnolinguistic studies regard language as a social and cultural instrument and try to uncover the cultural meaning behind linguistic structures and language use" (Foley 1997: 3). In this sense, "ethnolinguistics has variously been approached as the study of a group's experience of life as it is organized and expressed through the group's language tools and as a science whose aim is to examine the relationships between a language on the one hand and society and culture on the other" (Riley 2007: 8). However, the main focus of ethnolinguistics is the relationship between language and culture, communicative practices, and cognitive models of language and thought. It reaches many areas of linguistics, not only grammar and anthropological linguistics, but pragmatics and psycholinguistics as well: "The ethnolinguist tries to describe and understand the role of language in shaping the ways in which members of a group relate to the world, to one another and to others" (Riley 2007: 11).

Ethnosyntax is a closely related area of ethnolinguistics, but it presupposes a more abstract relation between culture, language and cognition. In our approach we used the term *culture* in its broadest sense: effects of environment-induced changes in

---

language use that can cause change in the grammar. In our project the term *ethnosyntax* does not refer only to syntactic structures but to any environment-induced characteristics of the grammar. Ethnosyntax in our approach covers also partly ecolinguistics, which explores the role of language in the life-sustaining interactions of humans, other species and the physical environment. It is obvious that effect of culture and environment on grammar is more abstract that effect to the lexicon or language use and practice and the "exploration" of the direct motivation is more complicated. So, in this paper we argue that the ethnolinguistic (and, in a narrower sense, ethnosyntactic) approach to Uralic languages in the light of the most recent research can add new results in the study of Uralic languages.

2. Uralic traditions in ethnolinguistics

Our ethnolinguistic research uses the earliest works of Uralic linguistics as its antecedents whose outlook defined Finno-Ugric linguistics for a long time. Research into the syntax of endangered Uralic languages in Russia has recently acquired a focus on contact linguistics, on the effects of language loss on syntax, primarily from a generative perspective (e.g. publications of RIL HAS, see http://www.nytud.hu/oszt/elmnyelv/urali/publ.html), and the number of syntactic descriptions carried out with a typological perspective has increased (cf. Wagner-Nagy et al. 2015; de Groot 2017, Bíró and Sipőcz 2017). These studies have yielded important findings and, in several cases, have provided new insights into significant historical relationships as well (cf. É. Kiss 2014). However, they also lacks a number of approaches which are present in mainstream linguistic research today, such as the newest trends in anthropological linguistics, the use of cognitive linguistic perspectives in comparative linguistics, or that of specialized ethnolinguistic research.

The roots of ethnolinguistic research in Uralic linguistics go back to the life works of Pál Hunfalvy, Bernát Munkácsi, and Antal Reguly (Gulya 1970/1978). At the same time, a lot of research was conducted also by Finnish and German scholars as well, e.g. Artturi Kannisto, Toivo Lehtisalo, Wolfgang Steinitz, Yrjö Wichmann. Following Hymes (1965), Gulya overviews the areas of ethnolinguistic research stating that ethnolinguistics "[is a linguistic] method of analysis which, while focusing on language, includes in the research the investigation of the people, society, and culture using the language and their histories in a complex and comprehensive manner" (Gulya 1978: 134). In Uralic linguistics, ethnolinguistic research has been successfully carried out on the lexicon (cf. Steinitz's research on kinship terms).

In the late 20th century, with the development of anthropological linguistics (cf. Duranti 1997, 2004) and the spreading of ethnolinguistics beyond the lexicon, new directions emerged, such as ethnopragmatics, ethnosemantics, and ethnosyntax. The term ethnosyntax has been widely known since Wierzbicka's 1979 study and became

a reference point through Enfield's 2002 volume. As a result of the latter, individual chapters in international handbooks of linguistics have been devoted to ethnolinguistic research (cf. Gladkova 2015).

In the present project we aim, on the one hand, to continue the (morpho)syntactic research tradition of Uralic linguistics (into syntactic change in endangered Uralic languages, contact effects in syntax etc.), and, on the other, to provide insight into new aspects of linguistic phenomena.

