FINNISCH-UGRISCHE FORSCHUNGEN

Inhalt

Band 66

Besprechungen

Berichte und Nekrologe

Jeremy Bradley & Christian Pischlöger
University of Vienna   University of Vienna

Converb constructions in Mari and Udmurt: Russian loanwords as a metric of productivity

Uralic languages of the Volga-Kama Region, especially Mari and Udmurt, show strong Turkic influence in the range of usages of converbial (gerundial, i.e. adverbial non-finite) forms. Converbs can be found in combination with syntactically superordinate verbs communicating different values, mirroring Turkic structures: modal ("swimming know" = 'know how to swim'), directional ("crawling leave" = 'crawl away'), benefactive ("baking give" = 'bake something for someone'), aspectual ("drinking send" = 'drink up'). It is debatable however to what extent one can speak of grammaticalized structures and to what extent one should speak of a body of loan translations in individual languages or varieties. The paper explores the prospect of using verbs borrowed from Russian as a metric of productivity: as these were borrowed after the phase of intense Turkic language contacts ended, their usage in Turkic-type structures can been seen as evidence for their grammaticalization, while their absence in such structures can be seen as evidence against it.

1. Introduction

The dividing line between a body of *loan translations* or *calques* – i.e. "word[s] or phrase[s] constructed by taking a foreign word or phrase as a model and translating it morpheme-by-morpheme" (Trask 1996: 21) – and grammatical structures resulting from contact-induced *grammaticalization* – i.e. "a shift from a more lexical meaning to more grammatical

content" (Campbell 2013: 297) – can be fuzzy and difficult to delimit in intense contact situations. At what point can the mirroring of grammatical structures from Language A in Language B no longer be considered just that, and must instead be considered a grammatical complex with a life of its own?

Diachronic changes in the contact situation can offer one metric in this regard: if the range of usage situations of said complex in Language B grows after language contacts with Language A have waned, this can be seen as evidence for the construction possessing a grammaticalized status. Particle verbs in Estonian serve as an excellent illustration of this principle: in these oftentimes highly idiomatic structures clearly calqued from German, a separable particle precedes a verb in the infinitive (e.g. German *vorwerfen* 'accuse (lit. ahead throw)' > Estonian *ette heitma* id.) but will jump to a later position in the sentence according to German (and due to language contacts, Estonian) word-order rules in inflection, as shown in (1).

Estonian[1]

(1) *Ma   heida-n   su-lle   ette,   et*   […]
      1SG   throw-1SG   2SG-ALL   ahead   that
      'I blame you (lit. throw ahead of you) for […]'
      (German: 'Ich werfe Dir vor, dass […]')

In his 1990 survey of these structures, Cornelius Hasselblatt identified 1679 distinct calques of German particle verbs, but also 164 particle verbs of dubious origin, and 951 particle verbs that cannot be traced back to German roots (Hasselblatt 1990: 205). In the decades since this monograph was published, additional particle verbs have entered usage in Estonian in connection with the digital revolution. Even when these are Estonian neologisms or loan translations from English, their usage mirrors the German structure, as shown in (2).

Estonian

(2) *Ma   laadi-n   fail-i   alla.*
      1SG   load-1SG   file-GEN   down
      'I'm downloading the file.' (German: 'Ich lade die Datei herunter.')

---

1. When not indicated otherwise, example sentences were provided by our native speaker informants; see Acknowledgements.

Irrespective of the low importance of German as a contact language in present-day Estonia, this grammatical structure triggered by German influence is alive and well in contemporary Estonian. It has been fully and thoroughly grammaticalized, as illustrated by usages that cannot possibly be German loan translations.

The paper at hand pertains to a similar conundrum that has not yet been answered as unambiguously: the status of Turkic-type converb constructions (see Section 4) in the Uralic languages of the Volga-Kama Region (see Section 3). A wide range of verb pairs consisting of a converb (gerund) and a superordinate verb can be found in Uralic languages of the region that clearly mirror their Turkic counterparts, e.g. (Southern) Udmurt *lobǯisa koškini*[2] 'fly away (flying leave)', Mari *ćoŋešten lektaš* id., Erzya *livťáź tujems* id., cf. Tatar *oçıp çıġu* id., Chuvash *vĕśse tuχ* id. (Isanbaev 1978: 66–67; Bereczki 1984: 312). Turkic-type verb pairs clearly have a stronger position in Mari (where they can be found in all varieties, cf. Bradley 2016a) than they do in Udmurt (where they are primarily documented in Southern varieties, cf. Horváth 2013: 121), and a stronger position in Udmurt than they do in Mordvin (where there are only sporadic accounts of such constructions, cf. Bereczki 1984: 312). In this survey, we aim to measure the vitality of different Turkic-type structures in two Uralic languages by using Russian loanwords as a metric (cf. also Sibatrova 2015): since Turkic contacts predate strong Russian contacts in the region, Russian loanwords used in Turkic-type verb pairs are an argument for the grammaticalization of the structure at hand, as they illustrate its usage with lexical material that was not present when the calquing of Turkic structures occurred. Section 5 will introduce the manner in which Russian verbs are borrowed into the languages under consideration; Section 6 will subsequently show how data on the usage of Russian verbs in Turkic-type converb constructions could be extracted from the nascent corpora of

2. Cyrillic language data in this paper was transcribed using the COPIUS Transcription & orthography toolset found at https://www.copius.eu/ortho.php. The Uralic Phonetic Alphabet (UPA) was used for Uralic languages and also for Russian, as Russian language data is only presented within the context of its borrowing into Uralic languages. For Tatar and Bashkir, the standardized Latin orthographies were used. The orthographic rendering of example sentences was homogenized where sources use conventions that differ from the standard Cyrillic or UPA rendering of language data.

Uralic languages of Russia currently at our disposal. Section 7 contains the data extracted from the corpora and Section 8 then reviews the implications this data has on estimations of productivity of individual structures in the respective languages.

2. Data sources

In addition to existing literature and consultations with native speakers, our survey is based on currently available nascent corpus infrastructures with rudimentary morphological annotation (which is mainly not needed for our queries, see below) for Mari and Udmurt. These allow us to examine the usage of Russian verbal borrowings in contemporary written language, both in literary sources and on social media. Our main sources were the Corpora of Uralic Volga-Kama Languages developed by Timofey Arkhangelskiy and his colleagues (Arkhangelskiy 2019a) which include literary and social-media corpora for both Udmurt and (Meadow) Mari and are summarized in Table 1.

Table 1: Scope of the Corpora of Uralic Volga-Kama Languages at time of survey (July 2021)

	Main	Social media
Mari	Tokens: 5.53 million	Tokens: 3.59 million
Udmurt	Tokens: 9.57 million	Tokens: 2.66 million

Thanks to the comparable size and choice of sources of these resources, the Mari and Udmurt results garnered from this study can be considered comparable. The literary corpora contain texts collected on the Internet from newspaper pages, blogs, Wikipedia, etc., while the social-media corpora contain postings gathered from the social-media platform VK (VKontakte), a service comparable to Facebook that is highly popular in Russia, also among minority communities (cf. Pischlöger 2016).

For Mari, two additional resources will be utilized outside of the direct comparisons:

- The Corpus of Literary Mari compiled by the international Mari Corpus Project workgroup, hosted by the Giellatekno research group at the University of Tromsø. Upon its initial publication in December 2020,

it contained 57.38 million tokens of Meadow Mari texts from different genres representing a century of Mari literacy.

- The totality of verb pairs found in lexical sources on Mari, primarily compiled in the 20th century and/or on the basis of texts from the 20th century. These were gathered during the compilation of the Mari-English dictionary (Riese et al. 2014).

In total, the Mari corpora thus include 69.61 million tokens, while the Udmurt corpora include 6.22 million tokens.

3. Historical background

The Turkic and Uralic languages of the Volga-Kama Region, in the area surrounding the confluence of the Volga and Kama rivers in Tatarstan roughly 750 kilometers east of Moscow, are oftentimes subsumed in the so-called Volga-Kama Sprachbund (e.g. Wintschalek 1993; Helimski 2003: 159) due to ample linguistic convergence between them. As is typical of areas of linguistic convergence, one can differentiate between a core in which convergence is strong and a periphery showing increasingly weak convergence. Tatar (Turkic > Common Turkic > Kipchak), Bashkir (Turkic > Common Turkic > Kipchak), Chuvash (Turkic > Oghur), Mari (Uralic), and Udmurt (Uralic > Permic) are frequently classified as core members, while Komi (Uralic > Permic) and Mordvin (Uralic) are classified as peripheral members (cf. Bradley 2016a: 7–9).

It has long been debated, and remains a matter of contention, when the current linguistic landscape of the Volga-Kama Region took shape. Archaeological evidence shows that the Bolgars, the linguistic ancestors of the Chuvash, migrated into the region from the south in the late 8th/ early 9th century (Róna-Tas 1988: 761; see Agyagási 2019 for a more detailed account), but it is less clear when actual language contacts between Bolgar and the indigenous Uralic languages commenced, with estimates ranging from the 9th century (Räsänen 1923: 94; Isanbaev 1989–1994: I: 28) to the 13th century (Wichmann 1924: 53; Bereczki 1992–1994: I: 16). Bolgar Turkic was supplanted by Kipchak Turkic as the dominant language of the region between the 13th century (Isanbaev 1989–1994: I: 28) and the early 15th century (Bereczki 1992–1994: I: 16).

As a function of these historical processes, linguistic convergence in the region differs not only between languages (with Mari showing more

Turkic features than Udmurt, and Udmurt showing more Turkic features than Mordvin and Komi), but also between different varieties of the languages spoken in the region:

- The Bolgar (Oghur) influence on Mari predated the split into the contemporary dialects. Bolgar influence is thus found in all varieties of Mari (Saarinen 1997a: 195; 1997b: 393); Hill Mari, spoken on the right bank of the Volga in immediate proximity to Chuvashia continued to be under Chuvash influence over the centuries and shows younger Chuvash loan elements (Ivanov 1981: 87). Meanwhile, a stronger influence of the Kipchak languages can be detected in the varieties of Mari spoken on the left bank of the Volga, especially in the varieties spoken by the Eastern Mari diaspora in Bashkortostan (Bereczki 1984: 311; 1992–1994: I: 26; Ilieva 2009: 3–8), which emerged as a result of the eastward migration of Maris in the 17th century (Pomozi 2004: 96).
- Likewise Proto-Permic, the ancestor of Komi and Udmurt, was spoken during the period of Bolgar dominance and was subject to Oghur Turkic influence (Róna-Tas 1988: 760). After the breakup of Proto-Permic, Komi was no longer subject to Turkic influence. In the case of Udmurt, the more recent Tatar influence is stronger in the Southern dialects (Kel′makov 1975: 95), especially the peripheral dialects (Csúcs 1998: 277). A notable exception to this trend is Beserman, a variety spoken in northern Udmurtia. One assumption is that Beserman speakers have a Turkic ethnic background (Róna-Tas 1988: 765).
- In addition to genealogical distance, a Mari substrate is assumed to be one of the reasons for the great structural difference between Chuvash and other Turkic languages (Agyagási 1998: 668). Mari influence is especially strong in Viryal Chuvash dialects (Berta 1998: 451; Johanson 2009a: 245) spoken in northern Chuvashia, in close proximity to Hill Mari.

Since the annexation of the Khanate of Kazan by Muscovy in 1552, the region has been under Russian control; Russian has increasingly become the dominant contact language for all languages under consideration. It is notable however that Russian only became truly dominant in the region in the twentieth century (Kangasmaa-Minn 1998: 220).

Language contacts in the region have long been observed and described on different levels of language: phonetics and phonology (e.g. Johanson 2000b), lexicon (e.g. Räsänen 1920; 1923; Isanbaev 1989–1994; Csúcs 1990;

Saarinen 1997a; 2010), loan translations (e.g. Saarinen 1997b, Hesselbäck 2005), and syntax (Wintschalek 1993). Turkic-type converb constructions (see Section 4 below) have long been noted as an example of structural borrowing in the region (cf. Čxaidze 1960; Keľmakov 1975; Isanbaev 1978; Bartens 1979: 143), though the time frame and origin of their borrowing remains unclear. In Udmurt, the Tatar origin of these structures seems clear. The picture is more complicated in the case of Mari. There these structures have historically been attributed primarily to Kipchak (e.g. Čxaidze 1967: 258), but this claim is problematic as Turkic-type converb constructions are ubiquitous in all varieties of Mari, while Kipchak language contacts in general only have a strong position in the eastern varieties of Mari. Given that Chuvash-type Turkic exerted influence on Proto-Mari before it broke up into its modern dialects and that Chuvash-type loans are evenly spread through all varieties of Mari, it is more likely that the system is of Chuvash origin in Mari. It is however highly likely that the usage of individual auxiliaries in eastern varieties of Mari has been colored by usage of auxiliaries in Kipchak languages at a later point in history.

4. Converb constructions

The morphological category of *converb* is understood here as "a non-finite verb form whose main function is to mark adverbial subordination" (Haspelmath 1995: 4); the term can here be considered equivalent to *gerund* in some linguistic traditions (e.g. Uralic studies) and its translations. *Converb constructions* (i.e. constructions consisting of a converb and some sort of superordinate verbs) are not to be confused with *serial verb constructions*, "a sequence of verbs which act together as a single predicate, without any overt marker of coordination, subordination, or syntactic dependency of any other sort" (Aikhenvald 2006).

Converbs find ample usage throughout the Uralic language family (Ylikoski, forthcoming). Numerous converbial endings are conventionally distinguished in the Uralic and Turkic languages of the Volga-Kama Region (e.g. Mari: 5, Udmurt: 4, Erzya: 3, Tatar: 6, Bashkir: 6, Chuvash: 7; Riese et al. 2019: 8, Csúcs 1998: 293; Zaicz 1998: 205; Landmann 2014a: VI; 2014b: VI; 2015: VI). Some of these, such as the Mari converb of prior action in *-meke*, have well-defined and clearly delimited functions that are unremarkable in a Uralic context, for example as instruments of temporal structuring of statements, see (3).

Mari

(3)   Урокым ыштымеке, Эчан телевизорым ончыш.

    Urok-ə̂m   ə̂štə̂-meke,   Ečan   televizor-ə̂m   ončə́-š.
    lesson-ACC   do-CVB.PRI   Echan   television-ACC   watch-PST1.3SG
    'After doing his homework Echan watched television.'

Other converbs such as the Mari converb in *-n*, the Udmurt converb in *-sa*, and the Tatar converb in *-p* have a wider range of usage.[3] They can be found in a range of situations where their usage is atypical for Uralic outside of this region (see Ylikoski 2004: 380 about isolated occurrences) but are commonplace in Turkic languages, where they are oftentimes referred to as "paired verbs" in the literature (cf. Bradley 2016a: 35–47 for an overview of terminology used in respect to these constructions). They can also be found in a number of Samoyedic languages – Kamas (Klumpp 2002), Selkup (Valijärvi 2008; Harder 2018), Nenets (Tereščenko 1981), Mator (Helimski 1997: 188) – that have likewise been subject to Turkic language contacts. Such a structure is illustrated in (4).

Tatar (Landmann 2014a: 99)

(4)   Агачны кисеп ташладылар.

    Ağaç-nı   kis-ep   taşla-dı-lar.
    tree-ACC   cut-CVB   throw-PST-3PL
    'They cut down the tree.'

Here *taşla-* 'throw' loses its lexical meaning but is used to indicate the abrupt and sudden execution of an action, i.e. it primarily conveys an aspectual meaning. A wide range of verbs are used as auxiliaries in such constructions where they fully or partly lose their lexical meanings in the relevant languages. Similar constructions can be found in genealogically diverse languages over a wide geographical area: Dravidian and Indo-Aryan languages of the Indian subcontinent, Turkic and Mongolic languages of northern Eurasia, Japanese, and Korean (Masica 1976: 141ff.). In fact, the individual auxiliaries often represent typologically common grammaticalization patterns (cf. Kuteva et al. 2019: 437 regarding verbs meaning 'throw' grammaticalized as perfect or completive markers) which has

---

3.   The most versatile converb in a respective language is glossed simply as CVB in this paper, while converbs with a more specific range of usages will be additionally tagged according to their function.

been used as an argument against the Turkic origin of these structures in Uralic (e.g. in Honti 2013). However, the appearance of such systems in Uralic languages and varieties strongly overlap with Turkic contacts and the auxiliaries strongly overlap in both their function and their syntax throughout the region. This leaves little doubt in the Turkic origin of the system in the mainstream view. Consequently, the converb constructions under consideration will be referred to as "Turkic-type" in this paper.

Auxiliaries used in converb constructions can have a range of functional values, which are oftentimes poorly differentiated in the literature:

- Modal (see Section 4.1 and Bradley 2016a: 268)
- Directional (see Section 4.2 and Bradley 2016a: 56–68, 263–264)
- Benefactive (see Section 4.3 and Bradley 2016a: 68–69, 265)
- Aspectual (see Section 4.4 and Bradley 2016a: 47–56, 265–268)

The following subsection will briefly introduce these subtypes and discuss their spread within the Volga-Kama Region. In all constructions under consideration here, Mari utilizes the affirmative instructive converb in *-n* (Riese et al. 2019: 267–270), while Udmurt uses the converb in *-sa* (Winkler 2011: 117–119). For the remainder of this paper, references to "converbs" in Mari and Udmurt refer to these forms, irrespective of the numerous other converbs founds in these languages.

4.1. Modal constructions

It is typical for modal auxiliaries denoting permission, necessity, or ability to govern converbs rather than infinitives in Turkic languages (Johanson 2009b: 498) – i.e. these converbs also function as infinitives of sorts (cf. Ylikoski, forthcoming). This is indeed the case in Kipchak Turkic: Tatar and Bashkir *al-* 'take; be able to' and *bel-* 'know; know how to' co-occur with a converb (Landmann 2014a: 93; 2015: 95), as shown in (5).

Bashkir (Landmann 2015: 95)

(5)   Һеҙ йөзә беләһеҙме?

Heź   yöz-ä   bel-ä-hegeź=me?
2PL   swim-CVB.MOD   know-PRS-2PL=INT
'Do you know to swim?'

Mari mirrors this, with *kertaš* 'be able to' and *moštaš* 'know to' both co-occurring with a converb (Riese et al. 2017: 173), illustrated in (6).

Mari

(6) Мый марла лудын ом мошто.

    Mə̂j   *mar-la*   *lud-ə̂n*   *o-m*   *mosto.*
    1SG   Mari-MOD   read-CVB   NEG-1SG   be_able.CNG

    'I can't read Mari.'

The auxiliary *kertaš* can also be found in combination with the infinitive. This usage is marked as non-standard in lexical resources and is associated with a slightly different meaning: 'be able to; carry out some kind of activity well; expert of' (Galkin et al. 1990–2005 s.v. *керташ*). In the Corpus of Literary Mari, co-occurrences with the converb vastly outnumber co-occurrences with the infinitive: *kertaš*[14] is immediately preceded by a converb in 42,964 cases, but only 199 times by an infinitive.[5]

    In other Uralic languages of the region, comparable modal auxiliaries govern the infinitive: Udmurt *bigatı̮ni* 'be able to' (Kirillova et al. 2008 s.v. *быгатыны*), Komi *vermı̮ni* (Beznosikova et al. 2000 s.v. *вермыны*), Erzya *maštoms* (Aasmäe 2012: 32). Interestingly, Chuvash *pultar* 'be able to' governs the infinitive (Landmann 2014b: 74). This is of especial note as the Turkic influence found in all varieties of Mari tends to be of the Oghur/Chuvash type, with Hill Mari showing little Kipchak Turkic influence. Yet, Hill Mari *kerdäš* 'be able to' couples with the converb (Krasnova et al. 2017: 163–164), rendering the classification of this pattern as borrowed from Turkic problematic and in need of further investigation.

4.2. Directional constructions

Following Leonard Talmy's (1985; 2007) typological classification of motion events, the Turkic languages show a strong propensity towards so-called verb-framed structures in which the *manner* of a movement (e.g. going, running, swimming, crawling, flying) can be expressed by a converb

---

4. The superscript Roman numeral indicates whether a verb belongs to the first or second conjugation in cases when the infinitive form is identical to that of a verb belonging to the other conjugation class.

5. This search pattern only found affirmative clauses with no deviations from the pragmatically neutral word order as in other cases, the infinitive or converb would not immediately precede the auxiliary verb. It is however a sufficient search pattern to determine the relative frequency of the converb and infinitive in combination with the auxiliary.

while the syntactically superordinate verb expresses the *path* (e.g. in, out, away, up, down) (cf. Slobin 2000: 109), as shown in (7).

Chuvash (Skvorcov & Skvorcova 2002 s.v. вылететь)

(7) Ҫĕкеҫ йăвинчен вĕҫсе тухрĕ.

    Čĕkeś   *jăv-in-ćen*   *vĕś-se*   *tuχ-r-ĕ.*
    swallow   nest-3SG-ABL   fly-CVB   exit-PST-3SG
    'The swallow flew out of (lit. flying exited) its nest.'

These structures are uncommon in Uralic, but they are the default manner of verbalizing motion events in Mari and can be sporadically found in other Uralic languages or varieties that have been in contact with Turkic languages (Bradley 2016b). Individual examples can be found in (presumably Southern) Udmurt, Beserman, and Erzya, as shown in (8–10).

Udmurt (Bereczki 1984: 312)

(8) лобӟыса кошкыны

    lobʒi̯-sa   *koški̯-ni̯*[6]
    fly-CVB   leave-INF
    'fly away (flying leave)'

Beserman Udmurt (Serdobol'skaja et al. 2012)

(9) Ву вылтӥ пичи пи уяса ваське.

    [V]u   *vəl-ti*   *piči*   *pi*   *uja-sa*   *[v]ašk-e.*
    water   surface-PROL   small   boy   swim-CVB   descend-3SG
    'Down the river swims a little boy.'

Erzya (Bereczki 1984: 312)

(10) Варака ливтязь тусь вирев.

    Varaka   *livťa-ź*   *tu-ś*   *viŕ-ev.*
    crow   fly-CVB   leave-PST1.3SG   forest-LAT
    'The bird flew away to the forest.'

---

6. Our transcriptions are based on literary Udmurt and on occasion standardized accordingly.

4.3. Benefactive constructions

Turkic languages make use of a globally common strategy (Kuteva et al. 2019: 192–194) where a verb meaning 'give' is used as an auxiliary indicating the presence of a beneficiary, i.e. "a participant that is advantageously affected by an event without being its obligatory participant (either agent or primary target, i.e. patient)" (Kittilä & Zúñiga 2010: 2), as shown in (11).

Chuvash

(11) Иван Верăна юрласа пачĕ.

Ivan Veră-na jurla-sa paḉ-ĕ.
Ivan Vera-DAT sing-CVB give-PST.3SG
'Ivan sang for Vera.'

This strategy is employed in Mari and in southern varieties of Udmurt (Bradley et al. 2019: 26), as shown in (12–13). Usage of *puaš* 'give' as a benefactive auxiliary has been described as uncommon in comparison to Turkic languages (Serebrennikov 1960: 198; Isanbaev 1978: 84).

Mari (Galkin et al. 1990–2005 s.v. *чыштыраш*)

(12) Ачай, пычалым налын пу.

Aḉa-j, pǝḉal-ǝm nal-ǝn pu.
father-VOC rifle-ACC buy-CVB give.IMP.2SG
'Daddy, buy me (lit. buying give) a rifle.'

Southern Udmurt (Kelʹmakov 1975: 102)

(13) кырӟаса сётыны

ki̯rʑa-sa śoti̯-ni̯
sing-CVB give-INF
'sing (for someone)'

In an autobenefactive construction, the agent in a clause also serves as the beneficiary (Creissels 2010: 2). Some, but by no means all, languages that use 'give' as a benefactive marker also use 'take' as an autobenefactive marker (Kittilä & Zúñiga 2010: 2). Autobenefactive 'take' is attested for Tatar, Chuvash (see 14), Mari, and Udmurt (Kelʹmakov 1975: 102; Isanbaev 1978: 83), though the interpretation of verb pairs of this type is difficult as 'take' verbs also serve as aspectual auxiliaries (see Section 4.4). In the case

of Mari, pairings with *nalaš* 'take' have been described as less common than their Turkic counterparts (Isanbaev 1978: 84).

Chuvash (Landmann 2014b: 90)

(14)   Сирён адреса çырса илтём.

    Sirĕn   *adres-a*   *śịr-sa*   *il-t-ĕm.*
    2PL.GEN   address-ACC   write-CVB   take-PST-1SG
    'I wrote down your address (for myself).'

Furthermore, in Tatar, Chuvash, and Mari, verbs meaning 'show' can be used to indicate that an action is carried out for illustrative purposes, similar to the verbal prefix *vor-* in German (e.g. *singen* 'sing' → *vorsingen* 'sing (for someone/an audience)'): Tatar *uqıp kürsät-* 'read (for someone/an audience) (lit. reading show)', Chuvash *vulasa kătart* id., Mari *ludân ońćâktaš* id. (Bradley et al. 2019: 28)). There are not currently any indications of this structure being used in Mordvin or Permic.

4.4. Aspectual constructions

The probably most-studied converb constructions of the Volga-Kama Region are constructions in which a converb is coupled with a syntactically superordinate verb that fully or partially loses its lexical meaning and primarily communicates an aspectual (or rather, *Aktionsart*) value. These pairings have been extensively studied in Tatar (Schönig 1984), Bashkir (Graščenkov 2012), Chuvash (Lebedev 2016), Mari (Čxaidze 1960; Isanbaev 1978; Bradley 2016a) and especially Hill Mari (Kashkin 2017; 2018a; 2018b; 2018c; 2019; 2020; Kashkin & Dyachkov 2018), and (Southern) Udmurt (Kel'makov 1975; Horváth 2011; 2012; 2013). In the following example, the Mari verb *šândaš* 'put, place' loses its lexical meaning completely and rather turns the atelic, static, imperfective act of 'loving' into a telic, transformative, perfective act of 'falling in love', as shown in (15).

Mari

(15)   Эчан Эвикам икымше ончалтыш гыч йöратен шынден.

    Ećan Eβika-m   *ik-âmše*   *onćaltôš*   *gôć*   *jörat-en*   *šând-en.*
    Ećan Evika-ACC   one-ORD   look   from   love-CVB   put-PST2.3SG
    'Ećan fell in love (lit. loving placed) with Evika at first sight.'

A closed set of a few dozen verbs can be found in the second position of such pairings serving as aspectual *auxiliaries* (in some nomenclatures *light verbs*, cf. Butt 2010; cf. Bradley 2016a: 45–47 for a discussion on the terminology). These verbs partially or completely lose their lexical meaning in these constructions and primarily or exclusively communicate an aspectual value; the converb communicates the lexical value in these pairings. The exact number of prospective auxiliaries is subject to debate. For Mari, we estimate that 45 different verbs can occur as aspectual auxiliaries (Bradley 2016a: 276–278). Verbs with the same lexical meanings generally communicate comparable aspectual values in the different languages of the region, though some variance between languages and varieties can be observed. Auxiliaries expressing the following aspectual values are grammaticalized from verbs with the following lexical meanings (Bradley 2016a: 265–268):

- **Ingressive (inchoative, inceptive) / initial-transformative auxiliaries** that "indicate the beginning of a situation" (Comrie 1998: 19): 'descend', 'go (away)', 'let go; send', 'become', 'look'.
- **Resultative auxiliaries** that indicate "the successful completion of a situation" (Comrie 1998: 20): 'lie down; fall', 'stay', 'sit down', 'stand up', 'leave something', 'put, place, stand', 'give', 'throw', 'stop'.
- **Exhaustive auxiliaries** where "the specified action is realized to the fullest possible extent and thus also extends to all possible subjects and objects, whether or not they are indicated" (Schönig 1984: 55–56): '(come to an) end', 'arrive; reach', 'become superfluous', 'take out', 'finish something', 'win', 'fill something up', 'strike', 'provide', 'be filled', 'manage'.
- **Delimitative auxiliaries** indicating a perfective situation occurs over a defined period of time (Comrie 1998: 22): 'take', 'carry out', 'leave'.
- **Durative auxiliaries** referring "to the fact that the given situation lasts for a certain period of time (or at least, is conceived of as lasting for a certain period of time)" (Comrie 1998: 41): 'live', 'lie', 'go/wander', 'sit', 'stand'.
- **Continuative auxiliaries** indicating "an event continuing to happen" (Binnick 1991: 146): 'endure'.
- **Iterative (frequentative) auxiliaries** denoting "the repetition of a situation, the successive occurrence of several instances of the given situation" (Comrie 1998: 27): 'set', 'turn', 'sit around', 'stand around'.
- **Gradual (incremental) auxiliaries** indicating "duration or repetition together with transformation" (Vinay & Darbelnet 1995: 78): 'come'.

There are clear differences within a language or variety as regards the frequency of individual auxiliaries in the individual languages. For example, while Mari *koltaš* 'send' is ubiquitous as an ingressive marker and can be found coupled with hundreds of different verbs in converb constructions (Bradley 2016a: 164–170), we have currently, in spite of the large size of our corpus (see Section 2), only found Mari *čĭtaš* 'endure' as a continuative marker in combination with three different verbs, making it highly unlikely that it can be reasonably called a grammaticalized marker. It remains unclear where one should draw the line between clearly grammaticalized markers and markers that only occur in individual idiomatic expressions.

4.5. Ambiguity of interpretation

The interpretation of individual verb pairs can be difficult as auxiliaries can have a broad functional range and as there are no clear syntactic differences between different converb constructions (beyond the fact that pragmatically motivated deviations from the standard word order are only permissible in some types of converb constructions, cf. Bradley 2016a: 73–77 – a fact that could be utilized in empirical research with native speakers, but of limited use in a corpus-based study). For example, Mari *puaš* 'give' (Bradley 2016a: 210) can occur as both a benefactive marker and as an aspectual marker with a resultative value; in fact these two readings are not incompatible with one another and pairings with *puaš* can allow for both a benefactive and an aspectual interpretation, as in (16).

Mari (Galkin et al. 1990–2005 s.v. *пуаш* II)

(16)   Технологлан пӧртым чоҥген пуэна.

> *Teχnolog-lan   pört-ĭm   čoŋ-en   pu-ena.*
> technologist-DAT   house-ACC   build-CVB   give-1PL
> 'We will build a house for the technologist.'

In other situations, though, the context would not permit a benefactive reading. Likewise, Mari *kajaš* 'go (away)' (Bradley 2016a: 149) is used both in directional constructions (indicating the path 'away') and as an aspectual marker with an ingressive value, and here again the two values are not incompatible with one another. In individual sentences, however, one of these values can be incompatible with the context.

Furthermore, the degree of semantic bleaching that auxiliaries experience can be quite variable. In some cases, it is clearly complete (e.g. Mari

jöraten šôndaš 'fall in love (lit. loving place)'), but in others it is debatable. For example, Mari *šińćaš*[II] 'sit', *šogaš* 'stand', and *kijaš* 'lie' are all three used as durative markers coupled primarily with intransitive verbs, but outside of highly idiomatic expressions their functional distribution seems to be determined by the verb's lexical meaning: *ludôn šińćaš*[II] 'lit. reading sit', *ludôn šogaš* 'lit. reading stand', and *ludôn kijaš* 'lit. reading lie' are all durative forms of *ludaš* 'read'; the difference between the pairings is the implied position in which the reading is carried out (Bradley 2016a: 260–263). Claus Schönig poetically refers to this phenomenon as *das Durchschlagen der Vollverbbedeutung* – the verb's lexical meaning breaking through (Schönig 1984: 73).

Consequently, a certain amount of ambiguity should be assumed in respect to verb pairs provided in the overview below.

5. Verbal borrowing strategies

Compared to the relatively straightforward process of nominal borrowings, verbal borrowings have long been known to be generally less common, typical of more intense language contacts, and when the recipient language is morphologically rich, requiring special mechanisms to accommodate them (Arkhangelskiy 2019b; 2020). Søren Wichmann and Jan Wohlgemuth distinguish four basic types of verbal borrowing strategies (Wichmann & Wohlgemuth 2008; Wohlgemuth 2009):

1. The **light verb strategy**, in which the borrowed element is coupled with a semantically light verb (usually 'do') which is inflected while the loan element remains inert (Wichmann & Wohlgemuth 2008: 93–96; Arkhangelskiy 2019b: 527), shown in (17).

Mari (Corpus of Literary Mari)

(17)   Туныктышо-влакым мобилизовать ыште.

    Tunôktôšo-βlak-ôm   *mobilizova-t́*   *ôšte.*
    teacher-PL-ACC   **mobilize-INF**[7]   do.IMP.2SG
    'Mobilize the teachers.' (< Russian *mobilizovat́*)

---

7.   In these glosses, bold indicates Russian verbal morphology.

2. **Indirect insertion**, in which an affix is added to the borrowed element and the resulting verb is then inflected (Wichmann & Wohlgemuth 2008: 97–99; Arkhangelskiy 2019b: 523–527), illustrated in (18).

Mari (Corpus of Literary Mari)

(18)   мобилизоватлыме пӧръеҥ-влак

   mobilizovat-lə̂-me   *pörjeŋ-βlak*
   mobilize-VRB-PTCP.PASS   man-PL
   'mobilized men' (< Russian *mobilizovat´*)

3. **Direct insertion** is when the loanword is inserted into the grammar of the recipient language without any morphological or syntactic accommodation (Wichmann & Wohlgemuth 2008: 99–102; Arkhangelskiy 2019b: 523–527), shown in (19).

Mari (Corpus of Literary Mari)

(19)   Ынде жарымын ӱпшыжымак шижын ом керт.

   Ə̂nde   *žarə̂-mə̂-n*   *üpš-ə̂ž-ə̂m=ak*
   now   fry-PTCP.PASS-GEN   smell-3SG-ACC=EMP

   šiž-ə̂n   *o-m*   *kert.*
   sense-CVB   NEG-1SG   be_able.CNG
   'Now I can't sense the smell of frying.' (< Russian *žarit´*)

4. **Paradigm transfer** is when a borrowed verb is inflected in accordance with the donor language's morphology rather than that of the recipient language (Wichmann & Wohlgemuth 2008: 102–104; Arkhangelskiy 2019b: 527–529), illustrated in (20).

Mari (Gavrilova 2014: 77)

(20)   Тый мо ден преподаёшь вара?

   Tə̂j   *mo*   *den*   **prepodaj-oš**   *βara?*
   2SG   what   with   **teach-2SG**   then
   'So what do you teach?' (< Russian *prepodavat´*, instead of *tunə̂kt-et* teach-2SG)

All four types can be encountered in both Mari (see above) and Udmurt (Arkhangelskiy 2019b), but different strategies receive different weight. As paradigm transfer can be considered a type of code mixing, it is not relevant within the context of the study at hand: one would not expect the

morphosyntax of the recipient language to co-occur in this situation. It will thus be disregarded in the following overview.

5.1. Mari

In literary Meadow Mari, the standard manner in which Russian verbs are integrated into Mari is indirect insertion via the derivational suffix *-l-*, which is in its primary function widely used denominally in Mari, e.g. *negâz* 'foundation' → *negâzlaš* 'found' (Riese et al. 2019: 387). The suffix is attached to the Russian infinitive ending *-tʹ* [8], which loses its palatalization, to form a Mari verb, e.g. Russian *filʹtrovatʹ* 'filter' > Mari *filʹtrovatlaš* id. (See Section 5.3. for analogous usage in Chuvash.) When reflexive, reciprocal, and intransitive Russian verbs with the ending *-śa ~ -ś* are borrowed into Mari, this suffix is replaced with the Mari valency-reducing derivational suffix *-alt-* (Riese et al. 2019: 391–392) which follows the derivational suffix *-l-*: Russian *filʹtrovatśa* 'be filtered' > Mari *filʹtrovatlaltaš* 'be filtered'.

 In Hill Mari, the dominant method is direct insertion: the non-past stem of a Russian verb (which oftentimes differs from the infinitive/past stem) serves as the stem of the Hill Mari verb (Krasnova et al. 2017: 48–49): Russian *filʹtrovatʹ* 'filter' → *filʹtruj-* (Russian non-past stem) > Hill Mari *filʹtrujaš*. As in Meadow Mari, the derivational suffix *-alt-* (~ *-ält-*) is used when Russian verbs with the ending *-śa ~ -ś* are borrowed: *filʹtrujaltaš* 'be filtered'.

 In Eastern varieties of Mari (speakers of which use the Meadow Mari literary norm in writing) subject to greater Turkic influence, *âštaš* 'do' is used in borrowings utilizing the light verb strategy: Russian *agitirovatʹ* 'agitate, campaign' > Eastern Mari *agitirovatʹ âštaš* (Sibatrova 2016). Cases can also be found of *âštaš* being used in combination with Russian verbs in *-śa ~ -ś*, e.g. (21).

Eastern Mari (Arkhangelskiy Social-Media Corpus, user lastochka0610[9])
(21) ала-ко" [sic] весе хуйня ден заниматься ышта, а мый огыл)))

ala-kö	*βese*	*χujńa*	*den*	***zanima-tʹ-śa***
INDF-who	other	bullshit	with	**handle-INF-REFL**

âšt-a,	*a*	*mâj*	*ogâl)))*
do-3SG	but	1SG	NEG

'Somebody else can deal with this bullshit, but not me :)))'

---

8. The rare Russian infinitives not ending in *-tʹ* (e.g. *idti* 'go') can be disregarded here.
9. A native of Kaltasy, Bashkortostan.

In summation, all three relevant strategies can be found in Mari, though the prevalence of strategies depends on speakers' dialectal background. There seems to be a spectrum ranging from direct insertion in the west (Hill Mari) over indirect insertion (Meadow Mari) to the light verb strategy in the east (Eastern Mari). Since the corpus resources currently at our disposal pertain to the Meadow Mari literary standard (used by speakers of Meadow Mari and Eastern Mari), when turning our attention to converb constructions using borrowed Russian lexemes, we will restrict ourselves to examining indirect insertion with *-l-* (including reflexive forms) and the light verb strategy, as direct insertion is not a productive process for speakers of these varieties.

5.2. Udmurt

While all strategies detailed above can be found in Udmurt (Arkhangelskiy 2019b), today two strategies dominate, with their productivity and acceptance among speakers subject to regional variance.

The standard strategy found in all dialects and the literary language is the light verb strategy, using *karịnị* 'do', an Iranian loanword (Holopainen 2019: 380–381). This verb is used as a light verb in combination with a wide range of words belonging to different parts of speech (cf. Tarakanov 2013), such as ideophones, adjectives, or nouns (including loanwords), e.g. *žur karịnị* 'murmur (lit. *žur* do)' *jegit karịnị* 'rejuvenate (lit. young do)', *murt karịnị* 'shun (lit. strange do)', *kịrś karịnị* 'pollute (lit. dirt do)' *keneš karịnị* 'consult (lit. advice do)' (Kirillova et al. 2008 s.v. карыны). It is also used in combination with Russian infinitive forms, e.g. Russian *vlijaťˊ* 'influence' > Udmurt *vlijaťˊ karịnị* id. (Arkhangelskiy 2019b: 527). The extensive usage of light verbs is typical of Turkic languages and also Tatar (Ganiev 1982; Arkhangelskiy 2019b: 546), making it unsurprising that this strategy has an especially strong position in southern varieties of Udmurt (Edygarova 2014: 395; Salánki 2015: 159) that show a stronger Turkic influence and are also those in which Turkic-type converb constructions have been primarily observed. Light verb constructions are however also typical of Iranian languages (Korn 2013), from where the primarily used Udmurt light verb was borrowed, raising the possibility that light verb constructions in Udmurt might predate Turkic contacts – though the seeming absence of these structures from Komi, Udmurt's sister language, complicates the picture. Some variance appears when Russian reflexive

verbs are borrowed. Sometimes the Udmurt passive suffix *-(i)śk-* is attached to the stem of the light verb, while on others it is not (mirroring Tatar, where the light verb *itü* 'do' is used regardless of voice (cf. Arkhangelskiy 2019b: 546)), and in yet others, *luįnį* 'be' is used as a light verb, as shown in (22a–c).

Udmurt (Arkhangelskiy 2019b: 543–544, 547; Cyrillic variants from corpus)

(22) a. Мон но тиледын [sic] фотографироваться карысал.

     *Mon   no   tiĺed-įn   **fotografirova-t́-śa**   karį-sal.*
     1SG   and   2PL-INS   **photograph-INF-REFL** do-COND
     'I'd also like to have a picture taken of me and you.'

   b. МИ НО ВНУЧКАЕНЫМ […] ФОТОГРАФИРОВАТЬСЯ КАРИСЬ-КИМ.

     Mi   no   vnućka-jen-įm   […]
     1PL   and   granddaughter-INS-1SG

     ***fotografirova-t́-śa**   kar-iśk-im.*
     photograph-INF-REFL do-REFL-PST.1PL
     'I also had a picture taken of me together with my granddaughter.'

   c. […] мировой экономикаен специализироваться луэ.

     […] *mirovoj ekonomika-jen **special'iźirova-t́-śa**   lu-e.*
     world   economy-INS   **specialize-INF-REFL** become-PRS.3SG
     '[…] [s/he] specializes in world economy.'

The main competing strategy in modern Udmurt is indirect insertion using the verbal derivational suffix *-t-* (Salánki 2015: 259; Horváth 2018; Arkhangelskiy 2019b: 520) which is added to the Russian infinitive: Russian *žarit́* 'fry' > Udmurt *žarit́įnį* id. This strategy is said to hold an especially strong position in northern varieties of Udmurt (Edygarova 2014: 395; Salánki 2015: 159), though empirical research has shown the opposite in some cases (Arkhangelskiy 2019b: 531).

5.3. Turkic

The three strategies under consideration are employed in verbal borrowings throughout the Turkic language family (Kincses Nagy 2006).

Direct insertion:

  Mongolian *čida-* 'be able, capable' > Tatar *çıda-* id. (ibid. 5)

Indirect insertion:

  Russian *žarit'* 'cook, fry' > Chuvash *šaritle-* id. (ibid. 2)

Light verb strategy:

  French *déchiffrer* 'decipher' > Turkish *deşifre et-* id. (*etmek* 'do') (ibid.)

Notable differences can be observed, however, in the frequency and universal applicability of the strategies. Direct insertion seems to mostly have a strong position in borrowings "between typologically very close, agglutinating languages in contact" (ibid. 6). It is for example employed in the borrowing of Mari verbs into Chuvash (ibid., e.g. Mari *vâlĝậžaš* 'flutter' > Chuvash *vĕlkĕš-* id., Fedotov 1990: 300) but does not seem to be documented in the borrowing of verbs from typologically distant Russian.

Meanwhile, the light verb strategy enjoys an especially strong position as a highly versatile strategy. It has for example been observed in Turkish as spoken in migrant communities in Europe, e.g. Dutch *opruimen* 'clean up' > Netherlands Turkish *opruimen yapmak* id. (*yapmak* 'make') (ibid. 2). This strategy also seems to be the dominant strategy in the borrowing of Russian loanwords into Turkic languages.

6. Russian borrowings in converb constructions: corpus data

6.1. Search queries: Mari

The first point of investigation will be Russian borrowings derived by means of the productive indirect insertion pattern – the default strategy in Meadow Mari – in which the derivational suffix *-l-* is attached to the Russian infinitive, which ends in *-t'* in Russian; Russian *-t'* is substituted with *-t-* in Mari (e.g. Russian *gladit'* 'iron' > Mari *gladitlaš* id.).

The aim in choosing an adequate search query to find a pattern – in this case, Russian borrowings realized as the converb in *-n* in combination with specific superordinate verb – is to establish a pattern that does not miss relevant sentences (i.e. it avoids false negatives) while on the other hand reducing the number of irrelevant sentences (i.e. false positives) in the search output. As the search output can be perused by eye and irrelevant search results removed manually, the first point is of greater importance.

The low amount of allomorphy in Mari morphosyntax alleviates the task at hand. The converb in *-n* (Riese et al. 2019: 267–270) without exception

has the form *-en* after the derivational suffix *-l-*. Consequently, every non-reflexive Russian loanword following the productive indirect insertion pattern in Mari will have the ending *-tlen*. After the reflexive *-alt-*, the converb either has the ending *-ǝn* or a zero ending (the alternation between these two endings is determined by non-trivial factors and cannot be addressed here, cf. Riese et al. 2019: 267), meaning that reflexive borrowings from Russian can either have the ending *-tlaltǝn* or *-tlalt*. A search for all words in a corpus with one of these three endings – *-tlen*, *-tlaltǝn*, *-tlalt* – will yield all converbs of Russian loans following the productive pattern under consideration. Needless to say, numerous false positives must be weeded out from the results, e.g. *süretlen* < *süretlaš* 'draw', derived from the Tatar loanword *süret* 'picture' (Moisio & Saarinen 2008: 648).

Next one must search for the syntactically superordinate verb that should occur in combination with the converb. The process is facilitated by the relatively rigid word order: in affirmative clauses, the superordinate verb almost always immediately follows the converb (Bradley 2016a: 73–75); deviations from this pattern are sufficiently rare that they can be disregarded in the study at hand. Only in negated sentences is the converb typically separated from its superordinate verb, by the negation verb, as shown above in (6). The negation verb is however immediately followed by the connegative form, meaning that the two elements of a verb pair are at most one word apart from one another. If a light verb pattern with *ǝštaš* 'do' (e.g. Russian *zanimatša* 'deal with' > (Eastern) Mari *zanimatša ǝštaš* id., cf. (21)) were to be used in a converb construction, the light verb would occur in its converb form *ǝšten* and would be followed by the superordinate verb (in negation preceded by a form of the negation verb).

6.2. Search queries: Udmurt

As the light verb strategy (e.g. Russian *vlijatʹ* 'influence' > *vlijatʹ kariṇi̥* id.) seems to be dominant in those southern varieties of Udmurt where Turkic-type converb constructions are best documented, this will be the first point of investigation here. The three prospective light verbs – *kariṇi̥* 'do', *kariśkiṇi̥* 'be done', and *luṇi̥* 'be' – have the respective converbs *kariṣa*, *kariśkiṣa*, and *luiṣa*. It would be possible to further restrict the search to only find pairings with individual auxiliaries, but as the body of results is already comparatively small at this point (see Section 7.2), the output of this search pattern can already be manually processed.

For Russian loanwords using indirect insertion (e.g. *žarit´* 'fry' > Udmurt *žarit̮ t̮ini̮* id.), the converbs always end in *-t̮t̮i̮sa*. Simply searching for all word forms ending in this combination of sounds finds all relevant converb constructions.

7. Data and implications on the productivity of auxiliaries

In this section we will present the totality of relevant converb constructions we could find in our corpora and our resulting judgments as regards the productivity of individual auxiliaries. For Mari, as the primary metric of productivity we are using type frequency, i.e. the number of distinct verbs with which an auxiliary co-occurs in our sources. An overview of all pairings is available in the appendix.

7.1. Mari

In our survey we are restricting ourselves to examining those verbs identified as auxiliaries in Bradley (2016a: 276–278).

Indirect insertion

An overview of Mari data can be found in Table 2. For each examined auxiliary, the following additional data points are taken from the source:

- the verb's lexical meaning,
- the page number in Bradley (2016a) of the section dedicated to the auxiliary,
- the functional meaning assigned to the verb as an auxiliary,
- the total number of distinct converbs, regardless of the etymology of the stem, found coupled with the auxiliary in Bradley (2016a) – i.e. the type frequency in the source – to give a baseline measurement of the auxiliary's productivity.

As regards new data, the number of distinct Russian loanwords found in the sources (i.e. the type frequency) detailed in Section 6 as well as the total number of occurrences within the 57.38 million tokens of the Corpus of Literary Mari (i.e. the token frequency) is provided; a listing of all pairings can be found in the appendix. When collecting these data, we have corrected obvious OCR mistakes and typos, but we have otherwise provided verbs as found in the corpus.

The auxiliaries are sorted by their type frequency, which can be seen as the best indicator of productivity yielded by this survey. Token frequency on its own, on the other hand, is a problematic measure of productivity as the exceeding frequency of certain pairings (e.g. *pečatlen luktaš* 'publish' < *pečatlaš* 'print' + *luktaš* 'take out' occurs 475 times in the corpus) distorts the picture.

For a number of Mari verbs treated as prospective auxiliaries in Bradley (2016a), we could not find any plausible pairings with Russian words. A list of these verbs is found in Table 2 on pp. 29–30.

From among the verbs described as prospective auxiliaries in Bradley (2016a), no pairings with Russian stems were found for *βoltaš* 'lower' (path 'down'), *βońćaš* 'go over' (path 'across, over'), *βozaš*[I] 'lie down' (resultative; path 'down'), *čâtaš* 'endure' (continuative), *čumâraš* 'gather' (path 'together'), *kâńelaš* 'get up' (path 'up'), *kudaltaš* 'throw' (resultative), *küzôktaš* 'raise' (path 'up(wards)'), *ojâraš* 'separate' (path 'apart'), *ojârlaš* 'go apart' (path 'apart'), *puraš*[II] 'come in' (path 'in, into'), *šütaš* 'pierce' (path 'through'), *šütlaš* 'wear through' (path 'through'), *temaš*[I] 'be filled' (exhaustive), *temaš*[II] 'fill sth. up' (exhaustive), *utaš* 'become superfluous' (exhaustive). The correlation between the overall type frequency of an auxiliary as given in Bradley (2016a) and the type frequency in combination with Russian borrowings is plainly visible if one plots these two values against each other, see Figure 1.
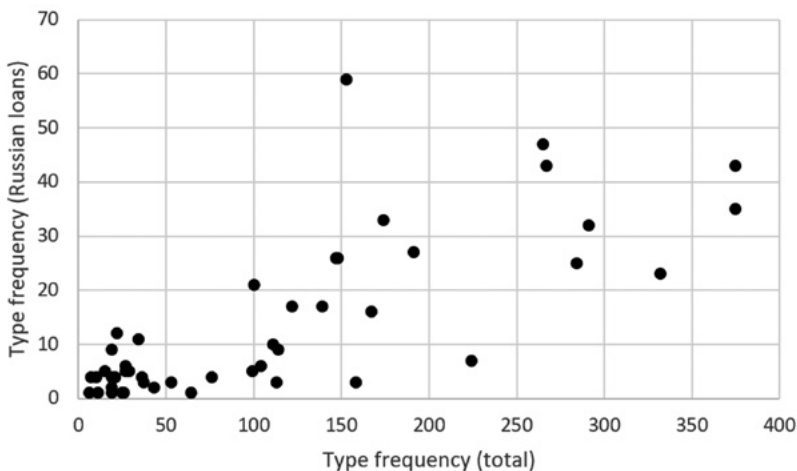
Figure 1: Scatter plot of type frequencies of different auxiliaries

Table 2: Overview of Mari data

Auxiliary	Lexical meaning	Page in Bradley (2016a)	Function as auxiliary	Type frequency in Bradley (2016a)	Type frequency (Russian borrowing)	Token frequency (Russian borrowings) in corpora
kertaš	be able to	154	modal (ability)	–[a]	–	100+
moštaš	be able to, know to	188	modal (ability)	–	–	100+
šuktaš	manage; lead to	241	exhaustive	153	59	158
pə̂taraš	finish	204	exhaustive	265	47	102
koštaš	go, wander	172	durative; 'at many locations'	267	43	104
šogaš	stand	232	durative	375	43	63
koltaš	let go; send	164	ingressive	375	35	104
puaš	give	210	benefactive; resultative	174	33	48
šə̂ndaš	put	219	resultative	291	32	75
tolaš[I]	come	248	gradual; path ('coming')	191	27	215
ončaš	look	196	ingressive; ('doing to try, test')	148	26	38
kodaš[II]	leave something	162	resultative	147	26	32
nalaš	take	189	auto-benefactive; delimitative	284	25	35
kajaš	go (away)	149	ingressive; path 'away'	332	23	49
kijaš	lie	158	durative	100	21	28
luktaš	take out	184	exhaustive; path 'out'	139	17	504
ilaš	live	147	durative	122	17	47
lektaš	go (out), leave	179	delimitative; path 'out'	167	16	315
tolašaš	try, strive	248	improper execution	22	12	18
seŋaš	win	216	exhaustive; 'manage to'	34	11	17
šińćaš[II]	sit	225	durative	111	10	24

a. As the modal auxiliaries *kertaš* 'be able to' and *moštaš* 'be able to, know to' are unambiguously productive, we did not collect type frequency data on these.

Auxiliary	Lexical meaning	Page in Bradley (2016a)	Function as auxiliary	Type frequency in Bradley (2016a)	Type frequency (Russian borrowing)	Token frequency (Russian borrowings) in corpora
šuaš[I]	arrive, reach	237	exhaustive	114	9	18
ońćə̑ktaš	show	199	benefactive ('in order to show')	19	9	13
pə̂taš	end	207	exhaustive	224	7	6
ertaraš	carry out, conduct	144	delimitative	27	6	10
pə̂štaš	put, place	202	resultative	104	6	9
naŋgajaš	take	193	path 'away'	27	5	42
kodaš[I]	stay	160	resultative	99	5	23
šińćə̑ltaš	sit around	227	iterative	15	5	6
sitaraš	gather, provide	218	exhaustive	29	5	4
kondaš	bring	170	path ('coming')	19	4	13
optaš	put, set	200	iterative	76	4	9
čarnaš	stop, cease	138	resultative	10	4	7
saβə̂rnaš	turn	215	iterative; path 'round'	21	4	5
lijaš	be; become	183	ingressive	7	4	3
šogə̂ltaš	stand around	236	iterative	36	4	2
šogaltaš	put, place, stand	230	resultative	37	3	52
šogalaš	stand up	228	resultative	113	3	27
šińćaš[I]	sit down	222	resultative	158	3	8
kə̂škaš	throw; scatter	156	resultative	53	3	4
purtaš	bring in	214	path 'in(to)'	19	2	6
mijaš	come, go	186	gradual; path 'up to'	43	2	2
töčaš	try, attempt	253	improper execution	11	1	3
βolaš	descend	131	path 'down'	19	1	3
namijaš	bring	192	path 'up to'	6	1	1
küzaš	climb, rise	177	path 'up'	25	1	1
ertaš	go by	145	path 'past'	26	1	1
šuaš[II]	throw	239	resultative	64	1	0

Light verb strategy

In contrast to the over 2000 pairings of Russian borrowings using indirect insertion with auxiliary verbs found within the 57.38 million tokens of the Corpus of Literary Mari, only three hapax forms of Russian borrowings using the light verb strategy can be found in combination with auxiliaries, two of which are given in (23–24).

Mari (Corpus of Literary Mari)

(23)   Доказать ыштен пу, Иван Иваныч!

    Dokaza-t́   *ə̂st-en*   *pu,*   *Ivan*   *Ivanə̂ć!*
    prove-INF   do-CVB   give.IMP.2SG,   Ivan   Ivanə̂ć
    'Prove it (to me/us), Ivan Ivanə̂ć!'

(24)   [Т]ендам раскулачить ыштен колтена!

    Tendam   ***raskulaći-t́***   *ə̂št-en*   *kolt-ena!*
    2PL.ACC   **dispossess_a_kulak-INF**   do-CVB   send-1PL
    'We'll dispossess you as kulaks!'

7.2. Udmurt

The data in this section was collected from two sources: the Corpora of Uralic Volga-Kama Languages' main and social-media corpora. Our survey uncovered only very few Turkic-type converb constructions with Russian borrowings, thus allowing only a qualitative but exhaustive (i.e. all examples we could find are given in this section) examination of our findings. Pairings of Russian borrowings as converbs with superordinate verbs do not generally show the same level of abstraction found in Mari, i.e. a lexical interpretation of the superordinate verb is more salient than a functional interpretation, leaving only few examples that can be considered auxiliary constructions. This sharp contrast with Mari is not commensurate with the sizes of the data sets under investigation: while we have roughly 10 times the tokens at our disposal for Mari (see 5.1), we could find over 2000 auxiliary constructions in our Mari sources, but only 12 distinct verbs in our Udmurt corpora.

    Among the examples we could find, *uļiņi* 'live' stood out as the most widely used in a clearly non-lexical manner. This verb is described as a

marker for "the incompleteness, duration, processual nature of an action or its periodicity" (Kelʹmakov 1975: 96) in Southern Udmurt and has also been observed in a comparable function in Beserman (Tepljašina 1970: 252–254). The interpretation of individual examples can be difficult as the lexical meaning of 'living' is generally compatible with actions by animate agents. Oftentimes however a functional reading compatible with those described in the sources is more transparent than a lexical reading and accepted as such by our native-speaker consultant, as in (25–26).

Udmurt (Corpora of Uralic Volga-Kama Languages, main)

(25)   […] и тани ми ку афишаос ошылймы гуртъёсы, мыным туннэ нуналын лумбыт звонить карыса улӥзы ни, лыкто-а, уг-а?

    […] *i*   *tańi*   *mi*   *ku*   *afiša-os*   *ošịl-i-mị*
         and so   1PL   when poster-PL   hang_up-PST1-1PL

gurt-jos-ị,   *mịnịm*   *tunne* *nunal-ịn* *lumbịt*
village-PL-ILL 1SG.DAT today day-INE   all_day

zvońi-tʹ *karị-sa* *ul-i-zị*   *ńi,*
call-INF do-CVB live-PST1-3PL already

lịkt-o-a,   *ug-a?*
come-FUT-1SG-INT, NEG-INT

'[…] and when we were hanging up posters in the villages, they already were calling me all day, will I come or won't I?'

(26)   Вдобавок берам пуке вал нылкышно, кудӥз ваньзэ комментировать карыса улӥз.

Vdobavok   *ber-am*   *puk-e*   *val*   *nịlkịšno,*
furthermore behind-1SG sit-3SG be.PST1.3SG woman

kudiz *vańze*   **_kommentirova-tʹ_** *karị-sa* *ul-i-z.*
which everything.ACC **comment-INF**   do-CVB live-PST1-3SG

'Additionally there was a woman sitting behind me who was (constantly) commenting on everything.'

Likewise from the main corpus: *služitʹ karịsa ulịnị* 'serve' *dokazịvatʹ karịsa ulịnị* 'demonstrate', *oχrańatʹ karịsa ulịnị* 'safeguard', *skanirovatṭịsa ulịnị* 'scan'. The last example is especially notable as it exemplifies indirect insertion, a strategy that is more typical of northern varieties of Udmurt in which Turkic-type auxiliary constructions are less widely used. Furthermore, the grammaticalization of a verb meaning 'live' as a habitual

or progressive marker is typologically common (cf. Kuteva et al. 2019: 261–262); a semantic shift of Russian *žiťˊ* 'live' > 'exist' has been observed in Russian dialects of Udmurtia (Martˊjanova 2004: 34; Ždanova 2021), showing semantic lability of verbs meaning 'live' as a regionally common phenomenon. It thus seems plausible that the functional meaning of *uḷiṇi* 'live' was fortified independently from the initial contact situation in which the usage of Turkic-type auxiliaries arose in southern varieties of Udmurt.

śotiṇi 'give' and *baśtiṇi* 'take' are described as benefactive and auto-benefactive markers, respectively, in the literature (e.g. Kelˊmakov 1975: 102). One example for each can be found in the main corpus, namely *romantizirovatˊ kariṣa śotiṇi* 'romanticize (something for someone)' and *rešitˊ kariṣa baśtiṇi* 'resolve (something for oneself)', as shown in (27).

(27) Нош писатель улонысь басьтэм […] частной конфликтэз ро-мантизировать карыса сётэ.

  Noš   *pisateľ*   *ulon-iś*   *baśt-em*   […]
  but   author   life-ELA   take-PTCP.PRF

  častnoj   *konfľikt-ez*   **romantiźirova-ť**   *kari-sa śot-e.*
  private   conflict-ACC   **romanticize-INF**   do-CVB give-3SG

  'But the author romanticized (for the readers) a private conflict taken from life.'

When examining those verbs described as auxiliaries in Kelˊmakov (1975), we could find one example each where the verb in question was prospectively serving as an auxiliary:

- *biḍtiṇi* 'end' as a marker of "completeness, finality" (Kelˊmakov 1975: 99): *otravitˊ kariṣa biḍtiṇi* (Rus. *otravitˊ* 'poison') (social-media corpus)
- *vetḷiṇi* 'go' as a marker of "the duration of an action, the non-directionality of a movement" (Kelˊmakov 1975: 100): *službitˊ kariṣa vetḷiṇi* (Rus. *službitˊ* 'serve') (social-media corpus)
- *voźiṇi* 'keep' as a maker of a "lengthy continuous state (process) consisting of separate, periodically repeated actions" (Kelˊmakov 1975: 103): *zaśśiśśatˊ kariṣa voźiṇi* (Rus. *zaśśiśśatˊ* 'defend') (social-media corpus)
- *kuštiṇi* 'throw' as a marker of "the meaning of exhaustiveness of an action" (Kelˊmakov 1975: 101): *redaktirovatˊ kariṣa kuštiṇi* (Rus. *redaktirovatˊ* 'edit') (main corpus).

8. Conclusions and outlook

Our survey supports the assumption that Turkic-type converb constructions have a stronger position in Mari than they do in Udmurt.

In Mari, Russian borrowings adapted into Mari using indirect insertion (where the derivational suffix *-l-* is attached to a Russian infinitive to form a Mari verb) are widely used in combination with auxiliaries in Turkic-type converb constructions. This serves as an argument for the vitality of the mechanism at hand, especially given how recent Russian borrowings such as *privatizirovatlaš* 'privatize' or *skanirovatlaš* 'scan' must be. Turkic-type converb constructions can be encountered both in literary texts and in social-media postings. A wide range of auxiliaries can be found to co-occur with Russian loanwords, and the type frequencies given in (Bradley 2016a) – i.e. the totality of distinct converbs previously found in combination with an auxiliary – were a good indicator of the type frequencies of Russian loanwords used with auxiliaries as determined in this survey. Nevertheless, the token frequencies are not exceedingly large.

It is noteworthy that not only were previously unknown pairings found in our sources, but also that previously known pairings from lexical sources were not found in the extensive corpora at our disposal. This seems to indicate a somewhat stochastic nature of verb pairs included in lexical sources: it seems that oftentimes, pairings found in lexical sources constitute random-usage examples of individual auxiliaries that lexicographers happened to be aware of rather than specific well-established collocations. Our personal experience from compiling the Mari-English dictionary (Riese et al. 2014) corroborates this suspicion. This again can be seen as a further argument for the productivity of these markers.

At first sight it seems curious how many Mari path markers that seemed fairly productive before this study (e.g. *puraš*[II] 'come in' > path 'in, into', 36 pairings in Bradley 2016a) cannot be found in combination with Russian loanwords at all in any of our sources. However, this can be seen as a function of an auxiliary's semantic constraints: even assuming full productivity of an auxiliary within semantic constraints, how many Russian verbs could form semantically sound pairings with it? A number of Russian verbs which can, in a highly figurative sense, be considered as markers of a means of movement or transportation can be found in combination with path markers, e.g. *mobilizovatlen naŋgajaš* 'mobilize (someone) away (lit. mobilizing take_away)', *evakuirovatlalt tolaš*[I] 'be evacuated to (lit. being_evacuated come)'. It

thus seems meaningful to consider the body of path-marking verbs as productive markers within semantic constraints, in analogy with the treatment of the Finnish prolative marker *-tse* which can be found in combination with recent loanwords, e.g. *mailitse* 'by e-mail' (Ylikoski 2018: 10).

Russian borrowings using the light verb strategy (i.e. the Russian infinitive is followed by the verb *ə̂štaš* 'do'), found in Eastern Mari, could only on three occasions be found in Turkic-type converb constructions. However, the general rarity of these structures within the data set at our disposal precludes strong conclusions from being made here.

In sharp contrast to the Mari data, the Udmurt data under examination yielded only 12 distinct Russian verbs in Turkic-type converb constructions. Only for *uli̮ni̮* 'live' could we find a significant number of examples in which it is used in accordance with its previously defined function as a durative marker, indicating that this auxiliary is uniquely productive in modern Udmurt. It should be noted that we could find this auxiliary in connection with a Russian verbal borrowing using indirect insertion, a strategy more conventional in northern Udmurt dialects in which Turkic-type converb constructions have a weak position: *skanirovatt̮i̮sa uli̮ni̮* 'scan'. This coupled with how typologically common it is for verbs meaning 'live' to become habitual or progressive markers and the observation that in local varieties of Russian in Udmurtia a semantic shift *žit´* 'live' > 'exist' has occurred (cf. Ždanova 2021) raises the prospect of areal processes affecting the functional meaning of verbs meaning 'live' independent from the original contact situation in which Turkic-type auxiliary constructions were borrowed from Tatar into Udmurt.

As regards other verbs that have been described as auxiliaries, some caution is advised before making overly bold conclusions based on negative data. Firstly, it was already established that Turkic-type converb constructions are typical of southern varieties of Udmurt (though some examples of Turkic-type verbalization of motion events having been described in Beserman), in contrast to Mari where Turkic-type converb constructions are ubiquitous in all varieties. The lower incidence of these structures could partly also be a function of the dialectal backgrounds of the writers, journalists, and users who produced the texts at our disposal. Furthermore, we are not fully certain that all strategies of verbal borrowing lend themselves equally to morphosyntactic modification, especially given the very few examples found of Russian borrowings using the light verb strategy used in Turkic-type converb constructions in (Eastern) Mari.

Acknowledgements

Non-standard abbreviations used in glosses

CNG	connegative	MOD	modal
EMP	emphatic	ORD	ordinal
ILL	illative	PRF	perfect
INDF	indefinite	PROL	prolative
INE	inessive	PST1	first past tense
INS	instructive	PST2	second past tense
INT	interrogative	VRB	verbalizer
LAT	lative		

Primary data sources

Corpora of Uralic Volga-Kama Languages:
    http://volgakama.web-corpora.net/index_en.html.
Corpus of Literary Mari:
    http://corpus.mari-language.com, https://gtweb.uit.no/u_korp/?mode=mhr.
Mari-English Dictionary:
    http://dict.mari-language.com, XML at http://source.mari-language.com.

References

Aasmäe, Niina. 2012. *E-kursuse „Ersa keel I" materjalid*. Tartu: University of Tartu.

Agyagási, Klára. 1998. On the characteristics of Cheremiss linguistic interference on Chuvash. In Johanson, Lars & Csató, Éva Ágnes (eds.), *The Mainz meeting: proceedings of the Seventh International Conference on Turkish Linguistics, August 3-6, 1994* (Turcologica 32), 667–682. Wiesbaden: Harrassowitz Verlag.

Agyagási, Klára. 2019. *Chuvash historical phonetics* (Turcologica 117). Wiesbaden: Harrassowitz Verlag.

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In Aikhenvald, Alexandra Y. & Dixon, Robert M. W. (eds.), *Serial verb constructions: A cross-linguistic typology* (Explorations in Linguistic Typology 2), 1–68. Oxford: Oxford University Press.

Arkhangelskiy, Timofey. 2019a. Corpora of social media in minority Uralic languages. In *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages*, 125–140. Tartu: Association for Computational Linguistics.

Arkhangelskiy, Timofey. 2019b. Russian verbal borrowings in Udmurt. *Folia Linguistica* 53(2). 519–552.

Arkhangelskiy, Timofey. 2020. Verbal borrowability and turnover rates. *Diachronica* 37(4). 451–473.

Bartens, Raija. 1979. *Mordvan, tšeremissin ja votjakin konjugaation infiniittisten muotojen syntaksi* (Mémoires de la Société Finno-Ougrienne 170). Helsinki: Suomalais-Ugrilainen Seura.

Bereczki, Gábor. 1984. Die Beziehungen zwischen den finnougrischen und türkischen Sprachen im Wolga-Kama-Gebiet. *Nyelvtudományi Közlemények* 86. 307–314.

Bereczki, Gábor. 1992–1994. *Grundzüge der tscheremissischen Sprachgeschichte (I-II)* (Studia Uralo-Altaica 34, 35). Szeged: University of Szeged.

Berta, Arpad. 1998. Tatar and Bashkir. In Johanson, Lars & Csató, Eva Agnes (eds.), *The Turkic languages*, 283–300. London & New York: Routledge.

Beznosikova et al. = Безносикова, Л. М. & Айбабина, Е. А. & Косынырева, Р. И. (eds.). 2000. *Большой коми-русский словарь*. Сыктывкар: ИЯЛИ Коми НЦ УрО РАН & Коми книжное издательство.

Binnick, Robert I. 1991. *Time and the verb: a guide to tense and aspect*. New York: Oxford University Press.

Bradley, Jeremy. 2016a. *Mari converb constructions: Productivity and regional variance*. Vienna: University of Vienna. (Doctoral dissertation.)

Bradley, Jeremy. 2016b. Verb-framed motion events in Uralic (with special attention to Mari). *Finnisch-Ugrische Forschungen* 63. 126–152. https://doi.org/10.33339/fuf.86122

Bradley, Jeremy & Hirvonen, Johannes & Milanova, Veronika. 2019. Benefaktív és malefaktív szerkezetek az uráliban (és azon túl): Előzetes vizsgálat. *Általános nyelvészeti tanulmányok* 30. 19–36.

BUTT, MIRIAM. 2010. The light verb jungle: still hacking away. In Amberber, Mengistu & Baker, Brett & Harvey, Mark (eds.), *Complex predicates: Cross-linguistic perspectives on event structure*, 48–78. Cambridge: Cambridge University Press.

CAMPBELL, LYLE. 2013. *Historical linguistics: An introduction*. [Third edition.] Cambridge, Massachusetts: The MIT Press.

COMRIE, BERNARD. 1998. *Aspect: an introduction to the study of verbal aspect and related problems* (Cambridge Textbooks in Linguistics 2). Reprinted. Cambridge: Cambridge University Press.

CREISSELS, DENIS. 2010. Benefactive and applicative periphrases: A typological approach. In Zúñiga, Fernando & Kittilä, Seppo (eds.), *Benefactives and malefactives: typological perspectives and case studies* (Typological studies in language 92), 29–70. Amsterdam: John Benjamins.

CSÚCS, SÁNDOR. 1990. *Die tatarischen Lehnwörter des Wotjakischen*. Budapest: Akadémiai Kiadó.

CSÚCS, SÁNDOR. 1998. Udmurt. In Abondolo, Daniel (ed.), *The Uralic languages*, 276–304. London: Routledge.

ČXAIDZE = Чхаидзе, М. П. 1960. *Спаренные глаголы в марийском языке*. Йошкар-Ола: Марийское книжное издательство.

ČXAIDZE = Чхаидзе, М. П. 1967. О происхождении и функциях марийских и удмуртских спаренных глаголов. Вопросы финно-угроведения 4. 247–259.

EDYGAROVA, SVETLANA. 2014. The varieties of the modern Udmurt language. *Finnisch-Ugrische Forschungen* 62. 376–398. https://doi.org/10.33339/fuf.86085

FEDOTOV = Федотов, М. Р. 1990. *Чувашско-марийские языковые взаимосвязи*. Саранск: Издательство Саратовского университета, Саранский филиал.

GALKIN, I. G. et al. (eds.). = Галкин, И. Г. et al. (eds.). 1990. *Словарь марийского языка (I–X)*. Йошкар-Ола: Марийское книжное издательство/МарНИИ.

GANIEV = Ганиев, Ф. А. 1982. *Образование сложных слов в татарском языке*. Москва: Наука.

GAVRILOVA = Гаврилова, В. Г. 2014. *Марийско-русский билингвизм: переключение и смешение кодов*. Йошкар-Ола.

GRAŠČENKOV = Гращенков, Р. В. 2012. Тюркские конструкции со вспомогательным глаголом и деепричастием на п (на материале языков кыпчакской группы). *Ural-Altaic Studies* 6. 55–77.

HARDER, ANJA. 2018. Converbal auxiliary constructions in Selkup. 7th International Conference on Samoyed Studies. Tartu, October 26–27, 2018. https://sisu.ut.ee/sites/default/files/samoyedology2018/files/harder_converbal_aux.pdf

HASPELMATH, MARTIN. 1995. The converb as a cross-linguistically valid category. In Haspelmath, Martin & König, Ekkehard (eds.), *Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds* (Empirical approaches to language typology 13), 1–55. Berlin: De Gruyter Mouton.

HASSELBLATT, CORNELIUS. 1990. *Das estnische Partikelverb als Lehnübersetzung aus dem Deutschen* (Veröffentlichungen der Societas Uralo-Altaica 31). Wiesbaden: O. Harrassowitz.

HELIMSKI, EUGEN. 1997. *Die matorische Sprache* (Studia uralo-altaica 41). Szeged: University of Szeged.

Helimski, Eugene. 2003. Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. *Nyelvtudományi Közlemé-nyek* 100. 156–167.

Hesselbäck, André. 2005. *Tatar and Chuvash code-copies in Mari* (Studia Urali-ca Upsaliensia 35). Uppsala: Uppsala Universitet.

Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of sound substitutions and distribution criterion*. Helsinki: University of Hel-sinki. (Doctoral dissertation.)

Honti, László. 2013. К вопросу о происхождении спаренных глаголов в финно-угорских языках Волжско-Камского региона. *Ural-Altaic Studies* 9. 109–113.

Horváth, Laura. 2011. Konverbumszerkezetek az udmurt nyelvben – a páros igék grammatikalizációjáról, aspektuskifejező funkciójáról. In Kiss, Zoltán & Ladányi, Mária & Petykó, Márton (eds.), *A pszicholingvisztikától a beszéd-technológiáig Tanulmányok az alkalmazott nyelvészet hagyományos és új té-maköreiből* (Segédkönyvek a nyelvészet tanulmányozásához 128), 39–49. Bu-dapest: Tinta Könyvkiadó.

Horváth, Laura. 2012. *Az úgynevezett páros igék aspektuális szerepe, grammati-kalizációja az udmurt (és a mari) nyelvben*. Budapest: ELTE BTK.

Horváth, Laura. 2013. On the aspectual markers of the Udmurt language – Ex-pressions of aspects in dialects. In Csepregi, Márta & Kubínyi, Kata & Sivo-nen, Jari (eds.), *Grammatika és kontextus – új szempontok az uráli nyelvek kutatásában III* (Urálisztikai tanulmányok 20), 108–123. Budapest: ELTE BTK Finnugor Tanszék.

Horváth, Laura. 2018. Igei kódváltási stratégiák. Orosz igék, infinitívuszok udmurt mátrixmondatokban. In Csepregi, Márta & Salánki, Zsuzsa (eds.), *A többnyelvűség dinamikája: a többnyelvűség megnyilvánulásai finnugor nyelvű közösségekben*, 33–80. Budapest: ELTE Bölcsészettudományi Kar.

Ilieva = Илиева, А. А. 2009. *Словарь балтачевского говора марийского язы-ка*. Йошкар-Ола: Марий государственный университет.

Isanbaev = Исанбаев, Н. И. 1978. Общее и отличительное в составных глаголах марийского и поволжско-тюркских языков. In *Вопросы марийского языка*, 59–90. Йошкар-Ола: Марийский научно-исследовательский институт.

Isanbaev = Исанбаев, Н. И. 1989–1994. Марийско-тюркские языковые контак-ты (I–II). Йошкар-Ола: Марийское книжное издательство & МарНИИ.

Ivanov = Иванов, И. Г. 1981. *Марий диалектологий*. Йошкар-Ола: Марий го-сударственный университет.

Johanson, Lars. 2009a. Chuvash. In Brown, Keith & Ogilvie, Sarah (eds.), *Concise encyclopedia of languages of the world*, 243–245. Amsterdam: Elsevier.

Johanson, Lars. 2009b. Modals in Turkic. In Hansen, Björn & de Haan, Fer-dinand (eds.), *Modals in the languages of Europe* (Empirical Approaches to Language Typology 44), 487–510. Berlin: De Gruyter Mouton.

Kangasmaa-Minn, Eeva. 1998. Mari. In Abondolo, Daniel (ed.), *The Uralic lan-guages*, 219–248. London: Routledge.

Kashkin = Кашкин, Е. В. 2017. К типологии грамматикализации глаголов перемещения: горномарийский глагол *keäš* „идти, уходить". In Кретов, А. А. (ed.), *Проблемы компьютерной лингвистики и типологии 6*, 36–47. Воронеж: Издательский дом ВГУ.

Kashkin = Кашкин, Е. В. 2018a. О семантике сложных глагольных комплексов в горномарийском языке: конструкции со значением достижения предела. In *Проблемы языка. Сборник научных статей по материалам Шестой конференции-школы „Проблемы языка: взгляд молодых ученых"*, 110–129. Москва: Институт языкознания РАН & Канцлер.

Kashkin = Кашкин, Е. В. 2018b. Грамматикализация горномарийских глаголов позиции. In *Языковые контакты народов Поволжья и Урала. XI Международный симпозиум (Чебоксары, 21-24 мая 2018 г.). Сборник статей*, 178–184. Чебоксары: Издательство Чувашского университета.

Kashkin = Кашкин, Е. В. 2018c. К типологии семантического развития глаголов со значением „бросать": данные горномарийского языка. *Ural-Altaic Studies* 4. 112–129.

Kashkin, E. V. 2019. Verbs of throwing and categorization of completive events: evidence from Hill Mari. *52nd Annual Meeting of the Societas Linguistica Europaea (Leipzig, 21-24 August 2019). Book of abstracts*, 152–154. Leipzig: Leipzig University.

Kashkin = Кашкин, Е. В. 2020. *Грамматикализация глагола anžaš „смотреть" в горномарийском языке*. Москва: Московский государственный университет.

Kashkin, E. V. & Dyachkov, V. V. 2018. Complex verb constructions in Hill Mari: semantics and event structure. *51st Annual Meeting of the Societas Linguistica Europaea (29 August – 1 September 2018)*, 129–130. Tallinn: Tallinn University.

Keľmakov = Кельмаков, В. К. 1975. Спаренные глаголы в удмуртском языке – на материале кукморского диалекта. In *Вопросы удмуртского языкознания – сборник статей*, 90–105. Ижевск: Удмуртский НИИ истории, экономики, литературы и языка при Совете Министров Удмуртской АССР.

Kincses Nagy, Éva. 2006. Verbal borrowings in Turkic languages. (Paper presented at the 13th. International Conference on Turkish Linguistics, Uppsala, 16–20 August 2006.)

Kirillova et al. = Кириллова, Л. Е. & Душенкова, Т. Р. & Егоров, А. В. & Ившин, Л. М. & Карпова, Л. Л. & Кириллова, Л. Е. & Титова, О. В. & Шибанов, А. А. (eds.). 2008. *Удмурт-ӟуч кыллюкам. Удмуртско-русский словарь*. Ижевск: Российская академия наук, Уральское отделение, Удмуртский институт истории, языка и литературы.

Kittilä, Seppo & Zúñiga, Fernando. 2010. Introduction. In Zúñiga, Fernando & Kittilä, Seppo (eds.), *Benefactives and malefactives: Typological perspectives and case studies* (Typological studies in language 92), 1–28. Amsterdam: John Benjamins.

Klumpp, Gerson. 2002. *Konverbkonstruktionen im Kamassischen* (Veröffentlichungen der Societas Uralo-Altaica 58). Wiesbaden: Harrassowitz in Kommission.

Korn, Agnes. 2013. Looking for the middle way: Voice and transitivity in complex predicates in Iranian. *Lingua* 135. 30–55.

Krasnova, Nadezhda & Riese, Timothy & Bradley, Jeremy & Yefremova, Tatiana. 2017. *Reading Hill Mari through Meadow Mari*. 1.0. Vienna: University of Vienna.

Kuteva, Tania & Heine, Bernd & Hong, Bo & Long, Haiping & Narrog, Heiko & Rhee, Seongha. 2019. *World lexicon of grammaticalization*. [Second edition.] Cambridge: Cambridge University Press.

Landmann, Angelika 2014a. *Tatarisch: Kurzgrammatik*. Wiesbaden: Harrassowitz.

Landmann, Angelika. 2014b. *Tschuwaschisch: Kurzgrammatik*. Wiesbaden: Harrassowitz.

Landmann, Angelika. 2015. *Baschkirisch: Kurzgrammatik*. Wiesbaden: Harrassowitz.

Lebedev = Лебедев, Е. Е. 2016. *Акционсартовые значения сложновербальных аналитических форм в чувашском языке*. Чебоксары: увашский государственный институт гуманитарных наук.

Mart'janova = Мартъянова, В. Н. 2004. *Слово в русских говорах Удмуртии*. Глазов: Глазовский государственный педагогический институт.

Masica, Colin P. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.

Moisio, Arto & Saarinen, Sirkka. 2008. *Tscheremissisches Wörterbuch*. Helsinki: Suomalais-Ugrilainen Seura.

Pischlöger, Christian. 2016. Udmurt on social network sites: A Comparison with the Welsh case. In Toivanen, Reetta & Saarikivi, Janne (eds.), *Linguistic genocide or superdiversity?*, 108–132. Bristol: Multilingual Matters.

Pomozi, Péter. 2004. The Mari – A historical overview. In Nanovfszky, György (ed.), *The Finno-Ugric world*, 95–99. Budapest: Teleki László Foundation.

Räsänen, Martti. 1920. *Die tschuwassischen Lehnwörter im Tscheremissischen* (Mémoires de la Société Finno-Ougrienne XLVIII). Helsinki: Suomalais-Ugrilainen Seura.

Räsänen, Martti. 1923. *Die tatarischen Lehnwörter im Tscheremissischen* (Mémoires de la Société Finno-Ougrienne L). Helsinki: Suomalais-Ugrilainen Seura.

Riese, Timothy & Bradley, Jeremy & Guseva, Elina. 2014. *Mari-English dictionary*. Vienna: University of Vienna.

Riese, Timothy & Bradley, Jeremy & Schötschel, Monika & Yefremova, Tatiana. 2019. *Mari (марий йылме): An essential grammar for international learners*. Draft. Vienna: University of Vienna.

Riese, Timothy & Bradley, Jeremy & Yakimova, Emma & Krylova, Galina. 2017. *Оҥай марий йылме: A comprehensive introduction to the Mari language*. 3.2. Vienna: University of Vienna.
https://www.univie.ac.at/maridict/site-2014/book/omj_2017.pdf

Róna-Tas, András. 1988. Turkic influence on the Uralic Languages. In Sinor, Denis (ed.), *The Uralic languages: Description, history, and foreign influences*, 742–780. Leiden: E. J. Brill.

Saarinen, Sirkka. 1997a. Borrowed vocabulary in Mari and Udmurt dialects. In Hahmo, Sirkka-Liisa & Hofstra, Tette & Honti, László & van Linde, Paul & Nikkilä, Osmo (eds.), *Finnisch-ugrische Sprachen in Kontakt*, 191–196. Maastricht: Shaker.

SAARINEN, SIRKKA. 1997b. Language contacts in the Volga region: loan suffixes and calques in Mari and Udmurt. In Viereck, Wolfgang & Wynne, Kenneth (eds.), *Language in time and space: studies in honour of Wolfgang Viereck on the occasion of his 60th birthday* (Zeitschrift für Dialektologie und Linguistik. Beihefte 97), 388–396. Stuttgart: Franz Steiner Verlag.

SAARINEN, SIRKKA. 2010. Marin sanaston alkuperästä. In Saarinen, Sirkka & Siitonen Kirsti & Vaittinen, Tanja (eds.), *Sanoista kirjakieliin – Juhlakirja Kaisa Häkkiselle* (Mémoires de la Société Finno-Ougrienne 259), 335–341. Helsinki: Suomalais-Ugrilainen Seura.

SALÁNKI, ZSUZSA. 2015. The bilingualism of Finno-Ugric language speakers in the Volga Federal district. In Stolz, Christel (ed.), *Language empires in comparative perspective* (Colonial and Postcolonial Linguistics 6), 237–264. Berlin: De Gruyter.

SCHÖNIG, CLAUS. 1984. *Hilfsverben im Tatarischen: Untersuchungen zur Funktionsweise einiger Hilfsverbverbindungen* (Veröffentlichungen der Orientalischen Kommission 35). Wiesbaden: F. Steiner.

SERDOBOL'SKAJA et al. = Сердобольская, Н. В. & Ильевская, А. А. & Минор, С. А. & Митева, П. С. & Файнвейц, А. В. & Матвеева, Н. С. 2012. Конструкции с сентенциальными актантами в финно-угорских языках. In Кузнецова, А. И. (et al., eds.), *Финно-угорские языки: фрагменты грамматического описания – Формальный и функциональный подходы.* (Studia Philologica), 382–475. Москва: Рукописные памятники Древней Руси.

SEREBRENNIKOV = Серебренников, Б. А. 1960. *Категории времени и вида в финно-угорских языках пермской и волжской групп.* Москва: Издательство Академии наук СССР.

SIBATROVA = Сибатрова, С. С. 2015. Грамматическое освоение русских заимствованных глаголов в марийском языке: выражение залоговых значений. In *Язык и социум: материалы Межрегиональной науч.-практ. конф., посвящ. Дню марийской письменности (9 декабря 2014 г.)*, 66–72. Йошкар-Ола: Марийский государственный универстиет.

SIBATROVA = Сибатрова, С. С. 2016. Грамматическое освоение русских заимствованных глаголов в марийском языке: выражение видовых значений. In *XV Игнатовские чтения. Материалы докладов и выступлений на Международной научно-практической конференции „Горные марийцы в культурно-истроическом ландшафте Урало-Поволжья" (Козьмодемясянк, 21-22 мая 2015г)*, 47–59. Йошкар-Ола: Марийский государственный универстиет.

SKVORCOV & SKVORCOVA = Скворцов, М. И. & Скворцова, А. В. (eds.). 2002. *Русско-чувашский словарь.* Чебоксары: Чувашское книжное издательство.

SLOBIN, DAN I. 2000. Verbalized events: A dynamic approach to linguistic relativity and determinism. In Niemeier, Susanne & Dirven, René (eds.), *Evidence for linguistic relativity* (Current Issues in Linguistic Theory 198), 107–138. Amsterdam: Benjamins.

TALMY, LEONARD. 1985. Lexicalisation patterns: semantic structure in lexical forms. In Shopen, Timothy (ed.), *Language typology and syntactic description. III. Grammatical categories and the lexicon*, 57–149. [First edition.] Cambridge: Cambridge University Press.

Talmy, Leonard. 2007. Lexical typologies. In Shopen, Timothy (ed.), *Language typology and syntactic description. III. Grammatical categories and the lexicon*, 66–168. [Second edition.] Cambridge: Cambridge University Press.

Tarakanov = Тараканов, И. В. 2013. Аналитические глагольные образования в удмуртском языке. *Вестник Удмуртского университета. Серия История и филология* 2013(2). 3–7.

Tepljašina = Тепляшина, Т. И. 1970. *Язык бесермян*. Москва: Наука.

Tereščenko = Терещенко, Н. М. 1981. Глаголы движения ненецкого языка и основные способы передачи в языке ненцев значение приставочных глаголов русского языка. *Вопросы финно-угорской филологии* 4. 80–96.

Trask, Robert Lawrence. 1996. *Historical linguistics*. London: Arnold.

Valijärvi, Riitta-Liisa. 2008. Converbs in Northern Selkup. In Vajda, Edward J. (ed.), *Subordination and coordination strategies in North Asian languages* (Current issues in linguistic theory 300), 167–178. Amsterdam: John Benjamins.

Vinay, J.-P. & Darbelnet, Jean. 1995. *Comparative stylistics of French and English: a methodology for translation* (Benjamins Translation Library 11). Amsterdam: John Benjamins.

Wichmann, Søren & Wohlgemuth, Jan. 2008. Loan verbs in a typological perspective. In Stolz, Thomas & Bakker, Dik & Salas Palomo, Rosa (eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes* (Empirical approaches to language typology 35), 89–122. Berlin: De Gruyter Mouton.

Wichmann, Yrjö. 1924. Etymologisches aus den permischen Sprachen. *Finnisch-Ugrische Forschungen* XIV. 29–63.

Winkler, Eberhard. 2011. *Udmurtische Grammatik* (Veröffentlichungen der Societas Uralo-Altaica 81). Wiesbaden: Harrassowitz.

Wintschalek, Walter. 1993. *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal* (Studia Uralica 7). Wiesbaden: Harrassowitz.

Wohlgemuth, Jan. 2009. *A typology of verbal borrowings* (Trends in linguistics. Studies and monographs 211). Berlin: De Gruyter Mouton.

Ylikoski, Jussi. 2004. Eine detaillierte Abhandlung über die aspektualen Verbverbindungen im Kamassischen. *Finnisch-Ugrische Forschungen* 58. 376–381.

Ylikoski, Jussi. 2018. Prolatiivi ja instrumentaali. *Sananjalka* 60. 7–27.

Ylikoski, Jussi. (forthcoming). Non-finites. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*. Oxford: Oxford University Press.

Zaicz, Gábor. 1998. Udmurt. In Abondolo, Daniel (ed.), *The Uralic languages*, 184–218. London: Routledge.

Ždanova = Жданова, Е. А. 2021. Семантика глагола жить в русских говорах Удмуртии. *Вестник Удмуртского университета, Серия История и филология* 31(2). 230–235.

Appendix

Abbreviations

Aux.   auxiliary

Lm.   lexical meaning

P.   page number in source

Am.   meaning as auxiliary

Tf.   type frequency in source

Tokens # tokens in corpora

Rv.   Russian verbs paired with auxiliary in corpora

Aux. *kertaš*, **Lm.** 'be able to', **P.** 154, **Am.** modal (ability), **Tf.** N/A, **Tokens** 100+

Rv. (not documented due to unambiguous full productivity of auxiliary)

Aux. *moštaš*, **Lm.** 'be able to, know to', **P.** 188, **Am.** modal (ability), **Tf.** N/A, **Tokens** 100+

Rv. (not documented due to unambiguous full productivity of auxiliary)

Aux. *šuktaš*, **Lm.** 'manage; lead to', **P.** 241, **Am.** exhaustive, **Tf.** 153, **Tokens** 158

Rv. 59: *agitirovatlaš* 'agitate, campaign', *analizirovatlaš* 'analyse', *arestovatlaš* 'arrest', *atakovatlaš* 'attack', *blokirovatlaš* 'block', *celitlaš* 'aim', *dressirovatlaš* 'train (animal)', *elektrificirovatlaš* 'electrify', *evakuirovatlaš* 'be evacuated', *evakuirovatlaš* 'evacuate', *formirovatlaš* 'form', *fotkatlaš* 'photograph', *fotografirovatlaš* 'be photographed', *gladitlaš* 'iron', *gruzitlaš* 'load', *izvinitlaš* 'apologize', *javitlaš* 'inform', *kačatlaš* 'pump', *kleitlaš* 'glue', *komplektovatlaš* 'complete', *maskirovatlaš* 'mask', *meχanizirovatlaš* 'mechanize', *montirovatlaš* 'mount', *motivirovatlaš* 'justify', *oformitlaš* 'put into shape', *okopatlaš* 'dig in', *organizovatlaš* 'organize', *pasterizovatlaš* 'pasteurize', *pečatlaš* 'be published', *pečatlaš* 'print', *peredavatlaš* 'broadcast', *pererabotatlaš* 'process', *prinimatlaš* 'take up (a post)', *privatizirovatlaš* 'privatize', *projektirovatlaš* 'design', *raskulačivatlaš* 'dispossess (a kulak)', *registrirovatlaš* 'be registered', *registrirovatlaš* 'register', *rešatlaš* 'solve', *rešitlaš* 'solve', *restavrirovatlaš* 'restore', *sdatlaš* 'hand over', *sdavatlaš* 'hand over', *skanirovatlaš* 'scan', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *suditlaš* 'sentence', *trenirovatlaš* 'train', *ut'užitlaš* 'iron', *vakcinirovatlaš* 'vaccinate', *vooružatlaš* 'arm', *vospitatlaš* 'bring up', *zagotovitlaš* 'lay in stock', *zakazatlaš* 'order', *zaŕaditlaš* 'load (tr.)', *žaritlaš* 'fry', *zavoditlaš* 'start, wind up', *zoritlaš* 'ravage', *zvonitlaš* 'ring'

Aux. *p̂taraš*, **Lm.** 'finish', **P.** 204, **Am.** exhaustive, **Tf.** 265, **Tokens** 102

Rv. 47: *betonirovatlaš* 'concrete', *bintovatlaš* 'bandage', *bombitlaš* 'bomb', *britlaš* 'shave', *čertitlaš* 'draw, trace', *deklamirovatlaš* 'recite', *dežuritlaš* 'be on duty', *diktovatlaš* 'dictate', *diskovatlaš* 'harrow', *gladitlaš* 'iron', *grimirovatlaš* 'apply theatre makeup', *gruzitlaš* 'load', *kleitlaš* 'glue', *komplektovatlaš* 'complete', *kritikovatlaš* 'criticize', *lišitlaš* 'deprive of rights', *lunkovatlaš* '?', *meχanizirovatlaš* 'mechanize', *minirovatlaš* 'mine', *orkestrovatlaš* 'orchestrate', *pečatlaš* 'print', *perepisatlaš* 'rewrite', *perevoditlaš* 'translate', *povtoritlaš* 'repeat', *pressovatlaš* 'compress', *putatlaš* 'tangle', *rastaskivatlaš*

'drag apart', *redaktirovatlaš* 'edit', *remontirovatlaš* 'repair', *repressirovatlaš* 'repress', *rešatlaš* 'solve', *restavrirovatlaš* 'restore', *rozmitlaš* '?', *sdatlaš* 'hand over', *sdavatlaš* 'hand over', *skanirovatlaš* 'scan', *skirdovatlaš* 'stack', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *sortirovatlaš* 'sort', *stroitlaš* 'construct', *štukaturitlaš* 'plaster', *vistupatlaš* 'perform', *zapravitlaš* 'fill up (tr.)', *zaŕaditlaš* 'load (tr.)', *zoritlaš* 'ravage', *zvonitlaš* 'ring'

Aux. *koštaš*, **Lm.** 'go, wander', **P.** 172, **Am.** durative; 'at many locations', **Tf.** 267, **Tokens** 104

Rv. 43: *agitirovatlaš* 'agitate, campaign', *batračitlaš* 'work as a farm labourer', *bespokoitlaš* 'bother', *bomževatlaš* 'be like a homeless person', *broditlaš* 'roam', *buntovatlaš* 'rebel', *figuŕatlaš* 'show off', *gastrolirovatlaš* 'tour', *gipsovatlaš* 'put in a cast', *grabitlaš* 'rob', *gul'atlaš* 'go for a walk', *instruktirovatlaš* 'instruct', *kalịmitlaš* 'moonlight', *kaznitlaš* 'execute', *komandovatlaš* 'command', *konfiskovatlaš* 'confiscate', *konvoirovatlaš* 'escort', *maitlaš* 'suffer', *maskirovatlaš* 'be masked', *maskirovatlaš* 'mask', *oformitlaš* 'put into shape', *patrulirovatlaš* 'patrol', *putešestvovatlaš* 'travel', *revizirovatlaš* 'requisition', *šabašatlaš* 'moonlight', *šabašitlaš* 'moonlight', *šalitlaš* 'play tricks', *šatatlaš* 'rock', *slavitlaš* 'extol', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *sormitlaš* '?', *šutitlaš* 'joke', *tancevatlaš* 'dance', *veselitlaš* 'have a good time', *vistupatlaš* 'perform', *vojevatlaš* 'wage war', *zagotlaš* '?', *zakazatlaš* 'reserve', *žalovatlaš* 'complain', *zvonitlaš* 'ring', *χlopotatlaš* 'make an effort', *χuliganitlaš* 'behave like a hooligan'

Aux. *šogaš*, **Lm.** 'stand', **P.** 232, **Am.** durative, **Tf.** 375, **Tokens** 63

Rv. 43: *dejstvovatlaš* 'function', *diktovatlaš* 'dictate', *distillirovatlaš* 'distil', *dokladịvatlaš* 'report', *doložitlaš* 'report', *eksportirovatlaš* 'export', *formirovatlaš* 'be formed', *fotografirovatlaš* 'photograph', *gladitlaš* 'iron', *golosovatlaš* 'vote', *kalitlaš* 'be heated', *kaznitlaš* 'execute', *koordinirovatlaš* 'coordinate', *kurịrovatlaš* 'curate', *maskirovatlaš* 'be masked', *nabl'udatlaš* 'observe', *obespečitlaš* 'guarantee', *obrabatịvatlaš* 'process', *organizovatlaš* 'organize', *pečatlaš* 'be published', *pečatlaš* 'print', *podderžatlaš* 'support', *podderživatlaš* 'support', *podpisatlaš* 'sign', *propagandirovatlaš* 'propagandize', *razoblačitlaš* 'reveal', *recenzirovatlaš* 'review', *rešatlaš* 'solve', *revizovatlaš* 'inspect', *rifmovatlaš* 'rhyme', *služitlaš* 'serve', *šutitlaš* 'joke', *travitlaš* 'poison', *trenirovatlaš* 'train', *uvažatlaš* 'respect', *veselitlaš* 'have a good time', *vịpisatlaš* 'subscribe', *vistupatlaš* 'perform', *vospitatlaš* 'bring up', *zakazatlaš* 'order', *žaritlaš* 'fry', *zavisitlaš* 'depend on', *zvonitlaš* 'call'

Aux. *koltaš*, **Lm.** 'let go; send', **P.** 164, **Am.** ingressive, **Tf.** 375, **Tokens** 104

Rv. 35: *blagoslovitlaš* 'bless', *dreseirovatlaš* [sic] 'train (animals)', *evakuirovatlaš* 'evacuate', *formirovatlaš* 'form', *fotografirovatlaš* 'photograph', *grabitlaš* 'rob', *gruzitlaš* 'load', *kalịmitlaš* 'moonlight', *komissovatlaš* 'give a fitness test for military service', *lišitlaš* 'deprive of rights', *loslovitlaš* '?', *mečtatlaš* 'dream', *obsluživatlaš* 'serve', *oformitlaš* 'put into shape', *organizovatlaš* 'organize', *perepisatlaš* 'rewrite', *podpisatlaš*

'sign', *pudritlaš* 'powder', *putatlaš* 'tangle', *raskulaćitlaš* 'dispossess (a kulak)', *registrirovaltaš* 'register', *rešatlaš* 'solve', *šabašitlaš* 'moonlight', *sdatlaš* 'hand over', *snimatlaš* 'be photographed', *suditlaš* 'sentence', *šuritlaš* '?', *tancevatlaš* 'dance', *veselitlaš* 'have a good time', *vipisatlaš* 'subscribe', *zakazatlaš* 'order', *žalovatlaš* 'complain', *žaritlaš* 'fry', *zavoditlaš* 'start, wind up', *χuliganitlaš* 'behave like a hooligan'

Aux. *puaš*, **Lm.** 'give', **P.** 210, **Am.** benefactive; resultative, **Tf.** 174, **Tokens** 48

Rv. 33: *britlaš* 'shave', *ćertitlaš* 'draw, trace', *citirovatlaš* 'cite', *deklamirovatlaš* 'recite', *deval'virovatlaš* 'devaluate', *dirižirovatlaš* 'conduct', *dokazatlaš* 'demonstrate', *gladitlaš* 'iron', *kalitlaš* 'heat', *kaznitlaš* 'execute', *kočegaritlaš* 'work as a stoker', *oformitlaš* 'put into shape', *paitlaš* 'solder', *paritlaš* 'steam', *pečatlaš* 'print', *perepisatlaš* 'rewrite', *podpisatlaš* 'sign', *pressovatlaš* 'compress', *rešitlaš* 'solve', *skanirovatlaš* 'scan', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *sortirovatlaš* 'sort', *točitlaš* 'sharpen', *ut'užitlaš* 'iron', *vipisatlaš* 'subscribe', *zapravitlaš* 'fill up (tr.)', *zaŕaditlaš* 'load (tr.)', *žaritlaš* 'fry', *zaveritlaš* 'attest', *zavoditlaš* 'start, wind up', *zvonitlaš* 'ring', *χlopotatlaš* 'make an effort'

Aux. *šôndaš*, **Lm.** 'put', **P.** 219, **Am.** resultative, **Tf.** 291, **Tokens** 75

Rv. 32: *arestovatlaš* 'arrest', *betonirovatlaš* 'concrete', *bintovatlaš* 'bandage', *ćertitlaš* 'draw, trace', *dressirovatlaš* 'train (animal)', *fantazirovatlaš* 'dream about', *gipnozirovatlaš* [sic] 'hypnotize', *gipsovatlaš* 'put in a cast', *gladitlaš* 'iron', *grafitlaš* 'plot on a graph', *gruzitlaš* 'load', *kleitlaš* 'glue', *kommentirovatlaš* 'comment on', *konservirovatlaš* 'conserve', *koptitlaš* 'smoke', *lepitlaš* 'sculpt', *maskirovatlaš* 'mask', *obšivatlaš* 'clad', *paitlaš* 'solder', *pečatlaš* 'print', *plombirovatlaš* 'seal', *registrirovatlaš* 'register', *rifmovatlaš* 'rhyme', *saχaritlaš* 'sugar', *šnurovatlaš* 'lace up', *štukaturitlaš* 'plaster', *trambovatlaš* 'trample', *valkovatlaš* 'put into windrows', *varitlaš* 'weld', *zaŕaditlaš* 'load (tr.)', *žaritlaš* 'fry', *zubritlaš* 'learn by rote'

Aux. *tolaš*[I], **Lm.** 'come', **P.** 248, **Am.** gradual; path '(coming)', **Tf.** 191, **Tokens** 215

Rv. 27: *demobilizovatlaš* 'be demobilized', *doložitlaš* 'report', *evakuirovatlaš* 'be evacuated', *ironizirovatlaš* 'speak ironically', *komissovatlaš* 'retire from armed service', *kontuzitlaš* 'be contused', *marširovatlaš* 'march', *maskirovatlaš* 'be masked', *naśtupatlaš* 'advance', *opravdatlaš* 'be discharged', *pasovatlaš* 'pass', *pečatlaš* 'be published', *pečatlaš* 'print', *perevospitatlaš* 'be reeducated', *propagandirovatlaš* 'propagandize', *ranitlaš* 'be injured', *razoblaćitlaš* 'reveal', *razoritlaš* 'be ruined', *razvivatlaš* 'delevop (intr.)', *razvoditlaš* 'separate', *reformirovatlaš* 'reform', *rešatlaš* 'solve', *sdatlaš* 'hand over', *služitlaš* 'serve', *suditlaš* 'be sentenced', *zanimatlaš* 'study', *zvonitlaš* 'ring'

Aux. *ońćaš*, **Lm.** 'look', **P.** 196, **Am.** ingressive; (doing to try, test), **Tf.** 148, **Tokens** 38

Rv. 26: *dežuritlaš* 'be on duty', *diskovatlaš* 'harrow', *golosovatlaš* 'vote', *gruzitlaš* 'load', *kaćatlaš* 'pump', *obrabotatlaš* 'process', *pećatlaš* 'print',

perevoditlaš 'translate', *prinimatlaš* 'take up (a post)', *probịtlaš* 'stay somewhere', *sdatlaš* 'hand over', *sdavatlaš* 'hand over', *snimatlaš* 'photograph, film', *suditlaš* 'sentence', *tancevatlaš* 'dance', *trenirovatlaš* 'train', *vịpisatlaš* 'subscribe', *vịstupatlaš* 'perform', *vkľučitlaš* 'switch on', *vospitatlaš* 'bring up', *zakazatlaš* 'order', *zapravitlaš* 'fill up (tr.)', *zaŕaditlaš* 'load (tr.)', *žaritlaš* 'fry', *zavoditlaš* 'start, wind up', *zvonitlaš* 'ring'

Aux. *kodaš*[II], Lm. 'leave something', P. 162, Am. resultative, Tf. 147, Tokens 32

Rv. 26: *blagoslovitlaš* 'bless', *bombitlaš* 'bomb', *britlaš* 'shave', *elektrizovatlaš* 'electrify', *fotografirovatlaš* 'photograph', *kristallizirovatlaš* 'crystallize', *kritikovatlaš* 'criticize', *mobilizovatlaš* 'mobilize', *oborudovatlaš* 'equip', *oformitlaš* 'put into shape', *oskorbitlaš* 'insult', *pečatlaš* 'print', *prostitlaš* 'pardon', *rozmitlaš* '?', *sdatlaš* 'hand over', *skanirovatlaš* 'scan', *snimatlaš* 'photograph, film', *stramitlaš* 'disgrace', *veselitlaš* 'have a good time', *vịkľučatlaš* 'turn off', *zapravitlaš* 'fill up (tr.)', *zaŕaditlaš* 'load (tr.)', *zaveššatlaš* 'bequeath', *zavoditlaš* 'start, wind up', *želatlaš* 'wish', *zoritlaš* 'ravage'

Aux. *nalaš*, Lm. 'take', P. 189, Am. auto-benefactive; delimitative, Tf. 284, Tokens 35

Rv. 25: *britlaš* 'shave', *fotografirovatlaš* 'photograph', *golosovatlaš* 'vote', *kalitlaš* 'heat', *keseritlaš* '?', *kritikovatlaš* 'criticize', *lečitlaš* 'treat', *paritlaš* 'steam', *pečatlaš* 'print', *podpisatlaš* 'sign', *privatizirovatlaš* 'privatize', *skandalitlaš* 'brawl', *skanirovatlaš* 'scan', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *sortirovatlaš* 'sort', *tancevatlaš* 'dance', *truditlaš* 'burden with work', *učastvovatlaš* 'participate', *veselitlaš* 'have a good time', *vospitatlaš* 'bring up', *zakazatlaš* 'order', *zanimatlaš* 'study', *žaritlaš* 'fry', *zvonitlaš* 'ring'

Aux. *kajaš*, Lm. 'go (away)', P. 149, Am. ingressive; path 'away', Tf. 332, Tokens 49

Rv. 23: *bombitlaš* 'bomb', *brakovatlaš* 'reject as defective', *buksovatlaš* 'skid, slip', *dokladịvatlaš* 'report', *evakuirovatlaš* 'be evacuated', *fotografirovatlaš* 'photograph', *klejmitlaš* 'brand', *markirovatlaš* 'mark', *nomerovatlaš* 'number', *prošśatlaš* 'forgive', *redaktirovatlaš* 'edit', *rifmovatlaš* 'rhyme', *snimatlaš* 'be photographed', *snimatlaš* 'photograph, film', *štampovatlaš* 'stamp, press', *štrobitlaš* 'cut a wall plate', *suditlaš* 'be sentenced', *tancevatlaš* 'dance', *uťužitlaš* 'iron', *verbovatlaš* 'enlist', *veselitlaš* 'have a good time', *vịpisatlaš* 'be subscribed', *zaritlaš* 'excite'

Aux. *kijaš*, Lm. 'lie', P. 158, Am. durative, Tf. 100, Tokens 28

Rv. 21: *dokladịvatlaš* 'report', *doložitlaš* 'report', *grabitlaš* 'rob', *gruzitlaš* 'load', *kalitlaš* 'be heated', *konspektirovatlaš* 'summarize', *koptitlaš* 'give off soot', *maitlaš* 'suffer', *maitlaš* 'suffer', *marinovatlaš* 'marinate', *pečatlaš* 'print', *perevaritlaš* 'overdo', *projavitlaš* 'display', *rozoritlaš* '?', *verbovatlaš* 'enlist', *vreditlaš* 'injure', *zanimatlaš* 'study', *zapravitlaš* 'fill up (tr.)', *zaŕaditlaš* 'load (tr.)', *zavoditlaš* 'start, wind up', *zvonitlaš* 'ring'

Aux. *luktaš*, **Lm.** 'take out', **P.** 184, **Am.** exhaustive; path 'out', **Tf.** 139, **Tokens** 504

Rv. 17: *čertitlaš* 'design', *dublirovatlaš* 'duplicate', *izvinitlaš* 'apologize', *kačatlaš* 'pump', *oformitlaš* 'put into shape', *paritlaš* 'steam', *pečatlaš* 'print', *pel'atlaš* '?', *pererabatıvatlaš* 'process', *predpolagatlaš* 'presume', *pressovatlaš* 'compress', *redaktirovatlaš* 'edit', *štampovatlaš* 'stamp, press', *suditlaš* 'sentence', *turitlaš* 'urge', *žaritlaš* 'fry', *χlopotatlaš* 'make an effort'

Aux. *ilaš*, **Lm.** 'live', **P.** 147, **Am.** durative, **Tf.** 122, **Tokens** 47

Rv. 17: *ekspluatirovatlaš* 'exploit', *grabitlaš* 'rob', *klevetatlaš* 'slander', *kompleksovatlaš* 'have a complex', *maitlaš* 'suffer', *maitlaš* 'suffer', *mečtatlaš* 'dream', *obižajaltlaltaš* [sic] 'be insulted', *otravitlaš* 'poison', *pečatlaš* 'be printed', *pereživatlaš* 'endure', *pitatlaš* 'feed', *snimatlaš* 'photograph, film', *veselitlaš* 'have a good time', *vooružatlaš* 'arm', *vreditlaš* 'harm', *zoritlaš* 'ravage'

Aux. *lektaš*, **Lm.** 'go (out), leave', **P.** 179, **Am.** delimitative; path 'out', **Tf.** 167, **Tokens** 315

Rv. 16: *analizirovatlaš* 'analyse', *dežuritlaš* 'be on duty', *kandužitlaš* '?', *paritlaš* 'take a steam bath', *prostitlaltaš* 'take one's leave', *pečatlaš* 'be published', *pečatlaš* 'print', *proveŕatlaš* 'check', *redaktirovatlaš* 'edit', *repetirovatlaš* 'rehearse', *revizirovatlaš* 'audit', *rozmitlaš* '?', *sdatlaš* 'hand over', *služitlaš* 'serve', *tancevatlaš* 'dance', *vıstupatlaš* 'perform'

Aux. *tolašaš*, **Lm.** 'try, strive', **P.** 248, **Am.** improper execution, **Tf.** 22, **Tokens** 18

Rv. 12: *buksovatlaš* 'skid, slip', *dokazatlaš* 'demonstrate', *duritlaš* 'play pranks', *golosovatlaš* 'vote', *pečatlaš* 'print', *rifmovatlaš* 'rhyme', *sdatlaš* 'hand over', *suditlaš* 'have legal proceedings', *tipizirovatlaš* 'typify', *zavoditlaš* 'start, wind up', *zvonitlaš* 'ring', *χodatajstvovatlaš* 'solicit'

Aux. *seŋaš*, **Lm.** 'win', **P.** 216, **Am.** exhaustive; 'manage to', **Tf.** 34, **Tokens** 17

Rv. 11: *agitirovatlaš* 'agitate, campaign', *dežuritlaš* 'dežuritle', *dokazatlaš* 'demonstrate', *maskirovatlaš* 'be masked', *obižatlaš* 'offend', *prostitlaš* 'pardon', *rešatlaš* 'solve', *rešitlaš* 'solve', *sdatlaš* 'hand over', *sofinansirovatlaš* 'co-finance', *zavoditlaš* 'start, wind up'

Aux. *šińćaš*[II], **Lm.** 'sit', **P.** 225, **Am.** durative, **Tf.** 111, **Tokens** 24

Rv. 10: *lepitlaš* 'sculpt', *nabiratlaš* 'engage', *pečatlaš* 'print', *podpisatlaš* 'sign', *pudritlaš* 'powder', *suditlaš* 'sentence', *sufleritlaš* 'prompt (in theatre)', *veselitlaš* 'have a good time', *zanimatlaš* 'study', *zubritlaš* 'learn by rote'

Aux. *šuaš*[I], **Lm.** 'arrive, reach', **P.** 237, **Am.** exhaustive, **Tf.** 114, **Tokens** 18

Rv. 9: *adaptirovatlaš* 'adapt (intr.)', *evakuirovatlaš* 'be evacuated', *formirovatlaš* 'be formed', *oformitlaš* 'take shape', *pečatlaš* 'be published', *rešitlaš* 'be solved', *rusificirovatlaš* 'be Russified', *služitlaš* 'serve', *žaritlaš* 'fry (intr.)'