3. The languages and corpora

The project focuses on two related languages spoken in Siberia: Mansi, an Ob-Ugric language, and Nganasan, a Samoyedic language. Despite the fact that they both belong to the same language family (Uralic) and are spoken in the same linguistic area (Northwestern Siberia), the two languages exhibit great differences in their structure as well as in the ways of life and environment of their speakers. The latter include geographical and climatic differences, which also result in differences in traditional lifestyle, culture, and religion. The investigated languages are severely endangered, their speakers typically Russian–Mansi and Russian–Nganasan bilinguals with Russian as their dominant language.

The two investigated languages differ greatly in the amount of linguistic resources available. Mansi has been documented much more extensively (due to long time research traditions), whereas Nganasan has been documented mostly in the past few decades only. Our research will be carried out on the material included in databases and corpuses compiled in recent years, partly by the project participants:

- the NSL Corpus (a corpus available through the University of Hamburg, containing 294 texts, with 40,235 sentences from 55 speakers, of which about 20,279 sentences have been glossed so far) Version 0.2 of the corpus is available online (Brykina et al. 2018).
- OUL and OUDB: glossed Mansi texts available online (http://www.babel.gwi.uni-muenchen.de/index.php?abfrage=NM_corpus&subnavi=corpus_pub)

4. The focus of the analysis

The following issues are investigated in the project.

4.1. The ethnolinguistic background of linguistic means of perspectivization

Perspectivization and subjectification can be expressed with a variety of linguistic means: lexically, syntactically, morphologically, and in a combination of these ways. (How a language expresses it is often dependent on categories such as epistemic modality and evidentiality.) A typical and most explicit way of perspectivization is direct quotation/speech. Descriptions of types of reported speech are abundant in the typology literature (cf. for instance Aikhenvald 2011, Li 1986), but data on small Uralic languages is completely missing from these. The works cited above also refer to the fact that there are language specific systems whose development is affected by cultural factors and/or contact effects. Our research partly focuses on uncovering such connections, namely, whether the development of reported speech (or lack thereof), the use of evidential morphological systems, and of markers of epistemic modality has been affected by such factors. We have already carried out investigations along these lines (Szeverényi 2017, Szeverényi and Sipőcz 2019), and have examined reported speech structures of Nganasan and Mansi. We also examined whether areal – primarily Russian – influence can be found in these structures. Indirect reported speech structures are rare in both languages, but sporadic data can be found in the corpus for them. In Mansi and Nganasan the direct reported speech structure is much more common than the indirect one. It can be observed that in Finno-Ugric languages the direct speech structure consisting of two clauses requires a subordinate conjunction. This is missing from these languages, although there are examples in our corpus for the adoption of the Russian conjunction *što* 'that'. In Mansi, in addition to the direct structure we can find the indirect structure without a conjunction, and a non-finite structure can be used, too. In Nganasan the evidential (reportative) mood serves as an alternative.

There are many studies in the linguistics literature that can serve as models for a complex analysis of perspectivization and passive structures, however, their connection with evidentiality has not been explored yet. Evidentiality as a grammatical category has appeared in the description of Ob-Ugric languages only in recent years. Our research into Mansi quotative structures also indicated a need to carry out further research into evidentiality.

It is a significant fact that participles are used with an evidential function in the Northern dialect of Mansi, and there is such data only for this dialect in the literature.

A similar observation has been made about Khanty, the closest related language of Mansi (Csepregi 2014). Analysis of the dialect data is ongoing, with texts from the Southern dialect not containing use of the evidential (Eastern dialect data is under analysis). All this is indicative of interesting historical, areal and cultural interconnectedness: the question arises whether the development of evidentiality is an areal phenomenon and/or internal development, given the fact that it happened only in the Northern dialects of Ob-Ugric languages. Bernárdez's cultural linguistic connections are important in both explanations, both in Nganasan and Mansi.

4.2. The event of giving in an ethnolinguistic framework

Investigating the event of giving in an ethnolinguistic framework has been an integral part of our previous research, in which ditransitive constructions in Mansi were investigated primarily from a typological perspective (K-101652). The event of giving has been studied in the literature on ethnosyntax as well. In one of his papers Newman (2002) distinguishes three categories of 'give' phenomena crosslinguistically from an ethnosyntactic perspective.

Ditransitive constructions in the Mansi language (and, in general, in Ob-Ugric languages, and some of the Samoyedic languages) display several characteristics which can be considered language specific: the alternation of indirective and secundative as well as of active and passive constructions (cf. Bíró and Sipőcz 2017). This is not rare typologically and can be considered a language specific feature that is independent of culture or of the environment. A more specific feature of Ob-Ugric ditransitive constructions is the full scale alternation of the two types of constructions (indirective and secundative) practically independently of lexical or semantic limitations. Moreover, a unique language specific feature of these languages (not discussed in the literature at all) is that the events of both giving and receiving are ditransitive and expressed with the verb 'give'. It is the secundative (R passive) construction involving the verb 'give' that expresses this event:

(1) *Kitit*   *mesta-l*   *Nižnewartowskij ūs-t*
    second   place-INSTR   Nizhnevartovsk   town-LOC
    ōl-ne   *xantə-t*   *maj-we-s-ət.*
    live-PTCP.PRS   Khanty-PL   give-PASS-PST-3PL
    'The Khanty people from Nizhnevartovsk got the second place.' ('The second place was given to the Khanty people from Nizhnevartovsk.')

Our more recent research into language use demonstrates that, under the influence of the Russian language, certain verbs – such as 'take' and 'carry' – are beginning to acquire the meaning of 'receive'. Such usage frequently occurs in, for instance, translations from Russian. Such influence can fundamentally affect the system of rules

underlying the alternation of ditransitive constructions.) (For more on this, see Sipőcz in this volume.)

4.3. Spatial expresssions from the perspective of ethnosyntax

The aim of our investigations into spatial expressions is to analyze the cultural background of a unique spatial orientation system found in Khanty, Mansi and Selkup: an adverb/preverb pair meaning 'toward/away from the river(bank)'; 'up to/down from the hill' and 'toward/away from the fire(place)'. This research is partly based on earlier literature on this topic, but our main aim is to also present new data from native speakers and recently assembled language corpora. Additionally, we make an attempt to identify similar phenomena in other languages of the area in order to give an explanation of this lexical orientation system. Several sources have been used for this investigation. In the case of Selkup, beside the lexicographical sources, the Selkup Spoken Language Corpus was also used. In the case of Khanty and Mansi, in addition to dictionaries, an Ob-Ugric Database was used. In addition, we collected data from the Mansi newspaper *Lūimā Sēripos* and from native speakers of Mansi. We can state that the investigated Khanty, Mansi, and Selkup orientation subsystems are results of independent but contact induced parallel development, however, following Fortescue (2011), language contact cannot be regarded as the only explanation, because similar systems can be detected in North American languages as well. The development of these orientation systems was motivated by the particular geographic circumstances of the speakers, namely, living on steep river banks, and also by the fact that fire and water play a central role in their way of thinking. The orientation system is in correlation with the traditional way of life of the indigenous peoples of Western Siberia. The source concepts of the system are connected to the rivers and the dwellings, namely, the ancient cave dwellings on the upper part of the river bank.

Our research into this topic was summarized in a paper published in 2019 (Bíró et al. 2019). We plan to augment it with a description in the cognitive metaphor theoretical framework of cultural linguistics.

4.4. Diminutives

Diminutive derivational suffixes are used in many languages and present interesting examples of language phenomena which encode cultural rules and meanings. Extensive research has been done on such aspects of Slavic languages (Gladkova 2015, Wierzbicka 1979). In our research we use such investigations as our starting point, which is motivated especially by the fact that Siberian Uralic languages are under the continuous and heavy influence of the Russian language. In Mansi, two diminutive suffixes are used: *-riś~rəś* and *-kwe* (*-ke*). They are very productive and