Aux. *ońćôktaš*, **Lm.** 'show', **P.** 199, **Am.** benefactive (in order to show), **Tf.** 19, **Tokens** 13

Rv. 9: *dokazatlaš* 'demonstrate', *idealizirovatlaš* 'idealize', *ill'ustrirovatlaš* 'illustrate', *kopirovatlaš* 'copy', *kritikovatlaš* 'criticize', *pečatlaš* 'print', *poetizirovatlaš* 'poeticize', *rulitlaš* 'taxi', *tipizirovatlaš* 'typify'

Aux. *pôtaš*, **Lm.** 'end', **P.** 207, **Am.** exhaustive, **Tf.** 224, **Tokens** 6

Rv. 7: *assimilirovatlaš* 'be assimilated', *koptitlaš* 'be smoked', *paritlaš* 'take a steam bath', *pečatlaš* 'be published', *ranitlaš* 'be injured', *žaritlaš* 'fry (intr.)', *zoritlaš* 'dawn'

Aux. *ertaraš*, **Lm.** 'carry out, conduct', **P.** 144, **Am.** delimitative, **Tf.** 27, **Tokens** 10

Rv. 6: *maitlaš* 'suffer', *pečatlaš* 'print', *služitlaš* 'serve', *snimatlaš* 'photograph, film', *veselitlaš* 'have a good time', *zoritlaš* 'ravage'

Aux. *pôštaš*, **Lm.** 'put, place', **P.** 202, **Am.** resultative, **Tf.** 104, **Tokens** 9

Rv. 6: *bintovatlaš* 'bandage', *koptitlaš* 'give off soot', *oformitlaš* 'put into shape', *pečatlaš* 'print', *podpisatlaš* 'sign', *žaritlaš* 'fry'

Aux. *naŋgajaš*, **Lm.** 'take', **P.** 193, **Am.** path 'away', **Tf.** 27, **Tokens** 42

Rv. 5: *arestovatlaš* 'arrest', *buksirovatlaš* 'tow', *konvoirovatlaš* 'escort', *mobilizovatlaš* 'mobilize', *verbovatlaš* 'recruit'

Aux. *kodaš*, **Lm.** 'stay', **P.** 160, **Am.** resultative, **Tf.** 99, **Tokens** 23

Rv. 5: *britlaš* 'shave oneself', *pečatlaš* 'be published', *snimatlaš* 'be photographed', *veselitlaš* 'have a good time', *želatlaš* 'wish'

Aux. *šińćôltaš*, **Lm.** 'sit around', **P.** 227, **Am.** iterative, **Tf.** 15, **Tokens** 6

Rv. 5: *kleitlaš* 'glue', *nastroitlaš* 'tune', *pečatlaš* 'print', *varitlaš* 'weld', *veselitlaš* 'have a good time'

Aux. *sitaraš*, **Lm.** 'gather, provide', **P.** 218, **Am.** exhaustive, **Tf.** 29, **Tokens** 4

Rv. 5: *bombitlaš* 'bomb', *fotografirovatlaš* 'photograph', *veselitlaš* 'have a good time', *vojevatlaš* 'wage war', *zapravitlaš* 'fill up (tr.)'

Aux. *kondaš*, **Lm.** 'bring', **P.** 170, **Am.** path '(coming)', **Tf.** 19, **Tokens** 13

Rv. 4: *arestovatlaš* 'arrest', *evakuirovatlaš* 'evacuate', *gruzitlaš* 'load', *mobilizovatlaš* 'mobilize'

Aux. *optaš*, **Lm.** 'put, set', **P.** 200, **Am.** iterative, **Tf.** 76, **Tokens** 9

Rv. 4: *gruzitlaš* 'load', *sortirovatlaš* 'sort', *varitlaš* 'weld', *vipisatlaš* 'subscribe',

Aux. *čarnaš*, **Lm.** 'stop, cease', **P.** 138, **Am.** resultative, **Tf.** 10, **Tokens** 7

Rv. 4: *bombitlaš* 'bomb', *veselitlaš* 'have a good time', *tancevatlaš* 'dance', *vistupatlaš* 'perform'

Aux. *saβ̂rnaš*, **Lm.** 'turn', **P.** 215, **Am.** iterative; path 'round', **Tf.** 21, **Tokens** 5
Rv. 4: *blagoslovitlaš* 'bless', *služitlaš* 'serve', *tancevatlaš* 'dance', *varitlaš* 'weld'

Aux. *lijaš*, **Lm.** 'be; become', **P.** 183, **Am.** ingressive, **Tf.** 7, **Tokens** 3
Rv. 4: *golosovatlaš* 'vote', *obespečitlaš* 'provide oneself', *proveŕatlaš* 'be checked', *rešitlaš* 'be solved'

Aux. *šogə̂ltaš*, **Lm.** 'stand around', **P.** 236, **Am.** iterative, **Tf.** 36, **Tokens** 2
Rv. 4: *skanirovatlaš* 'scan', *suditlaš* 'sentence', *točitlaš* 'sharpen', *zavoditlaš* 'start, wind up'

Aux. *šogaltaš*, **Lm.** 'put, place, stand', **P.** 230, **Am.** resultative, **Tf.** 37, **Tokens** 52
Rv. 3: *arestovatlaš* 'arrest', *betonirovatlaš* 'concrete', *stroitlaš* 'construct'

Aux. *šogalaš*, **Lm.** 'stand up', **P.** 228, **Am.** resultative, **Tf.** 113, **Tokens** 27
Rv. 3: *maskirovatlaš* 'be masked', *stroitlaš* 'draw up in a line', *tormozitlaš* 'brake'

Aux. *šińćaš*[I], **Lm.** 'sit down', **P.** 222, **Am.** resultative, **Tf.** 158, **Tokens** 8
Rv. 3: *kleitlaš* 'be glued', *lepitlaš* 'be sculpted', *maskirovatlaš* 'be masked'

Aux. *kô̂škaš*, **Lm.** 'throw; scatter', **P.** 156, **Am.** resultative, **Tf.** 53, **Tokens** 4
Rv. 3: *bombitlaš* 'bomb', *goŕačitlaš* 'excite', *χuliganitlaš* 'behave like a hooligan'

Aux. *purtaš*, **Lm.** 'bring in', **P.** 214, **Am.** path 'in(to)', **Tf.** 19, **Tokens** 6
Rv. 2: *kaćatlaš* 'pump', *pečatlaš* 'print'

Aux. *mijaš*, **Lm.** 'come, go', **P.** 186, **Am.** gradual; path 'up to', **Tf.** 43, **Tokens** 2
Rv. 2: *evakuirovatlaš* 'be evacuated', *sdatlaš* 'hand over'

Aux. *töćaš*, **Lm.** 'try, attempt', **P.** 253, **Am.** improper execution, **Tf.** 11, **Tokens** 3
Rv. 1: *suditlaš* 'have legal proceedings'

Aux. *βolaš*, **Lm.** 'descend', **P.** 131, **Am.** path 'down', **Tf.** 19, **Tokens** 3
Rv. 1: *pikirovatlaš* 'dive (when flying)'

Aux. *namijaš*, **Lm.** 'bring', **P.** 192, **Am.** path 'up to', **Tf.** 6, **Tokens** 1
Rv. 1: *koptitlaš* 'give off soot'

Aux. *küzaš*, **Lm.** 'climb, rise', **P.** 177, **Am.** path 'up', **Tf.** 25, **Tokens** 1
Rv. 1: *pikirovatlaš* 'dive (when flying)'

Aux. *ertaš*, **Lm.** 'go by', **P.** 145, **Am.** path 'past', **Tf.** 26, **Tokens** 1
Rv. 1: *stročitlaš* 'stitch'

Aux. *šuaš*[II], **Lm.** 'throw', **P.** 239, **Am.** resultative, **Tf.** 64, **Tokens** 0
Rv. 1: *gladitlaš* 'iron'

Josefina Budzisch
University of Hamburg

Marking strategies of attributive possession in Selkup: A study of frequency and types of possession

This paper deals with attributive possession in North, Central and South Selkup and focuses on a quantitative analysis of the frequency with which marking strategies are used in Selkup dialects. In Selkup, attributive possession can be head marked (with a possessive suffix), dependent marked (with genitive or adessive marking), and double marked (both combined), but close study shows that while dependent marking with genitive is most commonly used for lexical possessors, for non-lexical possessors the most common usage is head marking with a possessive suffix. The paper also illustrates the usage of different types of possession (e.g. inalienable/alienable) and shows that they are rarely treated differently with regard to their marking.

1. Introduction

This paper aims at providing an overview of different marking strategies for attributive possession and quantifying their frequency, as well as examining whether the type of possession plays a role in choosing a strategy. Marking strategies for attributive possession in Selkup are well established and described (among others, Kuznecova et al. 1980; Kim 1985; Bekker et al. 1995a; Budzisch 2015; Kim-Maloney & Kovylin 2015; Vorobeva et al. 2017), but in those descriptions the strategies are often presented as being equal,

and only a few notes on usage are provided. The objective of this study is to present a corpus-based approach and expand the picture of attributive possession in Selkup. This article also seeks to highlight dialect diversity and to place Selkup within the broader context of Samoyedic languages.

Thus, new insights into possessive relations, especially concerning third-person possessors, can be won. This mainly applies to the distribution of the use of pronouns and possessive suffixes. Grammatical descriptions, both for North (Kuznecova et al. 1980: 187) and for Central and South Selkup (Bekker et al. 1995a: 80), have claimed that third-person possession is most commonly marked with pronouns, but as this paper will explore, the corpus data paints a different picture: here, too, possessive suffixes are mainly used.

The paper is structured as follows. In Section 2 a brief introduction will be given to the notion of possession used in this paper. In Section 3 the Selkup dialects studied with regard to the topic of possession and the corpus data used in this paper are presented. In Section 4 lexical possessors are investigated in the three main dialects of Selkup, then Section 5 examines non-lexical possessors again in all three main dialects. In Section 6 the findings are summarized.

2. Possession

In this paper, a possessive relationship is defined in a rather narrow sense as a relationship of ownership (*Peter's house*), a kin relationship (*my grandmother*) or a part–whole relationship (most typically body parts: *his nose*). In that way, the findings on Selkup are comparable to other previous works in typology (see among others Koptjevskaja-Tamm 2003 and Haspelmath 2017). A widespread distinction made between these kinds of possessive notion is alienable vs. inalienable possession. Inalienable possession indicates that the possessor and the possessed are inseparable (kinship, body parts), while alienable possession means they are separable or the relationship is not permanent (ownership). It must be emphasized, however, that this distinction is a scale and that there are no clear boundaries. For the sake of clarity, in this paper body parts and kin relationships are seen as inalienable, while ownership of some kind of material possession is regarded as alienable.

Possessive relationships can be expressed as attributive (*my leg*) or predicative (*I have a leg*), but only attributive possession is taken into account in the following study. In attributive possession the two components of the possessive relation, namely the possessor and the possessed, form

a noun phrase in which the possessed is the head of the phrase. The possessor can be expressed either lexically (*Martin's father*) or non-lexically (*his father*). In some languages, the possessive relationship between the possessor and the possessed is expressed merely through juxtaposition, but the relationship is rather often overtly marked on either one of the elements (Koptjevskaja-Tamm 2003). In head-marked constructions (marking on the possessed), a possessive affix is usually used, while in dependent-marked constructions (marking on the possessor), the genitive case is commonly used to mark the relationship (Koptjevskaja-Tamm 2001: 961). In double-marked constructions (marking on both components), a combination of both elements is used (see also Nichols 1986).

3. Selkup dialects and corpus data

Selkup is a Samoyedic language spoken by approximately 1,000 speakers in Western Central and North West Siberia. Its closest relatives were Kamas and Mator, together they formed the South Samoyedic branch, but both of these languages are now extinct. The North Samoyedic[1] languages Nenets, Enets and Nganasan are still spoken today. Selkup can be divided into at least three dialect groups (with several subdialects), namely North, Central and South Selkup, with North Selkup behaving significantly differently from the latter two in many respects. Table 1 shows the dialect groups with the respective subdialects.

Table 1: Selkup dialects (following Gluškov et al. 2013: 50ff.)

North	Central	South
Taz	Vakh	Middle Ob
Laryak	Tym	Chaya
Karasino	Vasyugan	Ket
Turukhan	Narym	Upper Ob
Baikha		Chulym
Yelogui		

1. The term North Samoyedic languages is used here as a *terminus technicus* for any kind of summary of the languages Nenets, Enets and Nganasan. As the internal classification of the Samoyed languages has not been answered conclusively, it is not clear whether the North Samoyed languages are a genetic unit or rather an areal grouping.

The dialects show a slightly different consonant and vowel inventory and phonetic isoglosses can be depicted. For example, Central Selkup shows an *s ~ h* variation: while it is *suːrip* 'wild animal, beast' in North and South Selkup, the variant most commonly used in Central Selkup is *huːrip* id. Noteworthy, also for the present topic, is the nasal–plosive alternation. The alternation occurs between *m ~ p* (~ ∅) and *n ~ t* (~ ∅) and is to be found in free morphemes (*suːrip ~ suːrim ~ suːri* 'wild animal, beast') as well as in bound morphemes (genitive: *-n ~ -t*). It sometimes functions as a dialectal isogloss,[2] but it can also be found as a free variant within a dialectal group.

Nouns in Selkup are declined for number, case and possession, where the order of inflectional morphemes is usually as follows: stem [+ derivational suffix] – number – case – possession. In all dialects, there are three numbers: singular, dual and plural. While the singular is always unmarked and the dual is marked in all dialects with the ending *-qi*, there are some variants that are only found in some (sub)dialects: *-j(a)* in Taz, Narym and Middle Ob, *-qäːqɨ* in North and Central Selkup, and *-štja(qi)* in Central and South Selkup. The plural marker shows dialectal variation as well: in all dialect groups *-t* marks the plural, but in Central and South Selkup, the suffix *-la* is also used.

The case systems differ slightly depending on the dialect as well. Here, only the cases playing a role in attributive possession will be examined: the nominative, the genitive and the adessive. The nominative is unmarked and the genitive is marked with the suffix *-n* or *-t* (nasal–plosive alternation in free distribution) in all dialects. The adessive (ending *-nan*) on the other hand is only present in South and Central Selkup; North Selkup lacks this case completely. Table 2 presents the case suffixes used in marking possession in the Selkup dialects.

Table 2: Case suffixes in Selkup dialects (possessive marking)

	North	Central	South
NOM	*-∅*	*-∅*	*-∅*
GEN	*-n ~ -t*	*-n ~ -t*	*-n ~ -t*
ADE	–	*-nan*	*-nan*

2. The dialectal isogloss can be seen, for example, in the following: in the Ket dialect of South Selkup, there is only the form *qup* 'person', in the Tym dialect (Central Selkup) only *qum* id. Meanwhile, in the Taz dialect (North Selkup) a free distribution of *qup ~ qum* id. is documented (Kuper 1986: 103).

While the genitive (as mentioned in Section 2) is commonly used to mark possession, the adessive in Selkup needs more illustration. The adessive denotes a location adjacent to the referent of the noun. The case is named differently in most descriptions of Selkup, e.g. *locative-personal* in Bekker et al. (1995) and *locative II* (with the note that it is only used for animate referents) in Bykonja et al. (2005), both indicating that the case suffix can only be attached to animate nouns, though this is imprecise, as example (1) shows:

(1) *Känaŋ-dɨ   päa-nin   saːra-mba*
     dog-POSS.3SG tree-ADE   bind-PST.REP.3SG.S
     'He tied his dog to the tree.'
     (South: Ket, SVG_1964_IitekaPineweldju_flk.026)

Nonetheless, due to the nature of the adessive, it is far more frequently attached to animate nouns overall. It is mostly used in marking predicative possession (as e.g. in Finnish), but as will be shown in the following sections, in Central and South Selkup it is also used in attributive possession. In example (2), the use in a possessive sentence is presented.

(2) *Ma-nan   ando-m   ɛ-ja.*
     1SG-ADE   boat-POSS.1SG be-AOR.3SG
     'I have a boat.'
     (Central: Vasyugan, ChDN_1983_HerosDaughter_flk.035)

Possession can also be marked on the noun with possessive suffixes. In Selkup, there is no distinction for the number of the possessed, only the number and person of the possessor are taken into account. In Table 3, the possessive suffixes in the nominative for Central and South Selkup are presented.

Table 3: Possessive suffixes (nominative) in Central and South Selkup (Bekker et al. 1995a: 65)[a]

	SG	DU	PL
1P	*-mɨ*	*-miː*	*-min ~ -mɨt*
2P	*-lɨ*	*-liː*	*-lin ~ -lɨt*
3P	*-tɨ*	*-tiː*	*-tin ~ -tɨt*

a. Selkup shows high dialectal variance, so the suffixes listed here may appear in the data in slight modifications and abbreviations.

While Selkup is, as can be expected from a Uralic language, an agglutinative language, it shows fusional tendencies when it comes to possessive suffixes in several cases, e.g. in most local cases. There is one possessive ending for locative, illative and ablative, which is *-qinti* 'LOC/ILL/ABL.POSS.3SG' in the third-person singular; the case and possessive suffix cannot be separated in this fused ending (a full paradigm of possessive suffixes in different cases can be seen in Kuznecova et al. 1980: 185ff.). Furthermore, it is important to note that in Selkup as in other Samoyedic languages, possessive suffixes are not only used to mark possessive relations but are also used to track reference and mark semantically unique referents like the sun (e.g. Kim-Maloney & Kovylin 2015: 41; Budzisch 2017, 2021). However, in Selkup only the third-person singular possessive suffix is used in that regard, while in other Samoyedic languages first- and second-person suffixes can also be used in a non-possessive function; see e.g. Körtvély (2010) for (Tundra) Nenets, Siegl (2013: 371ff.) for (Forest) Enets and Zayzon (2015) for Nganasan.

There are no crucial differences between the inflection of nouns and pronouns in Selkup. Also, there are no specific possessive pronouns in Selkup, but the genitive form of the personal pronoun is used in this function in all dialects; for first and second person, nominative and genitive pronouns are homonymous, but in the third person nominative and genitive are clearly distinguished. In Central and South Selkup, the adessive pronoun can also be used to mark possessive relations. In Table 4, personal

Table 4: Personal pronouns in Central and South Selkup (Bekker et al. 1995b: 71, 79–82)[a]

	NOM	GEN	ADE
1SG	*man ~ mat*	*man ~ mat*	*manan*
2SG	*tan ~ tat*	*tan ~ tat*	*tanan*
3SG	*tep*	*tepin ~ tepit*	*tepinan*
1DU	*me*	*me*	*menan*
2DU	*te*	*te*	*tenan*
3DU	*tepqi*	*tepqin ~ tepqit*	*tepqinan*
1PL	*me*	*me*	*menan*
2PL	*te*	*te*	*tenan*
3PL	*tepit*; *tepla*	*tepitin ~ tepitit*; *teplan ~ teplat*	*tepitinan*; *teplanan*

a. As already mentioned for the possessive suffixes, it also applies to the pronouns that they can appear in variants that cannot all be listed here.

pronouns in Central and South Selkup are presented in the nominative, genitive and adessive. The aforementioned nasal–plosive alternation can be seen in both lexemes themselves (e.g. *man ~ mat*) and in the genitive endings for third persons; also in Central and South Selkup, two endings (-*t* and -*la*) for marking the plural can be found. Note that in North Selkup, there is no adessive case and only the first mentioned plural variants are in use, but apart from that the pronouns are the same.

Also worth mentioning, even if not a marking strategy on its own, is the Selkup intensifier pronoun with the base *on*-, which can be used to amplify pronouns, also in possessive constructions as shown in the following example (3), but is not used to mark possession itself.

(3) […] *pone*   *čanni-mba*   *onže*
          outwards   go.out-PST.REP.3SG.S   INT.3SG

     ad'uka-m-de   *pone*   *i-mba-d.*
     grandmother-ACC-POSS.3SG   outwards   take-PST.REP-3SG.O

     '[...] he went out and took his own grandmother out.'
     (Central: Vasyugan, ChDN_1983_Nikita_flk.044)

Overall, Selkup is in many respects a rather well-described language even though many materials are not yet published. Two grammars are worth mentioning explicitly, which also serve as a reference throughout this paper. In 1980, the grammar *Očerki po sel'kupskomu jazyku: Tazovskij dialekt* by Ariadna Kuznecova, Eugen Helimski and Elena Grushkina was published, and as the title suggests the book only takes North Selkup (Taz dialect) into account, but it is still the only full grammatical description of Selkup. In 1995, Erika Bekker, Larisa Alitkina, Valentina Bykonja and Irina Il'jašenko published *Morfologija sel'kupskogo jazyka: Južnye dialekty* (two volumes), in which they describe the phonology and morphology of Central and South Selkup in detail. Both works provide many helpful insights into the language and its grammatical peculiarities, but with the new possibility of examining digital corpora, new findings can be detected. This specifically holds true for Central and Southern Selkup, as North Selkup has often been the main focus in studies about Selkup. The quantifying study presented here, however, is based on the Selkup Language Corpus (SLC, Budzisch et al. 2019), a corpus containing 144 texts covering North, Central and South Selkup. The corpus is based on texts already published (but not made digitally available before), recorded in 1846–2014 by various researchers, with most texts having been recorded between the 1960s

and the 1980s. The corpus data and the data in the two grammars mentioned are well comparable, since for North Selkup the text basis covers the same dialect as the grammars, and most texts are conducted by the same researchers. This is also the case for the Central and South Selkup parts of the corpus, which largely take into account fieldwork records that have been used to compile the grammar by Bekker et al. (1995). The focus of the corpus is on Central and South Selkup indicated also by the text distribution: there are 26 North, 48 Central and 66 South Selkup texts in the data set. A more detailed overview of the data distribution is presented in Table 5.

Table 5: Data distribution in the SLC corpus

	North Selkup	Central Selkup	South Selkup
Texts	26	48	66
Sentences	1,140	3,426	4,018
Tokens	7,814	21,856	22,417

The corpus mainly covers three genres: folklore texts (109 texts), stories about daily life (16 texts) and translations (19 texts), providing an overview of different structures in the language. The reference given for each example refers to the name of the text as used in the corpus, the number at the end of each reference indicates the sentence number, and thus every example is traceable in the data. For more details on the corpus and the naming of texts, see Behnke & Budzisch (2021).

4. Lexical possessors in Selkup

In the Samoyedic languages in general, dependent marking with the genitive is the most common way to mark attributive possession with a lexical possessor. It is described for the North Samoyedic languages (Tundra) Nenets (Nikolaeva 2014: 143), (Forest) Enets (Siegl 2013: 234f.; Ovsjannikova 2020) and Nganasan (Wagner-Nagy 2014; 2019: 317), but also for the extinct South Samoyedic languages Kamas (Künnap 1999b: 16) and Mator (Helimski 1997: 137). For the North Samoyedic languages, it is also possible to use double marking with both the genitive marking on the possessor and a possessive suffix on the possessed, even though the possessive suffix is never obligatory (Nikolaeva 2014: 143, Wagner-Nagy 2019: 317). Künnap

(1999a: 18) believes the suffix to be used emphatically in Enets, but Siegl (2013: 234ff.) sees the need for further research in this area. The data set for Kamas and Mator is too small to give a more detailed account of the use of possessive suffixes with a lexical possessor.

Furthermore, Winkler (1913: 121) mentioned that in the Samoyedic languages in general the marking of a possessive relation between a lexical possessor and the possessed can be marked with juxtaposition, whereby the possessed might carry a possessive suffix. Sebestyén (1975: 41) calls this "a Uralic inheritance" as there are traces of this strategy in Finno-Ugric as well as Samoyedic languages (cf. Sebestyén 1975: 41ff.).

Both for North (Kuznecova et al. 1980: 173) and Central and South (Bekker et al. 1995a: 128) Selkup, dependent marking with the genitive is described. For North Selkup, double marking is thought to be rare, while in Central and South Selkup it is more commonly used (Bekker et al. 1995a: 83). Additionally, in Central and South Selkup the adessive can be used in the same way as the genitive (as mentioned before, the case does not exist in North Selkup) (Bekker et al. 1995a: 83).

In the following sections, a detailed view of different marking strategies in the Selkup dialect groups is presented to see whether the statements mentioned above hold true against the corpus data. In Section 4.4, the findings are summarized and quantified.

4.1. North Selkup

Dependent marking with the genitive is the default marking strategy in North Selkup for attributive possession, be it body parts (4), kinship relations (5) or the marking of ownership (6).

(4) *Ima*   *imaqota-t*   *üŋkɨlsaː-qɨt*   *mannɨ-mpa-tɨ.*
     woman   old.woman-GEN   ear-LOC   see-PST.REP-3SG.O
     'The woman looks into the old woman's ears.'
     (North: Taz, BVP_1973_East_flk.038)

(5) *Šettɨr-qɨt*   *qən-nɔː-tɨt*   *nätä-t*   *əsɨ-p*   *peːlä.*
     spring-LOC   go.away-AOR-3PL   girl-GEN   father-ACC   search-CVB
     'In spring they left to search for the girl's father.'
     (North: Taz, AVA_1973_Ichakicha_flk.075)

(6)   *Palna-n*   *mɔːt-tɨ*   *man*   *aššă*   *ičči-r-na-ŋ.*
      Palna-GEN   tent-ILL   1SG   NEG   go-FRQ-AOR-1SG.S
      'I do not go into the Palna's tent.' (North: Taz, APA_1977_Palna_flk.007)

According to Kuznecova et al. (1980: 187), double marking is ungrammatical, while Bekker et al. (1995a: 83) state that the combination of a genitive marking with a possessive suffix is possible but rarely used in North Selkup. The corpus data reflects the latter statement: there are a total of a mere four out of 75 occurrences (see also Table 6 below) showing double marking, two of them marking kinship, the remaining two marking ownership. Example (7) shows the marking of a kinship relation with the genitive on the possessor and the possessive suffix on the possessed.

(7)   *Iča-n*   *ima-tɨ*   *quː-mpɨ*
      Itja-GEN   woman-POSS.3SG   die-PST.REP.3SG.S
      'Itja's wife is dead.' (North: Taz, BEP_1977_Icha4_flk.026)

Apart from that, there are but two occurrences where the possessor is completely unmarked – one with ownership (8), the other within an unmarked kinship relation.

(8)   *Iča*   *nɪl'či-k*   *loːs-ira*   *kɪn'či-p*   *toː*   *imi-tɨ*
      Itja   such-ADV   devil-old.man   bolt-ACC   away   take-3SG.O
      'So Itja takes the devilish old man's bolt.'
      (North: Taz, BEP_1973_Itja2_flk.015)

4.2. Central Selkup

In Central Selkup the same holds true as for North Selkup: the most frequent marking strategy for lexical possessors is dependent marking with merely the genitive on the possessor. Also in these dialects, it can be used to link body parts to the respective person (9), to mark kinship relations (10) or ownership (11).

(9)   *Tüː-t*   *haj*   *paktɨ-mba*   *üčed'e-l'ika-t*   *kɨl-o-nd*
      fire-GEN   eye   jump-PST.REP.3SG.S   child-DIM-GEN   chest-EP-ILL

      al't'i-mba.
      fall-PST.REP.3SG.S
      'A spark flew off and fell onto the child's chest.'
      (Central: Vasyugan, ChDN_1983_MistressOfFire_flk.014)

(10) *Mat*   *tö-mba-k*   *amdɛlgu-n*   *nɛ-p*
      1SG   come-PST.REP-1SG.S   tsar-GEN   daughter-ACC

      i-hu   *pika-nd.*
      take-INF   bull-ILL
      'I have come to take the tsar's daughter for the bull.'
      (Central: Narym, MNS_NN_BullSon_flk.027)

(11) […] *naːɣur* *kjöt* *aːmdeɣo-t* *maːt-kit* […]
          three   ten   tsar-GEN   tent-LOC
      '[…] in the house of the thirtieth tsar […]'
      (Central: Tym/Narym, TTD_1964_Frog_flk.078)

Double marking with the genitive on the possessor and a possessive suffix on the possessed is far more frequent in Central Selkup than in North Selkup. There are 29 occurrences with this combination for kinship relations (12) and body parts (13) – the prototypical inalienable possession – but only two occurrences for ownership (see Table 7 below).

(12) *Paja-n*   *iː-t*   *šot-qəndo*
      old.woman-GEN   son-POSS.3SG   forest-ABL.POSS.3SG

      tö-mba.
      come-PST.REP.3SG.S
      'The son of the old woman comes from the forest.'
      (Central: Tym, MNN_1977_VillageKuleevo_nar.013)

(13) *Ara-n*   *ol-t*   *aj*   *tɛ*   *pača-nna-t.*
      old.man-GEN   head-POSS.3SG   again   away   chop-AOR-3SG.O
      'They cut off the old man's head again.'
      (Central: Narym, NS_NN_BullSon_flk.056)

Apart from the genitive marking, the marking of the possessor with the adessive in Central Selkup has been described in detail (e.g. Kim 1985; Budzisch 2015; Kim-Maloney & Kovylin 2015; Vorobeva et al. 2017). Bekker et al. (1995a: 83) claim that it is the most widely used construction in the Tym (Central) and Ob (South) dialects of Selkup and is used in these dialects to mark all types of possession. Kim (1985: 50) on the other hand declares that it is only used in the Tym and Narym dialects (both Central Selkup). The corpus shows findings for the Narym, Tym (both Central), Ket and Ob (both South) dialects, but overall the adessive is rarely used in attributive possessive constructions in the corpus data, though it is very

commonly used in marking predicative possession. In predicative posses-
sive constructions, the possessor in the adessive and the possessed are not
part of the same nominal phrase, but in attributive possessive construc-
tions they indeed are, as they cannot be split up: the possessor marked
with the adessive always precedes the possessed. It is therefore clear that
when used to mark the possessor with the adessive in these constructions,
it indeed is an internal and not an external possessor (see König 2001).
Adessive marking without a possessive suffix is only used for body parts in
the Central Selkup data under investigation here (14).

(14) *Ara-lžiga-nan   ol   tɛp   ača-l-ba.*
      old.man-DIM-ADE   head   away   chop-RES-PST.REP.3SG.S
      'The old man's head is chopped off.'
      (Central: Narym, MNS_NN_BullSon_flk.057)

The adessive in combination with a possessive suffix, however, is not only
used for body parts (15) but also for kinship terms (16), hence all of the
prototypical inalienable possession, but there is no occurrence for alien-
able possession in the data.

(15) *Ara-nnan   olo-m-d   tɛp   ača-l-ba-t.*
      old.man-ADE   head-ACC-POSS.3SG   away   chop-RES-PST.REP-3SG.O
      'They cut off the old man's head.'
      (Central: Narym, MNS_NN_BullSon_flk.037)

(16) *Waryə ne:ndə-nan   i:m-də   neptə-me-mba:-t*
      big   sister-ADE   son-ACC-POSS.3SG   name-TRL-PST.REP-3SG.O

      Stepan   Sareič'.
      Stepan   son.of.the.tsar
      'The son of the oldest sister was called Stepan Sareič'.'
      (Central: Narym, SDP_1964_FairytaleBlackZar_flk.060)

As in North Selkup, there are also occurrences with an unmarked pos-
sessor in the data for Central Selkup. They are only attested for inalienable
possession in Central Selkup: kinship terms (17) and body parts.

(17) […] *tab-ɨ-p   šedɛgut   amdɛlgup   nɛ-he*   […]
            3SG-EP-ACC   together   tsar   daughter-COM
      '[…] he together with the tsar's daughter […]'
      (Central: Narym, KIA_2014_GiantZobel_trans.044)

The combination of an unmarked possessor with a possessed marked with a possessive suffix can also be found, but only with kinship relations (and again, these finds are very rare, only 7 occurrences). It is not clear whether these really are unmarked possessive relations or rather compounds:

(18) *Iga*   *i-m-d*   *kwe-r-i-m-dɨt*   *Kitka.*
      child   son-ACC-POSS.3SG   call-FRQ-EP-PST.REP-3PL   Kitka
      'They called her grandson [lit. her child's son] Kitka.'
      (Central: Vasyugan, ChDN_1983_Nikita_flk.002)

4.3. South Selkup

Dependent marking with only the genitive on the possessor (i.e. no additional possessive suffix on the possessed) is the marking strategy most frequently used for body parts (19) and ownership (20) in South Selkup.

(19) *Tɨːɣoŋ*   *Iːd'ä-n*   *udə-nd*   *kwat'-o-n.*
      Tatar.prince   Itja-GEN   hand-ILL   catch-EP-3SG.S
      'The Tatar prince grabs for Itja's hands.'
      (South: Chaya, NN_1913_Itja_flk.040)

(20) *Iːt'e-n*   *tɨsse.*
      Itja-GEN   arrow
      'Itja's arrow.' (South: Ket, SVG_1964_Stars_flk.018)

These findings show that the preliminary analysis by Vorobeva et al. (2017: 57) stating that "affixation is obligatory in Southern Selkup for expressing inalienable possession" does not hold true, as there are regular occurrences for exactly the opposite (here shown in example (19), a total of 83 occurrences out of 141 for inalienable possession, see also Table 8). However, to mark kinship relations, the corpus data suggests that it is more common to use double marking: the genitive marking on the possessor and a possessive suffix on the possessed (21). Bekker et al. (1995a: 83) claim that double marking is the most common marking strategy in the Ket dialects of South Selkup for kinship terms and ownership. This holds true for kin relationships in all of South Selkup dialects, but there are only 7 occurrences for ownership marking.

(21) *Qɨba_nejd'e-n*   *n'eː-t*   *nɨ-nka-n*
forest.woman-GEN   daughter-POSS.3SG   stand-AOR-3SG.S

qəː-n-paːr-o-kən.
coast-GEN-top-EP-LOC

'The forest woman's daughter stands on the bank.'
(South: Middle Ob, PMP_1961_ForestWoman_flk.074)

Regarding the use of the adessive in South Selkup, there are more occurrences for the adessive with the possessive suffix than without: the adessive on the possessor and a possessive suffix on the possessed are used to mark body parts and ownership (22), but there are no instances found where kinship terms are marked with that strategy. The marking with only the adessive is attested for only two occurrences, both marking ownership (23).

(22) *Loɣa-nan*   *mat-tə*   *warɣ*   *e-k.*
fox-ADE   house-POSS.3SG   big   be-3SG.S

'The house of the fox is big.'
(South: Ket, TET_1979_TheHaresHouse_flk.009)

(23) [...] *sü-n*   *paja-ndɨ-nan*   *tul'd'ö-qɨn*   *wes*
dragon-GEN   woman-OBL.POSS.3SG-ADE   box-LOC   all

ündɨ-de-t.
hear-IPFV-3SG.O

'[...] in the box of the dragon's wife and hears everything.'
(South: Middle Ob, PMP_1961_Fairytale_flk.264)

Overall the occurrences of attributive possession marked with the adessive are again few in number (9 occurrences), especially when compared to the use of the genitive (172 occurrences). Therefore, the corpus data contradicts the claims of Bekker et al. (1995a: 83): the adessive (with or without an additional possessive suffix) is in fact not commonly used in the Ket and Ob dialects in attributive possessive constructions and it is also not used for all types of possession. On the contrary, the findings by Budzisch (2015) and the statement made by Vorobeva et al. (2017: 58) seem to hold true also with a larger data sample: "The occurrences with the lexical possessor marked for the locative [here: adessive] are very limited, the genitive case is by far the most prominent and common marker for nominal PNPs [here: lexical possessive NPs]" (Vorobeva et al. 2017: 58).

As in the other dialects, there are also a few occurrences with an un-marked possessor. Findings show two examples for body parts (24) and one for ownership, while the combination of an unmarked possessor with a possessive suffix on the possessed is found only for inalienable posses-sion: kinship (25) and body parts; it is again very low in frequency.

(24) *Tap as qunte-ŋ manni-mpa-t qiba_ned'eŋ oli-m.*
     3SG NEG long-ADV look-PST.REP-3SG.O forest.woman head-ACC
    'She did not look at the forest woman's head for long.'
    (South: Middle Ob, PMP_1961_ForestWoman_flk.065)

(25) *Tiːɣoŋ paja-t istol nʼakke-lže-t.*
    Tatar.prince woman-POSS.3SG table undress-TR-3SG.O
    'The Tatar prince's wife sets the table.'
    (South: Chaya, NN_1913_Itja_flk.039)

4.4. Overview of lexical possessors in Selkup

Regarding North Selkup, the findings in the data set corroborate the claims mentioned above by Kuznecova et al. (1980) about North Selkup: for lexical possessors, the picture is rather unequivocal – North Selkup mostly uses de-pendent marking. The possessor is marked with the genitive. Double mark-ing is indeed very rarely used, as mentioned by Bekker et al. (1995a: 83). Com-pletely unmarked possessive relations are very uncommon in North Selkup for lexical possessors, it seems to be the case in the corpus data that the pos-sessive relation has been introduced before and hence the need to mark it again explicitly is lowered, though this needs further research. The corpus data does not suggest that the type of possession plays a role in the selection of a marking option. In Table 6, the frequencies of the marking strategies found in the SLC corpus for lexical possessors in North Selkup are presented.

Table 6: Frequency of marking strategies for attributive possession with a lexical possessor in North Selkup

Type of possession	Dependent marking with genitive	Double marking with genitive and possessive suffix	Unmarked
Body parts	21	–	1
Kinship	23	2	–
Ownership	25	2	1

Central Selkup exhibits more options to mark lexical attributive possessive relations, but this dialect, too, mostly uses the marking with genitive on the possessor, sometimes with an additional possessive suffix. The other marking strategies are very low in terms of frequency, which is noteworthy (even though it has been mentioned in other studies as well) because the adessive marking strategy is commonly assumed to be of greater importance in Central and South Selkup than it is. One reason for that might be that it plays an important role in expressing predicative possession; it is therefore closely linked to the marking of possessive relations which might have led to the assumption that it is commonly used in attributive possession as well. Furthermore, in the qualitative analysis of possession in Selkup as carried out in the past, the unusual use of the adessive pronoun might have seemed interesting and therefore been mentioned. That might have led authors to overstate the use of the adessive pronoun slightly, as it is only now possible to search through a larger data set digitally, which makes statistical analysis much easier than was possible for descriptions such as Bekker et al. (1995a). The frequency of marking strategies for lexical possessors in Central Selkup is summarized in Table 7.

Table 7: Frequency of marking strategies for attributive possession with a lexical possessor in Central Selkup

Type of possession	Dependent marking with genitive	Double marking with genitive and possessive suffix	Dependent marking with adessive	Double marking with adessive and possessive suffix	Unmarked	Head marking with possessive suffix (+ unmarked possessor)
Body parts	48	17	3	4	8	–
Kinship	54	12	–	4	3	7
Ownership	14	2	–	–	–	–

The statements made above for Central Selkup also hold true for South Selkup: dependent marking with the genitive or double marking with the genitive and a possessive suffix are most commonly used in South Selkup as well. Again the adessive is of low importance, even though it is stated otherwise in Bekker et al. (1995a: 83). Interestingly, when looking at Table 8, it becomes clear that in South Selkup body parts and ownership are usually marked with only the genitive, while for kin relationships the additional use

of the possessive suffix is more common. South Selkup differs from Central and North Selkup in that respect. This is especially noticeable because one would expect an alienability split to appear between inalienable (body parts and kinship) and alienable (ownership), but not to separate these two categories. From a typological viewpoint, it is expected that if there are different strategies of marking, alienable possession would be more overtly marked (e.g. double marked) than inalienable possession (e.g. Haspelmath 2017). The assumption about an alienability split applies both to assuming that it is triggered by the iconic principle (see Haiman 1983) and to assuming that frequency of possessive marking is the deciding factor (see Haspelmath 2017). Kin relationships are commonly marked for possession in Selkup, and these expressions are relational, i.e. within them the relationship to another entity is expressed, therefore double marking is typologically remarkable because ownership is mostly dependent and therefore lesser marked in South Selkup. It can be concluded that the notion of alienability is not the decisive or sole factor here, but further research is needed.

Table 8: Frequency of marking strategies for attributive possession with a lexical possessor in South Selkup

Type of possession	Dependent marking with genitive	Double marking with genitive and possessive suffix	Dependent marking with adessive	Double marking with adessive and possessive suffix	Un-marked	Head marking with possessive suffix (+ unmarked possessor)
Body parts	75	7	–	4	2	3
Kinship	8	39	–	–	–	3
Ownership	36	7	2	3	1	–

In all three dialects, some occurrences of unmarked possessors (sometimes with a possessive suffix on the possessed) could be detected. It was briefly mentioned that it is quite unclear in some cases whether they are to be analyzed as attributive possession or compounds. But in most occurrences, there are no indications whatsoever that these expressions are commonly expressed with compounds in Selkup. Winkler (1913) and Sebestyén (1957) both have also reported that unmarked possession (or head-marked possession with an unmarked possessor) is to be expected in Samoyedic languages. Therefore, even though not frequent in use, it is fair to conclude that this does happen in Selkup.

Apart from the aforementioned strategies, there are constructions in which a "quasi" owner is marked with the adjective suffix *-l'*. It is however only rarely possible to read these occurrences as strictly attributive possession: examples such as *nu-l mat* [god-ADJZ house] are not interpreted as 'the house of god' here but rather as 'the godly house' or, simply, 'the god house' and hence are not seen as attributive possessive constructions. Some finds represent attributive possession less ambiguously, example (26) being one of them:

(26) […] *aj   midɨ-mba   Kat-man-puč-e-l   ando-nd.*
          and   reach-PST.REP.3SG.S Kat-Man-Puch-EP-ADJZ   boat-ILL
      '[…] and came to Kat-Man-Puch's boat.'
      (Central: Narym, KIA_2014_KatManPuch_trans.017)

These are very rare in the data under investigation in this study and mostly attested for a single speaker; all her texts are also written translations of Russian texts. Orlova (2018: 31) investigates the denominal *-l'* forms in North Selkup and finds that "in this [attributive] function they denote a kind of relation to the head of the NP such as: properties, possession, material etc.", but she does not provide any more detail on how much it is actually used for marking possession. A larger data set for North Selkup should be investigated for this. Under the present circumstances, this cannot be seen as a common way to mark possessive relations in the sense applied in this study.

Overall, the results of the quantitative evaluation show that Selkup dialects show agreement in many points but also behave differently in some respects, e.g. Central and South Selkup use adessive marking, and South Selkup tends to double mark kin relationships with a lexical possessor while the other dialects do not. They also show that a corpus-based approach is worthwhile even for topics that seem to be well studied. As mentioned in the introduction, attributive possession in Selkup is not a "new" topic, but the findings with the new possibilities of involving digital corpora differ from older descriptions.

Comparing the outcome of this analysis to the other Samoyedic languages mentioned above, it becomes clear that Selkup fits the picture rather well: it too mostly uses dependent marking with the genitive like the related Samoyedic languages do. Double marking with additional possessive suffixes is possible, especially in Central and South Selkup; but the use of the suffix never seems to be obligatory. In Koptjevskaja-Tamm (2003: 641f.), it is shown that many other Uralic languages also show a similar pattern between dependent marking, double marking and juxtaposition.

5. Non-lexical possessors

For non-lexical possessors, there are several ways to express them in all Samoyedic languages. In North Samoyedic languages, they are marked with an obligatory possessive suffix (head marking) but a pronoun is optionally possible, making double marking an option. There are no special possessive pronouns, but personal pronouns in the nominative are used (see Nikolaeva (2014: 142ff.) for (Tundra) Nenets, Siegl (2013: 233f.) for (Forest) Enets and Wagner-Nagy (2019: 317) for Nganasan). The extinct South Samoyedic language Kamas shows head marking, dependent marking and double marking (Künnap 1999b: 14–16, 33), while for Mator, the use of possessive suffixes is reported as well as a form of possessive pronouns, but there is not enough data to speak with certainty about the use in attributive possessive constructions (Helimski 1997: 141). For all Selkup dialects, a split between the marking strategies for first- and second-person possessors, on the one hand, and third-person possessors, on the other hand, is reported, with head, dependent and double marking in different combinations. Whether these claims can be validated against the corpus data will be examined in the following sections.

Additionally, Tauli (1966: 65) suggested that there might be a process taking place in all Uralic languages, but Samoyedic languages especially, from head marking to double marking to dependent marking; but while it is true that the use of pronouns in Samoyedic languages is very rare in early data (Sebestyén 1957: 45), according to the recent descriptions, it still is rare in the more recent data. The general idea of a grammatical process seems to be out of place here, as possessive suffixes are still widely used in the North Samoyedic languages and also in Selkup, as the following sections will show.

5.1. North Selkup

For North Selkup, Kuznecova et al. (1980: 187) describe that for first- and second-person possessors the possessive relation is always marked with a possessive suffix on the head of the phrase. The corpus data shows indeed that head marking is prominently used in these regards (29 occurrences out of 39 in total), with kinship (27), body parts (28) and ownership (29).

(27) *Onäk   qos   qu-lä-k,   iːja-m   ili-ŋiː.*
INT.1SG   DEF   die-OPT-1SG.S   child-POSS.1SG live-AOR.3SG.S
'I shall die, my child shall live.' (North: Taz, AAI_1973_Okyle_flk.057)

(28) *Kətsat,   saja-l   ıllä   sɔːr-ät!*
grandson   eye-POSS.2SG   down   bind-IMP.2SG.O
'Boy, blindfold your eyes!' (North: Taz, NN_197X_YoungBoy_flk.044)

(29) *Mat   or-qil-sa-p   ınnä   čatti-sa-p   qɔːsi-l'*
1SG   force-MULO-PST-1SG.O   up   throw-PST-1SG.O crust-ADJZ
anta-qäk.
boat-ILL.1SG
'I grabbed him and threw him into my boat made of bark.'
(North: Taz, BEP_1973_IchaAndNenets_flk.012)

Furthermore, Kuznecova et al. (1980: 187) describe that additionally a pronoun might be used, and indeed double marking is attested for all types of possession, but far less frequent (10 occurrences) than head marking. The following examples show this for body parts (30) and ownership (31).

(30) *Man   untal-tɛnta-p   tat   ola-mtɨ?*
1SG   look.for.lice-INFER.FUT-1SG.O 2SG.GEN head-ACC.2SG
'When shall I search your head for lice?'
(North: Taz, BVP_1973_East_flk.014)

(31) *Tan   man   kapkat-qäk   tuːtɨra-ntɨ.*
2SG   1SG.GEN trap-ILL.1SG get-2SG.S
'You fell into my trap.' (North: Taz, MIV_1977_Icha_flk.011)

The use of only the genitive pronoun is not attested in the data for first and second person, as is expected from the aforementioned descriptions in the grammar by Kuznecova et al. (1980).

For third-person possessors, the situation allegedly differs: Kuznecova et al. (1980: 187, 288) state that only dependent marking is used, double and head marking is ungrammatical. The data in this study shows that dependent marking is very rarely used in these cases: there are two occurrences of the expression of ownership with merely the genitive pronoun (32), but there are no cases for inalienable possessions.

(32) Təp-a-n   mɔːt-tɨ   tü-ŋɔː-qɩ   šittɨ   qum-oː-qɩ.
      3SG-EP-GEN   tent-ILL   come-AOR-3DU.S   two   person-EP-DU
      'There come two people into her house.'
      (North: Taz, BEP_1973_Fat1_flk.003)

Contrary to the grammatical description offered, inalienable possession
for third person is solely head marked with just the possessive suffix as
shown for kinship (33) and body parts (34). Ownership (35) is also more
frequently marked this way (29 occurrences).

(33) Nɨːnɨ   iːja-tɨ   or-i-m-mintɨ.
      then   child-POSS.3SG   force-EP-TRL-PST.INFER.3SG.S
      'His son has grown up.' (North: Taz, AAI_1973_Okyle_flk.043)

(34) Üŋkɨlsa-tɨ   ukoːt   kuttar   ɛː-sɨ   nɨlči-ŋ
      ear-POSS.3SG   earlier   how   be-PST.3SG.S such-ADV

      ej   ɛː-ŋa.
      and   be-AOR.3SG.S
      'Her ears were again there where they had been earlier.'
      (North: Taz, AVA_1973_Ichakicha_flk.072)

(35) Ɨtɨ-t   inni-m-t
      evening-LOC.ADV   bow-ACC-POSS.3SG

      äti-tɨ-tɨ.
      lay.arrow.at.bowstring-INFER-3SG.O
      'In the evening he sets an arrow in his bow.'
      (North: Taz, BIV_1941_KonMytyke_flk.019)

Double marking for third-person possessors is not found in the data.

5.2. Central Selkup

Non-lexical possessors in Central Selkup also behave differently with re-
spect to the person of the possessor. First- and second-person possessors
are, according to Bekker et al. (1995a: 71), double marked by default. The
analysis of the corpus showed that indeed the combination of a genitive
pronoun and a possessive suffix on the second part of the construction is
widely used (74 occurrences) to mark inalienable (kinship (36) and body
parts (37)) and also alienable (38) possession for first and second person.

(36) *Tat   mi   iː-ut.*
2SG   1DU.GEN   son-POSS.1PL
'You are our son.' (Central: Narym, MNS_NN_BullSon_flk.188)

(37) *Əgə   pacʼa-naː-də   miː   oll-ut.*
NEG   chop-AOR-3PL   1PL.GEN   head-POSS.1PL
'Do not chop our heads off.'
(Central: Narym, SDP_1964_FairytaleBlackZar_flk.1067)

(38) *Tat   tolʼžʼe-nd-ə   konne   čanžɨ-gu   aː*
2SG.GEN   ski-OBL.POSS.2SG-INS   upwards   go.out-INF   NEG

tan-wa-k.
know-AOR-1SG.S
'I cannot go up with your skis.'
(Central: Narym, MNS_1984_BrotherSister_flk.045)

Head marking is also accounted for in 74 occurrences; for kinship terms (39) this marking strategy is even the most frequently used (over 45%), but body parts (40) and ownership (41) are frequently only head marked as well.

(39) *Amba-ut   mʼiɣenɨt   čenča   […]*
mother-POSS.1PL   1PL.DAT   say.3SG.S
'Our mother tells us [...]'
(Central: Vasyugan, ChDN_1983_MistressOfFire_flk.003)

(40) *Куза-t   nebe-l   ak-t   [...]*
when-LOC.ADV   breast-POSS.2SG   mouth-ILL
'If your breast is in my mouth […]'
(Central: Narym, MNS_NN_BullSon_flk.187)

(41) *A   tüšše-m   qošhatelʼ.*
but   rifle-POSS.1SG   bad
'But my rifle is bad.'
(Central: Tym, PAV_NN_HowIBearCatch_nar.112)

Even though mentioned by Bekker et al. (1995a: 71) as being unusual, dependent marking with the genitive is indeed also found for all types of possession (34 occurrences) and it is the strategy used most often for alienable possession (42), though for inalienable possession it is used very rarely.

(42) *M'i*   *tedomi-p*   *ab-a-l-de.*
      1PL.GEN   thing-ACC   burn-EP-RES-3SG.O
      'She is going to burn our stuff.'
      (Central: Vasyugan, ChDN_1983_MistressOfFire_flk.005)

As with lexical possessors, the marking with the adessive is rarely used outside of predicate possession (only 8 occurrences for adessive possessors in attributive possession). The use of only the adessive pronoun is, at least in the data set, reserved for alienable possession (43), while inalienable possession (44) is additionally marked with a possessive suffix; the use of the possessive suffix was described as facultative by Bekker et al. (1980: 82).

(43) *Ma-nan*   *tüšše*   *swa*   *e-za.*
      1SG-ADE   rifle   good   be-PST.3SG.S
      'My rifle was good.'
      (Central: Tym, PAV_NN_HowIBearCatch_nar.113)

(44) *Ma-nan*   *aǯ'a-m*   *quː-mba.*
      1SG-ADE   father-POSS.1SG   die-PST.REP.3SG.S
      'My father had already died.'
      (Central: Narym, SAI_1984_StoryAboutLife_nar.005)

Bekker et al. (1995a: 80) also make statements about the third-person possessor: it is said that for alienable possession only the genitive pronoun is to be used, while with inalienable possession the pronoun is usually paired with the possessive suffix of the according person. In the data, however, head marking is in the vast majority of cases attested (89% in total) for all types of possession with a third-person possessor, i.e. there is no sign of an alienability split (see kinship (45), body parts (46) and also with ownership (47)).

(45) *Tɨmn'a-d*   *oqo-nči-la*   […]
      brother-POSS.3SG   ask-IPFV-OPT.3SG.S
      'His brother asked […]' (Central: Tym, JIF_1968_Kamacha_flk.012)

(46) *Tab pah-e*   *pargɨ-m-de*   *mal*   *koro-mba-d.*
      3SG knife-INS stomach-ACC-POSS.3SG separately   cut-PST.REP-3SG.O
      'He cut his belly open with a knife.'
      (Central: Vasyugan, ChDN_1983_Nikita_flk.044)

(47) *Poqqo-qɨntɨ*   *mukka-dɨ-l'čə.*
     net-LOC.POSS.3SG   press-TR-INT.PFV.3SG.S
     'He puts him into his net.'
     (Central: Vasyugan, ChDN_1983_ItjasTown_flk.009)

The second most used strategy which is far less used than head marking in the data is, again for all types of possession, dependent marking, i.e. only with the pronoun in the genitive, as shown here for kinship (48) and the expression of ownership (49). The latter one is expected from the grammatical description quoted before, but the marking of inalienable possession on the possessor alone, not the possessed, is not expected in Bekker et al. (1995a).

(48) […] *ku*   *qwan-ni-mbaːdɛt*   *n'aːb-la*   *tab-ɨ-t*
          where   go.away-TR-HAB-3PL   duck-PL   3SG-EP-GEN

     čemn'a-lika-p.
     brother-DIM-ACC
     '[...] where the wild ducks had taken her brother.'
     (Central: Tym, TTD_1964_WildDucks_flk.019)

(49) *Ög*   *mannö-mbeːd*   *täb-ö-t*   *porg!*
     NEG.IMP   look-HAB-IMP.2SG.S   3SG-EP-GEN   fur.coat
     'Don't look at his coat!'
     (Central: Tym, KAO_1912_SmartPeople_trans.010)

Double marking with the genitive pronoun and the possessive suffix is very rarely attested in the data set for a third-person possessor, but if it is used, then it is mostly for kinship terms (50).

(50) *Nača-t*   *tab-ə-t*   *n'eweːsta-d*   *warke-špa.*
     there-LOC.ADV   3SG-EP-GEN   bride-POSS.3SG   live-IPFV.3SG.S
     'His bride lived there.' (Central: Tym, TTD_1964_Frog_flk.010)

Adessive pronouns combined with a possessive suffix are used only for inalienable possession (51), but without the possessive suffix the picture is more ambiguous, as there are also examples for body parts (52) and ownership (53).

(51) *Tab-ɨ-nnan*   *fa*   *paja-d*   *šandɛ*   *ku-mba.*
     he-EP-ADE   good   woman-POSS.3SG   young   die-PST.REP.3SG.S
     'His beautiful wife died young.'
     (Central: Vasyugan, ChDN_1983_HerosDaughter_flk.002)

(52) *Tab-la   haže-špa-dɛt   ku-la-p   huru-la-p*
      3SG-PL   sting-IPFV-3PL   person-PL-ACC   wild.animal-PL-ACC

      i-r-i-lža-t   tab-ɛ-nan   kap.
      take-FRQ-EP-INT.PFV-3SG.O   3SG-EP-ADE   blood

      'They bite people, animals, suck their blood.'
      (Central: Narym, KIA_2014_ItjaKet_trans.138)

(53) *Tab   tab-i-nnan   nodɨk   i-mba-d.*
      3SG   3SG-EP-ADE   sterlet   take-PST.REP-3SG.O

      'He took his sterlet [*Acipenser ruthenus*].'
      (Central: Vasyugan, ChDN_1983_Nikita_flk.029)

5.3. South Selkup

For South Selkup, the same statements are made as for Central Selkup: first- and second-person possession should be double marked, dependent marking is not common (Bekker et al. 1995a: 71). Putting the corpus data to a test shows that, despite the claims, the most widely used strategy is head marking (112 occurrences) as exemplified here for kinship (54), body parts (55) and ownership (56).

(54) *Me   pal'd'u-z-ot   qottɨja-m-ni.*
      1PL   go-PST-1PL   grandmother-POSS.1SG-DAT

      'We went to my grandmother.'
      (South: Ket, KMS_1966_TwoSisters_flk.086)

(55) *Ol'o-l'   nano   warsapa-ŋ   as   je-ŋ.*
      head-POSS.2SG   then   shaggy-ADV   NEG   be-3SG.S

      'Your head is not shaggy.'
      (South: Middle Ob, PMP_1961_ForestWoman_flk.153)

(56) *I   labo-m   töt-qan   qala.*
      and   oar-POSS.1SG   shit-LOC   stay.3SG.S

      'And my oar remains stuck in the shit.'
      (South: Ket, TFF_1967_ItjaAldigaScale_flk.031)

Double marking with the genitive pronoun and the corresponding possessive suffix is the second most used strategy. It is mostly used to mark inalienable possession (53 occurrences, see 57); alienable possession is not often marked that way (7 occurrences).