can be attached to any word with the exception of conjunctions (Rombandeeva 1973: 176). A curious feature of Mansi is that these diminutive suffixes occur also in verb conjugation as well, attaching to any verb stem, with the possibility that then genus, tense and mood inflections can also attach to them. When attached to verbs, they serve to express subjectivity: *-kwe* (*-ke*) expresses the speaker's positive stance and adds a meaning of affection and politeness to the verb, while *-riś~rəś* expresses regret and scorn (Rombandeeva and Vahruseva 1989: 140, Riese 2001: 59), e.g. *toti-ke-m* 'I gladly bring', *toti-ke-n!* 'bring some, dear!', *toti-riś-əm* 'I bring poor (it is hard for me)', *tājə-riś-en*! 'eat (you poor, pitiable hungry wretch)!' (Munkácsi 1984: 40, Kálmán 1989: 56, 61). Some authors (Munkácsi 1894, Kálmán 1975) treat these forms as a separate mood ("affectionate" and "precative" or "kedveskedő" and "Präcativ" mood, respectively, in the originals), while others (e.g. Rombandeeva 1973, Rombandeeva and Vahruseva 1989, and Riese 2001) treat them as simple derived forms. In the case of nouns derived with the two diminutive suffixes, the positive and negative meanings are not conveyed necessarily, cf. *piɣriś* 'little boy', *piɣkwe* 'little boy'.

The aim of our analysis is to establish exactly what principles govern the use of diminutive suffixes, and to show what cultural phenomena are behind these principles, as well as to ascertain whether the positive and negative meanings associated with the two derivational suffixes are really so clearly separable. (For more details, see Bíró in this volume).

5. An example: cultural explanation of Nganasan evidentiality

In this section we illustrate a detailed description of a grammatical feature with a possible cultural explanation, namely, the interaction of a subsystem of the grammar and the environment. Our hypothesis is that the complex evidentiality system in Nganasan has been formulated only in the northernmost Arctic area. We apply Bernárdez's hypothesis (2017) on a cultural explanation of evidentiality. Our research questions here are as follow:

 (a) Does the investigation strengthen Bernárdez's cultural interpretation of evidentiality?
 (b) What is the role of evidentiality in perspectivization?

Evidentiality in linguistics concerns how the source of knowledge is expressed in linguistic communication, whether grammatically coded, lexically coded, or merely inferred (Ekberg and Paradis 2009: 5). Bernárdez has a hypothesis (2017) about the cultural interpretation of evidentiality. Previously, some scholars have already referred to the possible connections between culture and evidentiality, e.g. Aikhenvald wrote that "[s]peaking a language with obligatory evidentials implies adhering to strict

cultural conventions. Beliefs, mental attitudes, and patterns of behaviour appear to correlate with these" (Aikhenvald 2004: 361). Bernárdez was the first who found some correlations between environment, culture and evidentiality. His main findings are as follows:

A complex evidential system can be developed in connection with the following factors:

- "Small groups living in isolated environments enhance the probability of developing evidentials;
- Difficulties in accessing the world around enhance the probability of developing evidentials;
- Very tight relations within the group and with neighbouring groups also enhance the probability of developing evidentials." (Bernárdez 2017)

He established a small set of culturally determined principles:

(1) Every member of the community knows – to a greater or lesser degree – all, or most other members.
(2) Members of the community trust each other – except perhaps in a few cases.
(3) Sincerely telling (what one believes to be) the truth is a basic principle of behaviour in the community.
(4) Whenever someone cannot say that something has been directly experienced, s/he will say that what is being told is indirect experience, inference, etc.

In Uralistics and language typology it is a well-known fact that the Nganasan language (and languages closely related to it, such as Nenets and Enets) has a complex evidential system with the following basic distinctions:

direct evidential:	unmarked indicative (the speaker perceives the event)
indirect evidentials:	inferential
	sensitive (the speaker perceives by hearing, touching or smelling)
	narrative/reportative (hearsay or reported by other poeple)

These categories are inflectional and expressed by suffixes.

This kind of evidentiality is most typical in South America (e.g. Tucanoan languages) and among the languages of Papua New Guinea, but it has never been investigated in the Arctic. We argue that maintaining the culturally determined viewpoint is the crucial point in the development of evidential system, and its effect on other parts of the grammar that cover the means of perspectivization. "Perspectivization is a mode of perception and representation constitutes a basic disposition of the human mind, the relationship between perspectivization and narrativity is particularly close as narratives are characterized by the fact that a

narrator functions as an instance of mediacy which refracts the story" (Zeman 2016: 2). The most frequent approach refers to perspectivization in terms of the focalization models and the question 'who sees' vs. 'who perceives'. Perspectivization means how the information is connected to concrete or abstract cognisant entities other than the speaker, while subjectification is about how the information in an utterance is connected to the speaker by the use of expressive predicates (Sanders and Spooren 1997). Evidentials are the typical means of perspectivization that can combine perspectivization and subjectification:

- First, they refer to a subject of consciousness other than the present speaker, to whom the responsibility for the utterance must be attributed (perspectivization).
- Evidentials are perspectivizing in that they present an event from the point of view of an evidential origo (the perspective holder).
- Second, they express some degree of certainty with respect to the utterance in the current speaker (subjectification).

Means of perspectivizing can be reported speech constructions, evidentials, deictic elements (e.g. spatial, temporal), lexically coded elements (e.g. antonyms: come/go, give/get etc.), passivization (derivational suffix), and deontic modals.

Reported speech is the most explicit form of perspectivization, because it inherently involves perspectivization: with the subject of consciousness positioned in the speaker of the embedded rather than the matrix utterance. In direct speech, the embedded speaker is made responsible for the form as well as the propositional content of the quoted utterance, in indirect speech s/he only accounts for the latter (Tátrai and Csontos 2009). In Nganasan only direct speech constructions exist without shifts. Principally, in reported speech the following shifts seem to be possible on the morphological level:

	shift (in Nganasan)
person	no
tense	no ("perfect aorist > deictic past" shift seems to be possible but more data are requied)
evidential:	rarely occurs in DS (by corpus data)
sensory	never
reportative	in special functions (e.g. as quotative)
inferential	in special functions (e.g. as mirative)
epistemic moods:	typical in direct quotation, never occur together with evidentials in the same clause

These features confirm Bernárdez's hypothesis on evidentiality. The current speaker is always the perspective holder who is not responsible for the utterance of

the original speaker. The current speaker can determine the type of source of information from his own point of view. Epistemic moods and lexemes in the speech of the actual speaker occur always verbatim as the part of the quoted sentence. Nganasan fits the cultural interpretation of evidentiality (Bernárdez 2017), where evidentiality plays a fundamental role in perspectivization. The first results show that there is a strong correlation among the investigated features. A Nganasan speaker almost always marks his/her "responsibility" for the truth of an utterance. This kind of responsibility mirrors in occurrence of epistemic and evidential elements in the speech of current vs. original speakers. With terms of functional cognitive pragmatics – the move of the referential center is the crucial point in the organization of a narrative and discourse. The current speaker marks his/her own perspective (evidential origo). The shiftless deictic elements serve this rigid system: the speaker keeps the referential center or moves it totally to the embedded speaker.

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2011. Speech reports: A cross-linguistic perspective. In: Alexandra Aikhnevald and Robert Malcolm Dixon (eds.), *Language at large*. Leiden and Boston: Brill, 290–326.

Bernárdez, Enrique 2017. Evidentiality – a cultural interpretation. In: Farzad Sharifian (ed.), *Advances in cultural linguistics*. Singapore: Springer, 433–460.

Bíró, Bernadett, Sipőcz, Katalin, and Szeverényi, Sándor 2019. Nyugat-szibériai uráli nyelvek térbeli orientációs rendszere [The spatial orientation system of Western Siberian Uralic languages]. *Folia Uralica Debreceniensia* 26: 29–44.

Bíró, Bernadett and Sipőcz, Katalin 2017. Ditransitivity in the Mansi language from a typological perspective. *Finnisch-ugrische Mitteilungen* 41: 25–43.

Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor, and Wagner-Nagy, Beáta 2016. *Nganasan Spoken Language Corpus (NSLC): Archived in Hamburger Zentrum für Sprachkorpora.* Version 0.1.

Brykina, Maria, Gusev, Valentin, Szeverényi, Sándor and Wagner-Nagy, Beáta 2018. *"Nganasan Spoken Language Corpus (NSLC)." Archived in Hamburger Zentrum für Sprachkorpora.* Version 0.2. http://hdl.handle.net/11022/0000-0007-C6F2-8.

Busser, Rik De 2015. The influence of social, cultural, and natural factors on language structure. In: Rik De Busser and Randy J. LaPolla (eds.), *Language structure and environment.* Amsterdam and Philadelphia: John Benjamins Publishing, 1–28.