(57) *Tan*   *eze-l*   *qo-u-mba-n.*
      2SG.GEN   father-POSS.2SG   rich-be-HAB-3SG.S
      'Your father is rich.' (South: Middle Ob, PMP_1961_Fairytale_flk.029)

Dependent marking with the genitive pronoun is, as also stated by Bekker et al. (1995a: 71), not often found in the data (21 occurrences), but used for all types of possession; an example of the marking of ownership is shown in (58).

(58) *Mat*   *ma:t-qɨt*   *warga.*
      1SG.GEN   house-LOC   live.3SG.S
      'He lives in my house.' (South: Ket, TET_1979_TheHaresHouse_flk.019)

However, what is more often used in South Selkup is the marking with the adessive pronoun (53 occurrences), which appears in two forms in South Selkup dialects: with just the adessive suffix *-nan* and with an additional adjective ending *-nan-i*. Especially the latter form is often used in combination with a possessive suffix to mark kin relationships (59); without the suffix it is rarely used. The plain adessive pronoun is, as in Central Selkup, seldom used with or without an accompanying possessive suffix, but there are some occurrences for all types of possession, e.g. for ownership as in (60).

(59) *Man-na:n-i*   *i:-m*   *pe:ge.*
      1SG-ADE-ADJ   son-POSS.1SG   hazel.grouse
      'My son is a hazel grouse.'
      (South: Middle Ob, SEV_1980_HazelGrouse_flk.042)

(60) *Nännɨ tat i:-ča-l*   *meŋ-nan so:*   *kündä qorrä-m.*
      then   2SG take-FUT-2SG.O 1SG-ADE   good horse stallion-ACC
      'Then you will take my good stallion.'
      (South: Ket, AGS_1968_FairytaleSnake_flk.057)

For third-person possessors, the claim by Bekker et al. (1995a: 80) is that alienable possession is marked with only the genitive pronoun, while inalienable possession is double marked with the pronoun and possessive suffix. But as for North and Central Selkup, most possessive constructions are only head marked with the possessive suffix: examples for kinship (61), body parts (62) and ownership (63) are given.

(61) *Paja-la-t*   *as*   *wesa-q-wa-t.*
      old.woman-PL-POSS.3PL   NEG   get.up-ITER-AOR-3PL
      'Their women do not get up.'
      (South: Ket, KKN_1971_FiveCarpBrothers_flk.080)

(62) *I:t'a   wanč'ə-m   ora-nnə-t   olo-u-n-tə.*
     Itja   sheefish-ACC   hold-AOR-3SG.O   head-EP-ACC-POSS.3SG
     'Itja grabs the sheefish on its head.'
     (South: Chaya, NN_1879_Iitja_flk.027)

(63) *Nu   poŋgi-la-m-t   qwädi-t.*
     now   net-PL-ACC-POSS.3SG   leave-3SG.O
     'Now he places his nets.'
     (South: Middle Ob, PMP_1967_AboutItja_flk.028)

With a large margin in terms of frequency, double marking with the genitive pronoun and possessive suffix is the second-most used strategy, again attested for all types of possession, be it inalienable (64) or alienable (65).

(64) […] *täb-ɨ-n   n'äryu-mbədi   wando-yindi*   […]
          3SG-EP-GEN   redden-PTCP.PST   face-ILL.POSS.3SG
     '[…] at his red face […]' (South: Ket, KMS_1967_Antosja_nar.030)

(65) *Täp-a-n   ma:ki-t   ä-sa-n   piwu-pti-mbiti*
     3SG-EP-GEN   stick-POSS.3SG   be-PST-3SG.S   entangle-CAUS-PTCP.PST

     t'ümbu   pata-j   n'ütš-se.
     long   green-ADJZ   grass-INS
     'His stick was entangled in the long green grass.'
     (South: Ket, MMP_1964_Lgov_trans .129)

The marking with only the genitive pronoun is rarely used. Also, the adessive pronoun is, as stated previously, very rarely used to mark attributive possession. If accompanied by a possessive suffix, it is used for inalienable (66) and alienable possession, while if it appears without a possessive suffix, it is used for only the latter category (67).

(66) *Täb-ə-sta-yə-nan   i-ttə   ne-t   warya-n*
     3SG-EP-CR-DU-ADE   son-POSS.3DU   daughter-POSS.3DU   big-ADV

     azu-ba-ɣe.
     become-PST.REP-3DU.S
     'Their son and daughter grew up.'
     (South: Upper Ob, PVD_1961_FarmAssault_flk.028)

(67) *A   tab-u-nan   mat-qən   qiba i:-t*
     but   3SG-EP-ADE   house-LOC   small son-POSS.3SG

     t'eli-mpa-n.
     give.birth-PST.REP-3SG.S
     'In his house his little son had been born.'
     (South: Middle Ob, PMP_1966_BoyDevil_flk.017)

5.4. Overview of non-lexical possessors in Selkup

While, as shown in Section 4.1, in North Selkup, there were no surprises in the data with regard to the possessive constructions with a lexical possessor, constructions with a non-lexical possessor differ from what was expected from descriptions such as Kuznecova et al. (1980: 187), mostly concerning third-person possessors. Concerning first- and second-person non-lexical possession, the data shows that no matter what kind of possessive relation is expressed, head marking is the most commonly used option with an optional pronoun. For third-person possession, the claim of Kuznecova et al. (1980: 187) that it is marked with only the genitive pronoun and that double and head marking is ungrammatical, presents itself as a statement which only partly holds true: double marking is not attested in the data. However, by far the most frequent – and for inalienable possession the only – option is the marking with solely the possessive suffix attached to the head of the phrase. In Table 9, the findings are summarized, separated for person, type of possession and marking strategy.

Table 9: Frequency of marking strategies for attributive possession with a non-lexical possessor in North Selkup

Person	Type of possession	Dependent marking with genitive pronoun	Double marking with genitive pronoun and possessive suffix	Head marking with possessive suffix
1	Body parts	–	–	3
	Kinship	–	1	8
	Ownership	–	4	6
2	Body parts	–	3	3
	Kinship	–	1	7
	Ownership	–	1	2
3	Body parts	–	–	16
	Kinship	–	–	55
	Ownership	2	–	29

When looking at non-lexical possessors in Central and South Selkup, the picture is less clear than for North Selkup. One reason for this is simply that there are more marking strategies: the adessive pronoun (the plain adessive pronoun and the one with the additional adjective ending are

combined in Table 11 for clarity) with or without a possessive suffix added to the possessed can additionally be used to express attributive possession.

In Central Selkup, Bekker et al. (1995a: 71) claim that double marking is the default strategy for first- and second-person possessors, but this does not hold true: both head and double marking (with genitive) are the most commonly used strategies; for body parts and ownership with first-person possessors, dependent marking seems to be a widely-used option as well, but the data set is too limited to give a definite answer on the importance of that strategy. For third-person possessors, even though not as widely manifested as in North Selkup, head marking is by far the most extensively used strategy to mark attributive possession in Central Selkup as well. Again, as for North Selkup, this runs contrary to the descriptions which claim that a pronoun (genitive or adessive) has to be used for every type of possession. The findings for the SLC corpus data are summed up in Table 10 for Central Selkup.

Table 10: Frequency of marking strategies for attributive possession with a non-lexical possessor in Central Selkup

Person	Type of possession	Dependent marking with genitive pronoun	Double marking with genitive pronoun and possessive suffix	Dependent marking with adessive pronoun	Double marking with adessive pronoun and possessive suffix	Head marking with possessive suffix
1	Body parts	11	14	–	2	4
	Kinship	11	40	–	2	43
	Ownership	8	4	2	–	5
2	Body parts	1	2	–	1	5
	Kinship	2	13	–	1	17
	Ownership	1	1	–	–	–
3	Body parts	14	1	4	2	128
	Kinship	15	7	–	1	178
	Ownership	8	3	1	–	126

For South Selkup, the same claim has been made as for Central Selkup: first- and second-person possessors are to be double marked. But the data showed that for these persons combined, head marking is the most prominently used strategy to mark attributive possession, while double marking mainly comes in only second place in terms of frequency.

With regard to the marking of third-person possessors, South Selkup shows the same interesting phenomenon: the corpus data deviates significantly from the grammatical description, although both are based on comparable data. For all types of possession, the majority is marked with only the possessive suffix, while double marking or dependent marking are very low in frequency. The results for South Selkup are shown in Table 11.

Table 11: Frequency of marking strategies for attributive possession with a non-lexical possessor in South Selkup

Person	Type of possession	Dependent marking with genitive pronoun	Double marking with genitive pronoun and possessive suffix	Dependent marking with adessive pronoun	Double marking with adessive pronoun and possessive suffix	Head marking with possessive suffix
1	Body parts	7	11	4	5	13
	Kinship	2	27	5	12	43
	Ownership	10	6	3	3	11
2	Body parts	1	4	–	–	9
	Kinship	–	11	2	17	22
	Ownership	1	1	–	2	14
3	Body parts	6	8	–	2	119
	Kinship	1	12	–	1	355
	Ownership	4	4	4	1	125

In summary, it can be said that the dialects show differences (e.g. the adessive in Central and South), but they have in common that regardless of person or type of possession, head marking is most frequently used, followed by double marking. The split in first-/second- and third-person possessor pronouns does not seem to be necessary as they do not show significantly different behaviors. Still, for all dialects, one has to keep in mind that there are not as many occurrences in the corpus for first and second person as for the third person. This is due to fact that the corpus is mostly compiled from folklore texts, i.e. texts that are told about another person, and therefore evidence is lacking for first and second possessors. This is especially true when trying to detect if alienability is a key factor in choosing a marking strategy. For example, the data presented in Table 10 for Central Selkup and also in Table 11 for South Selkup suggests that ownership for first-person possessors

has a tendency to be marked only with a genitive pronoun, but more occurrences would be needed to be sure of that. As of now, there is no evidence that the type of possession plays a role in picking a marking strategy.

It can therefore be stated that Selkup fits rather well into the Samoyedic framework in which possessive suffixes are to be used and pronouns can optionally be used, even if the grammars suggested differences. Obviously, there are exceptions, but the use of possessive suffixes in Selkup is very frequent.

6. Conclusion

This study is based on a closed data set and the absolute numbers presented in Tables 6–11 obviously only hold true for this specific corpus, but the corpus is big enough to give some ideas about the quantity of marking strategies in general and also shows that some are used extensively, even though grammatical descriptions deemed them to be ungrammatical, while others are less frequent than suggested.

For lexical possessors, the most frequently used marking strategy is cross-dialectal dependent marking: The genitive is attached to the possessor. In most of these cases, the possessed is not marked with a possessive suffix, except in South Selkup, where kinship terms are rather often double marked.

Non-lexical first- and second-person possessors are in a large majority only head marked: a possessive suffix is attached to the possessed, while double marking with the genitive pronoun is the second-most used marking, except for marking ownership in Central and South Selkup where the second-most used marking is dependent marking with the genitive pronoun, but the data set is rather small for that. The most interesting findings concern third-person possessors, because the study could show that these are also predominantly head marked and that the use of pronouns is of no importance in the marking regardless of the type of possession, even though it has been claimed for both North (Kuznecova et al. 1980: 187) and for Central and South Selkup (Bekker et al. 1995a: 80) that dependent marking is the most used strategy. The data presented here shows that it is not necessary to split first-, second- and third-person possessors, as they behave the same.

Adessive marking is used in Central and South Selkup for lexical and non-lexical possessors, but very rarely. It takes a larger role in marking predicative possession, which is not part of this study.

In context with the other Samoyedic languages, the Selkup data suggest that they all exhibit nearly the same features: dependent marking for

lexical possessors, head marking for non-lexical possessors; an alienability split cannot be detected for any Samoyedic language. If one looks not only at the Samoyedic languages but at the Uralic languages as a whole, Koptjevskaja-Tamm (2003: 642) showed for e.g. Mari that the syntactic function is of importance when picking a marking strategy, while in Mordvin the definiteness of a referent is taken into account. The Samoyedic languages in general and also Selkup show no indication that these factors are crucial, at least in the attributive constructions presented here.[3] Altogether, Selkup's features are not too surprising and fit with the typological overviews as presented in Koptjevskaja-Tamm (2003) as well as Nichols and Bickel (2013) with regard to lexical possessors well.

Overall, the evaluation of the corpus has shown that even though attributive possession is a topic rather well studied for Selkup, there are still interesting finds and also still open questions that need further research. There is also a need to incorporate more data, especially for North Selkup in general and for first and second possessors in Central and South dialects, in order to get a broader picture.

Non-standard abbreviations used in glosses

ADE	adessive	INFER	inferential
ADJZ	adjectivizer	INT	intensifier
AOR	aorist	ITER	iterative
CR	connective-reciprocal	MULO	multiobjective derivation
CVB	converb	O	objective conjugation
DIM	diminutive	OPT	optative
EP	epenthetic vowel	REP	reportative mood
FRQ	frequentative	S	subjective conjugation
HAB	habituative	TRL	translative
ILL	illative		

References

Behnke, Anja & Budzisch, Josefina. 2021. *Selkup Language Corpus* (Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology 6). Szeged: Department of Finno-Ugric Studies. https://doi.org/10.14232/wpcl.2021.6

Bekker, Erika & Alitkina, Larisa & Bykonja, Valentina & Il'jašenko, Irina. 1995a. *Morfologija sel'kupskogo jazyka. Južnye dialekty. Čast' 1.* Tomsk.

---

3. See e.g. Ovsjannikova (2020) for nominative and oblique (genitive) possessors in Enets. The first type is mostly used in predicative possession (and to some extend in "describing the state of the possessor's body parts"; Ovsjannikova 2020: 93).

Bekker, Erika & Alitkina, Larisa & Bykonja, Valentina & Il'jašenko, Irina. 1995b. *Morfologija sel'kupskogo jazyka. Južnye dialekty. Čast' 2.* Tomsk.

Budzisch, Josefina. 2015. Possession in Southern Selkup. *Tomsk Journal of Linguistics and Anthropology* 4(10). 45–50.

Budzisch, Josefina. 2017. On the non-possessive use of possessive suffixes in Central and Southern Selkup. *Ural-Altaic Studies Scientific Journal* 25(2). 58–66.

Budzisch, Josefina. 2021. *Definitheit im Selkupischen* (Studia uralo-altaica 55). Szeged: Finnisch-Ugrisches Institut.

Budzisch, Josefina & Harder, Anja & Wagner-Nagy, Beáta. 2019. *Selkup Language Corpus (SLC).* Archived in Hamburger Zentrum für Sprachkorpora. Version 1.0.0. Publication date 2019-02-08.
http://hdl.handle.net/11022/0000-0007-D009-4.

Bykonja, Valentina & Kuznecova, Nadežda & Maksimova, Natal'ja. 2005. *Sel'kupsko-russkij dialektnyj slovar'.* Tomsk: Izdatel'stvo Tomskogo gosudarstvennogo pedagogičeskogo universiteta.

Gluškov, Sergej & Bajdak, Aleksandra & Maksimova, Natal'ja. 2013. Dialekty sel'kupskogo jazyka. In Tučkova, Natal'ja & Gluškov, Sergej & Košeleva, Elena & Golovnëv, Andrej & Bajdak, Alexandra & Maksimova, Natal'ja (eds.), *Sel'kupy. Očerki tradicionnoj kul'tury i sel'kupskogo jazyka*, 49–63. Tomsk: Izdatel'stvo Tomskogo politechničeskogo universiteta.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59. 781–819.

Haspelmath, Martin. 2017. Explaining alienability contrasts in adpossessive constructions: Predictability vs. iconicity. *Zeitschrift für Sprachwissenschaft* 36(2). 193–231. https://doi.org/10.1515/zfs-2017-0009

Helimski, Eugen 1997. *Die matorische Sprache* (Studia uralo-altaica 41). Szeged: Finnisch-Ugrisches Institut.

Kim, Alexandra. 1985. Ausdrucksweisen der Possessivfügungen im Samojedischen. *Linguistics* 1. 50.

Kim-Maloney, Alexandra & Kovylin, Sergej. 2015. Taksonomija jazykovych sredstv vyraženija posessivnosti v sel'kupskom jazyke. *Tomsk Journal of Linguistics and Anthropology* 4(10). 36–44.

König, Ekkehard. 2001. Internal and external possessors. In Haspelmath, Martin & König, Ekkehard & Oesterreicher Wulf, & Raible, Wolfgang (eds.), *Language typology and language universals* (HSK 20.2), 970–978. Berlin: de Gruyter. https://doi.org/10.1515/9783110171549.2

Koptjevskaja-Tamm, Maria. 2001. Adnominal possession. In Haspelmath, Martin & König, Ekkehard & Oesterreicher Wulf, & Raible, Wolfgang (eds.), *Language typology and language universals* (HSK 20.2), 960–970. Berlin: de Gruyter. https://doi.org/10.1515/9783110171549.2

Koptjevskaja-Tamm, Maria. 2003. Possessive noun phrases in the languages of Europe. In Plank, Frans (ed.), *Noun phrase structure in the languages of Europe*, 621–722. Berlin: Mouton de Gruyter.

Körtvély, Erika. 2010. On the function of possessive suffixes in Tundra Nenets: Possession and semantic-pragmatic definiteness. *Finnisch-Ugrische Mitteilungen* 32/33. 321–343.

Künnap, Ago. 1999a. *Enets* (Languages of the World 186). München: Lincom.

Künnap, Ago. 1999b. *Kamass* (Languages of the World 185). München: Lincom.

Kuper, Shimon. 1986. Mesto ketskogo dialekta v sisteme selʹkupskogo jazyka. In Nadeljaev, V. (ed.), *Fonetika jazykov Sibiri i sopredelʹnych regionov*, 103–106. Novosibirsk: Nauka.

Kuznecova, Ariadna & Helimski, Eugen & Gruškina, Elena. 1980. *Očerki po selʹkupskom jazyku. Tazovskij dialekt*. Tom 1. Moskva: Izdatelʹstvo Moskovskogo universiteta.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62. 56–119.

Nichols, Johanna & Bickel, Balthasar. 2013. Locus of marking in possessive noun phrases. In Dryer, Matthew & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/24) (Accessed 2021-01-13.)

Nikolaeva, Irina. 2014. *A grammar of Tundra Nenets* (Mouton Grammar Library 65). Berlin: De Gruyter. https://doi.org/10.1515/9783110320640

Orlova, Svetlana. 2018. Denominal *lʼ*-forms in Northern Selkup: syntax, morphology and semantics. *Tomsk Journal of Linguistics and Anthropology* 3(21). 31–40. https://doi.org/10.23951/2307-6119-2018-3-31-40

Ovsjannikova, Maria. 2020. Oblique and nominative nominal possessors in Forest Enets. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 11. 57–98. https://doi.org/10.12697/jeful.2020.11.2.03

Sebestyén, Irene. 1957. Die possessiven Fügungen im samojedischen und das Problem des uralischen Genitivs (I.). *Acta Linguistica Academiae Scientiarum Hungaricae* 7(1/2). 41–77.

Siegl, Florian. 2013. *Materials on Forest Enets, an indigenous language of Northern Siberia* (Memoires de la Société Finno-Ougrienne 267). Helsinki: Finno-Ugrian Society.

Tauli, Valter. 1966. *Structural tendencies in Uralic languages* (Uralic and Altaic series 17). The Hague: Mouton.

Vorobeva, Victoria & Novitskaya, Irina & Girfanova, Ksenia & Vesnin, Vitalij. 2017. Adnominal possessive constructions in Narym, Vasjugan and Middle-Ob dialects of Selkup. *Linguistica Uralica* 53. 54–64. https://doi.org/10.3176/lu.2017.1.05

Wagner-Nagy, Beáta. 2014. Possessive constructions in Nganasan. *Tomsk Journal of Linguistics and Anthropology* 1(3). 76–82.

Wagner-Nagy, Beáta. 2019. *A grammar of Nganasan* (Grammars and Language Sketches of the World's Languages 6). Leiden: Brill. https://doi.org/10.1163/9789004382763

Winkler, Heinrich. 1913. Samojedisch und finnisch. Zweite Mitteilung. *Finnisch-ugrische Forschungen* 13. 120–178.

Zayzon, Réka. 2015. Observations on non-possessive usages of personal markers (possessive suffixes) in Nganasan. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6(2). 259–278.

Chris Lasse Däbritz
University of Hamburg

Typology of number systems in languages of Western and Central Siberia

This paper investigates the linguistic expression of number in seven languages from Western and Central Siberia. In a first step the number system of each language is described in detail, and afterwards the most relevant convergences and divergences of the languages are dealt with. Three particularly interesting phenomena are discussed in more detail: First, it is shown that the concept of general number, denoting noun forms underspecified for number, is able to account for a range of related phenomena (unmarked noun forms after numerals, nouns denoting paired objects). Second, singulatives in Selkup, Ket and partly Eastern Khanty are analyzed, whereby it is argued that their similar morphosyntactic and grammaticalization patterns allow for analyzing them as a contact phenomenon. Third, two splits on the animacy hierarchy between the first and second person in Dolgan as well as Chulym Turkic are presented. Finally, the results are evaluated against a broader areal-typological background, whereby it is shown that the category of number does not support any larger areal groupings within Western and Central Siberia, but that the analyzed languages rather adhere to patterns of number marking present all over Northern Eurasia.

1. Introduction

In his much remarked paper on the usage of number markers in the Uralic languages, Paavo Ravila makes the following remarkable statement:

> Das Numerussystem der finnisch-ugrischen und samojedischen Sprachen ist unstreitig entwickelter als das der altaischen […]. So ist der Gebrauch des Singulars als absoluter Numerus in allen uralischen Sprachen ziemlich allgemein, z. B. nach dem Zahlwort erscheint der Singular, die Numeruszeichen der Nomina und Verben sind meist dieselben usw.

> 'The number system of the Finno-Ugric and Samoyedic languages is undoubtedly more developed than that of the Altaic [languages] […]. Thus, the usage of the singular as absolute number is rather common to all Uralic languages, e.g., after numerals the singular appears, the number markers of nouns and verbs are mostly the same and so on.' (Ravila 1941: 2)

This implies on the one hand that the category of number appears to be more than a bare opposition of singular and plural (and dual) in the Uralic languages, and on the other hand that number marking in the Uralic and Altaic languages may show parallels but does not function identically. Surely much work on number marking in both Uralic and Altaic languages has been done since then, but typologically oriented approaches are rare, if available at all (see Klumpp et al. 2018 on the role of typology in Uralic studies). The paper at hand aims to fill this gap by describing the number systems of seven languages from Western and Central Siberia systematically, and evaluating them against a typologically informed background. The choice of languages is certainly coincidental to a certain extent, but can nevertheless be motivated from both a geographical and a methodological perspective. It has been observed in many works that Western and Central Siberia – more precisely the Ob and Yenisei basins – form a region where language contact and contact-induced linguistic convergences are more the norm than the exception (e.g. Hajdú 1979, Pusztay 1987, Helimski 2003). Helimski (2003: 160) even proposes an "Ostyak (Ob-Yeniseic) Sprachbund" with its center being Eastern Khanty and Selkup, and possibly including Nenets, Mansi, Ket, Southern Ewenki and Chulym Turkic. The choice of languages in this paper – see Figure 1 and the list below – largely follows these approaches, albeit focusing rather on the Yenisei basin instead of the Ob basin. Given the limits of space, e.g. Nenets, Mansi as well as Northern and Southern Khanty varieties are not included, though

they certainly would fit equally well into the sample. A further motivation of the choice of languages was to represent each (sub)branch of all named language families present in the given region.

The following list presents the languages under investigation here, including their genetic relationships and numbers of speakers according to the latest Russian census from 2010 (VPN 2010).

- Eastern Khanty (< Finno-Ugric[1] < Uralic; 1000 speakers)
- Nganasan (< Northern Samoyedic[2] < Samoyedic < Uralic; 100 speakers)
- Selkup (< Southern Samoyedic < Samoyedic < Uralic; 1000 speakers)
- Dolgan (< North-Siberian Turkic < North-Eastern Turkic < Turkic; 1000 speakers)
- Chulym Turkic (South-Siberian Turkic < North-Eastern Turkic < Turkic; 50 speakers)
- Ewenki (< Northern Tungusic < Tungusic; 5000 speakers)
- Ket (< Yeniseian; 200 speakers)

The data forming the empirical basis for this study come from two different sources. First, descriptive grammars were consulted in order to get an overview of the presence of the number category in each language system. Later, also corpus data as well as data from text collections was taken into consideration, especially when it comes to such complex syntactic issues as number agreement or the behavior of different types of nouns such as mass nouns, abstract nouns and the like. The used corpora and text collections are listed in what follows. It must be noted that the database is surely not balanced regarding the scope of the corpora and text collections included. This holds particularly true for Eastern Khanty, Chulym Turkic and Ket, in the case of which published digital corpora are still lacking – all Eastern Khanty, Chulym Turkic and Ket material comes from published texts in the series *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area* edited in Tomsk.

---

1. The binary division of the Uralic languages is today considered doubtful by many. I however stick to the traditional view, i.e. that Proto-Samoyedic and Proto-Finno-Ugric are the immediate successors of Proto-Uralic.
2. It is still a matter of debate whether Northern and Southern Samoyedic represent genetic units. Here, the labels are used as mere technical terms, leaving this question open.

Figure 1: Languages of the Ob-Yenisei basin[3]

- Khanty texts (Filchenko & Potanina 2010, 2012, 2013; Filchenko et al. 2015; Filchenko 2017) – 1,434 utterances
- *Nganasan Spoken Language Corpus* (Brykina et al. 2018) – 21,723 utterances
- *INEL Selkup Corpus* (Brykina et al. 2020) – 7,887 utterances
- *INEL Dolgan Corpus* (Däbritz et al. 2019) – 11,329 utterances
- Chulym Turkic texts (Lemskaya 2010a, 2012, 2013, 2015, 2017) – 1,252 utterances
- *INEL Evenki Corpus* (Däbritz & Gusev, in preparation) – 8,052 utterances
- Ket texts (Kryukova & Glazunov 2010; Kryukova 2012, 2013, 2015; Kryukova & Nefedov 2017) – 669 utterances

---

3. The map was created on the base of OpenStreetMap (https://www.openstreet-map.de/), which is published under the Open Database License (ODbL) 1.0.

Despite the database's imbalance, the given empirical data can surely support reaching the aims of the study at hand. Therefore, the corpora and text collections were searched electronically for the relevant feature, taking into account all data included. This ensures that no "false positives" are taken as evidence for an otherwise unattested phenomenon. Moreover, the analysis of corpora and text collections made it possible to detect phenomena that formerly were not described in grammars or other descriptive work.

Given the wide range of empirical data used in this study, some comments on their representation are in order here. As for transcription issues, I rely on the sources and do not attempt to unify transcription systems, although this might be against the Uralist tradition using FUT. However, since the paper deals with morphosyntax, the transcription itself can be regarded as secondary from my point of view. At some important points, I, however, comment on this. Coming to glossing, I basically also rely on the consulted sources. Nevertheless, when important for the problems discussed, some unifications have been made; these unifications are always mentioned explicitly.

The paper is structured as follows. In Section 2, some theoretical preliminaries as well as the typological framework adopted are touched upon. Section 3 deals with the description of number systems in the languages named above, dealing with each language separately. In Section 4, the results from the previous section are combined and evaluated: In Section 4.1, the main convergences and divergences are summed up. Section 4.2 deals with general number and its entailments, whereby it is shown that the concept of general number may be a key for understanding unmarked noun forms after numerals as well as unmarked noun forms denoting paired objects in the languages of the sample. In Section 4.3, singulatives in Ket, Selkup and Eastern Khanty are dealt with, and it is shown that their grammaticalization patterns are very similar. Section 4.4 investigates the interaction of number and animacy, demonstrating two splits on the animacy hierarchy, namely (1) between kin referents and non-kin human referents with respect to the availability of general number forms and (2) between first and second person referents in Dolgan and Chulym Turkic with respect to the availability of dual forms and the (ir)regularity of plural marking, respectively. Section 5 evaluates the results from an areal-typological perspective and comes to the conclusion that the category of number gives no clues for establishing an Ostyak or Ob-Yeniseic Sprachbund, regardless

of how many and which languages are potentially included. However, the singulative markers in Eastern Khanty, Selkup and Ket will be analyzed partly as contact phenomena, yielding a common Uralic-Yeniseian feature. Finally, Section 6 concludes the paper and gives a further outlook on upcoming related questions.

2. Theoretical preliminaries and typological background

As it is widely known, the nominal category[4] of *number* is concerned with the amount of extra-linguistic entities, i.e. referents, included into a certain state of affairs. Although intuitively it looks like a rather simple category, Corbett (2000: 1) claims that it is "the most underestimated of the grammatical categories". This can be easily illustrated: The number opposition *hippo* vs. *hippos* and its Kiswahili counterpart *kiboko* vs. *viboko* (cf. Hillewaert & de Schryver 2004) does not seem to pose any problems in describing number in English and Kiswahili respectively. Nevertheless, even in English – apparently the most well-known and well-studied language in the world – there are many instances of number marking which are much more intricate, cf. e.g. singular *sheep* vs. plural *sheep* or the possibility of non-agreement of subject and predicate in British English *The committee has/have decided* (Corbett 2000: 6). Thus, often one must take a closer look in order to understand the patterns of number marking in a given language properly.

Having settled the understanding of the category and the term *number*, it has to be clarified which meaning distinctions do occur within this category. These meaning distinctions can best be described as *number values* (Corbett 2000: 19–20). The number values important in this study are singular, dual and plural. According to Greenberg's (1966: 94) universal 34 they are hierarchically ranked, stating that a language can only have a dual when it has a plural. Additionally, in many languages, there are unmarked noun forms that are underspecified with respect to the category of number as seen in (1).

---

4. Note that also (person-)number endings of verbs denote nominal number, since they point to entities referred to by nouns or pronouns. Verbal number, in turn, is concerned with the event structure of verbs, e.g. Rapanui (< Austronesian) *ruku* 'dive' vs. *ruku ruku* 'go diving' (Corbett 2000: 6). Here, only nominal number is dealt with.

Dolgan

(1) **D'ie͡**   *da*   *tup-put-a.*
     house   and   build-PST2-3SG
     'And he built a house.' ~ 'And he built houses.' ~ 'He was house-building.'
     (Däbritz et al. 2019; KiPP_XXXX_2009_Family_nar.KiPP.040)[5]

Here, only the lexical meaning of the noun is conveyed, without reference
to its number. In such cases, a singular and a plural reading of *d'ie͡* 'house'
may be available, but these are not mandatory. In the given example, refer-
ence to number is not important for the speaker, since she talks about her
husband's occupations in the past decades. The number of houses is not
relevant for the interpretation of the utterance, though the context surely
implies that more than one house was built. According to Corbett (2000:
9–10), this phenomenon is called *general number*, implying that *general
number* stands outside the number system itself, since it can possibly be
disambiguated into one or the other number value inherent to the system.
Another term used for the same phenomenon is *transnumeral* (Itturioz-
Leza & Skopeteas 2004: 1054–1055); I however will stick to *general number*
in this paper. In this context, also the phenomenon and term of *genericity*
is worth mentioning, which has to be clearly distinguished from general
number. Surely there are generic expressions like **the frog** *is an amphibian*,
which seem to function like general number. However, reference to a kind
is established here (see Krifka et al. 1995 for a detailed analysis), and no
reference to a certain group of frogs is possible, as it is in the case of general
number forms. Therefore, both phenomena and terms have to be kept apart.

Finally, the formal expression of general number differs across lan-
guages. There are languages where general number forms differ from
all other number forms, e.g. Bayso (< Cushitic < Afro-Asiatic) *lúban*
'lion.GN' vs. *lubán-titi* 'lion-SG' vs. *luban-jaa* 'lion-PAUC' vs. *luban-jool*
'lion-PL' (Corbett 2000: 11). In many languages, on the other hand, gen-
eral number forms are homonymous to another number form (Corbett
2000: 13–16). Most frequently, general number forms are homonymous
to singular forms, as it is the case in the Dolgan example (1), too. This
leads to a discussion of *markedness* and its connection to number mark-
ing. Haspelmath (2006) rightfully advises caution when using this term
due to its many connotations and various uses in linguistic research. Here,

---

5.   When data from corpora is cited, the reference to the respective utterance is given
     as follows: SpeakerCode_DateOfRecording_Title_Genre.NumberOfUtterance.

I will understand *markedness* as an indication of the low frequency of a given extra-linguistic concept resulting in higher prominence in linguistic representation. Thus, e.g. dual is a number value more marked than plural.

When it comes to number values, Haspelmath and Karjus (2017) argue that the least marked number value is not necessarily the singular, but it varies depending on the semantics of the given linguistic item. Therefore, nouns such as *day* tend to be used more frequently denoting a single entity, whereas nouns such as *pea* tend to be used more frequently denoting many entities (Haspelmath & Karjus 2017: 2). While in languages like English this does not make any difference for their encoding, e.g. Welsh (< Indo-European) has *pys* 'peas' as the basic form, from which the singulative form *pys-en* 'peas-SNGL' = 'pea' is formed (ibid.). Also from the perspective of markedness, Alexiadou (2019) claims that there are two types of languages with respect to number marking: In type I languages, such as English, the plural is the formally marked number value, but semantically unmarked, whereas in type II languages – e.g. Turkish – the plural is semantically marked. The main argument for this assumption is that in type I languages the use of the singular necessarily entails a singular interpretation, whereas the use of the plural – as in *Mary saw horses* – also allows the number-underspecified reading that Mary was "horse seeing"; in type II languages – as in the Dolgan example (1) – the latter reading is available with singular number (Alexiadou 2019: 126–128). Although in my view Alexiadou's (2019) binary division of languages is too strict, the underlying observations surely have to be acknowledged in number typology and will be relevant for the paper at hand, especially when discussing general number.

Coming to the formal properties of number marking, the most frequent strategies to express number are the following: use of number words, inflection of the relevant item (morphological strategy), agreement and lexical means (Corbett 2000: 133). Number words designate unbound number markers, e.g. Tagalog (< Austronesian) *mga bahay* 'PL house' = 'houses' (Corbett 2000: 134). The morphological strategy includes, roughly speaking, everything that happens to the morphological form of the relevant lexical item. This may include affixation, stem alternation, reduplication and zero expression. Agreement means that number is not expressed on the relevant lexical item itself, but somehow and somewhere else in the clause (Corbett 2000: 136). Typically, this includes subject (and object) agreement on the predicate, but also agreement between nouns and adnominal modifiers (adjectives, demonstratives or the like) is relevant. Finally, number

can be expressed by lexical means, called *suppletion* (Corbett 2000: 155), a good example being Russian *rebënok* 'child' vs. *deti* 'children', whereby, however, the latter also includes the plural marker -*i*. Here, two etymologically non-cognate stems form the base for expressing singular and plural respectively. Additionally, it is important to note that the described strategies are often hard to distinguish from one other and that often more than one strategy is used at the same time. Therefore, a careful and precise analysis also of the formal marking of number is in order.

The languages and language families under investigation here are not unstudied with respect to the number category. First of all, they are partly represented in the *World Atlas of Language Structures* (WALS), namely in Chapters 33A (Coding of nominal plurality, Dryer 2013), 34A (Occurrence of nominal plurality, Haspelmath 2013), and 35A (Plurality in independent personal pronouns, Daniel 2013). Dryer (2013) shows that nominal plurality is expressed via suffixation in Khanty, Selkup, Ewenki and Ket. Haspelmath (2013) claims that nominal plurality is obligatorily expressed on all kind of nouns (humans, animates, inanimates) in Selkup, Ewenki and Ket, if a referential plural reading is intended: In consequence, he claims that general number is absent in these languages, but leaves the possibility open that singular forms may be used in generic contexts.

As for Uralic languages, Honti's (1997) "Numerusprobleme" sums up much of the research that had been conducted in Uralic studies until then. According to Honti (1997: 2–3), three number values (singular, dual, plural) appear in the Uralic languages, where the singular is homonymous to a "Numerus absolutus / Numerus indefinitus", that is, to general number forms in the terminology adopted in the present paper. General number forms in Uralic languages occur *inter alia* after (a) numerals indicating plurality, (b) with paired items as e.g. paired body parts and (c) with items that usually occur in groups such as teeth, fingers but also some animals and plants (Honti 1997: 6). In contrast to general number, singular, dual and plural forms mark individualizable referents, whereby the dual additionally may have coordinating functions (Honti 1997: 3, 7–8). From a diachronic point of view, Honti (1997: 102) concludes that also Proto-Uralic exhibited general number, the relicts of which can be traced in many modern Uralic languages. Since then, not much research has been conducted on number marking in Uralic languages from a comparative – be it typological or diachronic – point of view, though in emerging grammatical descriptions the category of number is surely covered.

Going farther east, in recent years the term *Transeurasian languages*, coined by Lars Johanson and Martine Robbeets (2010: 1–2), has arisen in order to cover the Turkic, Tungusic and Mongolic language families (also taken together under the notion *Altaic languages*) as well as Korean and Japanese. In a recently published handbook, Gruntov and Mazo (2020: 527–528) account for the regular occurrence of singular and plural in Turkic and Tungusic languages in contrast to mere relics of a dual. Moreover, Gruntov and Mazo (2020: 532) claim that nouns in Transeurasian languages, including Turkic and Tungusic, usually exhibit "[…] neutral forms unmarked for plurality/singularity, that can indicate both singular or indefinite plurality depending on the context". Obviously, this is in line with the term *general number* used here. Generally, Gruntov and Mazo (2020) is supported by much preceding work, see e.g. the comparative overviews by Johanson (1998) and Benzing (1955). Investigating nominal morphology in Northern Eurasian languages, Janhunen (2014: 311–313) includes both the Uralic and the Transeurasian languages into a continuum of languages spoken in Northern Eurasia, labeled "Ural-Altaic", emphasizing that this term must not be understood in terms of genetic relationship. As for the category of number, Janhunen (2014: 315–316) states that from Northwest to Southeast the obligatoriness of number marking is decreasing.

Finally, the Yeniseian family – with its sole remaining representative Ket – was not included into any of these studies. Nevertheless, the category of number has not been left unexplored. Generally it can be stated that the number values singular and plural are recognized for the Yeniseian languages, and furthermore the plural formation is described as highly complex (see Porotova 1990 as well as Werner 1995: 86–91 for details). Additionally – and apparently in contrast to surrounding languages – the occurrence of singulatives has long been reported for Yeniseian languages (Werner 1995: 88; Helimski 2016: 158).

In a nutshell, it can be stated that general tendencies of number marking in the language families under investigation are quite well known. Nevertheless, investigations taking into account up-to-date typological perspectives are still rare. The latter, however, is what this paper aims at, and therefore the descriptions in Section 3 as well as the comparative views in Section 4 lay the focus on phenomena which are rather unexpected from both a typological and a language-internal perspective. The areal approaches touched upon here will be taken up again in Section 5.

3. Number systems of the analyzed languages

3.1. Eastern Khanty

Since Khanty is dialectally diversified to such an extent that the different varieties may be mutually unintelligible, one might even speak of different languages (Filchenko 2010: 10, 15–16). Only one bundle of varieties is dealt with here, namely Vasyugan, Yugan, and Alexandrovo (forming part of Eastern Khanty) described by Filchenko (2010). The reason for choosing these dialects is their close geographical vicinity to Central and Southern Selkup as well as Chulym Turkic.

In nominal inflection, Eastern Khanty exhibits three number values (singular, dual, and plural) that are marked via suffixation: the singular is unmarked while the dual and the plural have the suffixes *-ɣən* and *-(ə)t*, respectively (Filchenko 2010: 90). The suffixes remain unchanged when case suffixes are added, but when possessive suffixes are attached to the form, the dual suffix is *-(ə)ɣl*, and the plural suffix *-l* (Filchenko 2010: 92):

Eastern Khanty

(2) ***ämpä-t*** *ʃəräɣ-wəl-t*
  dog-PL noise-PRS-3PL
  'The dogs are making noise.' (Filchenko 2017: 54)

(3) *män-nä* ***ämp-l-äm*** *we-s-im*
  1SG-COM **dog-PL-1SG** take-PST-1SG.OBC
  'I took my dogs with me.' (Filchenko 2017: 54)

In addition, there are noun forms homonymous to the singular that are underspecified with respect to the category of number.

Eastern Khanty

(4) […] *qo* *juɣ* ***loɣo-l*** *al-a-wəl.*
    where 3SG **bone-3SG** lie-EP-PRS.3SG
  '[Nobody knows] where his bones are.' ~ ?'[Nobody knows] where his bone is.' (Filchenko et al. 2015: 72)

In the given context (boys looking for their dead father), it becomes clear that a plural reading is intended here. Khanty thus exhibits general number according to Corbett (2000), which stands outside the number system and may be disambiguated into one or the other number value.

Additionally, in the analyzed material there are few instances of the lexeme *sem* 'eye' used as a singulative marker, as displayed in example (5). The grammaticalized function of the lexeme can additionally be corroborated by the fact that it follows directly after the bare form of the noun it relates to, whereas in the case of the potentially similar item *səyəl'* 'tuft' an attributive form of the respective noun is used.

Eastern Khanty

(5) *pam-ə   səyəl' jomentʃʼəy* **sem** *əjqa   wəl-l-əyən.*
     grass-ATTR tuft   berry   **eye** together live-PRS-3DU
     'A tuft of grass and a berry live together.' (Filchenko et al. 2012: 56)

Mass nouns, group nouns, and abstract nouns only seldom take number suffixes due to their semantics (Filchenko 2010: 90). The unmarked form of nouns denoting paired objects refers to the respective pair and when referring to one single item, the quantifier *pələk* 'half' is used: *sem* 'pair of eyes' vs. *sem pələk* 'one single eye'. Plural forms of paired nouns are ambiguous, referring to either many single items of pairs or to many pairs as a whole (Filchenko 2010: 91).

In contrast to the highly regular number marking of nouns, personal pronouns exhibit a less transparent paradigm. Although the dual and plural forms may trace back to the same stem as the singular forms, they are irregularly formed from a synchronic point of view, since no dual or plural marker can be detected within them.

Table 1: Personal pronouns in Eastern Khanty (Filchenko 2010: 124)

	SG	DU	PL
1	*mä*	*min*	*məŋ*
2	*nöŋ*	*nin*	*nəŋ*
3	*joɣ*	*jin*	*jəɣ*

Within noun phrases, there is no number agreement of modifier and head noun as seen in (6).

Eastern Khanty

(6) a. **tʃu** *sart*   b. **tʃu** *sarta-***yəŋ**
       this pike   **this** pike-DU
       'this pike'   'these two pikes'

       (Filchenko 2010: 305)

After numerals and quantifiers nouns show no number marking (Filchenko 2010: 195) – whether these are general number or singular forms cannot be decided solely on the base of their form. This will be discussed in more detail in Section 4.2. In the case of the numeral *kät ~ kätqən* 'two'[6], however, the dual form of nouns occurs (ibid.).

Eastern Khanty

(7) […] ***kä*** *niŋ-**qən*** *män-nä* *ju-s-ɣən.*
      two woman-**DU** 1SG-LOC come-PST2-3DU

      '[While I was sitting,] two women came to me.' (Filchenko 2010: 195)

As for verbal paradigms, Eastern Khanty has two conjugations, traditionally called the subjective conjugation and the objective conjugation, respectively. The subjective conjugation establishes only subject agreement, while the objective conjugation establishes also subject agreement and partial (only number, no person) object agreement. Hence, in verbal paradigms in Eastern Khanty, number agreement of both subjects and objects may be indicated (Filchenko 2010: 271, 273). Subject reference is unambiguous for all persons and numbers, and the third person suffixes of the subjective conjugation (3DU *-ɣən*, 3PL *-(ə)t*) are identical to the nominal dual and plural suffixes (Filchenko 2010: 271). The number markers of object reference (dual *-(ə)ɣl*, plural *-l*) are identical to the number markers used together with possessive suffixes (Filchenko 2010: 273).

Number agreement in the verb phrase is obligatory for both nominal and pronominal subjects (Filchenko 2010: 328). In the case of objects, this holds true for pronominal objects but appears to vary in the case of nominal objects: Although Filchenko (2010: 274) accounts for object number agreement (8a), in the analyzed material many instances without object number agreement can be found as well (8b).

---

6. The forms *kät* and *kätqən* differ in their syntax: the former is used as modifier in noun phrases, the latter in all other positions. Note that the latter form *kätqən* itself is a combination of *kät* and the dual marker *-qən* (Honti 1984: 76–77). The final consonant of *kät* may additionally be omitted when there is a consonant onset in the following word like in example (7). This pattern has clear parallels in the other Ugric languages (cf. Mansi *kit ~ kitiɣ* (Keresztes 1998: 412) and Hungarian *két* vs. *kettő*). Additionally, in Mari the numerals up to 20 show the same distribution of forms (Alhoniemi 2010: 96–97).

Eastern Khanty

(8) a. *mä   wajaɣ-ət   wel-s-**əl-am***
    1SG   animal-PL   kill-PST-**PL-1SG.OBC**
    'I killed the animals.' (Filchenko 2010: 274)

  b. *Qolla šaj-at, qolla majlanta šaj-wal   patr-**il**.*
    all   thing-PL all   present   thing-PL.3PL put.PST-**3PL.OBC**
    'All the things, all the presents, they packed their things.'
    (Filchenko et al. 2015: 54)

Finally, a further usage of the dual should be mentioned. The dual suffix can be used in order to coordinate two nominal constituents, and in doing so it is attached to both constituents. A corresponding predicate also shows dual number marking then.

Eastern Khanty

(9) *Kimtaki   qotl-a   **im-ki**   **rätʃ-ken**   tʃu*
    second   day-ILL   **old.woman-DU**   **old.man-DU**   this

    *puɣl-a   joɣto-**qon**.*
    village-ILL   reach-**PST.3DU**
    'On the second day the wife and the husband reached that village.'
    (Filchenko et al. 2015: 57)

3.2. Nganasan

In nominal inflection, Nganasan exhibits the three number values singular, dual and plural. The singular is zero-marked. The dual has the suffix *-kəj ~ -gəj* in the nominative case, *-ki ~ -gi* in oblique cases and *-kəi ~ -gəi* in possessive forms. The plural has the suffix *-ʔ* in the nominative case and *-j* in oblique and possessive forms; the latter often is fused with case and possessive suffixes. Furthermore, in oblique and possessive plural forms, portmanteau morphemes occur.

Nganasan

(10) a. *kümaa*   b. *kümaa-**ʔ***   c. *kümaa-**j***
     knife   knife-**PL**   knife-**PL.ACC**
     (Wagner-Nagy 2019: 185, 192–193)

According to Wagner-Nagy (2019: 186), however, Nganasan exhibits nouns unmarked for number "if the NP has no specific referent". This certainly

qualifies for generic contexts, but it may possibly also hint at general number in the sense of Corbett (2000), as demonstrated in (11). Here, not the number of reindeer included into the state of affairs is important for the speaker, but the fact itself that they will be visible. The form *babi* 'wild. reindeer.ACC' may be analyzed as underspecified for the category of number and, thus, as a general number form.

Nganasan

(11) **Babi** *ŋəndīaiʔ* *katʼəmə-suðə-ŋ.*
     wild.reindeer.ACC probably see-FUT-2SG
     'You'll probably see a wild reindeer.' ~ 'You'll probably see wild reindeer.' (Brykina et al. 2018; MVL_090807_Hungabtadja_flks.306)

However, it has to be stated that in Nganasan such contexts are less frequent than in the other Uralic as well as Turkic languages under investigation here.

An important trait of number marking in Nganasan is the incompleteness of the dual paradigm: out of the eight cases of Nganasan only the three grammatical cases nominative, genitive and accusative exhibit synthetic dual marking. In the other cases a postpositional construction is used, the base noun in the genitive preceding it. This can be exemplified with the genitive and locative case forms of *kümaa* 'knife'.

Nganasan

(12) a. *kümaa*
        knife.SG.GEN
        'of a/the knife'

   b. *kümaa-tənu*
      knife-SG.LOC
      'on a/the knife; with a/the knife'

   c. *kümaa-gi*
      knife-DU.GEN
      'of two/both knives'

   d. *kümaa-gi* *na-nu*
      knife-DU.GEN near-LOC
      'on two/both knives; with two/both knives'

   e. *kümau-ʔ*
      knife-PL.GEN
      'of (the) knives'
      (Wagner-Nagy 2019: 191–193)

   f. *kümaa-tinü*
      knife-PL.LOC
      'on (the) knives; with (the) knives'

Abstract nouns exhibit only singular forms; nouns denoting paired objects behave like other nouns, i.e., singular number refers to one part of the pair, and dual number refers to the entire pair (Wagner-Nagy 2019: 186).

Nganasan

(13) *Mənə*   *ŋuə-l'aa*   **s'ejmɨ-mə.**
      1SG.GEN   one-LIM   **eye-1SG**
      'I have only one eye.'
      (Brykina et al. 2018; MVL_080304_NjomuKamleguNy_flks.395)

(14) **S'ejmɨ-gəj-t'ə**   *t'il'i-s'itə.*
      eye-DU-ACC.2SG   peck-FUT.3SG
      'It'll peck your eyes out.'
      (Brykina et al. 2018; PKK_71_OneTent_flkd.033)

Additionally, the dual forms *d'esɨ-gəj* 'father-DU' and *ńemɨ-gəj* 'mother-DU'
have the meaning 'parents' (Wagner-Nagy 2019: 186–187). In combination
with dyadic kinship terms such as wife/husband or sister/brother, the dual
also may refer to the given referent together with his/her opposite kin.
However, in this case also a concomitant dyadic marker is used (Wagner-
Nagy & Arkhipov 2020: 435–436).

Nganasan

(15) […] *nɨ-səd'ə-**gəj***   *n'iri-l'i͡ai-ʔ*   *küðü-ʔü͡a-gəj,*
          woman-DYA-**DU**   together-LIM-ADV   get.up-AOR-3DU

      *n'emɨ-**gəi**-t'i.*
      mother-**DU**-3SG
      '[…] wife and husband got up together, his parents.'
      (Brykina et al. 2018; TKF_990819_SomatuShaman_flkd.094)

Therefore, nominal number marking is quite regular in Nganasan. Note
that number marking is largely identical in non-possessive oblique forms
on the one hand and possessive forms on the other hand (see Däbritz 2017
for details). Personal pronouns can be split into first as well as second per-
son pronouns, and third person pronouns. The first and second person
pronouns behave irregularly from a synchronic point of view; the third
person pronoun is a combination of the noun *siʔ* 'portrayal; picture;
image' and the respective possessive suffix (Wagner-Nagy 2019: 106).

Table 2: Personal pronouns in Nganasan

	SG	DU	PL
1	*mənə*	*mii*	*mɨŋ*
2	*tənə*	*tii*	*tɨŋ*
3	*sɨtɨ*	*sɨtɨ*	*sɨtɨŋ*

Within the noun phrase, number agreement is obligatory, as demonstrated in (16).

Nganasan

(16) a. *aniʔka   maʔ*   b. *aniʔka-gəj ma-kəj* c. *aniʔka-ʔ   maðə-ʔ*
        big   tent   big-**DU** tent-**DU**   big-**PL**   tent-**PL**
        'a/the big tent'   'two/both big tents'   '(the) big tents'
        (Wagner-Nagy 2019: 309)

After quantifiers and numerals greater than one, usually singular forms occur. As a variant, the numeral *s'iti* can be followed by the dual form of the noun, and numerals greater than two can be followed by plural forms.

Nganasan

(17) a. *s'iti   səmu*   ~   *s'iti   səmu-gəj*
        two   cap   two   cap-**DU**
        'two caps'

   b. *nagür   səmu*   ~   *nagür   səmu-ʔ*
        three   cap   three   cap-**PL**
        'three caps'
        (Wagner-Nagy 2019: 158)

In verbal paradigms, number agreement of both subject and object is expressed. The third person suffixes of the subjective conjugation are zero (singular), *-kəi ~ -gəi* (dual) and *-ʔ* (plural) (Wagner-Nagy 2019: 229), the latter two being identical to the dual and plural marker in nominal inflection. In the objective conjugation, dual and plural objects are referred to with *-kəi ~ -gəi* and *-j* (which is often fused with person markers here), thus, the marking is similar to possessees in nominal inflection, as it is the case in Eastern Khanty, too.

Number agreement in the verb phrase is obligatory in Nganasan for both nominal and pronominal subjects (Wagner-Nagy 2019: 364). If the verb is inflected in the objective conjugation, number agreement is obligatory for both nominal and pronominal third person objects, too (Wagner-Nagy 2019: 230). Example (18) shows number agreement of both subject and object.

Nganasan

(18) *Bənsə-**gəj**   tə   ŋəm-hū͡að̄u-**kəi-t'ü**.*
       all-**DU**   well   eat.up-INFER-**DU-3SG.OBC**
       'She has apparently eaten both of them.'
       (Brykina et al. 2018; MVL_080226_TwoHorses_flks.132)

3.3. Selkup

Selkup exhibits considerable dialectal variation, the main divisions being between Northern, Central and Southern dialects. Since – unlike in the case of Khanty – all varieties are potential contact varieties to other languages included in the sample here, all three dialect groups are described.

Selkup has three number values in nominal inflection (singular, dual, and plural), exhibiting considerable dialectal variation. Table 3 summarizes the number markers in the main dialect groups (Kuznecova et al. 1980: 167–170; Bekker 1995: 46).

Table 3: Number marking in Selkup dialects[a]

	SG	DU	PL
Northern	-∅	*-qɪ*	*-t ~ -n*[b], *-iː*
Central	-∅	*-štja*	*-t, -la, (-iː)*
Southern	-∅	*-qi, -staɣi, -ja*	*-t, -la*

a. For the sake of comprehensiveness, not all (mor)phonological variants and allomorphs are given here.
b. *-t ~ -n* is a seemingly free plosive–nasal alternation (Kuznecova et al. 1980: 143–144).

Additionally, Kuznecova et al. (1980: 167) point out that nominal forms unmarked for number are not necessarily singular forms but can be underspecified. According to the terminology used here, these are general number forms, as seen in (19) from Northern Selkup. Here, it becomes clear only from the following context that the speech is about three knives, but in the given sentence, number reference is apparently not important for the understanding of the utterance.

Northern Selkup

(19) **Paŋi-mit**   qaj   əmilti-mni-mit.
      knife-1PL   whether   forget-PST.NAR-1PL
      'We forgot our knife, though.' ~ 'We forgot our knives, though.'
      (Brykina et al. 2020; SAI_1965_Palna_flk.074)

Not surprisingly, abstract and mass nouns hardly occur in non-singular forms. In order to single out a referent from a mass noun, there is the singulative suffix *-laka ~ -laga*, e.g. *čoš-i-laka* 'fat-EP-SNGL' = 'a piece of fat' ← *čoš* 'fat' (Kuznecova et al. 1980: 168; example from Brykina et al. 2020, KR_1969_RavensAndHares1_flk.008). Additionally, the lexeme *saj ~ haj* 'eye' can be used in this function (Beáta Wagner-Nagy, p.c.), e.g. *tü haj* lit. 'fire eye' = 'spark' (example from Brykina et al. 2020, YIF_1965_Kamadzha1_flk.050). Paired objects in the singular refer to the whole pair; if one item of the pair is referred to, an adjectivized form of the quantifier *pɛläk* 'half' is used, e.g. *pɛlä-l' tol'cɨ* 'half-ADJZ ski' = 'one ski' (Kuznecova et al. 1980: 168).

Like in Nganasan, the dual – together with a special dyadic marker – is involved in expressing dyadic kinship terms, i.e. the given referent and her/his opposite kin are referred to (Wagner-Nagy & Arkhipov 2020: 436–439). Interestingly, the term for 'parents' is exclusively formed from the term for 'mother' in the analyzed material, e.g. *ämi-si-qäqɪ* 'mother-DYA-DU' (example from Brykina et al. 2020, KR_196X_RavensAndHares2_flk.013).

In personal pronouns, a clear distinction between first and second person forms on the one hand and third person forms on the other hand is observed: The former have irregular dual and plural forms, whereas the latter add the usual dual and plural markers to the third person singular pronoun (Kuznecova et al. 1980: 288). Here it can be additionally said that the third person pronouns diachronically can be traced back to a demonstrative, cf. Selkup *tap ~ tam* 'this' (Kuznecova et al. 1980: 293). Table 4 shows the personal pronouns of Northern Selkup.