Csepregi, Márta 2014. Evidencialitás a hanti nyelvjárásokban In: Havas, Ferenc, Horváth, Katalin, Kugler, Nóra and Vladár, Zsuzsa (eds.), *Nyelvben a világ. Tanulmányok Ladányi Mária tiszteletére* [The world in language: Studies in honor of Mária Ladányi]. Segédkönyvek a magyar nyelvészet tanulmányozásához 160. Budapest: Tinta Könyvkiadó, 99–109.

De Groot, Casper (ed.) 2017. *Uralic essive and the expression of impermanent state*. Amsterdam: John Benjamins.

Deutscher, Guy 2010. *Through the language glass: Why the world looks different in other languages*. New York: Picador.

Duranti, Alessandro 1997. *Linguistic anthropology*. Cambridge: Cambridge University Press.

Duranti, Alessandro (ed.) 2004. *A companion to linguistic anthropology*. Oxford: Blackwell.

É. Kiss Katalin (ed.) 2014. *Magyar generatív történeti mondattan* [A generative historical syntax of Hungarian]. Budapest: Akadémiai Kiadó

Ekberg, Lena and Carita Paradis. 2009. Evidentiality in language and cognition. Editorial paper to special issue. *Functions in Language* 16(1): 5–7.

Enfield, Nick J. (ed.) 2002. *Ethnosyntax: Explorations in grammar and culture*. Oxford: Oxford University Press.

Foley, William 1997. *Anthropological linguistics: An introduction*. Oxford: Blackwell.

Gladkova, Anna Nikolaeva 2015. Ethnosyntax. In: Sharifian, Farzad (ed.), *The Routledge handbook of language and culture*. Abingdon Oxon UK: Routledge, 33–50.

Gulya, János 1970. Az etnolingvisztika magyar előfutárai [Hungarian precursors of ethnolinguistics]. In: Szathmári, István (ed.), Tanulmányok a magyar és finnugor nyelvtudomány történetéből [Studies on the history of Hungarian and Finno-Ugric linguistics] Budapest: Tankönyvkiadó, 131–135.

Gusev, Valentin 2007. Эвиденциальность в нганасанском языке. In: *Эвиденциальность в языках Европы и Азии. Сборник статей памяти Н. А. Козинцевой*. СПб.: Наука, 415–444.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. New York: Oxford University Press.

Kálmán, Béla 1989. *Chrestomathia Vogulica*. Budapest: Tankönyvkiadó.

Kulonen, Ulla-Maija 1985. A perspektíva és a vogul szenvedő szerkezet [Perspective and the Mansi passive construction]. *Nyelvtudományi Közlemények* 87: 217–224.

Kulonen, Ulla-Maija 1989. *The passive in Ob-Ugrian* [SUST 203]. Helsinki: Suomalais-Ugrilainen Seura.

Li, Charles N. 1986. Direct and indirect speech: A functional study. In: Florian Coulmas (ed.), *Direct and indirect speech.* Trends in linguistics studies and monographs 31, New York: Mouton de Gruyter, 29–45.

Miestamo, Matti, Tamm, Anne and Wagner-Nagy, Beáta (eds.) 2015. *Negation in Uralic languages*. Amsterdam and Philadelphia: John Benjamins.

Munkácsi, Bernát 1894. *A vogul nyelvjárások szóragozásukban ismertetve* [Inflection in Vogul dialects]. Budapest: MTA.

Newman, John 2002. Culture, cognition, and the grammar of 'give' clauses. In: N. J. Enfield (ed.). *Ethnosyntax: Explorations in grammar and culture*. Oxford: Oxford University Press.

Riese, Timothy 2001. *Vogul*. Languages of the World/Materials 158. Muenchen: Lincom Europa.

Riley, Philip 2007. *Language, culture and identity: An ethnolinguistic perspective*. London and New York: Continuum.

Rombandeeva, Evdokia Ivanovna, and Vahruseva, M. P. [Ромбандеева, Евдокия Ивановна – Вахрушева, М. П.] 1989. *Mansijskij jazyk* [Мансийский язык]. Ленинград: Просвещение Ленинградское отделение.

Rombandeeva, Evdokia Ivanovna [Ромбандеева, Евдокия Ивановна] 1973. *Mansijskij (vogulskij) jazyk* [Мансийский (вогульский) язык.] Москва: Издательство «Наука», 65–79.

Sanders, José and Wilbert Spooren 1997. Perspective, subjectivity, and modality from a cognitive linguistic point of view. In: Liebert, Wolf-Andreas, Gisela Redeker and Linda R. Waugh (eds.), *Discourse and cerspective in cognitive linguistics*. Amsterdam and Philadelphia: John Benjamins, 85–112.

Szeverényi, Sándor 2017. Egy nganaszan episztenciális finit ige lexikalizációja [The lexicalization of a Nganasan epistential finite verb]. In: Forgács, Tamás, Németh, Miklós and Sinkovics, Balázs (eds.), *A nyelvtörténeti kutatások újabb eredményei* IX [New findings in historical linguistics, vol. 9]. Szeged: SZTE, 303–317.

Szeverényi, Sándor and Sipőcz, Katalin 2019. Szibériai uráli nyelvek idéző szerkezetei [Quotative structures in Uralic languages in Siberia]. In: É. Kiss, Katalin, Hegedűs, Attila, and Pintér, Lilla (eds.) *Nyelvelmélet és kontaktológia* 4 [Linguistic theory and contact linguistics, vol. 4]. Budapest and Piliscsaba: PPKE BTK. 127–146.

Tátrai, Szilárd and Csontos, Nóra 2009. Perspectivization and modes of quoting in Hungarian. *Acta Linguistica Hungarica* 56(4): 441–468.

Wierzbicka, Anna 1979. Ethno-syntax and the philosophy of grammar. *Studies in Language* 3(3): 313–383.

Zeman, Sonja 2016. Introduction. In: Igl, Natalia and Zeman, Sonja (eds.) 2016. *Perspectives on narrativity and narrative perspectivization.* Amsterdam and Philadelphia: John Benjamins. 1–14. https://doi.org/10.1075/lal.21

Studia uralo-altaica

No. 1   Róna-Tas, A. – Fodor, S. Epigraphica Bulgarica. (1973)

No. 2   Die erste sölkupische Fibel aus dem Jahre 1879. Eingeleitet von P. Hajdú. (1973)

No. 3   Новицкий, Гр. Краткое описание о народе остяцком (1715). (1973)

No. 4   Paasonen, H. Tschuwaschisches Wörterverzeichnis. Eingeleitet von A. Róna-Tas. (1974)

No. 5   A. Molnár, F. On the History of Word-Final Vowels in the Permian Languages. (1974)

No. 6   Hajdú, P. Samojedologische Schriften. (1975)

No. 7   N. Witsens Berichte über die uralischen Völker. Aus dem Niederländischen ins Deutsche übersetzt von T. Mikola. (1975)

No. 8   Ph. Johann von Strahlenberg Das Nord und Ostliche Theil von Europa und Asia. With an Introduction by I. R. Kruger. (1975)

No. 9   Kiss, J. Studien zur Wortbildung und Etymologie der finnisch-ugrischen Sprachen. (1976)

No. 10   U. Kőhalmi, K. Chrestomathia Sibirica. Auswahl aus der Volksdichtung der sibirischen Urvölker. (1977)

No. 11   Mikola, T. Materialien zur wotjakischen Etymologie. (1977)

No. 12   Попова, Я. Н. Ненецко-русский словарь. Лесное наречие. (1978)

No. 13   Tardy, L. Beyond the Ottoman Empire. (1978)

No. 14   Clauson, G. An Etymological Dictionary of Pre-Thirteenth-Century Turkish, Index. Vol. 1, With a Preface by A. Róna-Tas. (1981)

No. 15   Clauson, G. An Etymological Dictionary of Pre-Thirteenth-Century Turkish, Index. Vol. 2. (1982)

No. 16   Honti, L. Nordostjakisches Wörterverzeichnis. (1982) No. 17 Studies in Chuvash Etymology. 1. Edited by A. Róna-Tas. (1982) No. 18 Domokos, P. Handbuch der uralischen Literaturen. (1982)

No. 19   M. Korchmáros, V. Definiteness as Semantic Content and its Realization in Grammatical Form. (1983)

No. 20   Tardy, L. Sklavenhandel in der Tartarei. (1983)358

No. 21   Berta, Á. Die russischen Lehnwörter in der Mundart der getauften Tataren. [Tatarische etymologische Studien I.] (1983)

No. 22   Helimski, E. The Language of the First Selkup Books. (1983)