Table 4: Personal pronouns in Northern Selkup

	SG	DU	PL
1	*man ~ mat*	*meː*	*meː*
2	*tan ~ tat*	*tɛː*	*tɛː*
3	*təp ~ tən*	*təpäː-qɪ*	*təpɨ-t ~ təpɨ-n*

Noun phrases exhibit no number agreement (Kuznecova et al. 1980: 267), as demonstrated in (20).

Southern Selkup

(20) *Tü-sa-ŋ* **arin** *qu-la-ne.*
come-PST-1SG **foreign** human-PL-ALL
'I came to the strangers.'
(Brykina et al. 2020; KMS_196X_Lifestory_nar.013)

After numerals and quantifiers, usually the unmarked form of nouns is used (Kuznecova et al. 1980: 167). After the numeral *šittɨ ~ šittə ~ šədə* 'two', however, also dual forms may occur (Kuznecova et al. 1980: 169). The material analyzed here (Brykina et al. 2020) shows that this is most frequent in Northern Selkup dialects when denoting human referents. The exact patterns of dual marking after *šittɨ ~ šittə ~ šədə* 'two', however, call for further research.

Northern Selkup

(21) *Toːna* **šittɨ** *qum-oː-qɨ* *ilɨ-mpɔː-qɨ.*
that **two** human-EP-DU live-PST.NAR-3DU
'Those two people lived.'
(Brykina et al. 2020; NEP_1965_OrphanBoyAndPanOldMan1_flk.136)

In verbal paradigms, person and number of the subject is expressed regularly. Though Selkup exhibits an objective conjugation, neither person nor number of the object is cross-referenced (Kuznecova et al. 1980: 234–235). Instead, only the existence of a direct object is indicated.[7] In verb phrases, subject and predicate nearly always agree in number in the analyzed material (Brykina et al. 2020), be the subject nominal or pronominal.

3.4. Dolgan

In the Dolgan number system, two number values are expressed: singular and plural. In nominal inflection, the singular is unmarked, and the plural has the suffix *-LАr*, which can clearly be separated from case and possessive suffixes in oblique forms (Artemjev 2013: 26–27, 51). Two lexemes exhibit

---

7. The usage of the objective conjugation is far from clear and will not be investigated further here. Supposedly, it is motivated by information structural phenomena.

an irregular stem extension in their plural form, namely *kɨrgittar* 'girls' (vs. *kɨːs* 'girl'), *u͡olattar* 'boys' (vs. *u͡ol* 'boy') (Artemjev 2013: 29); otherwise plural marking is completely regular.

Apart from that, there are instances described as singular forms that can have plural reference (Artemjev 2013: 26), cf. (1) repeated here for convenience:

Dolgan

(22) **D'i͡e**   *da*   *tup-put-a.*
house and build-PST2-3SG
'And he built a house.' ~ 'And he built houses.' ~ 'He was house-building.'
(Däbritz et al. 2019; KiPP_XXXX_2009_Family_nar.KiPP.040)

As was already discussed in Section 2, number reference is not important here for the speaker. Although the following context suggests a plural reading, the form can be regarded as underspecified, which indicates general number. From a broader perspective, this phenomenon is quite common in Turkic languages (Johanson 1998: 38, 51).

Mass nouns as well as nouns referring to paired objects occur in the singular (Artemjev 2013: 26, 28). When one single item of a pair is referred to, the lexeme *aŋaːr* 'half' is used as a quantifier:

Dolgan

(23) *Bu* *kɨːs kajɨh-ar,*   *onto* **aŋaːr karak-taːk** *e-bit.*
this girl look.around-PRS.3SG then **half** **eye-PROPR** be-PST2.3SG
'This girl looks around, and apparently she has one eye.'
(Däbritz et al. 2019; BaA_1930_OneEyedGirl_flk.013)

In nominal possessive inflection, it is noticeable that plural marking of possessor and possessee is identical in the third person, yielding homonymous forms (Artemjev 2013: 31–32). Only the context can disambiguate this homonymy.

Dolgan

(24) a. *taba-ta*   b. *taba-lar-a*
    reindeer-3SG   reindeer-PL-3SG
    'his/her one reindeer'   'his/her many reindeer' ~ 'their one reindeer' ~ 'their many reindeer'
    (Artemjev 2013: 31–32, own glossing and translation)

In the paradigm of personal pronouns, the plural forms of first and second person pronouns are irregularly formed, but the plural form of the third person pronoun is built with the usual plural marker *-LAr* (Artemjev 2013: 138).

Table 5: Personal pronouns in Dolgan

	SG	PL
1	*min*	*bihigi*
2	*en*	*ehigi*
3	*gini*	*giniler*

Within noun phrases, there is no agreement (Artemjev et al. 2013: 60–62); after numerals and quantifiers, unmarked forms of the noun occur, e.g. *bies kün* lit. 'five day' = 'five days' (Artemjev 2013: 125).

In verbal paradigms, person and number of the subject is expressed systematically. The third person plural marker is homonymous to the usual plural marker *-LAr* in one set of endings, and homonymous to the possessive suffix of the third person plural *-LArA* in the second set of endings (Artemjev 2013: 202).[8] Number agreement of both nominal and pronominal subjects with the predicate in verb phrases is obligatory according to the material analyzed here (Däbritz et al. 2019).

Finally, two traces of dual marking can be observed in Dolgan. First, there is the item *bihikki* 'I together with another person', whose categorical status is rather unclear. It occurs together with nouns and personal pronouns and is always postposed. Historically, it is a contraction of *bihigi* 'we' and *ikki* 'two' (Däbritz 2019: 9). Second, in the imperative paradigm, there is a separate dual form in the first person (Däbritz 2019: 10). Example (25) demonstrates both phenomena.

Dolgan

(25) *Biːr   d'ie-ge   d'ie-len-**iek**   en   **bihikki**.*
     one   house-DAT.LOC   house-VBZ-**IMP.1DU**   2SG   **we.two**
     'Let us [two] [= you and me] live together in one house.'
     (Däbritz et al. 2019; ErSV_1964_WarBirdsAnimals_flk.046)

---

8. As is typical of Turkic languages, there are two sets of person–number endings in Dolgan. One set of endings, the so-called "pronominal endings" are used together with nominal predicates as well as in one part of the tense–aspect–mood forms. The other ending set, the so-called "possessive endings" are used with possessed nouns as well as in the other part of tense–aspect–mood forms.

3.5. Chulym Turkic

Chulym Turkic has two number values, namely singular and plural. In nominal inflection, the former is unmarked while the latter has the marker *-LAr* (Birjukovič 1979: 68). In addition, Birjukovič (1978: 69) states that singular forms can have plural reading. According to the terminology adopted here, these are instances of general number, as seen in (26).

Chulym Turkic

(26)   *Aran   išt-ɨn-da*   **čɨlya.**
          yard   interior-3SG-LOC   **horse**
          'Inside the yard there [is] a horse.' ~ 'Inside the yard there [are] horses.' (Lemskaya 2015: 228)

Once more, only the context can disambiguate the clause, if necessary. Here, a plural reading is intended.

The nominal plural paradigm is principally regular but exhibits peculiarities in marking third person plural possessors and possessees: plural possessors are not regularly expressed in the third person, *palaː-zɨ* 'child-3' can mean either 'his/her child' or 'their child' (Birjukovič 1979: 12–13). This is a clear deviation from the standard Turkic pattern, since the possessive suffix of the third person plural usually contains the plural marker *-LAr* (Johanson 1998: 39), as e.g. in Dolgan (see above). Since plural possessees are expressed with the usual plural marker, the form *palaː-lar-ɨ* 'child-PL-3' means either 'his/her children' or 'their children'. Interestingly, the Upper Chulym dialect disambiguates these forms by using different allomorphs of the plural morpheme, e.g. *qaraq-**tor**-u* 'eye-PL.POSS-3', which, according to Anderson and Harrison (2006: 53–54), means 'their eye' but not *'his/her/their eyes' – the morphonologically expected form *qaraq-**tar**-ɨ* 'eye-PL-3', in turn, has the latter reading.

Personal pronouns have an irregular plural form in the first person but regularly built forms in the second and third person (Li et al. 2008: 37).

Table 6: Personal pronouns in Chulym Turkic

	SG	PL
1	*män*	*pis*
2	*sän*	*silär*
3	*ol*	*olar*

In this context it can be mentioned that the plural marker *-LAr* occurs also in the possessive suffix of the second person plural *-(I)ŋnAr*, but not in the possessive suffix of the third person plural *-(z)I* (Birjukovič 1979: 61; Li et al. 2008: 34).

Within noun phrases, there is no number agreement, i.e. adjectives or demonstratives modifying a noun show no plural marking (Li et al. 2008: 40, 47).

In verbal paradigms, person and number is expressed systematically. The plural marker *-LAr* is present in the forms of both the second and third person plural (Li et al. 2008: 68). In the first person non-singular imperative, there are the two forms *-(A)q* and *-(A)qtAr* (Li et al. 2008: 81). According to Lemskaya (2010b: 122–123) it is not yet settled whether dual and plural (like in Dolgan, see Section 3.4) or minimal and augmented inclusiveness are expressed here. In either case, *-(A)qtAr* can readily be seen as a combination of *-(A)q* and the plural marker *-LAr*.

In the verb phrase, number agreement of a nominal subject and the predicate is possible, but it is not obligatory in the analyzed material. In example (27), there is plural marking on the subject, but the predicate points to the zero-marked third person singular.

Chulym Turkic

(27) ***Kiži-l'är***   *anɨ*   *ištiːn-in*   *suɣr-up*
      human-PL   3SG.ACC   trousers-ACC.POSS.3   remove-CVB

      *sal-**ɣan***
      put.AUX-**PST.3SG**
      'People removed him his trousers.'
      (Lemskaya 2015: 231, glossing adapted)

Number agreement of pronominal subjects and the predicate, however, is obligatory.

3.6. Ewenki

Ewenki exhibits much dialectal variation. Here, only those varieties are discussed that have potential contacts with other languages in the scope of the paper, namely Northern Ewenki (contacts with Nganasan and Dolgan) as well as Southern Ewenki (contacts with Selkup and Ket), and which are represented in the analyzed material (Däbritz & Gusev, in preparation). As for number marking, the available grammatical descriptions do not point to dialectal differences. However, since Eastern Ewenki is not included, no empirical observations can be made here regarding potential divergences.

Ewenki has two number values, namely singular and plural. The former is unmarked, the latter has the suffix *-l ~ -r* (Bulatova & Grenoble 1999: 6; Nedjalkov 1997: 141). However, in the case of mostly kinship terms, also the suffixes *-sal ~ -hal*[9] as well as *-til* are used (Bulatova & Grenoble 1999: 7). According to Vasilevič (1958: 672), there are relicts of a dual in some Southern dialects (suffix *-ti*). The material analyzed here (Däbritz & Gusev, in preparation), however, does not support this. Case marking of plural forms is completely regular (Bulatova & Grenoble 1999: 8; Nedjalkov 1997: 142).

Mass and abstract nouns do not show plural forms (Vasilevič 1958: 672). Nouns denoting paired objects behave like other nouns, that is, one item of the pair is referred to with the singular form, both items are referred to with the plural form.

Southern Ewenki

(28) *Murịwul*   **ŋaːlạ-t-pị**   *kultu-rä-n,*   **ŋaːlạ-n**
     Muriwul   **hand-INS-REFL.POSS**   hit-AOR-3SG   **hand-3SG**

     laŋgara-ra-n.
     stick-AOR-3SG

     'Muriwul hit with [one] hand, his hand got stuck.'
     (Däbritz & Gusev, in preparation; KS_1930_Muriwul_flk.203)

According to Konstantinova (1964: 44), singular forms marked with the indefinite accusative case can have a plural reading. This applies if the given referents form a homogenous group, eventually standing in a row with other similar groups. The material analyzed here (Däbritz & Gusev, in preparation) shows such instances rarely, nevertheless some examples can be found.

Northern Ewenki

(29) *D'iktə-jə*   *d'əp-i-ŋnə-m.*
     berry-ACC.INDF   eat-EP-HAB.AOR-1SG

     'I used to eat berries.'
     (Däbritz & Gusev, in preparation; KN_192X_Pereden_flk.019)

Applying Haspelmath and Karjus' (2017) markedness approach, berries and similar items (nuts, herbs, but also animals that are hunted) are more likely to appear in semantically plural contexts than in semantically singular contexts (see the discussion of *pea* in Section 2). That makes their linguistic expressions prone to be unmarked, when a plural reading is

---

9. *s ~ h* is due to dialectal variation, the former occurring in Eastern and Southern dialects, the latter in Northern dialects.

intended. Given the fact that berries and e.g. fish are more likely to correspond to proto-patient roles (see Dowty 1991), it is not surprising that direct objects appear in the given contexts rather than e.g. subjects. However, it cannot ultimately be said whether this is a matter of frequency, or whether unmarked nominative forms with a plural reading are impossible in Ewenki. In any case, it can be concluded that there are instances of general number in Ewenki, but they are far more restricted than in other languages investigated here.

The plural personal pronouns of the first and second person are formed irregularly. In the first person, there is an inclusive and an exclusive form. The third person plural pronoun is formed on the basis of the third person singular pronoun, to which the plural marker *-r* (a regular allomorph of *-l*) as well as the possessive suffix of the third person plural *-tin* is added. Interestingly, in the unpublished Ewenki material collected by K. M. Rychkov[10] the possessive suffix is sometimes missing from this form. Table 7 shows the personal pronouns in Ewenki (Nedjalkov 1997: 196).

Table 7: Personal pronouns in Ewenki

	SG	PL	
1	*biː*	EXCL	*buː*
		INCL	*mit*
2	*siː ~ hiː*	*suː ~ huː*	
3	*nuŋan*	*nuŋartin*	

According to Bulatova and Grenoble (1999: 57) and Nedjalkov (1997: 277), there is obligatory number agreement within noun phrases. The material analyzed here, however, does not exhibit clear patterns in this respect. In the case of adjectives, the pattern exhibiting agreement (30a) is more frequent than the pattern without agreement (30b) in the analyzed material. In the case of demonstratives, however, the pattern exhibiting agreement (31a) is less frequent than the pattern without agreement (31b).

---

10. The material is stored at the Institute of Oriental Manuscripts of the Russian Academy of Sciences (IVR RAN) in Moscow. Parts of it will be published in the INEL Evenki Corpus (Däbritz & Gusev, in preparation).

Northern Ewenki

(30) a. [NP *Aja-**l**-du*   *ilə-**l**-dụ*]   *bụ:-čə:-s*   *minə* […].
         good-PL-DAT.LOC   person-PL-DAT.LOC   give-PST-2SG   1SG.ACC

        'You gave me to good people […].'
        (Däbritz & Gusev, in preparation; KI_1931_Woman_flk.025)

    b. [NP *Omakta-du*   *dundə-**l**-du*]   *lawikta*   *kətə*   *bi-so-n.*
          new-DAT.LOC   place-PL-DAT.LOC   lichen   many   be-PST-3SG

        'At the new places, there were many lichens.' (Däbritz & Gusev, in
        preparation; BTV_20190822_ReindeerRuns_flk.018)

(31) a. *Hụrụ-rə*   *d'u:-la:-wər*   [NP *tar-ị:-**l***   *bəjə-**l***].
         leave-AOR.3PL   house-LAT-REFL.POSS.PL   that-EP-PL   man-PL

        'Those men went home.'
        (Däbritz & Gusev, in preparation; MN_1931_Ogre_flk.033)

Southern Ewenki

    b. [NP *Tar*   *bəjə-**l***]   *gụ-l-dˈrə:*   […].
          that   human-PL   say-INCH-AOR.3PL

        'Those humans said: […].'
        (Däbritz & Gusev, in preparation; KSh_1930_LowerWorld_flk.019)

Both Vasilevič (1948: 13) and the material analyzed here account for a dialectal distribution of (non-)agreement in Ewenki noun phrases: Northern Ewenki dialects – especially those dialects on the Taimyr Peninsula which are in close contact with Dolgan – tend to exhibit no agreement, whereas Southern Ewenki dialects tend to exhibit agreement. Probably, the latter pattern can be regarded as inherited, since other Tungusic languages exhibit it as well (Benzing 1955: 149). Consequently, the former non-agreeing pattern in Northern Ewenki dialects appears to be influenced by the surrounding non-agreeing patterns in Dolgan and Sakha. Given the close contacts between Ewenki with Dolgan and Sakha, this scenario seems to be entirely plausible. Nevertheless, the agreement patterns of Ewenki definitely call for further research in order to explain the observed variation in more detail.

     In verbal paradigms, both person and number are expressed systematically. Number marking in third person verbal forms is worth commenting on, since there are two sets of person–number endings in Ewenki. One ending set, the so-called "pronominal endings" occur in one part of the tense–aspect–mood forms. The other ending set, the so-called "possessive

endings" occur on possessed nouns as well as in the other part of tense–aspect–mood forms. In the former set, the third person singular is marked with *-n* while the third person plural has a zero morpheme, cf. *əmə-rə-n* 'come-AOR-3SG' vs. *əmə-rə* 'come-AOR.3PL' (Bulatova & Grenoble 1999: 33; Konstantinova 1964: 170–171), something that is counterintuitive and typologically uncommon given the higher markedness of the plural compared to the singular. In the latter set of endings, the third person forms are homonymous with the corresponding possessive suffixes (*-n* and *-tin* respectively). In verb phrases, number agreement of target and controller is obligatory in the analyzed material (Däbritz & Gusev, in preparation). This holds true for both nominal and pronominal subjects.

3.7. Ket

The Ket language is usually divided into three dialects (Northern, Central and Southern Ket), and these dialects seem to differ mostly in phonetics and lexicon (Werner 1997a: 1–2; Vajda 2004: 4; Georg 2007: 20–22). All mentioned grammatical descriptions are based on the Southern Ket dialect, which this paper also follows.

Ket nouns exhibit two number values, namely an unmarked singular and a marked plural (Werner 1997a: 96; Vajda 2004: 19; Georg 2007: 91). Plural marking is complex, including suffixation with *-(V)n ~ -(V)ŋ*[11], tonal changes[12] (e.g. *qóqbun* 'cuckoo' : *qòqbun* 'cuckoos'), stem alternations ("Ablaut") and in a few cases suppletion (Werner 1997a: 96–99; Vajda 2004: 19–20; Georg 2007: 93–100). A few nouns, e.g. *bəˀn* 'duck(s)' or *sūj* 'mosquito(s)', have no morphologically marked plural form, but nevertheless they do distinguish singular from plural, on the one hand via subject and object agreement in the verb phrase (see below), on the other hand via "plural" case suffixes (Vajda 2004: 20; Georg 2007: 97). Therefore, this phenomenon does not qualify for general number in the sense of Corbett (2000).

---

11. The distribution of the two suffixes is not entirely settled. However, there seems to be the interesting tendency that animate nouns take *-n*, whereas most inanimate nouns take *-ŋ* (Vajda 2004: 19; Georg 2007: 93).
12. Ket is a tone language. The notation of tones here follows mostly Georg (2007: 47, 56). The following tones may occur: V̄ (even or slightly rising tone), Vˀ (shortly rising tone, accompanied by pharyngeal constriction), VV (rising–falling tone), V̀ (sharply falling tone), and V́ (rising tone).

Several nouns denoting paired objects distinguish three number forms, cf. *ókde* 'ear' vs. *ókde-n* 'ear-PL' = 'a pair of ears' vs. *ókde-n-iŋ* 'ear-PL-PL' = 'many ears'. According to Georg (2007: 92), these are traces of a "natural dual, or paral". In my view, this is misleading, since these terms would stand to reason having the paired object as base form, and not one part of it. The double occurrence of the plural suffix rather points to composed number (see Corbett 2000: 36), whereby, however, it remains open whether *ókde-n-iŋ* 'ear-PL-PL' can also have the meaning 'many pairs of ears'. This is even more complicated by the possibility of referring to one single part of the pair by using the word *qóleb* 'half', as seen in (32).

Southern Ket

(32) *Bú-da*   *qi:m*   **qóleb-des**   *òbilda.*
       he-GEN.M   wife   **half-eye**   COP.PST.SG
       'His wife had only one eye.' (Georg 2007: 92)

This pattern, in turn, implies that the entire pair is referred to by the un-marked noun form. Thus, a closer investigation into the semantics of the forms would be necessary, especially regarding the meaning distinction "many pairs of X" vs. "many single items of X".

Besides singular and plural, Ket exhibits the possibility of forming singulatives from mass nouns. The suffixes used here are *-dis* and *-lamt*, which are grammaticalized from *dēs* 'eye' and *lámt(a)* 'piece' respectively, e.g. *hónaŋdis* 'grain of sand' ← *hónaŋ* 'sand' and *sújiŋlamt* 'rag' ← *sújiŋ* 'fabric' (Georg 2007: 132–133; Helimski 2016: 158–159).

The plural paradigm of nouns is, apart from assimilation process-es (*d > n* after the plural marker *-(V)n ~ -(V)ŋ*, e.g. *hík-da* 'man-GEN' : *hík-en-na* 'man-PL-GEN'), regular (Werner 1997a: 104; Vajda 2004: 21; Georg 2007: 104–105). The plural forms of personal pronouns are formed regularly in Ket, too. In all three persons the plural marker *-(V)n ~ -(V)ŋ* is – besides regular stem alternations – clearly detectable (Georg 2007: 164). Table 8 shows the personal pronouns in Ket.

Table 8: Personal pronouns in Ket

	SG	PL
1	*ād*	*ótn*
2	*ū(k)*	*ókŋ*
3	*bū*	*būŋ*

According to Vajda (2004: 80) and Georg (2007: 138) a small class of adjectives, e.g. *qà* 'big', may agree with their head noun. As for demonstratives, Vajda (2004: 80) states that they agree regularly with their head noun, whereas according to Georg (2007: 167) demonstratives agree only with animate head nouns (33).

Southern Ket

(33) a. *kíne   de'ŋ*　　　　b. *kíde báŋ-in*
　　　this.PL   human.PL　　　　　this place-PL
　　　'these people'　　　　　　　'these places'
　　　(Georg 2007: 167)

After numerals greater than one, the plural form of nouns occurs, e.g. *īn de'ŋ* 'two human.PL' = 'two persons' (vs. *qo'k ke'd* 'one human') (Vajda 2004: 80). After quantifiers, the plural form of count nouns but the singular form of mass nouns occurs, e.g. *òn de'ŋ* 'many human.PL' = 'many people', but *qómat na'n* 'few bread.SG' = 'few/little bread' (Georg 2007: 183).

　　The verbal morphology of Ket is very extensive, exhibiting both subject and object agreement marking (Vajda 2004: 44; Georg 2007: 184). Here, only the main tendencies can be shown. In most conjugation classes, subject person agreement markers are realized separately from subject number agreement markers, i.e. there are two separate morphemes in two different slots, as can be seen in example (34b). The singular is unmarked on the verb form while the plural is expressed via the marker *-n*, which is apparently a variant of the usual plural marker *-(V)n ~ -(V)ŋ* (see above) (Vajda 2004: 48–50; Georg 2007: 190–201, 263).

Southern Ket

(34) a. ***dí*-*n*-*ò***　　　　　　b. ***dí*-*n*-*ò*-*n***
　　　1.SUBJ-**PST**-die　　　　　　**1.SUBJ**-**PST**-die-**PL.SUBJ**
　　　'I died.'　　　　　　　　　　'We died.'
　　　(Vajda 2004: 48–50)

Object agreement markers, in turn, express both person and number of the object referred to within one suffix (Vajda 2004: 48–50; Georg 2007: 190–201).

Southern Ket

(35) a. ***k*-*in*-*di*-*bǝk***　　　　　b. ***k*-*in*-*daŋ*-*bǝk***
　　　2.SUBJ-**PST**-**1SG.OBJ**-find　　　**2.SUBJ**-**PST**-**1PL.OBJ**-find
　　　'You [SG] found me.'　　　　　'You [SG] found us.'
　　　(Vajda 2004: 48–50)

Leaving further conjugation classes aside, it can be concluded that both subject and object agreement is expressed on the Ket verb, whereby the former is expressed with two suffixes in each verb form, and the latter with one single suffix.

Within the verb phrase, target and controller exhibit obligatory number agreement in the case of objects. In the case of subjects, pronominal subjects agree obligatorily in number, but nominal subjects show variation: only nouns denoting animate referents (36a) agree with the predicate in number, whereas inanimate referents do not (36b) (Georg 2007: 101).[13]

Northern Ket

(36) a. […] ***ra-bis'nimin***   *qaj*   ***d**-aŋat-ol-i-bet-**n**.*
         3SG-brother.PL   PTCL   **3P.SUBJ**-wait-PST-EP-make-**PL.SUBJ**
        '[…] his brothers began waiting for him.'
        (Kryukova & Glazunov 2010: 189)

    b. *Sʲɛsʲ-talʲ*   *eke*   ***lu-ŋ***   *bara*   ***do:ɣ**-ol-evat.*
        larch-ABL.M only **chipping-PL** PTCL **3P.SUBJ.fly**-PST-ITER
        '[Only some] chippings fell down from the larch.'
        (Kryukova & Nefedov 2017: 192; sentence 39)

4. Comparison and typological implications

4.1. Convergences and divergences

As can be seen from the discussion in Section 3, the languages under investigation exhibit many common traits as well as typologically common phenomena, which will be dealt with here shortly.

All languages under discussion systematically exhibit singular and plural number, the Uralic languages (Eastern Khanty, Selkup, Nganasan) additionally dual number. In Dolgan, there is dual number in two minor domains (first person imperative; dual-marking item *bihikki*), while in Chulym Turkic, Ewenki and Ket, there is synchronically no dual number at all. In all three Uralic languages, the dual is to some extent involved in expressing coordination or accompaniment, especially together with kinship terms. From a broader Uralic perspective, this is nothing special, but it is often reported (e.g. Honti 1997: 46–47; Kulonen 2007: 51 for Eastern

---

13. Glossing and translations are slightly adapted and unified in (36).

Mansi, Siegl 2013: 241 for Forest Enets). Additionally, general number forms – which are homonymous with singular forms – appear regularly in Eastern Khanty, Selkup, Dolgan and Chulym Turkic, though in Nganasan and Ewenki they occur much less frequently, and in Ket they do not occur at all. General number and its entailments for related phenomena is discussed in detail in Section 4.2, as are the singulative forms in Eastern Khanty, Selkup and Ket in Section 4.3. Table 9 summarizes the number values found in the analyzed languages.

Table 9: Number values in the analyzed languages

Language	General number	Singular	Dual	Plural	Singulative
Eastern Khanty	+	+	+	+	(+)
Nganasan	(+)	+	+	+	–
Selkup	+	+	+	+	+
Dolgan	+	+	(+)	+	–
Chulym Turkic	+	+	–	+	–
Ewenki	(+)	+	–	+	–
Ket	–	+	–	+	+

As for the formal expression of number in the languages under investigation, again many common traits can be observed. In all languages, number on nouns is predominantly marked via affixation in a very regular manner. Nevertheless, this is not the whole story. While in Selkup, Dolgan, Chulym Turkic and Ewenki the stems remain largely unchanged, in Eastern Khanty, Nganasan and Ket several regular stem alternations occur. In Eastern Khanty, the stem alternations are concomitant and, furthermore, not restricted to number marking (e.g. *wajəɣ* 'animal; beast' : *wajk-at* 'animal-PL', *joɣəl* 'bow' : *joɣl-at* 'bow-PL'), thus, they hardly mark number themselves.

In Nganasan, the situation is more complex. Firstly, Nganasan exhibits a morphonological process called consonant gradation, which leads to regular stem alternations in various morphological positions (Wagner-Nagy 2019: 74–78). As for number marking, consonant gradation applies in the nominative plural as well as in some oblique plural cases (Wagner-Nagy 2019: 193), e.g. *maʔ* 'tent' : *maðə-ʔ* 'tent-PL' : *maðə-j* 'tent-PL.ACC' : *ma-tinü* 'tent-PL.LOC'. Secondly, in the genitive plural as well as in possessive plural forms, the stem of the noun and the plural marker *-j* amalgamate, leading to

vowel changes (Wagner-Nagy 2019: 193, 207), e.g. *kümaa* 'knife' : *kümaa-ʔ* 'knife-PL' : *kümau-ʔ* 'knife-PL.GEN' : *kümau-ńə* 'knife-PL.1SG' : *kümau-čə* 'knife-PL.2SG'. In the non-possessive forms exhibiting consonant grada-tion, plural marking is still established via the plural markers *-ʔ* and traces of *-j*, respectively, or via portmanteau plural case suffixes. But in possessive plural forms, no plural marker is visible any more – synchronically, only the stem alternation as well as the form of the possessive suffix itself tell whether a form is plural or not (see Däbritz 2017 for details). Thus, Ngana-san has two strategies to express number: affixation and to a lesser extent (in possessive plural forms) stem alternation, too.

In Ket, plural marking interacts with stem alternations in a very complex way. First, there are stem alternations that are concomitant with affixation, e.g. the loss of stem-final stops, cf. *lūk* 'splinter' : *luu-ŋ* 'splinter-PL' (Georg 2007: 99). Second, there are vowel changes, traditionally called "Ablaut". Also these vowel changes are concomitant with the plural suffix like in *diʔ* 'log' : *daʔ-n* 'log-PL' or *qīt* 'wolf' : *qát-n* 'wolf-PL' (Georg 2007: 97). Thirdly, there are tone changes in the stem. Tone changes often are also concomitant with plural marking via affixation, e.g. *àŋ* 'rope' : *áŋ-en* 'rope-PL' and *qām* 'arrow' : *qám-en* 'arrow-PL' (Georg 2007: 95). But in contrast to consonant deletion and vowel changes, tone changes can be the only indication that a form is plural, e.g. *ēj* 'tongue' : *èj* 'tongue.PL' and *qóqbun* 'cuckoo' : *qòqbun* 'cuckoo.PL' (Georg 2007: 98). This means that Ket uses two strategies for plural marking, namely affixation (with possible concomitant stem alterna-tions) and stem alternations, more precisely tone changes.

Zero expression of number in the narrow sense (like English *sheep* : *sheep*) does occur very rarely in the languages under investigation. Only a few lexemes in Ket like *sūj* 'mosquito.SG' ~ 'mosquito.PL' can count as an example (see Section 3.7). Moreover, in one set of verbal person-num-ber endings, Ewenki shows zero marking for the third person plural, e.g. *əmə-rə* 'come-AOR.3PL', while the third person singular is marked with *-n*, e.g. *əmə-rə-n* 'come-AOR-3SG'.

As for suppletion, the borderline between stem alternation, irregularity and suppletion is often fuzzy. In a narrow sense, suppletive number mark-ing on nouns is absent or very rare in all languages. Examples that can be noted are only Khanty *qu* 'man; person' : *jaɣ* 'people' (Filchenko 2010: 91) and Ket *keʔd* 'person' : *deʔŋ* 'people', *ōks* 'tree' : *aʔq* 'trees' and *qoʔd* 'way; road' : *qíneŋ* 'ways; roads' (Georg 2007: 100). Within the domain of personal pronouns, on the contrary, the issue is more complex. From a synchronic

point of view, many dual and plural forms can by no means be derived from the corresponding singular form by simply adding a number marker. Diachronically, however, most forms can be traced back to one and the same stem. A case in point is Eastern Khanty, whose first person pronouns *mä*, *min* and *məŋ* can be traced back to a single pronominal stem **m(V)-* (Rédei 1988: 294–295), the dual and plural forms, nevertheless, showing no traceable number morphology. Consequently, the Eastern Khanty personal pronouns are no instance of suppletion in the narrow sense, but surely exhibit irregular patterns of forming dual and plural forms. This is even more obvious in the case of Dolgan, cf. *min* vs. *bihigi* in the first person and *en* vs. *ehigi* in the second person. It is interesting to note that irregular plural (and dual) forms of personal pronouns are most common in the first person, common in the second person, but rare in the third person. This will be discussed in more detail in Section 4.4 in context of the category of animacy.

Number agreement within noun phrases is absent in Eastern Khanty, Selkup, Dolgan and Chulym Turkic, which is a frequent pattern in Northern Eurasian languages (Nevskaya & Amal 2020: 589; Rießler 2016: 170–171), regardless of the type of modifier included. In Nganasan, in turn, both adjectives and demonstratives obligatorily agree with their head noun (Wagner-Nagy 2019: 309), which is a clear divergence from the typical pattern in the Uralic languages (except for the Finnic and partly the Saami branches). In Ewenki, agreement seems to be the prevalent pattern for adjectives – though varying across dialects – while demonstratives do not tend to agree (according to the material in Däbritz & Gusev, in preparation). Finally, in Ket agreement of adjectives with their head noun may occur, but it is restricted, according to Vajda (2004: 80) by the semantics of the adjective.

Numerals greater than one (or greater than two, if there is dual number) are followed by unmarked forms of nouns without exception in Eastern Khanty, Selkup, Dolgan and Chulym Turkic. In Nganasan and Ewenki, both unmarked and plural forms can be observed, and in Ket plural forms are obligatory. These patterns will be discussed in more detail in Section 4.2 in the context of general number forms. In the case of the numeral 'two', the dual may occur on the modified noun – obligatorily in Eastern Khanty (Filchenko 2010: 195), and optionally in Nganasan (Wagner-Nagy 2019: 313) and Selkup (Kuznecova et al. 1980: 189) – otherwise the unmarked form is used, too. According to the material analyzed here (Brykina et al. 2018; Brykina et al. 2020), in both Nganasan and Selkup, dual forms occur mostly with animate and human referents.

Table 10 summarizes the agreement patterns in noun phrases in the investigated languages.

Table 10: Number agreement in noun phrases

Language	Adj N	Dem N	Num (> 1) N	
Eastern Khanty	–	–	+	(dual)
			–	(plural)
Nganasan	+	+	(+)	(dual)
			(+)	(plural)
Selkup	–	–	(+)	(dual)
			–	(plural)
Dolgan	–	–	–	
Chulym Turkic	–	–	–	
Ewenki	+	(+)	(+)	
Ket	(+)	(+)	+	

Number agreement in the verb phrase can concern subject agreement and object agreement. Subject number agreement is obligatory for pronominal subjects in the first and second person. For third person subjects, be they nominal or pronominal, this holds true for Eastern Khanty, Nganasan, Selkup, Dolgan and Ewenki. In Chulym Turkic, subject number agreement is optional for all third person referents (pronoun, human, animate, inanimate referents), as seen in (37) with the third person plural pronoun as subject and the predicate with third person singular morphology.

Chulym Turkic

(37) *Äšämändä ajt-wa-**an**   **olar**.*
      nothing   say-NEG-**PST.3SG**   **3PL**
      'They said nothing.' (Lemskaya 2015: 239)

From a Turkological point of view, it can be noted that Chulym Turkic, allowing for the lack of agreement within the verb phrase, behaves more typically than Dolgan, calling for agreement with the verb phrase (see Johanson 1998: 53). In Ket, first and second person subjects as well as animate third person subjects exhibit agreement with the predicate.

The Uralic languages Eastern Khanty, Nganasan and Selkup exhibit an objective conjugation, often referred to as "object agreement". In the case of Selkup, the appearance of this conjugation only indicates the existence

of a direct object in the clause, but neither number nor person of the object are marked on the verb (Kuznecova et al. 1980: 234–235). In Eastern Khanty and Nganasan, the number of the direct object is indicated when the objective conjugation is used (Filchenko 2010: 273; Wagner-Nagy 2019: 230), though in Eastern Khanty number agreement is not obligatory. The usage of the objective conjugation is, however, not mandatory if there is a direct object in the clause. It rather depends on largely information-structural criteria (Däbritz 2021: Ch. 9.3). From a Uralic point of view, the appearance of an objective conjugation governed by information-structural criteria is common to all Ob-Ugric and Samoyedic languages (Dalrymple & Nikolaeva 2011: 194–200) and the languages analyzed here form no exception in this respect. From a formal perspective the third person object number markers are homonymous with the nominal number suffixes used together with possessive suffixes in Eastern Khanty and Nganasan. This is surely expected from a broader Uralic perspective, since it can – at least in relicts – be observed in all other Uralic languages exhibiting object number agreement (i.e. Erzya and Moksha, Mansi, Northern and Southern Khanty, Nenets, Enets), and can diachronically be explained (e.g. Honti 1997: 102). From a more general point of view, this is in line with Siewierska's (1998: 35–36) observation that it is slightly more likely for object agreement paradigms to exhibit parallels with adnominal possessor marking than for subject agreement paradigms. A possible explanation is that both objects and possessed nouns tend to correspond to thematic patient roles, whereas subjects tend to correspond to thematic agent roles (see Dowty 1991: 575–576).

Finally, also Ket exhibits object agreement. In Ket, however, object agreement is not conditioned by information structure, rather it is obligatory if there is an object in the clause. This pattern is common within the whole Yeniseian language family (Werner 1995: 98, 100–102), Ket, thus, being a typical representative.

4.2. General number and its entailments

As was shown in Section 3, but rarely explicitly mentioned in grammatical descriptions, almost all the languages under investigation – with the exception of Ket – exhibit general number forms of nouns to a certain extent. Already in Section 2 it has been said that this is by no means uncommon, nor unknown in Northern Eurasian languages (see Benzing 1955, Honti 1997, Johansson 1998, Gruntov & Mazo 2020). However, the existence of

general number in a language appears to have two entailments which are worth commenting on, given the analyzed language data.

First, there is a striking parallel of general number on the one hand and the form of nouns following cardinal numerals on the other hand. One group of the languages investigated (Eastern Khanty, Selkup; Dolgan, Chulym Turkic) has general number and numerals greater than one or two (depending on the availability of dual forms) followed by number-unmarked forms. Ket at the other extreme has no general number and numerals greater than one are followed by plural forms. Ewenki and Nganasan take an intermediate position, showing variation: after numerals greater than two, Nganasan has mostly number-unmarked forms, but not always, while Ewenki, in turn, has mostly plural forms after numerals greater than one, whereby also number-unmarked forms are attested. This gives rise to the assumption that the observed parallel may be not concomitant but causal. In this case, the former group of languages would have the relevant numeral followed not by singular forms but by general number forms. In the given context, the numeral easily disambiguates the form of the noun underspecified for number. Thus, plural (or dual) marking of the noun itself is no longer needed. Honti (1997: 6–7, 41, 102) argues exactly the same way, and – given the wide distribution of general number forms in the Uralic languages according to him (ibid.) – reconstructs general number even for Proto-Uralic. From a more general point of view, the parallel of general number forms and unmarked nouns after numerals is in line with Alexiadou's (2019) observations mentioned in Section 2, since in either domain the underspecification of the bare noun form is a prerequisite. So, the languages under investigation as well as other Northern Eurasian languages imply that the observed parallel is indeed a causal one. Further investigation involving further languages and language families is, thus, highly desirable, but it goes beyond the scope of this paper.

Second, the seemingly irregular behavior of nouns denoting paired objects is very enlightening in this context. Given *inter alia* Haspelmath and Karjus' (2017) markedness- and frequency-based approach, the least marked denotation of these nouns is the whole pair (i.e. two pieces together), since eyes, ears or shoes seldom occur alone. As in the investigated languages with general number, the base form of a noun is not per default associated with a singular reading, this underspecified base form of paired items refers to the whole pair here. In contrast, in Nganasan and Ewenki – as well as in German or English – the bare noun form is associated with

a singular reading, whence for denoting pairs the dual or plural has to be used. Consequently, the former group of languages refers to one part of the pair with a quantifier meaning 'half', whereas in Nganasan and Ewenki simply the singular form of the noun is used. The Ket data, admittedly, behaves inconclusively in this respect. Ket exhibits no general number forms, but in the case of paired objects, the unmarked noun form may refer to either one part of the pair or to the whole pair. The former pattern can be proven by the complex plurals in *ókde* 'ear' vs. *ókde-n* 'ear-PL' = 'a pair of ears' vs. *ókde-n-iŋ* 'ear-PL-PL' = 'many ears' (see above), and the latter pattern can be proven by the usage of the quantifier *qóleb* 'half'. From a Yeniseian point of view, however, Ket is no exception in this respect, since for Yug and Kott the same is reported (Werner 1995: 87; 1997b: 69–70).

4.3. Singulatives

Abstract nouns and mass nouns hardly exhibit number marking in the investigated languages, which is neither surprising nor typologically uncommon (Corbett 2000: 78–82). In order to single out individual pieces or parts of a mass noun, Selkup and Ket have a morphological strategy, namely singulative words and suffixes. The following examples show their usage in context.

Central Selkup

(38) *Mat qiška-**xaj**-tko   me-ja-p.*
      1SG   star-**SNGL**-TRL   do-AOR-1SG.OBC
      'I turned them into [single] stars.'
      (Brykina et al. 2020, KFN_1967_BigBear2_flk.018)

Northern Selkup

(39) [...] *qaj-**lak**,   awsɨ-**laka**   orqɨl'-pa-t.*
             what-**SNGL**   food-**SNGL**   catch-PST.NAR-3SG.OBC
      '[The raven] grasped a piece of something, a piece of food.'
      (Brykina et al. 2020; KR_196X_RavensAndHares2_flk.029)

Northern Ket

(40) *Bu   bat   qa-jn-em   **qiʔt lɛmte**, [...].*
       3SG.M PTCL   take-PST-STEM **fat   SNGL**
      'He took a piece of fat [and threw it down in the cauldron].'
      (Kryukova 2015: 158)

The Selkup examples (38) and (39) show the singulative markers *-xaj* and *-lak(a)*, which are grammaticalized from the words *saj ~ haj* 'eye' and *laka* 'piece' respectively (Beáta Wagner-Nagy, p.c.). The Ket example (40) shows the singulative marker *lɛmte* 'piece' (a variant of *lámt(a)* id.), which can also occur as suffix *-lamt*; similarly, the lexeme *dēs* 'eye' and the suffix *-dis* can be used (Georg 2007: 132). In either language, a grammaticalization process is, thus, ongoing: free lexemes denoting small entities (piece, eye) are juxtaposed to lexemes denoting masses such as food and fat in (39) and (40), or referents that usually occur in groups such as stars in (38); this step can be seen in the Ket example (40). Then, both lexemes gradually amalgamate into one phonological word, whereby the singulative lexeme may change its phonological shape, e.g. via vowel apocope like in the Selkup example (39). Finally, the singulative marker is completely included into the morphological pattern of the noun modified by it, the Selkup example (38) being a good illustration, since the singulative marker is even followed by a case suffix here.

In Eastern Khanty, the lexeme *sem* 'eye' appears to be used in similar contexts as well, whereby, however, only one relevant type – displayed in (41) – could be found in the analyzed material.

Eastern Khanty

(41) *əjpä   jomentʃʼəɣ   **sem**   nuɣ   kül'-ɣən   […].*
      once   berry   **eye**   up   get.up-PST.3SG
      'Once the berry got up [and started to make a fire].'
      (Filchenko et al. 2012: 56)

Also Steinitz (1966–1993: 1339) mentions similar types for the relevant Khanty dialects, e.g. Vakh Khanty *jəŋ-sem* 'water drop; lit. water-eye' and *märən-sem* 'egg of roe; lit. roe-eye' as well as Vasyugan Khanty *tŏntəɣ-sem* '(small) stripe of birch bark; lit. birch bark-eye'. Especially the latter item points towards a grammaticalization of *sem* 'eye' as a singulative marker, since an elongated item is referred to, something that can hardly be reconciled with the semantics of 'eye'. Nevertheless, it has to be stated that the Eastern Khanty case calls for further research, since especially the morphosyntactic patterns of the given items are not fully understood yet.

Singulatives are not unattested in a global perspective, the most prominent examples probably being Welsh (< Indo-European), Maltese (< Semitic) and several Nilo-Saharan languages (Haspelmath & Karjus 2017:

1214–1217; Grimm 2018: 530–538). From a quantitative point of view, singulatives are rather rare; within Plank's (2003) sample of 205 languages, only 11 exhibit singulative marking.

Within the Uralic language family, grammaticalized singulative markers as described here are not attested at all, except for the Selkup case discussed in Section 3.3. However, Jussi Ylikoski (p.c.) has drawn my attention to the circumstance that in many lexicographic sources very similar examples are mentioned without further ado, as displayed in the probably not exhaustive Table 11 (data from Nielsen 1932, Itkonen 2011 [1958], Munkácsi & Kálmán 1986 and Lehtisalo 1956).

Table 11: Lexeme 'eye' marking singulatives in Uralic

Language	Item	Literal meaning	Meaning
Northern Saami	*muohtačalbmi*	snow-eye	snowflake
	káf(f)ečalbmi	coffee-eye	coffee bean
Skolt Saami	*mueʹrjjčââlmaž*	berry-eye	one single berry
Northern Mansi	*rakw-sam*	rain-eye	raindrop
Tundra Nenets	*sārʹun seβ*	rain-eye	raindrop

The given data stand to reason that the usage of successor forms of Proto-Uralic **ćilmä* 'eye' is much more widespread than it was formerly attested (Ylikoski 2021). However, morphosyntactic and empirically based studies on this topic are still lacking for the indicated languages, so that it remains an open question to what degree the relevant item is grammaticalized in each particular language. Additionally, the Hungarian classifiers *szem* 'eye', *fej* 'head' or *darab* 'piece' – e.g. *egy szem szőlő* 'one single grape' or *két fej karfiol* 'two heads of cauliflower' – appear to be a related phenomenon at first glance. However, Csirmaz and Dékány (2014) show convincingly that their occurrence is both morphosyntactically and semantically considerably restricted, whereby the latter is more important here. The combinations **egy fej szőlő* or **két szem karfiol* are not possible, something that can be explained by the fact that *szem* 'eye' is only applied to round, single and small objects, and *fej* to round, single but bigger objects. This contradicts analyzing these items as singulative markers, as does the fact that they can elliptically even be used without their head noun, which would not be possible if they were semantically bleached.

Within the Yeniseian language family, in turn, singulative markers appear to be rather common, since also Yug and Kott exhibit them, e.g. Yug *čəŋ-lap* 'hair-SNGL' = 'one single hair' vs. *čəŋ* 'hair' (Werner 1995: 87–88). Possible areal implications of this striking Uralic–Ket parallel will be discussed in Section 5.

4.4. Number and animacy

The semantic category of animacy is concerned, roughly speaking, with the distinction of certain groups of referents, such as humans, non-humans, kin, animates and inanimates, whereby the categorization of referents is surely language-dependent (Comrie 1989: 185). With respect to the linguistic behavior of the items included, several hierarchies can be established in this domain, namely person hierarchy (1st/2nd person > 3rd person), nominal hierarchy (pronoun > noun) and animacy hierarchy proper (human > animate > inanimate) (Comrie 1989: 197–198). The culmination of these single hierarchies leads to the well-known animacy hierarchy (see Silverstein 1976, Comrie 1989 for details):

1st/2nd person > 3rd person > kin > human > animate > inanimate

The hierarchy implies that items on its left are high on the hierarchy while items on its right are low on the hierarchy. Thus, the animacy hierarchy predicts the linguistic behavior of different classes of nominal referents, number marking being no exception. From a functional point of view, the existence of different number values in a language does not necessarily mean that they are equally accessible for all kinds of referents. The animacy hierarchy predicts that items high on the animacy hierarchy are potentially marked for more number values than items low on it, but never the other way around (Corbett 2000: 56–57). A simple example is Northern Saami (< Uralic), where personal pronouns have dual forms (*mon ~ mun* 'I' vs. *moai* 'we two' vs. *mii* 'we many') but common nouns do not (*guolli* 'fish' vs. *guoli-t* 'fish-PL') (Korhonen 1981: 206–210). In the languages under investigation here, two aspects are important and possibly meaningful for typological research in general.

First, general number forms are not available for personal pronouns, be they free or bound, and neither for nouns denoting kinship. The latter can be explained by the fact that reference to kinship terms is mostly definite

or at least specific. In the languages investigated here, the nearly obligatory possessive marking of kinship terms underlines this explanation, as seen in (42).

Northern Selkup

(42) *Ɔmtijqo* **näja-iː-m-tɨ**   *ponä*   *sɔːntir-qa* *aša*
      czar   **daughter-PL-ACC-3SG**   outwards   play-INF   NEG

      üːtɨ-mmi-mpa-t.
      let.go-DUR-PST.NAR-3SG.OBC
      'The czar didn't let his daughters play outside.'
      (Brykina et al. 2020; KNS_1966_Markincha_flk.003)

In contrast to that, non-kin human referents can be expressed via general number forms, as displayed in (43).

Central Selkup

(43) *Okkɨr* *na-l'-gu-t*   *üde-mba-t*   *šidaːro*
      one   woman-ADJZ-human-3SG   send-PST.NAR-3PL   twenty

      wersta-nd   *täbɨ-l'*   ***qup***   *qwär-gu.*
      verst-ILL   man-ADJZ   **human** call-INF
      'They sent one woman twenty versts away, to call the men.'
      (Brykina et al. 2020; KFN_1965_GirlAndBear1_nar.020)

This pattern can be observed in the case of the other languages exhibiting general number forms, too. Thus, there is a split on the animacy hierarchy between kin referents and non-kin human referents with respect to general number forms. At first glance, this seems to contradict the prediction made above that items high on the hierarchy potentially mark more number values than items low on the hierarchy. However, since general number is not a number value of its own but refers to the underspecification of certain number forms, the hierarchy rather predicts that items high on the hierarchy rather call for obligatory number marking, whereas items low on the hierarchy are rather available for underspecified forms (Corbett 2000: 70). Given this reading, the observed pattern completely adheres to the animacy hierarchy.

     Second, the role of pronouns (both free and bound) is special with regard to number marking: When referring to full nouns, the amount of referents of a plural form can be viewed as a sum of single referents. But when referring to pronouns, this is more complicated. The first person

pronoun *I* points to the speaker of an utterance, but its plural counterpart *we* seldom refers to many speakers, but rather to one speaker and some referent(s) associated with her/him. The second person pronoun *you* [i.e. *thou*] points to the hearer of an utterance, and its plural counterpart *you* [i.e. *ye*] may point either to many hearers or to a single hearer and some referent(s) associated with him/her. The third person pronouns *he/she/it* point to an entity being neither speaker nor hearer, and its plural counterpart *they* points to many such entities; see Jespersen (1924: 191–194) and Bhat (2004: Chapter 4.2) for a thorough discussion. Combining this with the person and animacy hierarchies, it appears to be highly expected that first person pronouns behave least regularly with respect to number marking, followed by second person pronouns, whereas third person pronouns may behave rather regularly. This expectation is confirmed by the material analyzed here.

A split between number marking of first/second person pronouns and third person pronouns could be observed in all languages but Eastern Khanty. Additionally, Dolgan as well as Chulym Turkic exhibit a split between the first and the second person each. In Dolgan, there are two instances of dual marking in the first person, but not in the second person and neither with items lower on the animacy hierarchy. Within the Turkic language family, this pattern is exceptional; seemingly parallel patterns in Sakha (Yakut) and South Siberian languages are rather instances of inclusive vs. exclusive first person (see Nevskaya 2005). In Chulym Turkic, both free and bound person markers are irregularly formed in the first person (e.g. *män* 'I' vs. *pis* 'we'), but regularly formed with the plural marker *-LAr* in the second and third person (*sän* 'you.SG' vs. *silär* 'you.PL' and *ol* 's/he' vs. *olar* 'they'). This is a shared phenomenon with other South Siberian Turkic languages such as Shor and Khakas, but a clear deviation from the standard Turkic pattern (Schönig 1998: 408–409). Especially these splits are important for the categories of animacy and person inasmuch as the hierarchy of the first and second person is highly debated in linguistic research (Corbett 2000: 64–66; Gildea & Zúñiga 2016). Dolgan and Chulym Turkic, thus, give independently from each other evidence for the first person being indeed hierarchically higher than the second person.

5. Areal implications

At the beginning of this paper, the choice of languages included in the investigated sample was partly motivated by observations that in Western and Central Siberia contact-induced linguistic convergences are rather frequent (e.g. Hajdú 1979, Pusztay 1987, Helimski 2003) and that there might be even an "Ostyak (Ob-Yeniseic) Sprachbund" (Helimski 2003: 160). Coming back to these observations, it has to be stated that the category of number can give no evidence in favor of assuming meaningful areal patterns. On the one hand, the languages included in the sample of this study do not exhibit conclusive patterns that would make it possible to treat them as an areal unit – e.g. only the Uralic languages included have dual number, only Chulym Turkic systematically exhibits no agreement of subject and predicate in the third person etc. On the other hand, including neighboring languages such as Mansi, Nenets or Yug would not change the picture dramatically, either: also Mansi and Nenets have dual number, and neither of those languages exhibits lack of subject agreement etc. However, it can be shown that the observed features at least account for a homogeneity in the sense of e.g. Janhunen's (2014) "Ural-Altaic" continuum within Northern Eurasia, since the linguistic expression of the category of number shows many parallels in Uralic and Altaic (Transeurasian) languages (general number with all its entailments, lack of number agreement in noun phrases, rather irregular number marking in 1st/2nd person vs. rather regular marking in 3rd person) and only few divergences (dual in Uralic languages, object agreement in Uralic languages). In contrast, when including e.g. Russian as dominant contact language, many divergences can be detected, the most important being the lack of general number and obligatory number agreement within noun phrases,[14] something also indicating that Northern Eurasian languages can be analyzed as forming a rather homogeneous unit. In this context, also Ravila's (1941) assumption of the "more developed" number system in Uralic languages compared to

---

14. The latter surely raises the question of whether number agreement in noun phrases in Northern Samoyedic languages as well as Ewenki may be due to or at least accelerated by Russian influence. In my view, no final statement can be made here, since in many similar cases – e.g. SVO patterns in Nganasan and Dolgan (see Däbritz 2020) – Russian influence can be excluded as the driving factor, if one analyzes language data as well as sociolinguistic patterns carefully.

Altaic languages has to be relativized: surely the former exhibit dual marking, which the latter lack, but in other domains, as could be shown, there are far more convergences than divergences.

Table 12 provides a summary of those features which were discussed in more detail in Section 4 – extended with data from Mansi, Nenets, Enets and Yug as well as Russian.[15]

Table 12: Comparison of number features

	General number	Dual	Singulatives	Subject number agreement	Object number agreement	Agreement in NPs
E. Khanty	+	+	(+)	+	(+)	–
Nganasan	(+)	+	–	+	+	+
Selkup	+	+	+	+	–	–
Dolgan	+	(+)	–	+	–	–
Chulym Turkic	+	–	–	(+)	–	–
Ewenki	(+)	–	–	+	–	(+)
Ket	–	–	+	+	+	(+)
Mansi	+	+	(+)	+	+	–
Nenets	(+)	+	(+)	+	+	+
Enets	(+)	+	–	+	+	(+)
Yug	–	–	+	+	+	(+)
Russian	–	–	(+)	+	–	+

Finally, one meaningful linking can be derived from this list, namely the occurrence of singulatives in relevant Uralic languages and Yeniseian languages – most clearly in Selkup and Ket – but their lack in further potential contact languages such as Chulym Turkic or Southern Ewenki. Already the occurrence of singulatives itself in two unrelated neighboring languages (and more generally: language families) appears to be suspicious inasmuch as singulatives are typologically rather uncommon (Plank 2003:

15. The given list solely relies on grammatical descriptions in the case of the added languages (Keresztes 1998, Nikolaeva 2014, Siegl 2013, Werner 1997b) and does not take into account corpus data.

259). Moreover, exactly the same lexemes (denoting 'eye' and 'piece') are involved, and the path of grammaticalization appears to be parallel, since in both Selkup and Ket the singulative markers can be free morphemes as well as suffixes. Finally, also the sociolinguistic situation makes a contact scenario entirely plausible, since Yeniseian people (both Kets and Yugs) settled together with Eastern Khanty and Selkup (Northern, Central and Southern) people for a long time (Werner 1997b: 3–4; Vajda 2009: 480). However, one single uniting element surely does not suffice for establishing a language area, regardless of the latter's size. Further distinctive features, which connect Selkup (as well as possibly Eastern Khanty) and Yeniseian, but separate them from surrounding languages, would additionally be needed. Besides that, it appears to be the case that in many further Uralic languages (e.g. Saami, Mansi, Nenets) the lexeme 'eye' occurs in very similar domains, possibly being grammaticalized as a singulative marker, too (Ylikoski 2021). Therefore, further research on this topic from both language-internal and comparative perspectives is highly needed. In any case, however, it can be stated that the Uralic and Yeniseian languages under observation clearly diverge from the surrounding Turkic and Tungusic languages, something which already represents a meaningful insight.