No. 23   Pusztay, J. Die Pur-Mundart des Waldjurakischen. Grammatikalischer Abriss aufgrund der Materialien von T. V. Lehtisalo. (1984)

No. 24   Radics, K. Typology and Historical Linguistics. Affixed Person-Marking Paradigms. (1985)

No. 25   Róna-Tas, A. Language and History. Contributions to Comparative Altaistics. (1986)

No. 26   Keresztes, L. Geschichte des mordwinischen Konsonantismus 2. Etymologisches Belegmaterial. (1986)

No. 27   Keresztes, L. Geschichte des mordwinischen Konsonantismus 1. (1987)

No. 28   Papers on Derivation in Uralic. Szegeder und Turkuer Beiträge zur uralischen Derivation. Edited by J. Vastag (1987).

No. 29   Schulze, B. Der Wortparallelismus als ein Stilmittel der (nord-) ostjakischen Volksdichtung. (1988)

No. 30   Tatarische etymologische Studien II. Edited by Árpád Berta (1988)

No. 31   Berta, Á. Lautgeschichte der tatarischen Dialekte. (1989)

No. 32   Zimonyi, I. The Origins of the Volga Bulgars. (1990)

No. 33   Róna-Tas, A. An Introduction to Turkology. (1991)

No. 34   Bereczki, G. Grundzüge der tscheremissischen Sprachgeschichte 2. (1992)

No. 35   Bereczki, G. Grundzüge der tscheremissischen Sprachgeschichte 1. (1994)

No. 36   Mikola, T. Morphologisches Wörterbuch des Enzischen. (1995)

No. 37   Doerfer, G. Formen der älteren türkischen Lyrik. (1996)

No. 38   Kakuk, Zs. Mischärtatarische Texte mit Wörterverzeichnis. (1996)

No. 39   Historical and linguistic interaction between Inner-Asia and Europe, Proceedings of the 39th Permanent International Altaistic Conference [PIAC], Szeged, Hungary: June 16–21, 1996. Edited by Árpád Berta. (1997)

No. 40   Rybatzki, V. Die Tońuquq-Inschrift. (1997)359

No. 41   Helimski, E. Die matorische Sprache. Wörterverzeichnis − Grundzüge der Grammatik − Sprachgeschichte, unter Mitarbeit von Beáta Nagy. (1997)

No. 42   Месарош, Э. Словообразовательные суффиксы глагола в эрзянском языке. (1999) No. 43 Wagner-Nagy, B. Die Wortbildung im Nganasanischen. (2001) No. 44 Ivanics, M. – Usmanov, M. A. Das Buch der Dschingis-Legende. Däftär-i Čingiz-namä 1. (2002) No. 45 Mikola, T. Studien zur Geschichte der samojedischen Sprachen, Aus dem Nachlass herausgegeben von Beáta Wagner-Nagy. (2004)

No. 46   Körtvély, E. Verb conjugation in Tundra Nenets. (2005) No. 47 Turcology in Turkey. Selected papers. Edited by László Károly, papers selected by Nurettin Demir and Emine Yılmaz. (2007) No. 48 Rassadin, V. I. Soyotica. Edited by Béla Kempf. (2010) No. 49 The Szeged Conference. Proceedings of the 15th International Conference on Turkish Linguistics held on August 20–22, 2010 in Szeged. Edited by Kincses-Nagy, É. & Biacsi, M. (2012)

No. 50   Kakuk, Zs. Krimtatarisches Wörterverzeichnis. Aufgrund der Sammlung von Ignác Kúnos zusammengestellt. (2012)

No. 51.   Uralic and Siberian Lexicology and Lexicography. Edited by Sándor Szeverényi and Bayarma Khabtagaeva. (2017)

No. 52.   Ideas behind symbols – languages behind scripts. Proceedings of the 60th Meeting of the Permanent International Altaistic Conference (PIAC) August 27 – September 1, 2017 Székesfehérvár, Hungary. Edited by Ákos Bertalan Apatóczky (2018)

No. 53. Competing Narratives between Nomadic People and their Sedentary Neighbours Papers of the 7th International Conference on the Medieval History of the Eurasian Steppe Nov. 9–12, 2018 Shanghai University, China Edited by Chen Hao (2019)