6. Conclusion and further outlook

In this paper seven languages from Western and Central Siberia (Eastern Khanty, Nganasan, Selkup; Dolgan, Chulym Turkic; Ewenki; Ket) were analyzed with respect to the linguistic expression of number. It was shown that all of them mark number grammatically, the number values expressed everywhere being singular and plural; dual number occurs regularly in the Uralic languages Eastern Khanty, Nganasan, and Selkup, and additionally in two minor domains in Dolgan. Besides that, there are underspecified forms, which are formally homonymous to singular forms, in Eastern Khanty, Selkup, Dolgan, Chulym Turkic, and partly in Nganasan and Ewenki. These forms qualify for *general number* in the sense of Corbett (2000), since they do not necessarily convey any reference to the number category. This is most noteworthy, since in most descriptions these forms are labeled as "singular with plural meaning" or the like. Furthermore, the existence of general number forms in these languages entails that nouns are underspecified after numerals greater than one or two (depending on the availability of dual forms) and that paired objects are referred to as a

whole by a bare noun form. In addition, Selkup, Eastern Khanty and Ket as well as further Yeniseian languages exhibit singulatives, singling out a referent from a mass or group of similar referents. Given the linguistic geography in the Yenisei basin, this can be analyzed as contact phenomenon, yielding a common Uralic-Yeniseian feature. Finally, the expression of number is tightly interwoven with the category of animacy from both a formal and a functional perspective. The Dolgan dual in the first person imperative and the formation of plural personal pronouns in Chulym Turkic (suppletive in 1st person, regular in 2nd and 3rd person) are especially noteworthy, since both phenomena show an otherwise seldom-attested split of the animacy hierarchy between the first and second person.

From an areal linguistic point of view, two conclusions can be drawn. First, the linguistic expression of the category of number cannot give evidence for assuming any kind of language area in the Ob-Yenisei region, regardless of which languages are included or not. Secondly, however, the observed phenomena and features adhere to the prevailing patterns in Siberia and Northern Eurasia in general, possibly being called a "Ural-Altaic" or "Ural-Transeurasian" continuum.

Finally, the investigation of genetically partly unrelated but typologically rather similar languages of Western and Central Siberia can contribute to the theoretical understanding of the number category, especially when it comes to the peculiarities discussed in Sections 4.2, 4.3 and 4.4. Similar investigations of both a broader set of languages and e.g. the closely related category of person, thus, appear promising for Uralic, Transeurasian and Siberian studies as well as for linguistic typology in general.

Acknowledgement

Non-standard abbreviations used in glosses

ADJZ	adjectivizer	LIM	limitative
AOR	aorist	NAR	narrative
ATTR	attributive	OBC	objective conjugation
DYA	dyadic kinship	PAUC	paucal
EP	epenthesis	POSS	possessor
GN	general number	PROPR	propriative
HAB	habitual	PTCL	particle
ILL	illative	SNGL	singulative
INCH	inchoative	STEM	stem formant
INFER	inferential	SUBJ	subject
ITER	iterative	TRL	translative
LAT	lative	VBZ	verbalizer

Data sources

BRYKINA, MARIA & GUSEV, VALENTIN & SZEVERÉNYI, SÁNDOR & WAGNER-NAGY, BEÁTA. 2018. *Nganasan Spoken Language Corpus (NSLC)*. Version 0.2. Publication date 2018-06-12. Archived in Hamburger Zentrum für Sprachkorpora. http://hdl.handle.net/11022/0000-0007-C6F2-8.

BRYKINA, MARIA & ORLOVA, SVETLANA & WAGNER-NAGY, BEÁTA. 2020. INEL Selkup Corpus. Version 1.0. Publication date 2018-06-30. Archived in Hamburger Zentrum für Sprachkorpora. In Wagner-Nagy, Beáta & Arkhipov, Alexandre & Ferger, Anne & Jettka, Daniel & Lehmberg, Timm (eds.), *The INEL corpora of indigenous Northern Eurasian languages*.
http://hdl.handle.net/11022/0000-0007-E1D5-A.

DÄBRITZ, CHRIS LASSE & GUSEV, VALENTIN. (in preparation). INEL Evenki Corpus. Version 1.0. Archived in Hamburger Zentrum für Sprachkorpora. In Wagner-Nagy, Beáta & Arkhipov, Alexandre & Ferger, Anne & Jettka, Daniel & Lazarenko, Elena & Lehmberg, Timm & Riaposov, Aleksandr (eds.), *The INEL corpora of indigenous Northern Eurasian languages*.

DÄBRITZ, CHRIS LASSE & KUDRYAKOVA, NINA & STAPERT, EUGÉNIE. 2019. INEL Dolgan Corpus. Version 1.0. Publication date 2019-08-31. Archived in Hamburger Zentrum für Sprachkorpora. In Wagner-Nagy, Beáta & Arkhipov, Alexandre & Ferger, Anne & Jettka, Daniel & Lehmberg, Timm (eds.), *The INEL corpora of indigenous Northern Eurasian languages*.
http://hdl.handle.net/11022/0000-0007-CAE7-1.

FILCHENKO, ANDREY. 2017. Khanty texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 5, 23–62. Tomsk: Tomsk Pedagogical State University.

Filchenko, Andrey & Kurganskaja, Ju. V. & Fellan, P. & Varda, V. E. & Potanina, O. S. 2015. Khanty texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 4, 27–107. Tomsk: Tomsk Pedagogical State University.

Filchenko, Andrey & Potanina, Olga. 2010. Khanty texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 1, 38–132. Tomsk: Veter.

Filchenko, Andrey & Potanina, Olga. 2012. Khanty texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 2, 24–71. Tomsk: Vajar.

Filchenko, Andrey & Potanina, Olga. 2013. Khanty texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 3, 26–152. Tomsk: Vajar.

Kryukova, Elena A. 2012. Ket texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 2, 101–142. Tomsk: Vajar.

Kryukova, Elena A. 2013. A Ket text. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 3, 202–294. Tomsk: Vajar.

Kryukova, Elena A. 2015. Ket texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 4, 150–184. Tomsk: Tomsk Pedagogical State University.

Kryukova, Elena A. & Glazunov, Pavel Ju. 2010. Ket texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 1, 185–210. Tomsk: Veter.

Kryukova, Elena A. & Nefedov, Andrey V. 2017. Ket texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 5, 177–209. Tomsk: Tomsk Pedagogical State University.

Lemskaya, Valeriya M. 2010a. Chulym Turkic texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 1, 262–314. Tomsk: Veter.

Lemskaya, Valeriya M. 2012. Chulym Turkic text. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 2, 184–237. Tomsk: Vajar.

Lemskaya, Valeriya M. 2013. A Chulym-Turkic text. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 3, 295–345. Tomsk: Vajar.

Lemskaya, Valeriya M. 2015. Chulym-Turkic texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 4, 217–291. Tomsk: Tomsk Pedagogical State University.

Lemskaya, Valeriya M. 2017. Chulym-Turkic texts. In Filchenko, Andrey (ed.), *Annotated folk and daily prose texts in the languages of the Ob-Yenisseic linguistic area*. Vol. 5, 103–176. Tomsk: Tomsk Pedagogical State University.

References

Alexiadou, Artemis. 2019. Morphological and semantic markedness revisited: The realization of plurality across languages. *Zeitschrift für Sprachwissenschaft* 38(1). 123–154. https://doi.org/10.1515/zfs-2019-0004.

Alhoniemi, Alho. 2010. *Marin kielioppi* (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten X). [2nd printing.] Helsinki: Suomalais-Ugrilainen Seura.

Anderson, Gregory D.S. & Harrison, K. David. 2006. Ös tili: Towards a comprehensive documentation of Middle and Upper Chulym dialects. *Turkic languages* 10. 47–72.

Artemjev, Nikolaj M. 2013. *Dolganskij jazyk. Čast′ 2. Morfologija*. Sankt-Peterburg: Almaz-Graf.

Artemjev, Nikolaj M. & Nazmutdinova, Tatjana S. & Spiridonova, Žanna P. 2013. *Dolganskij jazyk. Čast′ 3. Sintaksis*. Sankt-Peterburg: Almaz-Graf.

Bekker, Erika G. 1995. *Morfologija sel′kupskogo jazyka. Južnye dialekty*. Tomsk: Tomskij Gosudarstvennyj Pedagogičeskij Institut.

Benzing, Johannes. 1955. *Die tungusischen Sprachen: Versuch einer vergleichenden Grammatik* (Abhandlungen der geistes- und sozialwissenschaftlichen Klasse 1955(11)). Wiesbaden: Akademie der Wissenschaften und der Literatur in Mainz.

Bhat, Darbhe N.S. 2004. *Pronouns* (Oxford studies in typology and linguistic theory). Oxford: Oxford University Press.

Birjukovič, R. M. 1979. *Morfologija čulymsko-tjurkskogo jazyka. Čast′ 1. Kategorii imeni suščestvitel′nogo*. Moskva: Akademija Nauk SSSR.

Bulatova, Nadezhda & Grenoble, Lenore. 1999. *Evenki* (Languages of the world. Materials 141). Munich: LINCOM Europa.

Comrie, Bernard. 1989. *Language universals and linguistic typology: syntax and morphology*. 2nd edition. Oxford: Blackwell.

Corbett, Greville G. 2000. *Number* (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press.

Csirmaz, Anikó & Dékány, Éva. 2014. Hungarian is a classifier language. In Simone, Raffaele & Masini, Francesca (eds.), *Word classes: Nature, typology and representations* (Current issues in linguistic theory 322), 141–160. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.332

Däbritz, Chris Lasse. 2017. Zur Markierung von Numerus an Nomina in den samojedischen Sprachen – Synchronie und mögliche Implikationen für die Rekonstruktionen des Protosamojedischen und des Protouralischen. *Finnisch-Ugrische Mitteilungen* 41. 63–101.

Däbritz, Chris Lasse. 2019. First person imperative in Dolgan – clusivity or number distinction? *Finnisch-Ugrische Mitteilungen* 43. 1–12.

Däbritz, Chris Lasse. 2020. Focus position in SOV ~ SVO-varying languages – evidence from Enets, Nganasan and Dolgan. *Eesti ja soome-ugri keeleteaduse ajakiri* 12(2). 99–118. https://doi.org/10.12697/jeful.2020.11.2.04

Däbritz, Chris Lasse. 2021. *Topik, Fokus und Informationsstatus: Modellierung am Material nordwestsibirischer Sprachen* (Language, context and cognition 17). Berlin: De Gruyter. https://doi.org/10.1515/9783110716337

Dalrymple, Mary & Nikolaeva, Irina Nikolaeva. *Objects and information structure* (Cambridge studies in linguistics 131). Cambridge: Cambridge University Press.

Daniel, Michael. 2013. Plurality in independent personal pronouns. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/35 (Last access: 26.01.2021.)

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Dryer, Matthew S. 2013. Coding of nominal plurality. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/33 (Last access: 26.01.2021.)

Filchenko, Andrey. 2010. *Aspects of the grammar of Eastern Khanty*. Tomsk: Tomsk State Pedagogical University.

Georg, Stefan. 2007. *A descriptive grammar of Ket (Yenisey-Ostyak)* (Languages of Asia series 1). Folkestone: Global Oriental.

Gildea, Spike & Zúñiga, Fernando. 2016. Referential hierarchies: A new look at some historical and typological patterns. *Linguistics* 54(3). 483–529. https://doi.org/10.1515/ling-2016-0007

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of language*, 73–113. Cambridge [Mass.]: MIT Press.

Grimm, Scott. 2018. Grammatical number and the scale of individuation. *Language* 94(3). 527–574. https://doi.org/10.1353/lan.2018.0035

Gruntov, Ilya & Mazo, Olga. 2020. A comparative approach to nominal morphology in Transeurasian. Case and plurality. In Robbeets, Martine & Savelyev, Alexander (eds.), *The Oxford guide to the Transeurasian languages*, 522–553. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198804628.001.0001

Hajdú, Péter. 1979. Language contacts in North-West Siberia. *Finno-Ugrica Suecana* 2. 19–31.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70. https://doi.org/10.1017/S0022226705003683

Haspelmath, Martin. 2013. Occurrence of nominal plurality. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/34 (Last access: 26.01.2021.)

Haspelmath, Martin & Karjus, Andres. 2017. Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics* 55(6). 1213–1235. https://doi.org/10.1515/ling-2017-0026

Helimski, Evgenij A. 2003. Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: a survey. *Nyelvtudományi Közlemények* 100. 156–167.

Helimski, Eugene. 2016. S-singulatives in Ket. *Journal of Language Relationship* 14(3). 157–163. https://doi.org/10.31826/jlr-2017-143-404 (Post mortem published by Valentin Gusev.)

Hillewaert, Sarah & de Schryver, Gilles-Maurice. 2004. *Online Kiswahili (Swahili) – English dictionary*. http://goswahili.org/, last access: 26.03.2020.

Honti, László. 1984. *Chrestomathia ostiacica*. Budapest: Tankönyvkiadó.

Honti, László. 1997. Numerusprobleme (Ein Erkundungszug durch den Dschungel der uralischen Numeri). *Finnisch-Ugrische Forschungen* 54. 1–126.

Itkonen, T. I. 2011 [1958]. *Koltan- ja kuolanlapin sanakirja* (Lexica Societatis Fenno-Ugricae XV). Helsinki: Suomalais-Ugrilainen Seura.

Itturioz-Leza, José Luis & Skopeteas, Stavros. 2004. Numerus. In Booij, Geert E. & Lehmann, Christian & Mugdan, Joachim & Skopeteas, Stavros (eds.), *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung* (Handbücher zur Sprach- und Kommunikationswissenschaft 17, 2), 1053–1066. Berlin: De Gruyter.

Janhunen, Juha. 2014. The polygenetic origins of nominal morphology in the Transeurasian zone. In Robbeets, Martine & Bisang, Walter (eds.), *Paradigm change: in the Transeurasian languages and beyond* (Studies in language companion series 161), 311–336. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.161.21jan

Jespersen, Otto. 1924. *The philosophy of grammar*. London: Allen & Unwin.

Johanson, Lars. 1998. The structure of Turkic. In Johanson, Lars & Csató, Éva Á. (eds.), *The Turkic languages*, 30–66. London: Routledge.

Johanson, Lars & Robbeets, Martine. 2010. Introduction. In Johanson, Lars & Robbeets, Martine (eds.), *Transeurasian verbal morphology in a comparative perspective: genealogy, contact, chance* (Turcologica 78), 1–6. Wiesbaden: Harrassowitz.

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

Klumpp, Gerson & Mazzitelli, Lidia F. & Rozhanskiy, Fedor. 2018. Typology of Uralic languages: current views and new perspectives. *Eesti ja soome-ugri keeleteaduse ajakiri* 9(1). 9–30. https://doi.org/10.12697/jeful.2018.9.1.01

Konstantinova, Ol'ga A. 1964. *Èvenkijskij jazyk. Fonetika, morfologija*. Moskva: Nauka.

Korhonen, Mikko. 1981. *Johdatus lapin kielen historiaan* (Suomalaisen Kirjallisuuden Seuran Toimituksia 370). Helsinki: Suomalaisen Kirjallisuuden Seura.

Krifka, Manfred & Pelletier, Francis J. & Carlson, Gregory N. & ter Meulen, Alice & Chierchia, Gennaro & Link, Godehard. 1995. Introduction to genericity. In Carlson, Gregory N. & Pelletier, Francis J. (eds.), *The generic book*, 1–124. Chicago: University of Chicago Press.

Kulonen, Ulla-Maija. 2007. *Itämansin kielioppi ja tekstejä* (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten XV). Helsinki: Suomalais-Ugrilainen Seura.

Kuznecova, Ariadna I. & Chelimskij, Evgenij A. & Gruškina, Elena V. 1980. *Očerki po sel'kupskomu jazyku. Tazovskij dialekt.* Moskva: Izdatel'stvo Moskovskogo Universiteta.

Lehtisalo, Toivo. 1956. *Juraksamojedisches Wörterbuch* (Lexica Societatis Fenno-Ugricae XIII). Helsinki: Suomalais-Ugrilainen Seura.

Lemskaya, Valeriya M. 2010b. Middle Chulym: The state of the art. *Turkic Languages* 14(1). 113–126.

Li, Yong-Song & Lee, Ho-Young & Choi, Hyong-Won & Kim, Geon Sook & Lee, Dong Eun & Ölmez, Mehmet. 2008. *A study of the middle Chulym dialect of the Chulym language* (Altaic Languages Series 3). Seoul: Seoul National University Press.

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

Nedjalkov, Igor. 1997. *Evenki*. London: Routledge.

Nevskaya, Irina. 2005. Inclusive and exclusive in Turkic languages. In Filimonova, Elena (ed.), *Clusivity. Typology and case-studies of the inclusive-exclusive distinction* (Typological studies in languages 63), 341–358. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.63

Nevskaya, Irina & Amal, Lina. 2020. The nominal group, possessive agreement, and nominal sentences in the Transeurasian languages. In Robbeets, Martine & Savelyev, Alexander (eds.), *The Oxford guide to the Transeurasian languages*, 587–603. Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198804628.001.0001

Nielsen, Konrad. 1932. *Lappisk ordbok: grunnet på dialektene i Polmak, Karasjok og Kautokeino. Vol. 1: A–F.* Oslo: Aschehoug.

Nikolaeva, Irina. 2014. *A grammar of Tundra Nenets* (Mouton grammar library 65). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110320640

Plank, Frans. 2003. The selective elaboration of nominal or pronominal inflection. In Plank, Frans (ed.), *Noun phrase structure in the languages of Europe* (Empirical approaches to language typology, EUROTYP 20-7), 253–287. Berlin: De Gruyter Mouton.

Porotova, Teľmina I. 1990. *Kategorija množestvennosti v enisejskich jazykach*. Tomsk: Izdatel'stvo Tomskogo Universiteta.

Pusztay, János. 1987. *Analóg vonások az uráli és a paleoszibériai nyelvek névszói rendszerében* (A Magyar Nyelvtudományi Társaság kiadványai 173). Budapest: Magyar Nyelvtudományi Társaság.

Ravila, Paavo. 1941. Über die Verwendung der Numeruszeichen in den uralischen Sprachen. *Finnisch-Ugrische Forschungen* 27. 1–136.

Rédei, Károly. 1988. *Uralisches etymologisches Wörterbuch. 1. Uralische und finnisch-ugrische Schicht*. Budapest: Akadémiai kiadó.

Rießler, Michael. 2016. *Adjective attribution* (Studies in diversity linguistics 2). Berlin: Language Science Press. https://doi.org/10.17169/langsci.b19.294

Schönig, Claus. 1998. South Siberian Turkic. Johanson, Lars & Csató, Éva Á. (eds.), *The Turkic languages*, 403–416. London: Routledge.

Siegl, Florian. 2013. *Materials on Forest Enets, an indigenous language of Northern Siberia* (Suomalais-Ugrilaisen Seuran Toimituksia 267). Helsinki: Suomalais-Ugrilainen Seura.

Siewierska, Anna. 1998. Nominal and verbal person marking. *Linguistic Typology* 2. 1–53.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Dixon, R. M. W. (ed.), *Grammatical categories in Australian languages* (Linguistic Series 22), 112–171. Canberra: Australian Institute of Aboriginal Studies.

Steinitz, Wolfgang. 1966–1993. *Dialektologisches und etymologisches Wörterbuch der ostjakischen Sprache*. Berlin: Akademie-Verlag.

Vajda, Edward. 2004. *Ket* (Languages of the world. Materials 204). Munich: LINCOM Europa.

Vajda, Edward. 2009. Loanwords in Ket, a Yeniseian language of Siberia. In Haspelmath, Martin & Tadmor, Uri (eds.), *Loanwords in the world's languages: a comparative handbook*, 471–495. Berlin: De Gruyter Mouton.

Vasilevič, Glafira M. 1948. *Očerki dialektov èvenkijskogo jazyka*. Leningrad: Učpedgiz.

Vasilevič, Glafira M. 1958. *Èvenkijsko-russkij slovarʹ. S priloženijami i grammatičeskim očerkom èvenkijskogo jazyka*. Moskva: Gosudarstvennoe izdatelʹstvo inostrannych i nacionalʹnych slovarej.

VPN 2010 = *Vserossijskaja perepisʹ naselenija 2010. Tom 4. Nacionalʹnyj sostav i vladenie jazykami. Častʹ 5. Vladenie jazykami naseleniem Rossijskoj Federacii* [All-Russian census 2010. Vol. 4. National composition and command of languages. Part 5. Command of languages by the population of the Russian Federation]. http://www.gks.ru/free_doc/new_site/perepis2010/croc/Documents/Vol4/pub-04-05.pdf (Last access: 26.01.2021.)

Wagner-Nagy, Beáta. 2019. *A Grammar of Nganasan*. Leiden: Brill. https://doi.org/10.1163/9789004382763

Wagner-Nagy, Beáta & Arkhipov, Alexandre. 2020. Comitative constructions in Nganasan and Selkup. In Bartens, Hans-Hermann & Larsson, Lars-Gunnar & Mattsson, Katja & Molnár, Judit & Savolainen, Tiina (eds.), *Kīel joug om šīld. Festschrift zum 65. Geburtstag von Eberhard Winkler* (Veröffentlichungen der Societas Uralo-Altaica 94), 425–441. Wiesbaden: Harrassowitz.

Werner, Heinrich. 1995. *Zur Typologie der Jenissej-Sprachen* (Veröffentlichungen der Societas Uralo-Altaica 45). Wiesbaden: Harrassowitz.

Werner, Heinrich. 1997a. *Die ketische Sprache* (Tunguso-Sibirica 3). Wiesbaden: Harrassowitz.

Werner, Heinrich. 1997b. *Das Jugische (Sym-Ketische)* (Veröffentlichungen der Societas Uralo-Altaica 50). Wiesbaden: Harrassowitz.

Ylikoski, Jussi. 2021. Ice eyes, blood eyes: Remarks on the Uralic singulative marker **ćilmä* 'eye'. Poster presented at the conference Språkets funktion 17, 26–27 May 2021, Åbo. http://cc.oulu.fi/~jylikosk/filer/singulative.pdf

Niko Partanen & Alexandra Kellner
University of Helsinki

On the interplay between tense marking, aspect and temporal continuity in Udora Komi

The Udora dialect of Zyrian Komi lacks the morphological opposition between the present and future tenses that is found in other Komi dialects and the written standard. The morphemes corresponding to these tenses are, however, found in this dialect, with individual verbs showing a strong tendency to choose one of the two. This study shows that the two morphemes are not in free variation but rather carry various grammatical meanings, and that the variants are strongly connected to the lexical aspect of individual verbs.

Due to the rigidity of the system, the authors refer to the variants here as conjugation types. The *-as-* conjugation type, which corresponds to the Standard Komi future marker, occurs with all transitive verbs and a majority of intransitive verbs. However, the study also identifies a group of intransitive verbs occurring with the conjugation type *-ę-*. The verbs in the latter group can be analysed as temporally continuous. Additionally, there are other subgroupings that can be postulated, including verbs that describe involuntary actions. The system interacts in a predictable manner with Komi derivational morphology.

The study also corroborates the previously proposed historical connection between this characteristic of verbal morphology in the Udora dialect and Old Komi. The authors suggest that the verbal morphology seen in these Komi varieties must predate the contemporary tense system. The study provides a new direction for analysing the development of the tense system in the Permic languages, as it is shown that the factors underlying the variation extend beyond transitivity. As a previously undescribed phenomenon, the study describes the use of the Udora conjugation types in narrative tense structuring and demonstrates parallels with Standard Komi.

1. Introduction

The Udora dialect is a variety of the Zyrian Komi[1] language that is spoken in the westernmost corner of the Komi speaking area. In this region, Komi villages follow two large rivers, the Vaška and the Mezen. Geographically speaking, the area is relatively remote. The Udora region has 21 settlements that are included in the population censuses, but these can usually be divided into smaller local clusters of villages. According to the Russian population census conducted in 2010, there were 8,018 Komi living in the Udora region, comprising 40 percent of the region's population (Federal State Statistics Service, Russia 2010). There is no recent sociolinguistic research on the situation with regard to language maintenance in the region. Based on the authors' fieldwork experience, the traditional Komi villages in the area are largely Komi-speaking, but the younger inhabitants regularly move to larger cities, which has an impact on their language use. The official population statistics for the region paint a picture where approximately 75 percent of the inhabitants of the smaller villages, which include Glotovo, Koslan, Ćernut′evo, Ćuprovo, Pučkoma, Pyssa, Važgort and Jortom, are Komi-speaking (Federal State Statistics Service, Russia 2010). As such, the dialect has numerous unique features at all levels of language. Morphologically, the Udora dialect does not distinguish between the present and future tenses, which is unusual among the Permic languages and dialects.

In this study, we aim to describe the third-person verb marking in Udora and the functions of two distinct morphemes that are in variation. We propose that the use of these morphemes is lexically conditioned and connected to a lexical aspect system, so that intransitive, temporally continuous verbs receive a distinct morphological marking. The traditional description of this feature states that the Udora dialect employs formally identical (i.e. syncretic) present-tense and future-tense markers, but the temporal distinction between them is blurred (see Sorvačeva & Beznosikova 1990: 67; Lytkin 1961: 52–53; 1977a: 282; Cypanov 2005: 28, 141, 147).

---

1. Throughout the study, we use the terms *Zyrian Komi* and *Permian Komi* to refer to the two main Komi varieties and written standards. We find that this convention fits well with the terms *Зыран-Коми* and *Перым-Коми* used in Komi scientific research. The terms *Komi-Zyrian*, *Komi-Permyak* and *Perm Komi* can certainly also be used, and we do not consider the English conventions for referring to these languages to have fully emerged yet.

Following this pattern, a Udora dialect speaker would customarily inquire about the identity of a new person by asking:

(1)   Кудз тэнö шуасны?

     kuʒ́   *teṇe*   ***šu-asṇị***
     how   2SG.ACC say-FUT.3PL
     'What's your name? (lit. How will they call you?)'
     (Authors' field notes)

In Standard Komi, and to our knowledge in all other dialects, the idiomatic expression would almost without exception contain the present tense verb form *шуöны* /šuẹṇị/ 'they say', as the speaker's motive is to find out how someone is called at the moment so they will know how to refer to them. Of course, given the right context, the future form could also occur, as seen in (2). We note that the writer of this text, Aleksandr Matveev, was born in the Syktyvdin region:

(2)   Бöбöн тэнö шуасны, Владимир, татшöм ныв вылö кö он гöтрась!

     bẹb-ẹn *teṇe*   ***šu-asṇị***, *vlad´imir,* *tačẹm*   *ṇịv*
     fool-INS 2SG.ACC say-FUT.3PL Vladimir this_kind girl

     vịl-ẹ   *kẹ* *on*   *gẹtraś*
     on-ILL if NEG.2SG marry.CNG
     'You will be called an idiot, Vladimir, if you don't marry a girl like that!' (Matveev 1958: 48)

In examples such as this one, however, there is often a clear temporal or sequential interpretation. In fact, getting married in the future is the topic of the humorous play from which this example originates. Cypanov (2005: 148) characterises the Komi future tense as being used when the event described event after the moment of speech, in a one-time process that cannot be divided into phases, and with modal connotations about the certainty of the outcome. This is exactly what we see here. Most of the play is in the present tense, but there are a few instances where the future tense is used. We will use another example from the same play to illustrate the Standard Komi use of the present and future in Section 3.2.

    The system we find in Udora is strikingly different, with verbs marked with *-as-* in the third person being the most common, and with *-ẹ-* marking occurring with individual verbs. We see this in (3) and (4).

(3) Иван чаревич водас и унмовсяс. И узьö куим сутки. Сы дінö воасны став зверыс, коді тэсі вöйтча.

ivan tsarević **vod-as** *i* **unmovś-as.** *i*
ivan tsarevich lie_down-FUT.3SG and fall_asleep-FUT.3SG and

uź-ę *kujim sutki. sị din-ę* **vo-asnị**
sleep-PRS.3SG three day.PL 3SG at-ILL come-FUT.3PL

stav źver-ịs, kodi tes-i vęjćća.
all animal-3SG who meet-PST.3SG before

'Ivan Tsarevich lies down and falls asleep. And he sleeps for three days. All the animals he has met come to him.' (Fokos-Fuchs 1916: 160)

(4) Öти идзас-бынмöд пыртас и мöдла петö, а купечъес повзясны и пышйöны öти дöрöм кежö.

ęt´i iźas-bịnm-ęd **pịrt-as** *i mędla*
one straw-bunch-PROL take_in-FUT.3SG and elsewhere

pet-ę, *a kupeć-jes* **povź-asnị** *i*
go_out-PRS.3SG but merchant-PL frighten-FUT.3PL and

pịšj-ęnị *ęt´i dęręm kež-ę.*
escape-PRS.3PL one shirt middle-ILL

'He takes in a stack of straw and goes out. But the merchants are frightened and run into a thread shirt.' (Fokos-Fuchs 1916: 172)

In these examples, the majority of verbs are marked with what in Standard Komi would be the future tense, but individual verbs such as 'sleep', 'go out' and 'escape' are in the present tense. It seems clear to us that, in examples such as these, the intended function cannot be tense marking. There is, however, an underlying systematicity, and that is what we aim to describe.

Our study is structured so that we first discuss the position of Udora within the Komi dialects. From this, we move into a more general description of how the use of the present and future tenses is described in contemporary Komi grammar. After this, we discuss the dataset used for our study. In our analysis, we examine different features and hypotheses that have been connected to the Udora third-person verb forms in earlier literature: object conjugation (Lytkin 1977a: 283) and transitivity (for the transitivity and object conjugation hypothesis, see also Serebrennikov 1956: 68), imperfect aspect especially with specific derivations (Serebrennikov 1963: 255) and derivation types in general (Sorvačeva 1952: 46; Ponarjadov 2004: 112). We complement this analysis with an external comparison to

Old Komi, where similar variation to that found in the Udora dialect has been described before (Lytkin 1977a: 282; Cypanov 2005: 30). This external comparison serves to root our results in their historical context within Komi dialectology. We also examine and discuss our novel observation that the morphemes under investigation are used in the Udora dialect as a way to mark event structure in narratives. In that section, we provide a large number of complete example texts that also illustrate the use of tense in the Udora dialect more broadly.

The examples in this study are presented in the contemporary Komi orthography, with morphemic interlinearisation in Finno-Ugric Transcription and glosses in English. For the sake of clarity, we use the abbreviations PRS and FUT in our interlinear glosses to mark the verb forms under investigation, although these are not unproblematic labels for glossing occurrences of these forms in the Udora dialect. The level of transcription is phonemic, and examples taken from different sources have been harmonised into a comparable representation. The relevant third-person verb forms are marked in bold. For the original versions of the transcriptions, we encourage the reader to consult the works cited. All translations have also been revised and edited by the authors, and are not identical to those found in the original sources. Whenever numeric results are discussed, the observations obtained from our corpus are presented as contingency tables. We have also published the individual example dataset as an online appendix to the current study (https://doi.org/10.33339/fuf.97371) so that our results can be compared with different materials and to facilitate further work.

2. Related work

There have been several individual studies connected to Udora verbal morphology, but the examples from Udora are usually used to illustrate a broader argument, especially in the context of historical morphology. Studies that exclusively investigate the Udora dialect have remained rare. The earliest modern description of the Udora dialect is by Sidorov (1930), who does not discuss the behaviour of the present and future markers among the particularities of the dialect. A few decades later, Sorvačeva (1952) provides a more detailed treatment, which also addresses Udora verbal morphology. In her view, reflexive verbs formed with ś and verbs of action and movement employ the -ę- conjugation (Sorvačeva 1952: 46). She argues that at Upper Vaška, aspectual differences are marked as in Russian

(Sorvačeva 1952: 44). This idea is later supported by Serebrennikov (1963: 255) in connection with past-tense allomorphy in Udora, but it is rejected by Cypanov (2005: 87), who do not see this analysis as being supported by contemporary data. This mainly illustrates that the notion of aspect has been regularly present in discussions of the Udora dialect. We suggest in our study that aspect is a relevant category behind the variation in Udora, but this is primarily in the sense of lexical aspect, which is a viewpoint that has not been previously suggested.

The most important description of the Udora dialect is arguably the monograph by Sorvačeva and Beznosikova (1990). In it, they describe a non-differentiated present–future tense and state that the morphemes are used interchangeably (Sorvačeva & Beznosikova 1990: 68). The variation is not further explained, but the authors refer to Lytkin's explanation about the different stem vowels being reflected in the two endings (Sorvačeva & Beznosikova 1990: 67). Here, the argument is that the -*s*- would be a separate element that was added due to homonymy with other persons in the past tense (Lytkin 1961). Synchronically, the monograph does not offer more details about the use of these tense marking morphemes, but it is still the most thorough description of this dialect.

Lytkin (1961: 53–54) compares the conjugation systems in the Udora dialect and Mari, arguing that the systems are similar, and agrees with the early analysis of Sorvačeva (1952: 46) in that action and movement verbs and specific derivations are connected. He argues, however, that this does not explain the whole picture and that verbs cannot be categorised into only these groups in the contemporary Udora dialect. Lytkin (1961: 54) states that defining the exact parameters for the variation is not possible. Somewhat later, Lytkin (1969: 97) also connects the Udora conjugations to historical stem vowels, eventually extending the comparison to other Uralic languages.

The connection to Old Komi was also recognised at an early stage: Lytkin (1977a: 63–64) proposes that in Old Komi, the present and future tense were not distinguished, and that the same phenomenon is in some manner visible in the Udora dialect. Lytkin (1977b: 280, 283) connects the Zyrian Komi past tense allomorphy to the present–future variation in Udora, and also discusses it in the context of Old Komi. The development Lytkin proposes is that object conjugation was in the process of developing in Komi but never became fully established.[2] This underdeveloped object conjugation was an-

---

2. In Lytkin's words: *не пустило глубоких корней*.

alysed in the rest of the Komi speaking area as a present–future opposition, but has left various traces, especially in the Udora dialect (Lytkin 1977a: 283). The suggested development is that the forms marked with -s- used to express transitivity, and that the function of future tense emerged from this. Later, the use of this element also spread to intransitive verbs (Lytkin 1977a: 282–283; Bartens 2000: 191). Similarly, Cypanov (2005: 141) analyses the undifferentiated present–future marking in Udora as an archaism, where the functions of the -ę- and -as- morphemes were not yet entirely separate, but he does not find the historical scenario proposed above entirely convincing. The idea that Proto-Permic would not have had distinct present and future tenses is further complicated by the fact that, with the exception of the Udora dialect, the Permic varieties have been described as having remarkably similar present–future tense systems, with an essentially identical set of two morphemes, as has also been noted by Ponarjadov (2004: 106).

The study by Ponarjadov (2004) is one of the few recent studies that focuses only on Udora materials. His approach is etymological, and his results support Lytkin's idea that the two morphemes are selected by individual verbs (Ponarjadov 2004: 108). He further connects this to the different reconstructed stem vowels, and concludes that it is not possible to find a temporal distinction between these morphemes in the Udora dialect. In Section 5.5, we do report a distinctive usage in narratives that, to our knowledge, has not been described before. Ponarjadov (2004: 110–111) also finds that derivational morphology impacts the choice of morphemes but states that the reasons for this are unclear. Our analysis in Section 5.4 shows that the process in Udora is connected to the properties of different derivations, depending essentially on the semantics of individual verbs. The analysis of Ponarjadov (2004), especially with regard to derivations, is an important predecessor to our current work. We do not, however, examine the reconstructed stem vowels in more detail, primarily because this would require a very extensive investigation of the latest etymological literature, which is beyond the scope of this study. We hope, however, that our word lists and results will also be useful for further research in this direction. The most current research on Uralic historical morphology also suggests that we cannot reconstruct the Proto-Uralic tense system very well, and there are various conflicting possibilities as to how it could be reconstructed (Aikio, forthcoming).

The connection between these forms and object conjugation is, according to Csúcs (2005: 260), the generally accepted opinion of the field, and he discusses it in connection with a similar analysis by Rédei (1989: 199).

The diversity of opinions in the field, the complexity of the issue and our evolving understanding of the earlier language stages suggest that more research is needed.

In this study, we do not attempt to propose alternative developments, but rather try to analyse the phenomenon primarily on a synchronic level using attested language forms as our sources. It seems to us that earlier work on this topic has often proceeded in a bottom-up fashion, where the current state of the art in historical linguistics has been used to reason and explain the variation we see today in these language varieties. In our view, however, we should emphasise the best possible contemporary description, which should then impact the possible reconstructions.

3. Background

3.1. Udora within the Komi dialects

Zyrian Komi is traditionally described as having ten dialects. The written standard is based on the dialect from the Syktyvkar region, which is spoken at the boundary of two major dialect groups. Zyrian dialects are commonly divided into the Northern and Southern groups, with the dialects of Udora, Lower Vyčegda, Vym and Ižma belonging to the Northern (also known as North-Western) group (Popova & Sažina 2014: 8). There is also a history of connecting the North-Western dialects to Old Komi (Lytkin 1977a: 282), with which Udora shares other isoglosses (Popova & Sažina 2014: 103). These varieties appear to belong to the same historical dialect group (Ljašev 1980: 12). Lytkin (1952: 121) has also analysed Udora, Vym and Old Komi as belonging to the same dialect group and as having minor differences from one another, in contrast to the clearer difference they have from the Sysola dialects. For a thorough overview of Komi dialects, see Popova and Sažina (2014) and the monographs on individual dialects. A thorough review is beyond the scope of this paper, but it is worth noting that Udora exhibits a variety of features that do not have parallels in any other Komi dialect; for further details, see Sidorov (1930) and Sorvačeva (1952). At the same time, Vym and Ižma have features not shared with Udora, such as dative object marking (Klumpp 2008: 189; Ljašev 1977: 120), which would indeed support the idea, presented in Figure 1a, that Ižma would historically descend mostly from Vym and not from Udora.
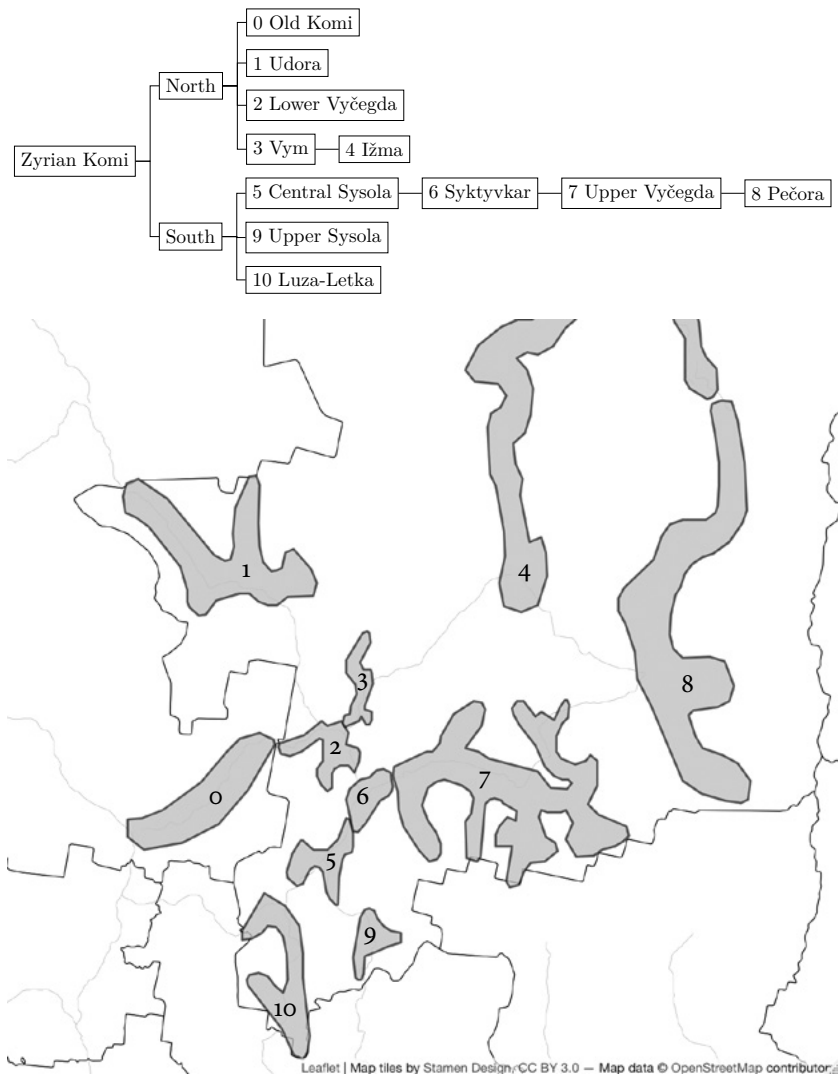
Figure 1: Tree (1a) and map (1b) of Zyrian Komi dialects

According to Sažina (2014: 82), there are six modern Zyrian Komi dialects that can be regarded as continuations of the earliest dispersal of Komi to its current speaking area. These are the Southern Zyrian dialects of Luza-Letka, Upper Sysola and Central Sysola, and the Northern Zyrian dialects of Vym, Lower Vyčegda and Udora. These are also the areas where Zyrian

settlements could already be found by the 15th century (Savel'eva 1997: 55, 61). These oldest dialects are marked in Figure 1a at the lowest level in the Zyrian dialect hierarchy with the extinct Old Komi. The other varieties, although relatively old, represent later areal expansions.

From the perspective of our study, the most central implication of the historical background of the Zyrian dialects is that Udora is among the oldest dialects, and none of the other dialects descend directly or solely from it, nor can we postulate any of the modern dialects to be its predecessor. Therefore, if a unique feature is described in the Udora dialect, even if it cannot be found in other modern Komi dialects, we cannot conclusively demonstrate that the development must be a Udora innovation. When we discuss a feature that occurs within the Permic languages in only one dialect, the historical relations of the dialects are central to what can be postulated.

From the point of view of population history, we can also take into account the suggestion of Žerebcov (1972) that Udora was populated from the south, which is the area where Old Komi was spoken (see Figure 1b). The Yarenga area, located to the south of Udora, was once Komi-speaking, and there were migrations to the Vaška area in the 1400s before the local language shift was complete and the areas in the south became Russian-speaking (Žerebcov 1972: 19). The earlier Komi presence in these regions, including the Pinega Basin, is also clear from toponymic evidence (Turkin 1984: 178). Vaška was populated first, and permanent settlements at Mezen appeared in the second half of the 16th century, but there were also migrations to Mezen from parts of Vym (Žerebcov 1972: 21–22). This supports the description of Sidorov (1930: 49), who states that the Udora subdialects differ such that Vaška is more similar to Lower Vyčegda and Mezen closer to Vym.

3.2. Third-person tense marking in Komi

Komi has an elaborated tense system with various functionally differentiated forms. The morphologically marked tenses are two past tenses, present and future. The rest of the tenses are constructed using auxiliaries and participles, and the second past tense also originates from a participle. The second past tense has additional meanings connected to evidentiality and indirectivity (Leinonen & Vilkuna 2000), or to the sense of non-involvement (Baker 1983: 76). For an analysis of the different functions, see also Cypanov and Leinonen (2009). In the glosses, the second past tense is

distinguished using the abbreviation PST2, whereas the first past tense is marked with PST. In all Komi dialects, the functional distinction between present and future appears in the inflectional paradigm only in the third-person forms. In the first and second person, the forms are identical. In the third person, the distinction is marked with the morpheme *-ẹ-* in the present tense and *-as-* in the future tense. Additionally, future meanings can be expressed by means of verbs used in a modal sense, such as *кутны* /kutni̱/ 'hold', *пондыны* /pondi̱ni̱/ 'begin' or *мӧдны* /mẹdni̱/ 'go (in a certain direction)' (Popova & Sažina 2014: 212). These two distinct grammatical strategies for expressing future meanings are customarily referred to in Komi grammaticography as the first and second future tenses, in analogy to the numbered past tenses.

The starting point for our investigation is that, as described in modern grammars, the present and future in third person should be distinguished in the written Zyrian Komi standard as two different tenses. We are not aware of any research that would argue that the main distinction is something else. The most detailed description of the uses of the Komi present and future tenses can be found in the grammar *Öнія коми кыв* (Fedjunëva et al. 2000). According to this grammar, the Komi present tense is used in the following situations (examples taken from Fedjunëva et al. 2000: 239–240) (translation by the authors, examples within text omitted):

1. To express the direct present, i.e. the actual, current moment when the described action is taking place at the exact time of speaking, without moving beyond it.

2. To indicate a reoccurring, abstract present tense, when the moment of speaking is understood widely and the action takes place more than once.

3. To express events that exist at all times and in all possible situations.

4. To indicate the relative, indirect present tense in the past, when the narrator is describing something happening, often in a relative clause introduced by main clause in a different tense.

5. In folklore, speech and literature, one encounters the historical present tense where the verb marks a distant event as if it would be taking place now.

6. To express a "transferred" tense, which is when the present tense is used in a context with another tense and takes on the meaning of that tense, as though the expression should perhaps have been in another tense.

The same grammar describes the use of the future tense as follows:

1. To express a specified future meaning.
2. To indicate a distant future meaning, which expresses an event that will take place at some unspecified time. […] This meaning can also be called clearly known future, which will take place in truly unknown time.
3. To express a gnomic, broad future tense meaning, which occurs in proverbs, in fixed phrases, […] [I]n proverbs, the present tense may also be used, […] the meaning does not change significantly. The gnomic meaning appears in future forms when the event discussed takes place always or happens from time to time.
4. As a modal future that expresses the speaker's opinion about the topic, how it is valued or evaluated. It was already mentioned that the future tense is used for a clearly known, certain action. […] Often with this form, the speaker displays their own belief, i.e. the possibility that the thing will happen. This is often used in situations where the speaker expresses their own loss, disbelief or hopelessness.
5. The "transferred" meaning of the future that is used in place of other tenses is very common. Future verbs are used in present and past tense contexts to bring the speaker closer to the events and to make the speech more vivid. […] The transferred meaning can surface in two contexts: a) with the present tense, and b) with the past tense. When used with the present tense, it expresses continuously happening events that are being discussed. This meaning is particularly common in everyday speech […]. When used together with the past tense, the future marking expresses things that have happened at some point; it often appears in legends, stories, literature and when recounting memories. Most commonly, the past tense expresses its own base meaning, marking phases in the events that take place in known times.

It is worth noting that there are numerous instances where the present tense and future tense are used in similar contexts to one another. Point 3 in both lists, always ongoing or gnomic events, is very similar for both tenses, and it is even noted that both tenses can be used here interchangeably. Fedjunëva et al. (2000: 237) point out that both the present and the future appear in past-tense narratives as stylistic devices. They also distinguish as one parameter between present and future tense the aspectual difference where the future tense marks events that will be completed (2000: 238).

The reader may recall that in Section 1, we provided a short example of a typical use of the Komi future tense (see Example 2 above). As can be gleaned from the descriptions in the current section, however, the use of the Komi present and future tenses is much more complicated. To help the reader fully appreciate this complexity, we present here another example from the same play, which was originally published in the journal *Войвыв Кодзув* in 1958. Example (5) is from the part of the text describing the typical and habitual behaviour of the protagonist Vladimir's mother. No example is perfect, but we believe this one shows many typical properties of Komi tense marking. In the beginning, where the action of hiding is described, the subordinate clause is in the present tense. This fits in with feature 4 in the list of uses of the present tense provided above. At the end of the example, the tense shifts back to the future, which can be compared to feature 5 in the listed uses of the future tense, as it is essentially connected to the stylistic use of the future in the narrative.

(5)   Локтасны сы дінö, а сийö дзебсьö. Öтчыд öти фотокорреспондент топöдöма вöлі, но спаситіс öш. Микулай нима. Кыдз тай Микулайыд сюръяснас потолокас лэбöдас снимайтчан аппаратсö!

lokt-asni̩	si̩	din-ẹ,	a	sijẹ	**ʒeb-ś-ẹ.**	ẹćći̩d	ẹt´i
come-FUT.3SG	3SG	at-ILL	but	3SG	hide-REFL-PRS.3SG	once	on

fotokorrespondent	*top-ẹd-ẹma*	*vẹl-i,*	*no*
photojournalist	press-CAUS-PST2.3SG	be-PST.3SG	but

śpasit-is	*ẹš.*	*mikulaj*	*ńim-a.*	*ki̩ʒ*	*taj*
release-PST.3SG	bull	mikulaj	name-ADJ	how	that

mikulaj-i̩d	*śur-jas-nas*	*potolok-as*	**leb-ẹd-as**
mikulaj-2SG	horn-PL-INS	ceiling-3SG.ILL	fly-CAUS-FUT.3SG

śńimajććan	*apparat-sẹ!*
photography	machine-3SG.ACC

'They come to her, but she hides. Once a journalist was approaching her, and she released a bull. It was called Mikulay. Oh how that Mikulay threw [the journalist's] camera up to the roof with his horns!' (Matveev 1958: 47)

The middle part of (5) employs the second and first past tenses. According to Leinonen and Vilkuna (2000: 501), the second past tense is used in Komi to express a break in the main line of narration, which may be what is happening here. We also encounter such uses in the Udora materials, where the second past tense is followed by a consistent use of a future-marked

tense. We will discuss this use further in Section 5.5. The reader can also consult Examples (21) and (25), where the second past tense is used in an arguably similar function at the beginning of the narrative to establish or distinguish a certain phase of the text.

Table 1: Komi third-person paradigm with the verb *мунны* /munnị/ 'go', dialectal variation included. The variants accepted in the Zyrian Komi written standard are marked in bold.

	Past II	Past I	Present	Future I
Singular	*mun-ẹm*	**mun-i**	**mun-ẹ**	**mun-as**
	mun-ẹma	*mun-is*		
Plural	*munẹ-maś*	**mun-i-nị**	**mun-ẹ-nị**	*mun-a-nị*
	munẹ-maaś	*mun-i-nịs*	*mun-ẹ-nịs*	**mun-as-nị**
	munẹ-ma-ẹś	**mun-is-nị**		*mun-as-nịs*
		mun-is-nịs		

When we look at the actual tense marking morphemes, we can see that there is extensive allomorphic variation in Komi dialects. The variants found in different dialects are presented in Table 1. Forms allowed within Standard Zyrian Komi are marked in bold. Our current study focuses on the variation between the present and future forms *-ẹ-* and *-as-* in both singular and plural. Especially the variation in the first past tense is so similar to the variation we see in Udora that one has to ask whether the allomorphs *Vs ~ V* are in variation that is similarly conditioned to what we find in the Udora present–future tenses. This question is beyond the scope of the current study, but it is connected to our research question and illustrates that there are related phenomena concerning various Komi dialects, not only Udora. We will also use the Standard Komi description of these tenses as a point of comparison when we analyse the phenomenon in Udora.

Although the present and future tense are contemporarily described in Komi, their status has not always been clear. In the earlier descriptions, there has been a great deal of variation as to whether the present and future tense are described separately, but contemporary grammars do so (Cypanov 2005: 123–125). Lytkin (1969: 96) also argued that the reason previous researchers, mainly Wiedemann (1884: 116) and Uotila (1938: 54), had analysed the Komi non-past tenses as an undifferentiated present–future tense was that the Komi future is frequently used in historical narratives.[3]

---

3. In Lytkin's terms, *elbeszélő jövő* in Hungarian and *будущее повествователь-ное* in Russian.

Lytkin stresses that these tenses are, in fact, differentiated in the Komi dialects besides Udora. However, Cypanov (2005: 146) mentions that the use of the present tense in contexts where the future would be expected is common in the Komi dialects. To our knowledge, there is no study devoted only to the historical narrative tense in Komi standard varieties or in dialects, but the phenomenon itself has been recognised. Essentially, the system is constructed so that the present tense is used as a *historical present* and the future tense as a *historical future*, which Serebrennikov (1960: 51, 83) also points out. Other grammatical descriptions of the Komi languages (i.e. Lytkin 1955: 213–214; 1962: 251) also mention the use of the present and future to express historical meanings.

4. Data

In our study, we use a Udora dialect corpus we have compiled from various earlier resources and our own fieldwork. The size of the resulting corpus is 52,081 tokens, which are relatively evenly distributed over a period of one hundred years. For illustration, the Udora texts published by Uotila (1989) result in 8,000 tokens, and our contemporary fieldwork from the 2010s is 17,975 tokens. Erik Vászolyi's published Udora texts total 13,558 tokens. All in all, there are 4,184 third-person verb forms that we have analysed, and they include 762 distinct verbs. Out of all the distinct verbs, 243 occur three times or more in the corpus. In most of the individual studies, we use three occurrences as a threshold for including a verb in the analysis. We determine that a verb belongs to one type or the other based on the most common third-person marker that occurs with the verb in the corpus. When we analysed the derivations, however, we also took rarer verbs into account. All verbs were initially extracted from the corpus using a morphological analyser for Komi (Rueter 2000) accessed using the Python package UralicNLP (Hämäläinen 2019), and then verified manually.

Several published text collections include samples from the Udora dialect. The earliest collected materials date back to the late 19th century and are published in Wichmann's 1916 work *Syrjänische Volksdichtung*, which contains three short texts from Udora. Another early source is Fokos-Fuchs' *Zürjén szövegek* from 1916, which contains a longer collection of narratives and other texts collected in 1913. *Syrjänische Texte III* (Uotila 1989) also includes a long section of Udora texts collected from a single speaker during the 1940s. There are no recordings for the earliest texts, but they are still

very important and sizeable records of this dialect. Even if all individual examples cannot be confirmed with audio recordings, it seems plausible that the systematic patterns they contain are accurate and correct. Of the more contemporary materials, we have used the text collection *Образцы коми-зырянской речи* by Žilina and Sorvačeva (1971). These texts are primarily based on recordings made between the 1940s and 1960s, and have also been used in other studies on the Udora dialect, such as Ponarjadov (2004).

In the late 1950s, Erkki Itkonen took a two-week-long trip to Syktyvkar, where he recorded several Komi speakers (Itkonen 1958: 70) including people from the villages along the river Vaška. Günter Johannes Stipa also travelled to Syktyvkar with the goal of making recordings (Stipa 1962: 65–66), and his consultants included at least one Udora speaker from Koslan. Further work around Koslan followed soon, when Muusa Vahros-Pertamo went on a fieldwork trip to that area accompanied by the young Komi writer Albert Vaneev. There, they recorded a large collection of narratives, conversations and songs (Vahros-Pertamo 1963), which to our knowledge have not been published. Later in the 1960s, Erik Vászolyi also worked in the area, and his texts have been published (Vászolyi-Vasse 1999). Vászolyi's work is particularly significant as it is one of the rare collections that includes recordings from the Upper Mezen river; most of the materials published from Udora have focused on Vaška and Lower Mezen.

As the audio collections of Itkonen, Stipa, Vahros-Pertamo and Vászolyi are archived in the Institute for the Languages of Finland, the majority of the audio recordings were available to us. These earlier materials are further supplemented with materials recorded by the authors in 2012 and 2013. These materials have been archived in the Language Bank of Finland and are available for research use (http://urn.fi/urn:nbn:fi:lb-2021111821). See also the online appendix for a complete listing of the examples used in this study.

There are also published sources that we have not used, especially folkloric materials that include segments in the Udora dialect and certain dictionaries that contain phrases in the dialect. Our main reason for leaving these sources out has been practical, as we feel that the current corpus is large enough for the current investigation. There are also methodological aspects, as we usually cannot know the process through which the folkloric texts were edited, whereas with linguistic transcriptions we should assume that exact linguistic representation has been the goal. At the same time, the dictionary examples may be too fragmentary to allow for fuller analysis. That said, dictionaries (i.e. Fokos-Fuchs 1959, Beznosikova et al.

2012) could be very useful additional sources and have also been used in previous works on this topic, for example, by Lytkin (1969: 99). Ideally, these sources could be used to verify or refine our results.

Additionally, as there are still untranscribed recordings in both the earlier and contemporary materials, transcribing more recordings would be one way to obtain additional information, especially about the areal and temporal nuances of different questions related to the Udora dialect. We aim, however, to provide currently available results using the existing materials, as materials can always be increased and refined.

5. Analysis

5.1. Distinct tense markers as two conjugation types in Udora

Earlier research indicates that most verbs in the Udora dialect systematically select one of the two forms under investigation in the third person (Ponarjadov 2004). Earlier descriptions have led to the conclusion that the use of these allomorphs in the Udora dialect is primarily a lexical question, representing essentially two conjugation types, and possibly originating from Proto-Uralic stem vowels (Lytkin 1969: 97). Our study attempts to provide a synchronic description, so we do not make broader comparisons between the Uralic languages here. According to common morphological models, a conjugation is defined as a class of verbs, all of which take the same set of inflectional allomorphs (cf., e.g. Dixon 2010: 334). In the case of the Udora dialect, these conjugations differ only in the third person. Bakró-Nagy, Laakso and Skribnik (2020: 33) emphasise that the term *conjugation* is used in very different ways in different linguistic traditions. What we mean here by conjugation is strongly connected to the Uralic tradition, where it is used in Mari and Udmurt research to mark the two main verbal inflection types.

The phenomenon discussed here can be illustrated with (6):

(6) Иван Царевич мӧдасылас сувтас, а кар сулалӧ.

*ivan   tsarević   mẹd-asịl-as   **suvt-as**,*
Ivan   Tsarevich   other-morning-3SG.INE   get_up-FUT.3SG

*a   kar   **sulal-ẹ**.*
but   city   stand-PRS.3SG

'Ivan Tsarevich gets up the next morning, and the city is [still] standing.' (Fokos-Fuchs 1916: 96)

In this example, the event described takes place the next day, and Ivan Tsarevich getting up and the city still standing are presented as consequently occurring events. Of course, it can be surmised that some more complex temporal distinction is marked here. In Section 5.3, we discuss this further. The next example comes from the monograph describing the Udora dialect:

(7)   Мунӧныс, мунӧныс, вартасны, вартасны дай тэсӧ и лягей.

    mun-ẹnis,   ***mun-ẹnis***,   ***vart-asni***,   ***vart-asni***   *daj*
    go-PRS.3PL   go-PRS.3PL   jump-FUT.3PL   jump-FUT.3PL   and

    tes-ẹ   *i*   *ľagej.*
    meet-PRS.3SG   also   frog

    'They walk and walk, gallop and gallop onward, and meet a frog.'
    (Sorvačeva & Beznosikova 1990: 68)

In the middle part of (7), the third-person marking switches from -*ẹ*- to -*as*-, and then back to -*ẹ*- again at the end. The durations of the activities described here may be different, as the general act of going can be understood as a longer process than that of jumping, but then the verb *тэсны* /tesni̯/ 'meet'[4] that follows is again marked with -*ẹ*-. In our corpus, all instances of this verb 'meet' occur with -*ẹ*-, as is overwhelmingly the case for the verb 'go', while 'jump' always occurs with -*as*-. This is not entirely without exception, but the system appears to be very regular. The same explanation also holds consistently for the verbs in (6). Table 2 lists the most frequently occurring verbs that appear with -*ẹ*-, while Table 3 lists the most frequent verbs that take -*as*-. For a more thorough listing, please refer to the online appendix.

    When looking at Tables 2 and 3 (on pp. 158–159), we can see that none of the very frequent verbs are particularly ambiguous when it comes to which form they use. It has been suggested in earlier literature that the verb *шуны* /šuni̯/ 'say' does not show a clear preference for one form over the other (Ponarjadov 2004: 108), but our data shows that with enough examples, a preference does emerge. All of the verbs in Table 2 are intransitive in Komi. We look into this phenomenon more closely in Section 5.2.

    Verbs that are marked in the third person with -*as*- are significantly more numerous than verbs marked with -*ẹ*-. This is noteworthy, as this

---

4. This is an intransitive verb in Udora dialect, so the frog is the subject, and the other participant, if expressed, would be marked with the dative case.

pattern already deviates very strongly from Standard Komi, where, as expected, the absolute number of present-marked verbs is higher than that of future-marked ones in a given corpus. The group of verbs that consistently occurs with -ę- is also significantly smaller than that of -as- verbs. Based on the data presented here, we can conclude that the analysis of the Udora system as lexically selected allomorphy holds relatively well. The exceptions are important, and will be analysed in detail, but the overall picture remains that the majority of occurrences of a given verb use either -ę- or -as- to mark the non-past tense. Next, we will try to analyse some of the properties of these conjugation groups.

5.2. Correlation with transitivity

As was discussed above, transitivity has traditionally been connected to the development of third-person verb marking in Komi. In the Old Komi materials presented by Lytkin, all of the verbs that occur with the contemporary present-tense marker seem to be intransitive, whereas those with the future marker include both transitives and intransitives (Lytkin 1952: 111–112; Bartens 2000: 118). According to Lytkin's analysis, Old Komi had undergone a development towards object conjugation, which subsequently ceased and developed into present and future markers (Lytkin 1977b: 63–64). Lytkin also states that this development is somehow still visible in the Udora dialect (see Section 5.6 for further discussion).

Given this background, it is crucial to evaluate how the two verb endings under investigation are connected to transitivity. To do so, we have classified the Udora Komi verbs according to their transitivity. We used two categories for the classification: transitive and intransitive. These were primarily defined by the verbs' ability to take a direct object, based on corpus data, and by the presence of valency-increasing or valency-decreasing derivations in specific verbs. We include the classification in the online appendix to this paper. Next, we describe the results of this investigation. As previously mentioned in the description of our dataset, we used three occurrences as a threshold for which verbs to include. As shown in Section 5.1, relatively few verbs absolutely always occur in one of the types, but with enough examples, a clear pattern emerges. Table 4 shows the relation between the conjugation markers and the proportion of transitive and intransitive verbs that use them.

Table 2: Most frequent verbs occurring with the conjugation *-ę-*

Verb	Transitivity	Translation	*-as-*	*-ę-*	Total
lonị	IV	be; become	14	160	174
munnị	IV	go	16	107	123
kovnị	IV	have to	1	112	113
loknị	IV	come	11	59	70
petnị	IV	exit	5	48	53
ovlịvlịnị	IV	happen (freq.)	0	52	52
lebnị	IV	fly, glide, rush	1	39	40
ovnị	IV	live	1	36	37
śurnị	IV	occur	1	32	33
pišjịnị	IV	flee	5	27	32
kolʹnị	IV	remain	12	15	27
sulavnị	IV	stand	0	26	26
kịvnị	IV	hear	7	18	25
vetlịnị	IV	go around	1	22	23
pukavnị	IV	sit	1	21	22
uśnị	IV	fall	3	17	20
poźnị	IV	can	2	16	18
uźnị	IV	sleep	5	11	16
kujlịnị	IV	be lying down	1	13	14
kuvnị	IV	die	1	11	12
tesnị	IV	meet	0	12	12
vetlẹdlịnị	IV	go around, walk around	1	9	10
vivnị	IV	be (used in a compound verb with *ovnị* 'live')	3	7	10
petavnị	IV	exit (freq.)	1	9	10
tịdavnị	IV	appear	0	10	10
bẹrdnị	IV	cry	0	9	9
kojnị	IV	mate (of birds)	0	9	9
vorsnị	IV	play	0	8	8
gorzịnị	IV	scream, yell	0	8	8
kịvlịvlịnị	IV	listen	0	8	8
tunị	IV	flood	1	7	8
udʒavnị	IV	work	1	7	8
dumajtćịnị	IV	ponder	1	6	7
śorńitnị	IV	speak	1	6	7
šuśisśịnị	IV	be called	0	7	7

Table 3: Most frequent verbs occurring with the conjugation *-as-*

Verb	Transitivity	Translation	-as-	-ę-	Total
šuni̯	TV	say	211	75	286
pondi̯ni̯	IV	start	152	0	152
mędni̯	IV	go (for)	103	2	105
voni̯	IV	come, arrive	96	3	99
bośtni̯	TV	take	85	1	86
pukti̯ni̯	TV	put	62	5	67
karni̯	TV	do	43	8	51
pukśi̯ni̯	IV	sit (down)	42	6	48
pi̯rni̯	IV	enter	44	2	46
ki̯jni̯	TV	hunt	28	11	39
suvtni̯	IV	stand up	36	0	36
teći̯ni̯	TV	gather	30	1	31
vęći̯ni̯	TV	make, do	25	5	30
leźni̯	TV	let	28	1	29
vajni̯	TV	bring or take (by carrying)	27	1	28
viźędni̯	TV	look	22	3	25
viśtavni̯	TV	tell, express oneself; give a speech	14	11	25
śetni̯	TV	give	20	5	25
aźźi̯ni̯	TV	see	24	0	24
kutni̯	TV	hold	22	2	24
dumajtni̯	TV	think	19	4	23
ki̯ski̯ni̯	TV	pull	17	6	23
korni̯	TV	ask	20	2	22
nuni̯	TV	bring or take (on foot)	17	4	21
povźi̯ni̯	IV	get frightened	19	0	19
śojni̯	TV	eat	12	7	19
juavni̯	TV	ask	15	4	19
vermi̯ni̯	TV	be able to	18	0	18
lepti̯ni̯	TV	rise	12	4	16
śijavni̯	TV	set a trap	16	0	16
viźni̯	TV	keep	12	3	15
lijni̯	TV	shoot	14	1	15
juni̯	TV	drink	14	1	15
domni̯	TV	bridle, rein	14	0	14
kani̯	IV	climb up	13	1	14

Table 4: Transitivity and conjugations

Conjugation	Intransitive	Transitive
-as-	56	114
-ę-	63	6

In our data, there are very few transitive verbs belonging to the *-ę-* conjugation, and none of them are frequent verbs, with all six examples belonging to the threshold category that has only three examples. Meanwhile, the *-ę-* type contains both transitive and intransitive verbs in fairly equal proportion. We can confirm with Pearson's Chi-squared test that this distribution is statistically highly significant (p < 0.001). Essentially, we are observing variation within the category of intransitive verbs, or, to put it more precisely, we can see that some intransitive verbs are marked distinctly from all other verbs. We can also state as one of our results that there are no frequent transitive verbs in the Udora dialect that would prefer the *-ę-* conjugation.

One of the exceptions in the data is the verb *повзьӧдлыны* /povźędli̭ni̭/ 'frighten sb. repeatedly', which occurs three times and is always marked with the conjugation *-ę-*. It is worth noting that this verb is also a frequentative. In the data, it occurs in a single text in which all verbs are similarly marked (Uotila 1989: text 206). In our classification, the verb *бӧрдӧдны* /bęrdędni̭/ 'cry a lament, lament sb./sth.' is another transitive verb that appears to fall into this group, but it also occurs only three times, two of which use the *-ę-* conjugation. These examples show that when we start to examine rarer verbs, it also becomes more difficult to establish which pattern they belong to. It is also noteworthy that both of these verbs are causatives derived with *-ęd-*.

Similarly, the reflexive derivation *корсисьны* /korśiśni̭/ 'search for a long time' is classified here as transitive, as we do have an example where it has a direct object, although most of the time the object is not present, and such a derivation could be expected to be intransitive. These examples illustrate how our transitivity classification scheme could be further elaborated. With this in mind, a fruitful direction for future research would be to tag all objects separately in the corpus. For our current results and investigation, however, the scheme used now appears to be sufficient. Another thing these individual exceptions illustrate is that when there are only a handful of examples, all occurrences of a verb are often present in

the same narrative, which may further influence the conjugation patterns, as we will discuss in Section 5.5.

Our analysis of the correlation between the conjugations and transitivity confirms the previous idea that the choice of morpheme is connected to transitivity, but with subtle differences, since transitivity alone does not explain the system we encounter in Udora. Instead, transitivity represents one clear boundary: the -ę- conjugation is not used with transitive verbs. This means that the factor determining which conjugation is selected must occur within the category of intransitive verbs. To provide a more comprehensive explanation of the factors at play here, we will next analyse questions of aspect and derivation.

5.3. Aspect and temporal continuity

As we can conclude from the previous analysis that all frequent transitive verbs belong to the -as- conjugation type, and that intransitive verbs can belong to either, we end up with two groups of intransitive verbs. We also find that the intransitive verbs in the Udora corpus are evenly split between the two patterns. The next question is which factors could be behind this alignment. As discussed in Section 2 on earlier research, it has been suggested that action verbs and verbs of movement would use the -ę- conjugation. Lytkin (1961: 53–54) was not convinced that these were the actual factors behind the situation in contemporary Udora. Our suggestion is that the verbs that use the -ę- conjugation appear to be temporally continuous. This contrasts with intransitive verbs marked with -as-, which can be analysed as being temporally bound. This definition is not perfect, but it does capture something in the difference between two groups of verbs. The lexical aspect of individual verbs would therefore appear to be one of the defining parameters. This distinction has previously been useful in categorising the use of other grammatical structures in Komi, for example in the study of augmentative verb forms by Todesk (2015), where the telicity of the event was shown to be central in determining the semantic reading of the comparative clitic -ǯik when used with verbs. In the case of Udora, the situation is not so straightforward, as the categories presented in Section 5.1 do not fall as cleanly into aspectual groups. Indeed, 'die' and 'sleep' are the exact verbs Dahl (1985: 26) uses as examples of lexemes that have different aspectual potential: in our data, these are both firmly -ę- conjugation verbs.

It is still worth looking more closely into whether lexical aspect could be a suitable determining factor. Concerning temporal continuity, verbs such as 'climb', in the sense of 'reach the top of the mountain', have a necessary endpoint, whereas verbs such as 'move' have no inherent end (cf. Filip 2012: 727). Similarly, 'stand' is a continuous process, but 'stand up' is temporally bounded. An opposition of this type could also be postulated between the verbs *воны* /voni̮/ 'come, arrive' and *локны* /lokni̮/ 'come'. The concept of telicity in itself may not be sufficient to describe the Udora phenomenon, as classifying individual verbs without context into telic and atelic may not capture their different uses (Comrie 1976: 45). Dahl (1985: 26–27) has also argued that the 'inherent aspectual meaning' of a verb can be very difficult to define based on different possible uses. In a similar vein, Croft (2012) emphasises that predicates have the potential to be conceptualised as different aspectual types. The reason we have not yet started to distinguish these types more carefully at the contextual level is the lack of variation we see in the Udora data: as we showed in Section 5.1, each verb is relatively strongly drawn to one conjugation. We will go through some key exceptions separately below. If we can show that lexical aspect is a decisive factor behind the choice of verb form in Udora, this would also help to explain why some verbs, namely those which are more temporally bound, behave similarly to transitive verbs. Indeed, Hopper and Thompson (1980) have associated transitivity with both higher telicity and punctuality cross-linguistically. As discussed in Section 3.2, the morphologically marked future tense we find in Standard Komi has also been described as having aspectual connotations referring to single events that have endpoints. When we look at the verb groupings in Tables 2 and 3, this explanation seems to hold some weight. Table 2 also shows several instances of derivations.

The majority of verbs that use the -*ę*- conjugation are verbs that describe an ongoing process without an obvious endpoint, such as 'stand', 'be', 'sit', and 'sleep'. That said, verbs such as *тэсны* /tesni̮/ 'meet (intr.)', *кувны* /kuvni̮/ 'die' and *петны* /petni̮/ 'exit' are bit at odds with this description and may call for a different analysis. There are also some relatively rare process verbs that use the -*ę*- conjugation, such as *быдмыны* /bi̮dmi̮ni̮/ 'grow (intr.)', *кульмыны* /kuľmi̮ni̮/ 'spawn', *сывны* /si̮vni̮/ 'melt', *чордны* /ćordni̮/ 'harden' and *туны* /tuni̮/ 'rise (of water)'. The verbs *пӧдны* /pędni̮/ 'drown; suffocate', *вӧйны* /vejni̮/ 'sink, go underwater' and *усьны* /uśni̮/ 'fall' can also be seen as belonging to this group. To explain these cases, some other parameter than telicity needs to be considered. One such

parameter could be the lack of volition, as the examples above include processes that do not have a conscious subject. This is also clear in the case of the -ę- conjugation verb *пеждыны* /pežḓịnị/ 'be released (of a trigger)'.

We can also suggest that among the rarer lexical items, verbs describing speech and vocalisations appear to occur commonly in the -ę- conjugation. This is true for both human and animal vocalisations, with verbs such as *серавны* /śeravnị/ 'laugh' and *горзыны* /gorzịnị/ 'scream, yell', but also *баксыны* /baksịnị/ 'moo; bleat (of cows or sheep)', *кӧкны* /kęknị/ 'cuckoo' and *шкоргыны* /škorgịnị/ 'roar' belonging to this group. Example (8) illustrates this use.

(8) Кырсизь кӧ тювкӧтчӧ — зэра лоӧ.

    kịrśiź   *kę*   **t'úvkęćć-ę**   – *zera* **lo-ę**.
    black_woodpecker if call-PRS.3SG – rainy be-PRS.3SG
    'If a black woodpecker calls, it is going to rain.' (Žilina & Sorvačeva 1971: 268)

The intransitive verbs marked with -as- form a few logical groups. One of these is momentaneous verbs, which occur in this form almost without exception. We will discuss these in more detail in Section 5.4. We also find a number of verbs that denote motion taking place along some defined dimension, such as *кайны* /kanị/ or /kajnị/ 'climb' (northern and southern dialectal variants both occur), *кывтны* /kịvtnị/ 'slide down' and *лэччыны* /leććịnị/ 'go down'. Verbs of posture change also fall firmly into the -as- conjugation.

Determining how productive the system truly is requires further research, especially in the form of elicitation. There are numerous examples where one could think that, especially when an -as- verb shifts to the -ę- conjugation, there is some kind of an extended temporal frame of reference. For instance, in (9), this is achieved using the frequentative derivation, which expresses that the otters regularly or repeatedly climb in a certain place.

(9) Вурд кыясны вӧралиссес берӧгӧ, вурд кытчы калывлӧ, лыаэс вылӧ.

    vurd **kịj-asnị**   *vęraliś-śes* *bereg-ę,* *vurd* *kịććị*
    otter hunt-PRS.3PL   hunter-PL beach-ILL otter where

    ka-lịvl-ę,   *lịa-es*   *vịl-ę.*
    climb-FREQ-PRS.3SG sand-PL top-ILL
    'The hunters catch otters at the beach, where they tend to climb, on top of the sands.' (Uotila 1989: 362)

The verb кыйны /kijn̦i̦/ 'hunt' itself displays wide variation between conjugations, although it is predominantly an *-as-* conjugation verb. The variation we see here is not easy to explain, but we believe it is connected to the fact that verbs of hunting and fishing are used to mean both 'capture an animal using a certain method' and '(regularly) practice some specific capturing method'. Usually, however, the verb for practicing the action would be marked with a reflexive derivation: indeed, the verb кыйсьыны /kij́și̦n̦i̦/ 'go hunting', for instance, belongs exclusively to the *-ɛ-* conjugation.

At times, we can also see the difference at the contextual level; Example (10) shows this well.

(10)   Буткыльтчылöны, буткыльтчылöны и некудз оз вермыны буткыльтны, и тшöж пондас буткыльтны и буткыльтчас.

butki̦l̦´-ćć-i̦l-ɛn̦i̦,   **butki̦l̦-ćć-i̦l-ɛn̦i̦**   i   ń e-kuź
roll-REFL-FREQ-PRS.3SG   roll-REFL-FREQ-PRS.3SG   and   not-how

oz   vermi̦-n̦i̦   butki̦l̦ț-n̦i̦,   a   čɛž   **pond-as**
NEG.PRS.3SG   be_able-INF   roll-CAUS-INF   but   duck   start-FUT.3SG

butki̦l̦´-t-n̦i̦   i   **butki̦l̦´-ćć-as**.
push-CAUS-INF   and   roll-REFL-FUT.3SG

'They push and push, and cannot push (it) in any way, and then the duck starts to push, and it rolls (over).' (Fokos-Fuchs 1916: 157)

In this story, several characters are trying to get a stone to roll. They try many times to roll it, but to no avail (the action is marked with an underived transitive verb form in the infinitive). Then, the duck tries to roll the stone and manages to roll it over. The final attempt is temporally bounded, as it has a clear endpoint and there is a change in state. This is closely connected to the derivational operations we will analyse in the next section.

Another peculiar type of exception we can distinguish is where verbs marked with the *-ɛ-* conjugation express some sort of property or characteristic of their grammatical subject. This is connected to our earlier observation that the lack of volition appears to have some effect. If we look at (11), the transitive verb 'shoot' is used here not to describe the occurrence of a single shooting event, but rather a specific property of the gun, i.e. with the meaning 'the gun shoots well'.

(11)   Писсяльыс вöлöма бур, лыйö бура.

piśśal̦´-i̦s   vɛl̦-ɛma   bur,   **li̦j-ɛ**   bura.
gun-3SG   be-PST2.3SG   good   shoot-PRS.3SG   well
'The gun was good, it shoots well.' (Vászolyi-Vasse 1999: 502)

In this case, the verb is used intransitively, and there is also a certain semantic ambiguity. However, describing the property of a gun is also temporally something very different from describing the action of shooting. Here we can present one more example where the temporal continuity is modified, as above, by the properties of the subject and the overall context, given the ambiguity of the verb. In (12), the verb 'hold' is used in the meaning of 'contain, be able to hold', with an inanimate subject 'trap'.[5]

(12)   Капканыс кутö, овлывлö, комын тув и пуд.

    kapkan-i̯s   **kut-e̦**,   **ov-li̯v-l-e̦**,   *komi̯n*
    trap-3SG   hold-PRS.3SG   be-FREQ-FREQ-PRS.3SG   thirty

    tuv   *i*   *pud.*
    pound   and   pood

    'The trap can hold, ordinarily, thirty pounds and a pood.' (Uotila 1989: 354)

These kinds of instances suggest that the conjugations are used to express certain grammatical distinctions. It is important to note that all examples presented here show variation wherein a verb ordinarily occurring with *-as-* is used with *-e̦-*. We saw in Section 5.1 that exceptions in the opposite direction also occur. In our materials, however, these exceptions are of a different type: they are related to narrative structure, which overrides the conjugation preferences set by the lexical aspect of individual verbs. We will discuss this in further detail in Section 5.5.

5.4. Derivations and conjugation classes

Komi has a complex system of verbal derivation. The most productive derivational categories are frequentative, reflexive, causative and momentaneous. These are able to combine with one another, which further enhances the complexity, and when it comes to frequentatives at least two types can be distinguished: *-li̯-* and *-al-*. We illustrate the derivations here using the verbs *гижны* /gižni̯/ (in Udora also /gežni̯/) 'write' and *мунны* /munni̯/ 'go'.

    We can see from Table 5 that not all intransitive verbs can be causativised. The meaning of the reflexive derivation also depends a great deal on

---

5.   It has been pointed out to us that this sentence is, in fact, a bit odd, as thirty pounds is roughly equivalent to a pood. Perhaps the original intention of the speaker was to say that the trap 'holds thirty pounds, or, in other words, a pood'.

the semantics of the stem verb: with intransitive verbs, the resulting verb can have a connotation of unintentional action; with verbs of posture, the meaning is usually a posture change; and when used with transitive verbs, the reflexive derivation renders them intransitive. Within the derivational system, the frequentative -*li̯*- is connected to the combining derivations -*vli̯*- and -*li̯vli̯*-, as these all are used to mark action that is regular, repetitive or continuous (Fedjunëva et al. 2000: 297). The nuances of the system are more complicated, and Fedjunëva et al. (2000: 299) emphasise that there is also an allomorph -*li̯*- used to express an action that occurs only once. At the moment, we have not differentiated these meanings beyond stating that they are two frequentative categories. In combined derivations, we use the last morpheme to determine the category. In connection with the Udora conjugations and derivations, Serebrennikov (1963: 255) and Sorvačeva (1952: 46) have suggested that reflexive derivations would always belong to the -*e̯*- conjugation.

Table 5: Komi verbal derivation system illustrated using a transitive and an intransitive verb.

Derivation	Form	Meaning	Form	Meaning
underived	*gižni̯*	write (trans.)	*munni̯*	go
reflexive	*gižśi̯ni̯*	sign up (intrans.)	*munśi̯ni̯*	go (unintentionally)
causative	*gižedni̯*	ask someone to write (trans.)	–	–
frequentative	*gižavni̯*	write often (trans.)	*munavni̯*	go (often)
frequentative	*gižli̯ni̯*	write sometimes (trans.)	*munli̯ni̯*	go (sometimes)

Here, we have decided to test the correlation between the derivations first by verb type and then with regard to all verbs occurring in the corpus, without using the threshold of three occurrences. Table 6 shows how the derivations are distributed across the verbs that occur more than three times.

Table 6: Derivations and verb types

Derivation	-as-	-ẹ-
causative	26	2
frequentative -lị-	8	5
frequentative -al-	17	12
momentaneous	2	0
reflexive	22	10
underived	95	40

Table 7: Derivations and verb occurrences

Derivation	-as-	-ẹ-
causative	285	31
frequentative -lị-	96	108
frequentative -al-	192	159
momentaneous	32	2
reflexive	265	114
underived	1785	1115

In the Udora data, there are a number of situations where derivations change the conjugation type of the verb. We can describe a system in which the derivations are closely and often predictably linked to conjugation types. This supports the analysis according to which the conjugations would mark lexical aspect, which would then be independent for each derivation.
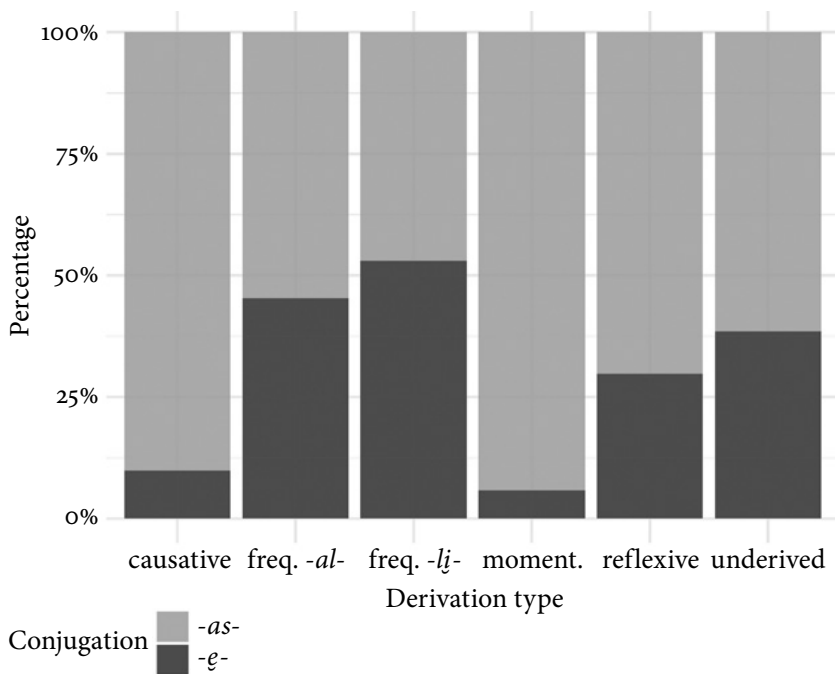

Figure 2: Relationship between derivations and conjugations used, based on data in Table 7

For example, there is an intransitive derivation of the transitive verb *кутны* /kutn̯i/ 'hold; start', which is *кутасьны* /kutaśn̯i/ 'try to catch' (used, for example, in children's games of tag or to describe when a child is learning to grasp items). The texts where this verb is used clearly describe the process of attempting to catch someone or something. Similarly, the derivation *кутчысьны* /kući̯śn̯i/ 'catch from somewhere', also intransitive, shifts the verb into the *-as-* conjugation type. The difference here appears to be connected to the changed telicity of the derived verb. At this point, we also wish to draw attention to (12), where the transitive verb *кутны* /kutn̯i/ 'hold; start' was used with the *-e̯-* conjugation under very specific circumstances.

Furthermore, we see that frequentative derivations do not change intransitives that occur with *-e̯-* into the *-as-* conjugation. The derivations *овны* /ovn̯i/ 'live' → *овлывлыны* /ovli̯vl̯in̯i/ 'happen'; *пышйыны* /pi̯šji̯n̯i/ 'escape' → *пышъявны* /pi̯šjavn̯i/ 'be on the run'; *петны* /petn̯i/ 'exit' → *петавны* /petavn̯i/ 'be going out' all remain in the same conjugation as the verbs from which they are derived. This is also true of verbs such as *ветлыны* /vetl̯in̯i/ 'go' → *ветлӧдлыны* /vetle̯dl̯in̯i/ 'go', and → *ветлывлыны* /vetli̯vl̯in̯i/ 'go around'. We argue that the main reason for this is that when combined with verbs that are already temporally continuous, frequentative derivations serve only to expand the duration of the event, they do not change the verb's lexical aspect.

Predictably, the opposite shift from one conjugation type to the other does occur when a new frequentative verb is derived, and the *-as-* conjugation becomes more common. This kind of use is illustrated in (13).

(13) Урыс чувалігӧныс кывт кузяыс ёна вартлӧ.

    ur-i̯s   *ćuval-ige̯n-i̯s*   *ki̯vt*   *kuźa-i̯s*
    squirrel-3sg be_in_heat-cvb.sim-3sg trapline along-3sg

    jona   **vart-l-e̯**.
    much   jump-freq-prs.3sg

    'When the squirrel is in heat, it often runs along the trapline' (Uotila 1989: 356)

Again, we see a change in conjugation type when the action becomes temporally more continuous. Similarly, the verb *дугдыны* /dugd̯in̯i/ 'stop' primarily occurs in the *-as-* conjugation. However, when it occurs in a frequentative derivation, we occasionally see a similar change, as in (14):

(14)   И тятейыс бöрдöмись дугдывлö.

    *i   t́at́ej-ı̣s   bẹrdẹm-iś   **dugd-ı̣vl-ẹ**.*
    and   baby-3SG   crying-ELA   stop-FREQ-PRS.3SG
    'And the baby stops crying.' (Uotila 1989: 396)

This is not always a clear-cut process, and we can see that there is wide variation in the conjugations used with verbs formed with frequentative derivations. This variation is also evident from Figure 2.

In some instances, reflexive derivations can change a verb from one type to the other in either direction. When a reflexive derivation is used to detransitivise the verb, the pattern also tends to change. This can be expected, as the derivation changes the transitive verb into a temporally longer-lasting intransitive verb. All of the following transitive verbs switch to using the -*ẹ*- conjugation in intransitive derivations: *пуны* /punı̣/ 'boil' → *пусивны* /puśivnı̣/ 'be boiling', *кыйны* /kıjnı̣/ 'fish or hunt (for something)' → *кыйсьыны* /kı̣jśı̣nı̣/ 'practice fishing or hunting', *шуны* /šunı̣/ 'say' → *шусьыны* /šuśı̣nı̣/ 'be called'. We can illustrate this with (15) and (16). In (15), we see the normal use of the transitive verb 'hunt'.

(15)   Кöч кыясны капканöн и лэчöн.

    *kẹć   **kı̣j-asnı̣**   kapkan-ẹn   i   leć-ẹn.*
    hare   hunt-FUT.3PL trap-INS   and snare-INS
    'The rabbit is hunted with traps and snares.' (Uotila 1989: 356)

When a corresponding intransitive verb is derived from this stem, as we see in (16), the conjugation type changes to -*ẹ*-.

(16)   Лэчöн кыйсьöны; сияласны лэчсö кöч кывт вылас.

    *leć-ẹn   **kı̣j-ś-ẹn ı̣**;   **śijal-asnı̣**   leć-sẹ   kẹć*
    snare-INS hunt-REFL-PRS.3PL set-FUT.3SG snare-3SG.ACC hare

    kı̣vt   vı̣l-as.
    track   on-3SG.ILL
    'This is how one hunts with a snare; the snare is placed on the rabbit's track.' (Uotila 1989: 356)

However, we often see the opposite change when the same derivation is applied to an intransitive verb. The reason for this is connected to various distinct functions of Komi reflexive derivations, as discussed above. The change from -*ẹ*- to -*as*- takes place with derived verbs such as *пукавны*

/pukavni̱/ 'sit' → *пуксьыны* /pukśi̱ni̱/ 'sit down'; on the other hand, there is also a causative derivation of the same verb, → *пуктыны* /pukti̱ni̱/ 'put', which occurs with the *-as-* conjugation, as would be expected of a transitive verb.

A similar conjugation shift with reflexive derivations can also take place in the other direction. In (17), we see a typical use of the verb *пукавны* /pukavni̱/ 'sit'.

(17) А Баба-Ягаыс пукалӧ море дорын […]

    a   *baba*   *jaga-i̱s*   **puk-al- e̱**   *more*   *dor-i̱n*
    and   Baba   Yaga-3SG sit-FREQ-PRS.3SG sea   at-INE
    'And Baba Yaga is sitting by the sea […]' (Fokos-Fuchs 1916: 160)

The outcome of the derivation with an intransitive verb is different, and the resulting meaning is 'sit down', a temporally bounded action.

(18) А юр вылас тшӧж пуксяс.

    a   *jur*   *vi̱l-as*   *če̱ž*   **puk-ś-as**.
    and   head on-3SG.ILL duck sit-REFL-FUT.3SG
    'And a duck sits down on the top of his head.' (Fokos-Fuchs 1916: 159)

In this light, it seems to us that the idea proposed earlier, that reflexive derivations would always occur with the *-e̱-* conjugation, is only part of the picture, and that the different outcomes of this derivation need to be examined based on the lexical aspect of the resulting verb and the multiple functions of reflexive derivations in Komi.

As noted by Ponarjadov (2004), momentaneous verbs regularly occur with the *-as-* conjugation. If we look simply at the percentages for which different derivations fall into the two conjugation types, as seen in Figure 2, we can observe that causatives and momentaneous verbs are the ones that most clearly avoid the *-e̱-* conjugation. Other conjugation types occur with other derivations in similar proportions, even though in the corpus we see some processes that are related to both frequentatives and reflexives as well.

However, there appear to be subtle differences between the behaviour of momentaneous and causative derivations. Example (19) shows a momentaneous verb derived from the verb 'pour', which is itself a transitive *-as-* verb.

(19) Мужик бӧра кисьтыштас и аслыс этша кольӧ.

*mužik   bęra   **kiś-t-i̯št-as**   i   aslį̣s*
man   again   pour-CAUS-MOM-FUT.3SG   and   3SG.REFL.DAT

*eča   **kol´-ę**.*
bit   remain-PRS.3SG

'The man pours again and leaves just a bit (of drink) for himself.'
(Fokos-Fuchs 1916: 167)

When -ę- conjugation verbs are derived further with this same derivation, their conjugation type does not change, as we can see in (20).

(20) Ӧдва кодзувъясянь кельыд югыдтор усьыштӧ.

*ędva   kožuv-jas-śań   kel´i̯d   jugid-tor   **uś-i̯št-ę**.*
hardly   star-PL-EGR   pale   light-thing   fall-MOM-PRS.3SG

'Just a pale light falls from the stars.' (SKNA 291:1a)

Creating a momentaneous derivation therefore does not change the conjugation type of the base verb, but these hardly ever appear to be formed from verbs that typically occur with the -ę- conjugation. One reason for this could be that these verbs are perceived as having a continuous temporal reading, which is not compatible with the punctual semantics of momentaneous derivations. In Standard Komi, however, they do occur regularly with momentaneous verbs, which points to a difference between the written language and the Udora dialect.

The fact that causative verbs typically belong to the conjugation type that is connected to transitive verbs is logical, as causative derivations are normally transitive (Dixon 2010: 169). Therefore we would not expect to find many such verbs with -ę-.

It has also been reported that Udora reflexive derivations carry an inchoative meaning. The examples provided in Žilina and Sorvačeva (1971: 234) include *локсьыны* /lokśį̣nį̣/ 'start to go', *петчыны* /pećći̯nį̣/ 'start to exit' and *сёйсьыны* /śojśį̣nį̣/ 'start to eat; eat'. Nekrasova (2000: 48) also describes these Udora forms as reflexives, pointing to the examples *мунсьыны* /munśį̣nį̣/ 'go' and *гижсьыны* /gižśį̣nį̣/ 'write' (see the meaning in Standard Komi in Table 5). In these descriptions, the verbs are said to occur with -ę-, but as such forms are extremely rare or missing in our corpus, this interesting phenomenon must await further investigation. Lytkin (1961: 57) has also suggested that some onomatopoetic verbs in Udora would use the -ę- conjugation. Unfortunately we cannot verify this using

the current data, although the fact that verbs of vocalisation employ this conjugation could possibly support this.

5.5. The narrative use of tense markers in the Udora dialect

In the Udora dialect, as previously discussed, the Standard Komi present and future markers appear to be a primarily lexical feature that is connected to lexical aspect. However, as very few verbs fall absolutely into one conjugation or the other, we have decided to examine the broader context of their occurrence at the textual level. The narrative use of the Komi future tense has been regularly mentioned in the literature, as described in Section 3.2, and we can also find specific narrative-related phenomena in the Udora data.

The Udora data contains numerous instances where a narrative that has been in the past tense suddenly switches to using forms in *-as-* for all verbs. This can be compared, for instance, to the use of the past and historical present tenses in English. There, the variation in tenses can be connected to opening the narrative, with the tense returning to present in the coda section, among other structural functions (Schiffrin 1981, see also Fludernik 1991). A similar usage has also been described in Russian and other Slavic languages (Comrie 1976: 75–76). To our knowledge, tense variation in Komi narratives has not been studied extensively, although it is known that both present and future can also be used in narratives that describe past events, as was discussed in Section 3.2. In the Udora texts, we often encounter the use of verbs with *-as-* in parts of the story describing actively occurring events, whereas background information in the story appears either in the present tense, with the verbs either in the *-as-* or *-ẹ-* conjugation, or in one of the past tenses. The system seems to be used in such a way that in the beginning, the tense becomes established, and the verbs marked with *-as-* indicate that the narrative is continuing in the same tense.

We can first discuss a series of examples from a complete published and recorded text. In the beginning, the fact that the story takes place in the past, and was not witnessed personally by the narrator, is marked with the second past tense in the verb 'be'. Otherwise, the tense use in the beginning of the story is according to the standard Udora system, i.e. the choice of form is lexical and does not indicate a tense difference between the verbs. As the first sentence of the story already situates it in the past tense, we must consider the whole following section to already represent some type of historical present tense:

(21) Вот тар кыйӧм эшшӧ вӧлӧма сэтшӧм жӧ случай. Ӧти вӧралісь
мунӧ вӧрӧ, ружйӧ сьӧрсис. Видзӧдас: пу вылын пукалӧныс
таръяс. Видзӧдас да лыддяс: сё кымын унджык гашкӧ вӧлӧма
сэні. И думайтӧ: […]

vot	*tar*	*kij-ẹm*	*eššẹ*	*vẹl-ẹma*	*sečẹm*
well	grouse	hunt-PST.PTCP	also	be-PST2.3SG	this_kind

žẹ	*slućaj.*	*ẹt′i*	*vẹraliś*	**mun-ẹ**	*vẹrẹ.*	*ružjẹ*
FOC	event	one	hunter	go-PRS.3SG	forest-ILL	rifle

śẹrś-is.	**viʒẹd-as:**	*pu*	*vil-in*	**pukal-ẹnis**
hold-PST.3SG	look-FUT.3SG	tree	top-INE	sit-PRS.3PL

tar-jas.	**viʒẹd-as**	*da*	**lid′d′-as:**	*śo*	*kimin*
grouse-PL	look-FUT.3SG	and	count-FUT.3SG	hundred	about

un-ʒik	*gaškẹ*	*vẹlẹma*	*seni.*	*i*	**dumajt-ẹ:**
lot-CMP	maybe	be-PST2.3SG	there	and	think-FUT.3SG

'Well there was also a story about grouse hunting. One hunter goes
to the forest, with a rifle. He sees: the grouses are sitting in a tree. He
takes a look and counts: around a hundred, maybe more [grouses]
there were. And he thinks: […]' (Vászolyi-Vasse 1999: 440)

This continues into the second part of the narrative, presented in (22),
where the protagonist ponders where to shoot:

(22) Кытчӧ-я лыя? Улӧ, вылӧ или шӧрас? Кысь унджык сюрӧ?
Думайтас да и: лыйла, медджӧ, вылӧ!

kiććẹ-ja	*lij-a?*	*ul-ẹ,*	*vil-ẹ*	*il′i*	*šẹr-as?*
where-Q	shoot-PRS.1SG	under-ILL	top-ILL	or	middle-3SG.ILL

kiś	*un-ʒik*	**śur-ẹ?**	**dumajt-as**	*da*	*i:*
where.ELA	lot-CMP	catch-PRS.3SG	think-FUT.3SG	and	and

lij-l-a,	*meʒʒẹ,*	*vil-ẹ!*
shoot-FREQ-PRS.1SG	FOC	top-ILL

'Where should I shoot? Below, above or in the middle? How do
I catch the most? He decides: I'll shoot, you see, from above!'
(Vászolyi-Vasse 1999: 440)

After this, however, comes the climax of the story. This is a sequence of
events, where the first grouse falls onto the one under it and so on, caus-
ing a chain reaction of collapsing grouses. Every verb in this sequence is
marked with the *-as-* conjugation. This can be compared to the descrip-
tion of the Standard Komi future tense use in Section 3.2, where the future

tense was described in point five as being used to bring the speaker closer to the events and to make the speech more vivid. Similar use connected to plot progression, or tellability, has also been described for the historical present tense in English (Fludernik 1991: 392–393). It could also be argued that (21) and (22) both contain background information, while (23) is distinctly marked. Distinguishing between narrative events has been suggested as one use of the historical present, and it has also been proposed that individual tenses have a tendency to cluster together for functional and discourse-level reasons (Schiffrin 1981: 52). The next segment is also similar to complicating action clauses, which Schiffrin (1981: 48) describes as being used to relay a series of temporally ordered narrative events.

(23) Муй нö, лыяс: вылыс тарыд усяс. Уськöдас мöдсö, мöдыс коймöдсö, нёльöдыс витöдсö, дасöдсö, ветымынöдсö и сёöдсö, ставыс уськöдас. Мунас пу дорад, вöвтö матöдас сэтчö. Сöвтас дойд тыр. Да коляс на мöд сымыттöм сэтчö. Сідз и сэтчö прöпадитас ставыс.

muj *nẹ,* **lịj-as:** *vịl-as* *tar-ịd* **uś-as.**
what FOC shoot-FUT.3SG above-3SG grouse-2SG fall-FUT.3SG

uś-kẹd-as *mẹd-sẹ.* *mẹd-ịs* *kojmẹd-sẹ.*
fall-CAUS-FUT.3SG second-ACC second-3SG third-ACC

ńol'ẹd-ịs *vitẹd-sẹ,* *dasẹd-sẹ,* *vetịmịnẹd-sẹ* *i*
fourth-3SG fifth-ACC tenth-ACC fiftieth-ACC and

śoẹd-sẹ, *stav-ịs* **uś-kẹd-as.** **mun-as**
hundreth-ACC everything-3SG fall-CAUS-FUT.3SG go-FUT.3SG

pu *dor-ad* *vẹv-tẹ* **matẹd-as** *seććẹ*
tree at-2SG.ILL horse-2SG.ACC bring_near-FUT.3SG there

sẹv-t-as *dojd* *tịr.* *da* **kol'-as** *na*
put-CAUS-FUT.3SG sledge full and remain-FUT.3SG still

mẹd *sịmịttẹm* *seććẹ.* *siʒ* *i* *seććẹ*
other so_much there so and there

prẹpadit-as *stav-ịs.*
get_lost-FUT.3SG everything-3SG

'Well, he shoots: the uppermost grouse falls. It falls onto the second one, the second onto the third, the fourth onto the fifth, tenth, fiftieth and the hundredth, they all fall. He goes to the tree, brings the horse closer. Piles his sledge full. And he still leaves so much there. And so all that got wasted.' (Vászolyi-Vasse 1999: 440)

The story ends with the laughter of the interviewer, Albert Vaneev, who is also a native Udora dialect speaker.

(24)   Vaneev: Ӧни на сисьмӧ. Ekimov: Ӧни на сисьмӧ, да.

ęni	na	**śiśm-ę.**	ęni	na	**śiśm-ę,**	da.
now	still	rot-PRS.3SG	now	still	rot-PRS.3SG	yes

'Vaneev: Still they rot. Ekimov: Still they rot, right.' (Vászolyi-Vasse 1999: 440; edited after the recording SKNA 13761: 2az)

At this point the story is already over, and we return to the expected Udora system, with the next verb marked with the -ę- conjugation. This is the verb сисьмыны /śiśmịnị/ 'rot', an example of a non-volitional process verb, which we earlier analysed as often belonging to this conjugation. Most importantly, this narrative shows several instances where verbs that typically belong to the -ę- conjugation occur with -as-. For example, this happens to the verbs мунны /munnị/ 'go', колльны /kolʹnị/ 'remain' and усьны /uśnị/ 'fall'. We can therefore show that in some narrative contexts, the default third-person marking for different verbs can be overridden.

This temporal shift is not obligatory, as the narrator always has various means available to express events in the past. There are also texts where every verb is marked with -ę- in the third person. We can illustrate this here with an example from *Syrjänische Texte III*.

(25)   Сірп, сія овлывлӧ вурӧма пом помӧ дӧра. Пасьтаыс дӧраыслӧн кык аршын. Сія овлывлӧ дас аршын кузя. Вӧдзӧ вийым сірп шӧрас мешӧк, шуӧны сірп сітанӧн. Вӧдзӧ вийым сірп вылас вурлӧма пуись карӧм табъес. Улысла-дорас вурлӧма из ротӧ. Вӧдзӧ вийым мӧд мӧдӧр сірп помас гез, метра куим кузя. Гессӧ вурӧма сірп помас улысладорас и вылысладорас. Сы гезійӧн ваӧдіс кыскӧны. Улысладор гессӧ гаровтӧны кокӧ а вылысладор гезнас кыскӧны кийӧн. Сірпӧны куим морт, кык морт кыскӧ ва кузяыс а коймӧд черисӧ сірпланьыс повзьӧдӧ бедйӧн.

sirp,	sija	**ov-lịv-l-ę**		vur-ęma		pom	pom-ę
sirp	3SG	be-FREQ-FREQ-PRS.3SG		sew-PST2.3SG		end	end-ILL

dęra.	paśta-ịs	dęra-ịs-lęn		kịk	aršịn.	sija
cloth	length-3SG	cloth-3SG-GEN		two	aršin	3SG

ov-lịv-l-ę			das	aršịn	kuźa.	vęžę	vijịm	sirp
be-FREQ-FREQ-PRS.3SG			ten	aršin	long	then	EX	sirp

šer-as		*mešek,*	***šu-ęnį***	*sirp*	*sitan-ęn.*	*vęžę*
middle-3SG.INE		*mešek*	say-PRS.3SG	*sirp*	bottom-INS	then

vijįm	*sirp*	*vįl-as*	*vur-l-ęma*		*pu-iś*
EX	*sirp*	top-3SG.ILL	sew-FREQ-PST2.3SG		wood-ELA

kar-ęm	*tab-jes.*	*ulįslador-as*		*vur-l-ęma*
do-PST.PTCP	bobber-PL	underneath-3SG.ILL		sew-FREQ-PST2.3SG

iz	*rot-ę.*	*vędźę*	*vijįm*	*męd*	*mędęr*	*sirp*
stone	mouth-ILL	then	EX	other	other	*sirp*

pom-as	*gez,*	*metra*	*kuim*	*kuźa.*	*ges-sę*
end-3SG.ILL	rope	meter	three	long	rope-3SG.ACC

vur-ęma	*sirp*	*pom-as*	*ulįslador-as*		*i*
sew-PST2.3SG	*sirp*	end-3SG.ILL	underneath-3SG.ILL		and

vįlįslador-as.	*sį*	*gezj-ęn*	*va-ęd-įs*	***kįsk-ęnį***.
above-3SG.ILL	3SG	rope-INS	water-PROL-3SG	pull-PRS.3PL

ulįslador	*ges-sę*	***garovt-ęnį***	*kok-ę*	*a*	*vįlįslador*
underneath	rope-ACC	roll-PRS.3PL	foot-ILL	but	above

gez-nas	***kįsk-ęnį***	*kij-ęn.*	***sirp-ęnį***	*kuim*	*mort,*
rope-3PL.INS	pull-PRS.3PL	hand-INS	*sirp*-PRS.3PL	three	man

kįk	*mort*	***kįsk-ę***	*va*	*kuźa-įs*	*a*	*kojmęd*
two	man	pull-PRS.3SG	water	along-3SG	and	third

ćeri-sę	*sirp-lań-įs*	***pov-ź-ęd-ę***		*bedj-ęn.*
fish-3SG.ACC	*sirp*-APPR-3SG	fear-REFL-CAUS-PRS.3SG		stick-INS

'A *sirp*, it may be (made so) that a cloth is sewn from one end to
another. It is two *aršins* long. It can be ten metres long. Then in the
middle of the *sirp* there is *mešek*, it is called the *sirp* bottom. Then,
to the top of the *sirp*, are sewn wooden bobbers. To the underside a
stone is sewn, to the mouth. Then, at the other end, there is a rope,
it is three metres long. The rope is sewn to the end of the *sirp* from
underneath and from above. With this rope, it is pulled along the
water. The lower rope is rolled by foot, and the upper rope is pulled
by hand. Three persons use the *sirp*, two pull it along the water and
the third frightens the fish toward the *sirp* with a stick.' (Uotila
1989: 340)

There are, however, crucial differences between this text and the previ-
ous example. In (23), the events described serve a very specific function
in advancing the narrative plot, and they have a clear temporal order. The
protagonist shoots and causes the birds to fall, then he gets his horse and
sledge, then packs the birds up in one sequence of actions. Example (25), by

contrast, is a more static description of a type of fishing equipment. In the very first phrase, we have the verb form /ovli̯vlẹ/, which is a frequentative derivation of 'be'. There are also numerous verbs in the second past tense formed with frequentative derivations. It is also noteworthy that, although the example includes numerous transitive verbs, they are used primarily without explicit objects. In the very end, we have an instance of a transitive verb with a direct object, in the phrase /ćerisẹ povźẹdẹ/ 'frightens the fish'. Still, the conjugation type does not change with each verb, but rather spans the entire text. This can be compared to the way the use of the Standard Komi present tense is described in Section 3.2, where the tense is described as also being used to mark events in the distant past, along with events that exist at all times. In this context, the narrator might have though of fishing with a *sirp* as an older practice, remote from the current situation (as the narrator was a prisoner of war in Finland when the narrative was being told), or as a general way of doing this activity that essentially exists at all times.

According to Cypanov (2005: 150), the historical future tense is used to some extent in Komi, but it is particularly common in Permian Komi, which has a stronger Russian influence. He points to parallels in Russian dialects where the historical future is used. As Udora is one of the Komi dialects with a relatively strong Russian influence, it does not seem impossible that Udora Komi speakers would have become accustomed to Russian narrative practices. As both different tense markers are used in Udora in these kinds of narrative functions, it is not clear whether we should assume contact influence or an independent development in the dialect. That said, this process in Udora clearly illustrates that there are at least some functions where the conjugations are not used to mark only verb-specific lexical aspect, and that they can acquire tense-like use in longer spans of text under certain conditions.

In Section 3.2, we mentioned that the Komi present tense is used to express events that are true at all times, as well as to mark distant events in folklore. This may offer a fruitful point of comparison between the contemporary Komi and Udora systems. In (21), we also see an instance where the second past tense is used to establish the timeframe of the narrative, after which the present and future tenses are used. A similar use has also been described in Standard Komi. In the next section, we provide a more detailed comparison between Old Komi and Udora, also with a view to contemporary Standard Komi.

5.6. Comparison to Old Komi

The Old Komi corpus is extremely small, consisting of short texts on icons and historical manuscripts in the Old Komi script. This language form is attested from the 14th century, and it represents an extinct Komi dialect that was closely related to Udora and the Lower Vyčegda dialects. All in all, these materials consist of 233 words of connected text written in the Old Komi alphabet (Sidorov 1962: 189). In addition to these fragments, there is essentially one longer text in the Cyrillic alphabet that is thought to represent the same Old Komi language form, possibly having been copied from Old Komi originals. There are three known copies of this text. Two were used by Lytkin in his description of Old Komi: Lepehin's text and the Evgenian text. One more version was found and published by Sidorov (1962). With all of these texts taken together, but counting the text that exists in three versions only once, the entire Old Komi corpus is just under one thousand words.

These materials are now becoming more accessible than ever, as the characters used are currently in the Unicode standard under the Old Permic character block. The authors of this study have collected various versions of Old Komi texts published by Lytkin (1952) and Sidorov (1962), and constructed a corpus that is available online (Partanen 2021). We have analysed this corpus for various features, including instances of parallel forms existing in both Udora and Old Komi in other aspects of the lexicon and morphology, but crucially for the question at hand, we have morphologically annotated all third-person verbs in the present and future tense. We examined the corpus, which, to our knowledge, now contains all known Old Komi texts, and manually extracted all verbs that occur in the third person in the present or future. Table 8 provides a comparison of these verbs in Old Komi and our Udora materials.

The systems are very similar in the majority of cases, with Old Komi and Udora verbs employing the corresponding morphemes most of the time. We do see some differences, and the total number of verbs in the Old Komi corpus is very small, but the correspondences are still interesting and the pattern is so strong that it appears to be more than coincidental. Since the number of examples in the Old Komi corpus was so small, no further statistical tests were carried out. Most of the verbs occur only once in the Old Komi texts, and some are also absent from our Udora materials. There is also an example of the verb *ползьыны* /lolźinˊi/

Table 8: Old Komi and Udora third-person verbs compared

Verb	UPA	Translation	Old Komi	Udora	-ę-	-as-	Comment
вайны	*vajnį*	bring	-as-	-as-	1	27	
вайсьыны	*vajśįnį*	bring (refl.)	-as-	–	0	0	not found in the corpus
воны	*vonį*	arrive	-as-	-as-	3	96	
вуджны	*vuʒnį*	cross	-as-	-as-	0	2	
кутны	*kutnį*	hold	-as-	-as-	2	22	
кывны	*kįvnį*	hear	-as-	-ę-	18	7	see Section 5.1
локны	*loknį*	come	-ę-	-ę-	59	11	
лолзьыны	*lolźįnį*	revive	-ę-	-as-	0	2	semantic difference
лыддьыны	*lįd´d´įnį*	read	-anį-	-as-	1	6	cf. Upper Sysola future
лэдзны	*leʒnį*	let	-as-	-as-	1	28	
петны	*petnį*	exit	-ę-	-ę-	48	5	
сетны	*śetnį*	give	-as-	-as-	5	20	
сулавны	*sulavnį*	stand	-ę-	-ę-	26	0	

(Old Komi) ~ *ловзьыны* /lovźįnį/ (Udora), where the meaning in Old Komi is 'be resurrected, born again', and which in our Udora examples is used to refer to dough rising. Earlier in our analysis, we stated that many of the process verbs like this would normally use the -ę- conjugation in Udora. Additionally, Old Komi shows examples of morphemes that are similar to the future forms encountered in Upper Sysola, formed in the plural with /-anįs/ (Žilina 1975: 118), which is another type of variation not present in Udora. As we have already analysed in the earlier sections, the system in Udora is not always clear-cut. There is variation, and some of it is strongly related to narrative structures that override the lexical aspect, which is the main factor behind the selection of allomorphs. A wider comparison would be necessary but it is not currently possible, at least with the Old Komi materials themselves.

As the Old Komi texts are of a religious nature, there are also newer editions of the same sentences. We will therefore take a brief look at the corresponding texts in contemporary Komi. If we look at the Old Komi

text in (26), we see a short fragment that is also present in the contemporary Bible. One modern translation of the corresponding lines is: […] *but has crossed over from death to life. Very truly I tell you, a time is coming and has now come when the dead will hear the voice of the Son of God and those who hear will live.* (*The Holy Bible* 2011: John 5:24–25). One modern Russian translation, on the other hand, is as follows: […] *но перешел от смерти в жизнь. Истинно, истинно говорю вам: наступает время, и настало уже, когда мертвые услышат глас Сына Божия и, услышав, оживут.* (*Biblija* 2010: 1475). This text has an interesting tense structure with a clear future reference, which makes it relevant for further examination. In (26) from Old Komi, each third-person verb occurs in a different form, essentially following the conjugation pattern we see in Udora today. The transitive verb 'hear' occurs with *-as-*, as does the temporally bounded intransitive verb 'cross'. In contrast to these, the intransitive verbs 'come' and 'be born again' occur with *-ę-* conjugation.

(26) […] чӧчкӧ вузас куламлысʼ олӧм пычкӧ, ина вэсʼтас вӧйпам тийанлы, вэвэсʼ локтӧ йэз и ӧнэ вим куʒʼкӧ кулӧмайас кыласныс горасӧ йэн пилон, сэсʼ кылмысʼ и лôлзʼӧныс.

čeč	*kę*	***vuʒ-as***	*kulęm-liś*	*olęm*	*pičk-ę.*	*ina*
also	if	cross-FUT.3SG	death-ABL	life	inside-ILL	truly

veśt-as	*vęjp-am*	*tijanli,*	*veveś*
straight-3SG.ILL	condemn-PRS.1PL	2PL.DAT	because

lokt-ę	*jęz*	*i*	*ęne*	*viim*	*kuʒkę*
come-PRS.3SG	people	and	now	EX	somehow

kulęma-jas	***kil-asnis***	*gora-sę*	*jen*	*pi-lęn*
dead-PL	hear-FUT.3PL	voice-3SG.ACC	God	son-GEN

seś	*kilm-iś*	*i*	***lolź-ęnis***
then	word-ELA	and	be_born_again-PRS.3PL

'… also crosses from death to life, very directly we say to you that, it will come, people who now are, in some time the dead will hear the voice of the Son of God, and be born again.' (adapted from Lytkin 1952: 66)

The translation is the authors' interpretation of these lines, based on the Old Komi text and the different existing translations. If we then take the contemporary Standard Komi version of the same fragment, presented in (27), we see that the tense marking is very different and is reminiscent of

the tense marking in the English and Russian versions of the text. The time that *will come* is marked with the Standard Komi future tense, while the background information is given in the past tense. This can be compared to the variation in tense use in narratives that we examined in Section 5.5.

(27) […] но кулöмсьыс сийö вуджис нин олöмас. Веськыда, веськыда висьтала тіянлы: воас кад да воис нин, кор кулöмяяс кыласны Ен Пилысь гöлöс шысö. Сэки Сійöс кылысьяс ловзясны.

no	*kulẹm-śịs*	*sijẹ*	*vuž-is*	*ńin*	*olẹm-as.*
and	death-3SG.ELA	3SG	cross-PST.3SG	already	life-3SG.ILL

veśkịda,	*veśkịda*	*vištal-a*	*tijanlị:*	**vo-as**	*kad*
directly	directly	say-PRS.1SG	2PL.DAT	come-FUT.3SG	time

da	*vo-is*	*ńin,*	*kor*	*kulẹma-jas*	**kịl-asnị**
and	come-PST.3SG	already	when	dead-PL	hear-FUT.3PL

jen	*pi-lịś*	*gẹlẹs*	*šị-sẹ.*	*seki*	*sijẹs*
God	son-ABL	voice	sound-3SG.ACC	then	3SG.ACC

kịlịś-jas	**lovź-asnị**
listener-PL	born_again-FUT.3PL

'[…] and from death he turned already to life. Truly, truly I say to you: the time will come, and has already come, when the dead will hear the son of God's voice. Then those who hear it will be born again.' (*The New Testament in Komi-Zyrian language* 2013: Иоан серти бур юöр 5:24–25)

Although this is an isolated comparison, we think it provides a relatively good picture of the differences between the Udora dialect and Standard Komi when it comes to the use of these morphemes. Based on our intuition, the contemporary Bible translation reflects the way many contemporary Komi texts would express these tense distinctions. We can even compare the temporal sequence in it to Example (2) in this study, where the future tense is used to refer into an event, i.e. Vladimir being called a fool, that is presumed to occur later. However, we feel that these examples support the idea that the choice of third-person morpheme must have been conditioned at an earlier historical stage by the lexical aspect of individual verbs, possibly so that one of the morphemes was reserved for a small group of intransitive verbs, and that eventually this has come to express a more temporal sequence of events elsewhere in Komi.

6. Conclusion

As we have demonstrated in our study, rather than displaying distinct present and future tenses, the Udora dialect of Zyrian Komi has one undifferentiated non-past tense. The morphemes that correspond to the Standard Komi present and future tenses are used in the dialect to distinguish two groups of verbs that differ in their lexical aspect. When the verbs are derived in ways that modify their lexical aspect, the conjugation type also changes in a predictable manner. The aspectual properties of verbs belonging to the *-ę-* conjugation are connected to telicity, expressing temporal continuity and boundedness. Additionally, various process verbs and verbs of vocalisation belong to the *-ę-* conjugation, but due to their small number in the corpus, their exact properties have not been analysed here in detail. All frequent transitive verbs and approximately half of all intransitive verbs occur with *-as-*, which can be considered the most common third-person marker in the non-past tense in the Udora dialect.

We have described the historical background of the Udora dialect and shown that the process we encounter here has strong parallels in Old Komi. This fits well with the wider historical context of the Northern Zyrian dialects, a group that includes both Old Komi and Udora. This connection has been proposed several times before, but we hope that our investigation will serve to clarify the situation and lead to further dialectological research covering a wider range of varieties.

From the point of view of derivational morphology, we can distinguish several active processes influencing the selection of conjugation type. Causative derivations make verbs transitive, which shifts the derived verbs into the *-as-* conjugation type following the Udora system. Momentaneous verbs, which must be considered temporally bound, also occur principally with *-as-*. When it comes to verbs derived using the polysemous derivational morpheme *-ś-*, the conjugation type changes in either direction in a predictable manner, depending on the type of derivation that is created.

One occasion in the Udora dialect where the two conjugation types show tense-like properties is in the tense structure of narratives, where the background information in a story is given in one of the past tenses, and the forms that correspond to the standard Komi present and future tenses are used in longer intervals of text. This use appears to override the lexical demands for a verb to fall into a specific conjugation type, thereby contributing to the variation we see in the corpus. We compare this use to historical narrative tense reported in other Komi dialects and propose that

there could be a connection between the Udora system, this kind of tense-like usage in the narratives, and the fully functional present–future opposition that has been described in Standard Komi and the other dialects. We believe that there is still more to investigate with regard to how the tenses are used in different Komi varieties, but we hope that our contribution provides a clearer picture of the situation in the Udora dialect and the Northern Zyrian dialects more generally.

We have combined archival materials, published texts and contemporary recordings to create a collection of Udora texts that spans the entire 20th century. The example sentences are available in the online appendix to this study and are also archived in the Udora Komi collection maintained in the Language Bank of Finland. We hope that this will allow for further verification, comparison and extension of this work whenever new material becomes available. Although numerous open questions remain concerning the verbal morphology of Komi and the Permic languages more broadly, we hope that the findings of our research offer new perspectives and directions for further work.

Acknowledgements

Nonstandard abbreviations used in glosses

APPR	approximative	EX	existential predicate
CMP	comparative	FOC	focus particle
CNG	connegative	PST2	second past
CVB.SIM	simultaneous converb	PROL	prolative
EGR	egressive	Q	question particle

References

Aikio = Luobbal Sámmol Sámmol Ánte (Ante Aikio). (Forthcoming). Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*. Oxford: Oxford University Press.

Baker, Robin W. 1983. Komi Zyryan's second past tense. *Finnisch-Ugrische Forschungen* 45. 69–81.

Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena. 2020. As we say it in Finno-Ugric. Some thoughts on making Uralic language studies more accessible to outsiders. In Holopainen, Sampsa & Kuokkala, Juha & Saarikivi, Janne & Virtanen, Susanna (eds.), *Ёмас сымыҥ нэ̄кве вӧ̄ртур э̄тпост самын патум. Scripta miscellanea in honorem Ulla-Maija Forsberg* (Mémoires de la Société Finno-Ougrienne 275), 27–41. Helsinki: Finno-Ugrian Society. https://doi.org/10.33341/sus.11.1

Bartens, Raija. 2000. *Permiläiskielten rakenne ja kehitys* (Mémoires de la Société Finno-Ougrienne 238). Helsinki: Société Finno-Ougrienne.

Beznosikova et al. = Безносикова, Л. М. & Айбабина, Е. А. & Забоева, Н. К. & Коснырева, Р. И. 2012. *Коми сёрнисикаса кывчукӧр. Словарь диалектов коми языка. В двух томах.* Более 78000 слов. Сыктывкар: Кола.

Biblia = *Святое писание. Библия. Книги Священного Писания Ветхого и Нового Завета. Синодальный перевод*. Русский Паломник, 2010.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Croft, William. 2012. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.

Csúcs, Sándor. 2005. *Die Rekonstruktion der permischen Grundsprache* (Bibliotheca Uralica 13). Budapest: Akadémiai Kiadó.

Cypanov = Цыпанов, Е. А. 2005. *Грамматические категории глагола в коми языке*. Сыктывкар: Институт языка, литературы и истории Коми НЦ УрО РАН.

Cypanov & Leinonen = Цыпанов, Евгений & Лейнонен, Марья. 2009. Эвиденциальность в коми языке (на материале модального перфекта). *Linguistica Uralica* XLV(1). 23–35.

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Basil Blackwell.

Dixon, Robert M. W. 2010. *Basic linguistic theory. Volume 1: Methodology*. Oxford: Oxford University Press.

Federal State Statistics Service, Russia = Федеральная служба государственный статистики. 2010. *База данных показателей муниципальных образования. Республика Коми. Национальный состав населения.* https://gks.ru/dbscripts/munst/munst87 (Accessed 3.8.2021.)

Fedjunёva et al. = Федюнёва, Г.В. & Некрасова, Г. & Лудыкова, В. & Цыпанов, Е. & Попова, Э. 2000. *Ӧнiя коми кыв: Морфология*. (Edited by Г. В. Федюнёва.) Сыктывкар: Коми небӧг лэдзанiн.

Filip, Hana. 2012. Lexical aspect. In Binnick, Robert I. (ed.), *The Oxford handbook of tense and aspect*, 721–751. Oxford: Oxford University Press.

FLUDERNIK, MONIKA. 1991. The historical present tense yet again: Tense switching and narrative dynamics in oral and quasi-oral storytelling. *Text* 11(3). 365–397.

FOKOS-FUCHS, DÁVID RAFAEL. 1916. *Zürjén szövegek*. Budapest: Magyar Tudományos Akadémia.

FOKOS-FUCHS, DÁVID RAFAEL. 1959. *Syrjänisches Wörterbuch: Über 20 000 syrjänische Wörter in genauer Umschrift und mit deutscher Übersetzung, nebst Angabe ihrer Formen, Verbreitung und Anwendungen. Anschliessend ein deutsches Wortregister*. Budapest: Akadémiai kiadó.

HÄMÄLÄINEN, MIKA. 2019. UralicNLP: An NLP Library for Uralic Languages. *Journal of Open Source Software* 4(37). 1345. https://doi.org/10.21105/joss.01345.

The Holy Bible. New International Version. Biblica, Inc. 2011. https://biblehub.com.

HOPPER, PAUL J. & Thompson, Sandra A. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.

ITKONEN, ERKKI. 1958. Komin tasavallan kielitieteeseen tutustumassa. *Virittäjä* 62(1). 66–66.

KLUMPP, GERSON. 2008. *Differentielle Objektmarkierung & Informationsstruktur in Dialekten des Komi*. Habilitation. München: Fakultät für Sprach- und Literaturwissenschaften der Ludwig-Maximilians-Universität München.

LEINONEN, MARJA & VILKUNA, MARIA. 2000. Past tenses in Permian languages. In Dahl, Östen (ed.), *Tense and aspect in the languages of Europe*, 495–514. Berlin: Mouton de Gruyter.

LJAŠEV = Ляшев, В. А. 1977. Некоторые дистинктивные признаки северного нуль-элового ареала коми языка. *Советское финно-угроведение* 13(2). 119–123.

LJAŠEV = Ляшев, В. А. 1980. *Диалектное членение древнекоми языка*. Сыктывкар: Академия наук СССР Коми филиал.

LYTKIN = Лыткин, В. И. 1952. *Древнепермский язык: Чтение текстов, грамматика, словарь*. Москва: Издательство академия наук СССР.

LYTKIN = Лыткин, В. И. (ed.) 1955. *Современный коми язык*. Часть первая. Сыктывкар: Коми книжное издательство.

LYTKIN = Лыткин, В. И. 1961. К вопросу о конечной гласной основы глаголов настоящего времени в финно-угорских языках. *Вопросы языка, литературы и фольклора*. Труды, выпуск 15, 49–57. Йошкар-Ола: Марийский научно-исследовательский институт языка, литературы и истории.

LYTKIN = Лыткин, В. И. 1962. *Коми-пермяцкий язык: введение, фонетика, лексика и морфология*. Кудымкар: Коми-пермяцкое книжное издательство.

LYTKIN = LITKIN, VASZILIJ. 1969. A finnugor alapnyelv szóvégi magánhangzóiról [Über den Auslautvokale der finnougrischen Grundsprache]. *Nyelvtudományi Közlemények* 71. 95–99.

LYTKIN = Лыткин, В. И. 1977a. Зачатки объектного спряжения в коми языка. *Советское финно-угроведение* 13(4). 280–283.

LYTKIN = Лыткин, В. И. 1977b. *Коми кывлӧн историческӧй морфология*. Пермь & Сыктывкар: А. М. Горький нима трудвӧй Краснӧй Знамя ордена Пермса государсвеннӧй университет & СССР-лы 50-во нима Сыктывкарса государственнӧй университет.

MATVEEV = Матвеев, Александр М. 1958. Невеста. *Войвыв кодзув* 2. 47–50.

Nekrasova = Некрасова, Г. А. 2000. *Коми кывлӧн историческӧй фонетика.* Сыктывкар: Россияса наукаяслӧн академия.

The New Testament in Komi-Zyrian language. 2013. Institute for Bible Translation. http: //finugorbib.com/bible/komi_zyrian.

Partanen, Niko. 2021. *Old Komi: Digital Edition.* https://github.com/langdoc/ writtenkomi-corpus-old-komi.

Ponarjadov = Понарядов, Вадим. 2004. О глагольных окончаниях -ӧ и -ас в удорском диалекте. In *Вопросы коми диалектологии* (Труды Института языка, литературы и истории Коми НЦ УрО РАН. Вып. 65), 104–112. Сыктывкар: Институт языка, литературы и истории Коми НЦ УрО РАН.

Popova & Sažina = Попова, Р. П. & Сажина, С. А. 2014. *Фонетические и морфологические особенности коми диалектов (сравнительный аспект исследования).* Сыктывкар: Издательство СыктГУ.

Rédei, Karoly. 1989. Über die finnougrische Konjugation unter besonderer Berücksichtigung der ungarischen Personalsuffixe. *Journal de la Société Finno-Ougrienne* 82. 93–209.

Rueter = Рютер, Джек М. 2000. Хельсинкиса университетын кыв туялысь Ижкарын перымса кывъяс симпозиум вылын лыддьӧмтор. In *Пермистика: Проблемы синхронии и диахронии пермских языков и их диалектов* (Пермистика 6), 154–158. Ижевск: Удмуртский университет.

Savel'eva = Савельева, Э. А. (ed.) 1997. *Историко-культурный атлас Республики Коми.* Москва: Дрофа & ДиК.

Sažina = Сажина, С. А. 2014. Исторические условия формирования коми-зырянских диалектов. In Зайцева, Н. Г. & Муллонен, И. И. (eds.), *V Всероссийская конференция финно-угроведов «Финно-угорские языки и культуры в социокультурном ландшафте России». Материалы,* 81–83. Петрозаводск: Институт языка, литературы и истории КарНЦ РАН.

Schiffrin, Deborah. 1981. Tense variation in narrative. *Language* 57(1). 45–62.

Serebrennikov = Серебренников, Б. А. 1956. Из истории образования глагольных времен в коми языке. *Историко-филологический сборник.* Выпуск 3, 57–69. Сыктывкар: Академии наук СССР Коми филиал.

Serebrennikov = Серебренников, Б. А. 1960. *Категории времени и вида в финно-угорских языках пермской и волжской групп.* Москва: Издательство Академии наук СССР.

Serebrennikov = Серебренников, Б. А. 1963. *Историческая морфологииа пермских языков.* Москва: Издательство Академии наук СССР.

SKNA = Suomen kielen nauhoitearkisto (Tape Archive of the Finnish Language), Institute for the Languages of Finland.

Sidorov = Сидоров, А. С. 1930. Характеристика удорского (вашского) говора. In Лыткин, В. И. (ed.), *Сборник комиссии по собиранию словаря и изучению диалектов коми языка.* Выпуск 1, 49–56. Сыктывкар: Комиссия по собиранию Коми словаря при Обществе Изучение Коми края.

Sidorov = Сидоров, А. С. 1962. Новые памятники древнекоми письменности (с комментариями, подстрочными примечаниями и заключением В. И. Лыткина). *Вопросы финно-угорского языкознания,* 178–211. Ленинград: Издательство Академии наук СССР.

Sᴏʀᴠᴀčᴇᴠᴀ = Сорвачева, В. А. 1952. Некоторые фонетические и морфологические особенности верхне-вашского говора удорского диалекта. In *Лингвистический сборник. Серия лингвистическая*. Выпуск 2, 37–48. Сыктывкар: Коми Государственное Издательство.

Sᴏʀᴠᴀčᴇᴠᴀ & Bᴇᴢɴᴏsɪᴋᴏᴠᴀ = Сорвачева, В. А. & Безносикова, Л. М. 1990. *Удорский диалект коми языка*. Москва: Наука.

Sᴛɪᴘᴀ, Güɴᴛᴇʀ Jᴏʜᴀɴɴᴇs. 1962. Käynti syrjäänien tieteen tyyssijassa. *Virittäjä* 66(1). 61–68.

Tᴏᴅᴇsᴋ, Tʀɪɪɴ. 2015. The verbal augmentative and the inherent properties of verbs in Komi. *Linguistica Uralica* 51(1). 28–43.

Tᴜʀᴋɪɴ = Туркин, А. И. 1984. О двух гидронимических формантах в коми топонимии. In *Вопросы лексикологии и словообразования коми языка* (Труды Института ЯЛИ Коми НЦ УрО РАН 32), 177–182. Сыктывкар: Академия наук СССР Коми филиал.

Uᴏᴛɪʟᴀ, T. E. 1938. *Syrjänische Chrestomathie: mit grammatikalischem Abriss und etymologischem Wörterverzeichnis* (Hilfsmittel für das Studium der finnisch-ugrischen Sprachen VI). Helsinki: Société Finno-Ougrienne.

Uᴏᴛɪʟᴀ, T. E. 1989. *Syrjänische Texte. Band III. Komi-Syrjänisch: Luza-Letka-, Ober-Sysola-, Mittel-Sysola-, Prisyktyvkar-, Unter-Vyčegda- und Udora-Dialekte*. Gesammelt von T.E. Uotila. Übersetzt und herausgegeben von Paula Kokkonen (Mémoires de la Société Finno-Ougrienne 202). Helsinki: Société Finno-Ougrienne.

Vᴀʜʀᴏs-Pᴇʀᴛᴀᴍᴏ, Mᴜᴜsᴀ. 1963. Syrjäänien asuinseuduilla. *Virittäjä* 67(1). 77–85.

Vásᴢᴏʟʏɪ-Vᴀssᴇ, E. 1999. *Syrjaenica: Narratives, folklore and folk poetry from eight dialects of Komi. Upper Izhma, Lower Ob, Kanin Peninsula, Upper Jusva, Middle Inva, Udora. Volume 1* (Specimina Sibirica 15). Szombathely: Seminar für Uralische Philologie der Berzsenyi Hochschule.

Wɪᴄʜᴍᴀɴɴ, Yʀᴊö Jᴏᴏsᴇᴘᴘɪ. 1916. *Syrjänische volksdichtung* (Mémoires de la Société Finno-Ougrienne 38). Helsinki: Société Finno-Ougrienne.

Wɪᴇᴅᴇᴍᴀɴɴ, Fᴇʀᴅɪɴᴀɴᴅ Jᴏʜᴀɴɴ. 1884. *Grammatik der syrjänischen Sprache mit Berückichtigung ihrer Dialekte und des Wotjakischen*. St. Petersburg: Kaiserlichen Akademie der Wissenschaften.

Žᴇʀᴇʙᴄᴏᴠ = Жеребцов, Л. Н. 1972. *Хозяйство, культура и быт удорских коми в XVIII - начале XX в*. Москва: Наука.

Žɪʟɪɴᴀ = Жилина, Т. И. 1975. *Верхнесысольский диалект коми языка*. Москва: Наука.

Žɪʟɪɴᴀ & Sᴏʀᴠᴀčᴇᴠᴀ = Жилина, Т. И. & Сорвачева, В. А. (eds.) 1971. *Образцы коми-зырянской речи*. Сыктывкар: Академия наук СССР Коми филиал.

Appendix

The online appendix "Tense and Aspect in Udora Komi" is available at https://doi.org/10.33339/fuf.97371.

Susanna Virtanen
University of Helsinki

A template approach to pragmatic constituent order variation in modern Northern Mansi

Mansi belongs to the Ob-Ugrian branch of the Uralic language family. Northern Mansi constituent order and its pragmatic variation have not been examined comprehensively until now. This lack is filled in this article, by syntactic-pragmatic template analysis, using a new model of 9+1 templatic slots, which are filled with syntactic or pragmatic functions. Thus, this study is also an attempt to combine both pragmatic and syntactic levels in the same template analysis. Moreover, Rombandeeva's (1979; 1984) earlier observations on Northern Mansi word order, and those of other scholars, are compared to those drawn here from contemporary data.

1. Introduction

In this article, I provide a syntactic-pragmatic linearization template for Northern Mansi constituent order. To do this, I integrate a pragmatic level into syntactic template analysis. Template analysis is a variable device which is used to analyse linearization and which has been defined, for example, by Good (2016). The idea of writing this article arose from the lack of sufficient knowledge regarding Mansi constituent order. Mansi is an

indigenous language spoken in Western Siberia. It belongs to the Ob-Ugrian branch of the Uralic language family. There are still approx. 1,000 living speakers of Mansi. Three of the four main dialects have practically vanished; only the Northern dialect is still spoken. Many basic details of Mansi grammar have still not been researched at all, or not comprehensively.

Like all Uralic languages, Mansi is an agglutinative language with a rich variety of inflectional and derivational suffixes. It also has postpositions and verbal preverbs. The Northern Mansi case system consists of an unmarked nominative case and five case endings: locative, lative, ablative, instrumental and translative. Unlike many Uralic languages, Mansi has no genitive case: possession and other genitive-related relations are expressed with possessive suffixes. There are three numeral categories in Mansi: singular, dual and plural. All three numbers occur both in verb and noun inflection.

Northern Mansi is a language with Differential Object Agreement (DOA), which is a phenomenon closely related to Differential Object Marking (DOM, see e.g. Bossong 1985; Aissen 2003), and this is conditioned by pragmatics (Dalrymple & Nikolaeva 2011; Virtanen 2014; Sipőcz 2016). According to some recent studies (Skribnik 2001; Virtanen 2015; Sosa 2017), in the information structure[1] of the Ob-Ugrian languages, pragmatic functions correlate with syntactic functions. In other words, in Mansi, information structure is primarily expressed by variation between different syntactic structures (active vs. passive inflection, indirective vs. secundative alignment, etc.), not by variation between word-order patterns. One of the aims of this study is to discover how information structure and constituent order interact with each other in Mansi.

Mansi word order is traditionally described in terms of syntactic functions. Mansi is regarded a language with a basic subject-object-verb (SOV) word order (see e.g. Kálmán 1989; Rombaneeva 1979; Riese 2001). Constituent order has been touched on and discussed by some scholars in the literature. Rombandeeva gives a description of Northern Mansi word order in her book on syntax, written in Russian (1979) and translated into German (1984). Rombandeeva can be regarded as the only native Mansi researcher and her contribution to Mansi linguistics is noteworthy.

Rombandeeva (1984: 58–60) mentions seven rules concerning the placement of syntactic functions: 1) the subject and its modifiers always precede

---

1. In this study, the phenomenon of information structure is used as a subcategory of pragmatics. Both terms are used and often they refer to the same features or functions.

the predicate, 2) the predicate is always in the sentence-final position, 3) the expression of time precedes the subject and is placed in the sentence-initial position, 4) the expression of location is immediately before the predicate, or in the sentence-initial position, 5) adverbials of manner, goal or reason precede the predicate, often also preceding the element they belong to, 6) the direct object is usually between the subject and the predicate, and 7) attributives always immediately precede the word they belong to. She also presents a description of placement of infinitives, question words, auxiliaries, causative verbs, conjunctions and particles (Rombandeeva 1984: 62–75). Further, she discusses the phenomenon of "logical emphasis", which is parallel to the phenomenon of information structure in the modern literature: she mentions that logical emphasis can also result in changes in word order (Rombandeeva 1984: 75–77). Rombandeeva's analysis is quite comprehensive, and one of my aims is to estimate its effectiveness by means of modern linguistics and with comprehensive corpus data.

After Rombandeeva, Mansi word order has been discussed on a narrower level by e.g. Riese (2001), Keresztes (1998), Kálmán (1989), Skribnik (2001), Bíró (2015), and Kulonen (2007). None of these authors gives a comprehensive description. Some mention the pragmatic perspective, but still the issue is not discussed thoroughly. The above-mentioned studies concern several different dialects and the differences between these dialects are relatively large, so features of one dialect cannot automatically be applied to other dialects. Riese (2001: 62–63) mentions that SOV is the basic word order, but he stresses that other orders are possible, if any constituent part bears a particular emphasis position. Kulonen (2007: 191–192) discusses Eastern Mansi, mentioning the basic SOV word order, the constituent order of verbless clauses, and the placement of modifiers and genitives. Skribnik (2001: 223), Keresztes (1998: 420) and Kálmán (1989: 63) all mention that the basic word order is SOV, but without any further elaboration.

Bíró (2015), referring to Riese (2001), notes that in Southern Mansi the basic word order can be changed to SVO due to pragmatic circumstances. Bíró also states that in Northern Mansi, the emphasized (i.e. the most focal) argument occupies the pre-verbal position, while the topic always occupies the sentence-initial position. Further, she presents the word order of indirect and secundative[2] three-participant constructions (see e.g.

---

2. The secundative construction is also known as *Secondary Object Construction* (see Heine & König 2010; Malchukov et al. 2010).

Haspelmath 2015) both in active and passive in Northern and Southern Mansi, and comes to the conclusion that there is a pragmatic motivation behind the word-order variation between the different three-participant constructions. The most topical argument appears before the focal one, and the most focal argument is placed immediately before the predicate (Bíró 2015: 55). Consequently, instead of the most frequent order, RTV (where R = recipient; T = theme; V = verb), the rarer order, TRV, is used only in indirect structures with objective conjugation. This variation is due to pragmatic reasons. (See Bíró 2015: 53.)

All of these comments are supported by my data, but none of them describes the situation thoroughly. This study fills in the gap between the partial studies mentioned above and offers a full description of Northern Mansi word order, including the variation within it. This article discusses Northern Mansi constituent order from two perspectives: 1) bringing a pragmatic level into the discussion, 2) comparing the author's contemporary data with that of other scholars – mainly Rombandeeva's (1979) observations – and elaborating them with new results and views. Thus, this is an attempt to develop a template approach to both the syntactic and the pragmatic level, and possible dependencies between them.[3] The aim of this study is to observe how stable Northern Mansi constituent order is, and what factors the possible variation is based on. My key questions are:

1. What kind of constituent-order variation caused by information structure exists in Northern Mansi?
2. Is the same model applicable to all syntactic structures? Are there differences between active and passive, Indirect Object Construction and the Secondary Object Construction, questions or imperatives and declarative sentences?
3. How does my data support Rombandeeva's observations? How can my analysis supplement or develop Rombandeeva's results?

Before proceeding, some terminology needs to be defined. My emphasis is on how the main syntactic constituents are organized inside the clause. One constituent may include several words; there may be modifiers or non-finite

---

3. At this point, I would like to express my gratitude to docent Maria Vilkuna for all consultation and discussions on this study. Her advice have been more than helpful during this process.

verb forms connected to the main word. If we go further and consider the order of the components of the constituent, we talk about word order. Here, however, the emphasis is on constituents. I use the term "word order" only when referring to studies discussing word order itself or using this term.

In this context, it is also important to distinguish between syntactic functions, semantic roles and pragmatic status. One part of this study is how these levels of language correlate with each other. Syntactic functions (constituents) are referred to with well-established syntactic terms: *subject* (S), *direct object* (DO), *indirect object* (IO), *temporal adverbial* (Temp), *locational adverbial* (Loc), *manner adverbial* (Man) and *agent adverbial* (AgA). The category of indirect objects include two different types: the lative-marked recipient (IO_{LAT}) and the instrumental-marked semantic patient (IO_{INSTR}). The category of adverbials is simplified to include four main categories: time, location, manner and agency. Most of the adverbials found in the data are classified among these four main types. In addition, there is a class of *other adverbials* for some less frequent cases like conditional or reason adverbials. Further, the phenomenon of *scene-setting adverbials* (ScSA) is applied: this is not an independent syntactic category, but a specific category including several types of adverbials and carrying a particular pragmatic status. A scene-setting adverbial is a temporal or locational adverb which is placed at sentence-initial position and sets a spatial or temporal framework within which the main predication holds (see e.g. Chafe 1976: 50).

Respectively, the semantic roles are named *agent*, *patient*, *theme*, *recipient*, *locative*, *goal* and *source*. For pragmatic status, I use the terms *primary topic*, *secondary topic* and *focus*. The primary topic is the most topical element in the utterance, it is the constituent the whole sentence is about. The secondary topic is also topical but less salient. Focus is new, unpredictable or contrastive information brought to the discussion. (See e.g. Lambrecht 1994: 118, 207; Dalrymple & Nikolaeva 2011; Virtanen 2014.) In a broader sense, when analysing corpus data, we can also talk about *narrative topic*, which is the most topical element in the whole text or spoken entity.

Further, typological features typical for Mansi have implications for terminology. In Mansi, both active and passive conjugation paradigms correlate with the syntactic subject of the clause (see Section 2.1). I distinguish the subjects of active and passive clauses by marking them SA (Subject, Active) and SP (Subject, Passive). I only pay attention to the opposition between SA and SP where it causes differences in the constituent order. Otherwise, active and passive clauses are treated equally. Secondly, we cannot ignore

the fact that Mansi is a language with a high frequency of zero anaphora: a subject or a DO is often not expressed explicitly with a nominal constituent, if it is a topical argument and referred to with a conjugation suffix.

In the following, I approach Mansi constituent order by describing how and why the placement of individual syntactic functions varies between two or more templatic slots. This is discussed at the clausal level, because each clause normally has its own verb and represents an independent meaningful unit. According to my data, this model works with both active and passive clauses. The structure of this article is as follows: In Section 2 I present the most remarkable typological features of Mansi and then, in Section 3, some general facts about word-order typology. Section 4 is devoted to my research data and the implementation of my analysis, and the detailed results of my data are given in Section 5. The main conclusions are gathered in Section 6.

2. Information structure in Mansi

In this section I briefly discuss some relevant information-structure features of Mansi. Information structure is primarily expressed by variation between different syntactic structures (see e.g. Skribnik 2001; Virtanen 2015). The data presented in this section is not to describe Mansi constituent order, but to demonstrate the pragmatic variation between different syntactic choices. To understand the details of Mansi constituent order, one has to be aware of the features presented in this section.

First of all, variation between the active and the passive voice is due to a correlation between syntactic and pragmatic functions: the higher in the pragmatic hierarchy (PT > ST > F[4]) an argument is, the higher the syntactic function (Subject > DO > Oblique) it occupies (see e.g. Virtanen 2015). The pragmatic status of a single argument is defined by examining the following features: 1) the argument's possible inherent topicality[5], 2) the argument's appearance in the same text, in the same paragraph, in previous sentences and in the immediately preceding sentence, and 3) any contextual factors affecting the status.

---

4. PT = Primary Topic, ST = Secondary Topic, F = Focus; see Section 1.
5. Inherent topicality is a property that indicates a concept being automatically more easily accessed than others, regardless of the discourse context (see e.g. Taylor 1996: 219–220). For example the 1st and 2nd person referents are very inherently topical, because they are always already accessible in the discourse situation.

If any semantic role other than the agent (i.e. patient, recipient, sometimes even locative[6]) is promoted to the subject (i.e. it is the most topical syntactic argument), the passive is used. Variation between active and passive is demonstrated in examples (1–4). In examples (1) and (2) the semantic agent of the sentence is the primary topic, so an active structure is chosen to place the agent in the syntactic function of subject. In (3) and (4), a passive structure is chosen, because the primary topic is not the semantic agent. In example (1) the chosen sentence, and indeed the whole text, is about an old man. He is referred to by a third-person singular pronoun, which represents the pragmatic primary topic. Because it represents the semantic agent as well, an active sentence is used:

(1) Тав Хӓльӯс   район   Саранпāвыл-т сам-ын пат-ыс.
     Taw Xālʼūs   rajon   Saranpāwəl-t sam-ən pat-əs.
     3SG Beryozovo district Saranpaul-LOC eye-LAT start-PST.3SG
     'He was born in Saranpaul village in Beryozovo District.'
     (LS 21/2018: 2)

The sentence in example (2) is from another article about Ekur, another old man, whose life story has been discussed for several preceding sentences. Ekur represents the primary topic of the sentence; the sentence gives more information about him. The primary topic of the sentence is also the semantic agent: when it occupies the syntactic function of subject, it triggers an active sentence structure:

(2) Екур кӯщаи-г   хульт-ыс.
     Ekur kūśai-γ   χulʼt-əs.
     Ekur head.person-TRANSL stay-PST.3SG
     'Ekur stayed as a head of the household.' (LS 9/2014: 13)

Example (3) is about the President of the Russian Federation being taken to a museum. The article is about him, and he is also the primary topic of this sentence. In this case, however, someone else is taking him to the museum. The third-person singular pronoun represents the semantic patient, but at the same time it is the primary topic of the sentence. A passive sentence structure is chosen, and the verb is inflected in the past tense and third-person singular:

---

6. Also the semantic role of locative can in some rare cases occupy the syntactic function of subject (e.g. Kulonen 1989: 152).

(3) Тав   тāра   музей-н   тот-ве-с.
     Taw   tāra   muzej-n   tot-we-s.
     3SG   tradition museum-LAT   take-PASS-PST.3SG
     'He was taken to the Museum of Traditions.' (LS 21/2018: 2)

Example (4) is a passive clause and includes a semantic agent, which has the syntactic function of agent adverbial and is marked with the lative case. The article in question discusses patients, who represent the primary topic of the sentence. The new information provided is that in certain cases doctors send them to their own hospital. The third-person plural pronoun (i.e. the patients) is the semantic patient of the sentence and occupies the syntactic function of subject. Therefore, the verb is inflected in the passive third-person plural:

(4) Тāн […] лēккар-ыт-ын тāнти   пӯльница-ныл-н
     Tān […] lēkkar-ət-ən tānti   pūl′nitsa-nəl-n
     3PL   doctor-PL-LAT 3PL.STRESS hospital-POSS.PL<3PL-LAT

     кēт-аве-т.
     kēt-awe-t.
     send-PASS.PRS-3PL

     'They […] are sent to their own hospitals by the doctors.' (LS 1/2018: 6)

In active transitive structures, information structure is also reflected in DOA/DOM. Skribnik (2001) has discussed this concerning Northern Mansi, as has Virtanen (2014) concerning Eastern Mansi. Mansi marks a topical DO by indexing on the verb, but not focal indexing. In other words, there are two verb-inflection paradigms in Mansi. The objective conjugation is the primary technique for indexing a topical DO, it is used when the verb is accompanied by a topical DO (see Virtanen 2013; 2014). An objective-conjugation ending indicates the person and number of the subject, as well as the number of the DO. The subjective conjugation is used when there is no DO (intransitive action, or transitive action without a specified object), or it is accompanied by a non-topical (focal) DO. Examples of these variations can be found here, for instance, in examples (5), (9), (27), (28), (29) and (39) for the objective conjugation and (1), (2), (6), (7) and many others for the subjective conjugation.

A good example is number (5) below, where the DO turns from focal to topical. When the book or report is first mentioned with its author,

the verb is in the subjective conjugation (*хансыс* 'wrote'), because of the focality of the DO. Right after that, the book is topicalized, and in the next clause it appears as a topical DO only referred to with an objective-conjugation verb ending (*кēтыстэ* 'sent [it]'). The DO is the secondary topic of the latter clause, while the subject remains the primary topic of the whole sentence.

(5) Тав   та   пāвыл-т   сāлы-ӈ   мāхум   ялпыӈ
     Taw   *ta*   *pāwəl-t*   *sāli-ŋ*   *māχim*   *jalpəŋ*
     3SG   that   village-LOC   reindeer-ADJ   people   sacred

     хōтал-аныл   урыл   проект+нēпак   ханс-ыс   ос   ты
     χōtal-anəl   *urəl*   *projekt+nēpak* *χans-əs*   *os*   *ti*
     day-3PL   about   project+book   write-PST.3SG   and   this

     касыл-н   кēт-ыс-тэ.
     kasəl-n   *kēt-əs-te.*
     competition-LAT   send-PST-SG<3SG

     'She wrote a project report about the sacred day of the reindeer
     keepers in the village, and sent it to a competition.' (LS 23/2019: 2)

Further, the most topical subjects and DOs are not expressed explicitly at all but only referred to with a verb ending. Examples (6) and (7) are equal main clauses of the same sentence. Both of them lack an overt subject, which is only referred to with a past tense ending, a zero ending referring to the singular third-person. The subject referent is a reindeer, but due to its high topicality and mentions in the immediately preceding sentences, the noun argument is not repeated here:

(6) Тӯр   вāта   хосыт   ēмант-ас […]
     Tūr   *wāta*   *χosət*   *jōmant-as* […]
     lake   coast   along   saunter-PST.3SG
     '[The reindeer] sauntered along the lakeside […]' (LS 15/2017: 15)

(7) […] пēлп   āньт-ӣ-тыл   мā   хилая-с.
     […] *pēlp*   *āńt-ē-təl*   *mā*   *χilaja-s.*
         sharp   horn-POSS.SG<3SG-INS   land   dig-PST.3SG
     '[…] and dug the ground with his sharp horn.' (LS 15/2017: 15)

In the same way, in (8) there is no overt third-person plural subject in the passive clause. The text is about patients and how they are taken care of and informed about illnesses; the phenomenon of patients of the hospital is very accessible in this context. For this reason, a highly topical subject can be recognized by the context, and it is referred to with a normal verb ending.

(8) Мāнь нэ̄пак+лō̄мт-ыт-ыл   май-вē-с-ыт […]
     Māń *nēpak+lōmt-ət-əl*   *maj-wē-s-ət* […]
     small book+piece-PL-INS   give-PASS-PST.3PL
     '[The patients] were given small booklets […]' (LS 20/2018: 4)

Besides the variation between active and passive described above, there is pragmatic variation between ditransitive structures (see 9–11). Northern Mansi ditransitive constructions have recently been examined by Bíró and Sípőcz (2017) from the typological point of view. Referring to the terminology of Malchukov et al. (2010), they state that the Northern Mansi ditransitive constructions are the *Secondary Object Construction* (SOC) (or *Secundative Alignment*) and the *Indirect Object Construction* (IOC) (or *Indirective alignment*) (Bíró & Sípőcz 2017: 44–45). Also in ditransitive clauses, both active and passive, the most topical element occupies the syntactic function of subject and is the one the verb correlates with.

Example (9) illustrates the active Secondary Object Construction, where the semantic recipient occupies the syntactic function of DO, while the semantic theme appears as an instrumental-marked IO. A structure like this is used when the semantic theme is the pragmatic focus of the sentence, and the recipient is the secondary topic. Example (9) is from a news article about administration workers. Local leaders are awarding some of them for their good work. These leaders have been mentioned previously, so the semantic agent (third-person plural pronoun) is the primary topic and occupies the syntactic function of subject. The semantic recipients have been mentioned and are textually topical (their exact number is mentioned for the first time, whereas previously an approximation was given). So, the recipient is the pragmatic secondary topic: it appears as a syntactic DO and is accompanied by the objective conjugation. The new information in the sentence is the award of Certificates of Merit: as a focal argument the semantic Theme (Certificate of Merit) occupies the syntactic functions of IO, and it is inflected in the instrumental case.

(9) Тāн   акв   организация   ос   налыман нупыл акв   хо̄тпа
Tān   *akw*   *organizatsija*   *os*   *naliman nupəl akw*   *χōtpa*
3PL   one   organization   and   31   person

Почётный грамота   нам-па   нэ̄пак-ыл   ми-с-аныл.
Potšotnij gramota   *nam-pa*   *nēpak-əl*   *mi-s-anəl.*
certificate.of.merit   name-PTCP   document-INS   give-PST-SG<3PL

'They gave [documents called] Certificates of Merit for [the aforementioned] one organization and 31 persons.' (LS 14/2018: 2)

Examples (10) and (11) are both from a newspaper article about the Russian president's visit to Khanty-Mansiysk: (10) is the title of the article. The most topical element of the clause is the president. The new information (focus) in the title line is that he was given a traditional Mansi belt. For this reason, the passive Secondary Object Construction is used: the semantic recipient occupies the syntactic function of subject, and the semantic theme is marked with the instrumental case (as in 6). The semantic agent, which is focal as well, is marked with the lative case.

(10) Россия   Президент   ма̄ньщи   хум-н   э̄нтап-ыл
Rossija   *prezid́ent*   *māńśi*   *χum-n*   *ēntap-əl*
Russia   president   Mansi   man-LAT   belt-INS

мӯйлупта-ве-с.
mūjlupta-we-s.
give.a.present-PASS-PST.3SG

'The Russian president was given a [traditional Mansi] belt by a Mansi man.' (LS 21/2018: 2)

The active sentence in example (11) illustrates the Indirect Object Construction: it is from the same article as example (10) and discusses the exact same action, but from a wholly different perspective. In (11), the semantic agent appears as the subject of the sentence: from the context we know that the man, called Aleksandr, is one of the performers in the visit program, as he has been mentioned along with two other persons in preceding sentences. Consequently, the semantic agent is the pragmatic primary topic and appears as the syntactic subject. An active sentence structure is chosen. The belt, the semantic theme, is focal, so it is accompanied by the subjective conjugation. The semantic recipient, the president, is marked with the lative case.

(11) Ӭрыг   оигпам   юи-пāлт   Александр   Президент-ын   хум
     Ērəγ   *oiγpam*   *jui-pālt*   *Aléksandr*   *prezidént-ən*   *χum*
     song   end   after   Aleksandr   president-LAT   man

     хōтпа   ӭнтап   мӯйлупта-с.
     χōtpa   *ēntap*   *mujlupta-s.*
     person   belt   give.a.present-PST.3SG

'After the song was sung, Aleksandr gave a male belt to the president.'
(LS 21/2018: 2)

To sum up, Table 1 gathers together correlations between different sentence
structures, semantic roles and syntactic functions (for more details, see
Virtanen 2015: 53–58). On the horizontal bar we can see the active and pas-
sive 1-participant (intransitive), 2-participant (monotransitive) and 3-par-
ticipant (ditransitive) structures. The vertical bar is for the semantic func-
tions, and the chosen syntactic functions are marked as matches between
the sentence structures and the semantic functions. These are not absolute
correlations, but on a large scale, these are reliable prototype situations.

Table 1: Summary of correlation between semantic and syntactic
functions in different syntactic structures

	Active 1-P	Passive 1-P	Active 2-P	Passive 2-P	Active 3-P Sec.	Active 3-P Indir.	Passive 3-P Indir.	Passive 3-P Sec.
Agent	Subject	–		Subject	Oblique (LAT)	Subject	Subject	Oblique (LAT)	Oblique (LAT)
Theme	–	Subject	DO	Subject	IO (INSTR)	DO	Subject	IO (INSTR)
Recip-ient	–	–	–	–	DO	IO (LAT)	IO (LAT)	Subject

As the table above shows, each semantic role is realized as multiple syntac-
tic functions. How does this variation influence the linear order of syntactic
functions in an utterance? The variation between different syntactic struc-
tures described in examples (1–11) above is dependent on the correlation
between syntactic and pragmatic functions, while the constituent order is
primarily in connection with syntactic and pragmatic functions of indi-
vidual arguments. As described in the following sections, the constituent
order is not dependent on the chosen syntactic structure: the same linear

order can be applied to any structure (e.g. active, passive, 2-participant and 3-participant). In Section 5, I demonstrate the pragmatic constituent order variation I found, and how it works in different syntactic solutions.

3. Word order and constituent order as a typological question

In this section, I briefly present some general features and tendencies of word-order typology, and then proceed to the pragmatic approaches (Section 3.1). Finally, I discuss the phenomenon of the syntactic template as a device for describing constituent-order variation (Section 3.2).

3.1. Perspectives on word-order typology

Word order can be examined from several perspectives, which are situation-dependent. In his word-order handbook, Song (2012: 3–4) sees four approaches as relevant to linguistics today: linguistic typology, generative grammar, optimality theory and performance-based theories. From the point of view of this study, optimality theory and performance-based theories are the most relevant. It is also worth mentioning linguistic typology, which was used also in the seminal typological analysis by Greenberg (1963). Greenberg's approach, the first attempt ever to create a comprehensive word-order typology, concentrates on *basic word order*, i.e. the order of a stylistically neutral, independent, indicative clause with full noun-phrase participants; it is a prototypical transitive clause (Siewierska 1988: 8). Basic word order is normally approached by examining the division of the syntactic core arguments Subject (S), Direct Object (O) and Verb (V). The frequency hierarchy of the six basic word-order patterns is SOV > SVO > VSO > VOS > OVS > OSV (Song 2012: 25).

Optimality theory has been derived from generative grammar and, as the name implies, aims at achieving optimal output selection in word order (see e.g. Costa 1997; 1998; 2001; Zepter 2003). In the terms of the theory, an *input* can be realized as different *outputs*, and violable constraints affect the decision between possible outputs. The aim is to find the optimal satisfaction within conflicting constraints. There are no language-specific restrictions on the input, only on the output. Similarly to linguistic typology, optimality theory is very focused on basic word order.

Within the performance-based approach, Song mentions the Principle of Early Immediate Constituent Theory. In this view, different word

orders reflect the way languages respond to the demands of rapid and efficient processing in real time (Hawkins 1994: 57; Song 2012: 237). Briefly, language use is social interaction, and it should be approached from the perspective of how the hearer is receiving it. The hearer's processing is maximized when the immediate constituents of a given syntactic domain are recognized as rapidly and as early as possible (Song 2012: 259). Song criticizes Hawkins for focusing on processing rather than producing language (Song 2012: 259–260). Considering both opinions, Hawkins's theory has something in common with information-based theories like the one used here; information-structure research deals with the question of what the hearer is expected to know and understand.

While Song intends to be comprehensive, he wholly ignores information-based theories, only briefly referring to some approaches. He justifies his choice by stating that information-based theories vary widely, that – referring to Hawkins (1994) – they do not play a remarkable role in word-order-related research, and they fail to address the issue of grammaticalized word orders and correlations (Song 2012: 6–7). Contesting these views, the pragmatic or information structural level is essential to the analysis in this study. The Mansi basic word order has been proven to be SOV (see Section 1). My analysis extends to variation beyond the basic word order. *Word order flexibility* is possible, as basic word order patterns in an individual language vary considerably (Siewierska 1988: 10–11). It is important to note that in the research material used here, the different placement possibilities of the other arguments (besides S, O and V) may also vary due to pragmatics.

Siewierska examines word-order variation using linearization hierarchies. Referring to Allan (1987), she classifies them into three groups: formal hierarchies (e.g. length, structural simplicity or complexity), dominance hierarchies (e.g. personal hierarchies, semantic roles) and familiarity hierarchies (Siewierska 1988: 29–103). The familiarity hierarchies include the *more familiar > less familiar hierarchy*, the *topic > comment hierarchy* and the *given > new hierarchy* (Siewierska 1988: 61–75). She also mentions definiteness and referentiality hierarchies. In the broadest sense, all of the above are connected to information structure. Giving examples from various languages, Siewierska demonstrates how the word order of a single language can be affected by familiarity, topicality and the givenness or definiteness of the referents. Whether any familiarity hierarchy affects the word order or not is language-dependent: this is not possible in languages with a fixed word order. Downing (1995: 15–16) discusses the question of

topicality and word order: using the traditional terms *theme* and *rheme* she grounds the need of theme-initial utterances with the natural need to present the thematic, known information first, and to share the most unpredictable information at clause-final position. As I show in Sections 5 and 6, my Mansi data supports Downing's statement.

The difference between languages with fixed and pragmatically conditioned word order can be described using the opposition of *configurational* vs. *non-configurational* languages (see e.g. Hale 1983; Baker 2003). In configurational languages, syntactic functions occupy fixed places within the sentence structure, while in non-configurational languages the placement of a single constituent is dependent on other factors. As my data show, Mansi is neither a purely configurational nor a non-configurational language: some syntactic functions have fixed positions inside the syntax while others are placed according to their pragmatic functions. Van der Wal (2009) has come up with a similar conception: languages are not purely configurational or non-configurational, but rather they lie on a continuum of different features. Van der Wal (2009: 134) also questions the whole phenomenon of non-configurational language:

> […] the term 'non-configurational' does not seem appropriate to refer to languages. There are striking differences between languages in terms of word order and constructions, so the question is: what determines the configuration of sentences in a language? For the 'configurational' languages, the most influential factor is the syntactic functions and argument relations.

How should we refer to a language which is not a configurational one? É. Kiss (1995) describes Hungarian – a close relative to Mansi – with the term *discourse configurational* language: in a discourse configurational language pragmatic functions occupy fixed places within the syntax. Similarly Vilkuna (1989; 1995) has described Finnish word order as a *discourse conditioned* system. She defines the three main parts of sentence structure as the K, T and V fields (Vilkuna 1989: 37–40): T is for the function of a constituent that immediately precedes the finite verb in a textually neutral sentence, K is for the function of a constituent that precedes T, and the V field is the part of sentence that follows T. Despite the slightly complicated definitions, in practice K, T and V refer to pragmatic functions that occupy fixed places within the linear order: they function as pragmatic domains and are not directly connected to syntactic functions.

É. Kiss (1995: 6) defines two properties of discourse configurational language: 1) the discourse-semantic function *topic* is expressed through a particular structural relation; 2) the discourse-semantic function *focus* is expressed through a particular structural relation. In Mansi, the first property means that the syntactic function of subject – with some natural exceptions – correlates with the most topical element of the sentence: variation between the active and the passive voice is due to correlation between syntactic functions and pragmatic status (see examples in Section 2). The second property is realized by placing focal elements in syntactic non-core functions, and as a natural consequence of the previous statement, often marking them with oblique cases. (See e.g. Skribnik 2001; Virtanen 2016.) However, these conditions are not fully met in Mansi, for as Skribnik (2004) describes, also subject foci are possible in particular conditions.

At this point it is worth mentioning that in connection with (di)transitive structures, the variation between constituent order has also been examined from the point of view of valency. It has been proven that, for example, in German the obligatory adverbials (in terms of the valency) follow the non-obligatory ones in the surface word order (see Flämig 1991). Valency factor is excluded from this current study, because the following template analysis proves very clearly that the adverbials are placed according to their functions (time, place and manner). Combining valency with this analysis would likely not bring any valuable further results. This does naturally not exclude the possibility of examining the effects of valency in further studies.

3.2. Syntactic template as a device for describing constituent-order variation

Observing Mansi data easily raises the question whether the language can be defined as purely configurational or non-configurational at all, since it demonstrates features of both. The closest definition would be discourse configurational, but this still needs elaboration. In the following sections I will demonstrate how Northern Mansi constituent order is dependent, on the one hand, on syntactic functions, and on the other, on the order of pragmatic status. This complex two-level system is described by template analysis. A template approach can be used to observe the variation in constituent order from several perspectives: this way a partly configurational or discourse configurational language can be described on both the syntactic and pragmatic levels simultaneously. This shows that what we

call variation above, actually only varies in terms of syntactic functions, because from the interdimensional point of view, constituent order is very consistent. My main aim here is to describe the factors affecting Mansi constituent order as fully as possible.

The term "template" is not easy to define, because it includes various kinds of analysis of different levels of language structure, and it is not meant to refer to a strictly delimited device. Good (2016: 7) defines a template as follows: "An analytical device used to characterize the linear realization of a linguistic constituent whose linear stipulations are unexpected from the point of view of a given linguist's approach to linguistic analysis." However, Good (2016: 22–23) points out that the citation above is a descriptive definition, not a prescriptive one. Further, Good amplifies his definition by specifying that when greater terminological precision is needed, the term *desmeme* should be applied to emphasize merely linear stipulation, and to avoid the problematic phenomenon of *unexpectedness* (Good 2016: 23). In other words, template analysis can be adapted to situations with expected stipulations as well.

A template can be used for describing morphological, morphophonological or – as is done here – syntactic linearizations. Morphophonological and morphosyntactic templates have a prominent place in the literature (Good 2016: 66), but that does not exclude the use of syntactic templates either. According to Good (2016: 17–18), templates have been successfully applied to German syntax (Höhle 1986; Kathol 1995; 2000; van Riemsdijk 2002) and Dutch syntax (Shannon 2000). Interestingly, the work of Vilkuna (1989; see Section 3.1) on Finnish syntax can also be regarded as a template analysis, as the three fields she applies to constituent order can be understood in the same way as the slots in my analysis (see Section 5). Mansi and Finnish are distantly related languages, and while they are very different, they share some common features.

The vast diversity of environments where templates are applied are well reflected in the unsettledness of the terminology and practices used. The aim of a template analysis is to identify and define the factors that influence the linear order of components, and in the following, these linear positions are called *slots*. There is no single correct way of doing template analyses. As Good humorously expresses it, templates are "a twice incoherent class of phenomena" (Good 2016: 22) or a "wastebasket" (Good 2016: 27). Though many kinds of approaches are taken, a common feature is that the final result is a linear representation of the components involved.

Good (2016: 40–103) makes an effort to provide a settled template terminology, which he calls a *description language for templates*. He describes in detail how the elements and features involved in the analysis should be named, and how the terminology varies. He starts by defining basic phenomena like *stricture*, *foundation*, *desmeme* and *component*, and describing their features. Every template analysis practically includes these four elements, even if they are not named so. According to Good's (2016: 53–54) terminology, *component* is an immediate subconstituent of a given templatic construction: the analysis normally discusses the occupancy of the components. In this study, components are syntactic (phrasal) constituents that occupy numbered slots inside the template. *Stricture* is a feature that classifies the nature of the linearization specifications in a given template: it is either the length or the order of the components (Good 2016: 66). In this study, the stricture is simply order: the main question concerns the linear order of the syntactic constituents. *Foundation* is a feature describing how components of a template are organized into an overall templatic form; the foundation can be either *span* or *arch* type. A span foundation includes *left-support components* and *right-support components*, and the remaining components (*Restkomponenten*) are placed between them. An arch foundation is built around a *keystone*, which is the topmost component at the centre (head) of the template. The components directly adjacent to the keystone are called *voussoirs*, and they are either left-support or right-support – depending on which side of the keystone they are placed. (See Good 2016: 75–77.)

As can be seen in Section 5, this current template analysis represents the span model with a right-support verb and a left-support primary topic. In other words, both syntactic and pragmatic functions are involved simultaneously. Finally, *desmeme* is used as a synonym for *template*: it refers to the whole linear pattern analysed (Good 2016: 65). The desmeme is here represented by a clause. Before proceeding, it is worth noting that this template analysis does not exactly adhere to any kind of "Good's model", although it can be described in the framework of Good's terminology. For the needs of this study, in the light of recent research and the nature of my data, the template analysis had to be applied to the pragmatic level. This kind of template analysis enables accurate and interdimensional observation of the linear order of constituents when the order is affected by both syntactic and pragmatic factors. This model results in partly very similar descriptions as Rizzi (1997) or É. Kiss (1995). Rizzi and É. Kiss, however,

focus on realization of the main pragmatic functions, while my template analysis gives an interdimensional description of the effects of both pragmatic and syntactic factors. Placement possibilities of individual functions are discussed as a part of a larger entity. The details of implementation of my template analysis are presented in Section 5.

4. Research data and implementation of the analysis

In this section, I will first briefly present the source and content of my research data, and then describe the classification of the data and implementation of data classification and the analysis.

4.1. Research data

My data are gathered from *Лӯимā сӭрипос* [Lūimā Sēripos], the only up-to-date Mansi newspaper, published twice a month in Khanty-Mansiysk. An average issue contains 15–20 pages. Although the data is from one single source, the genres published within it vary, e.g. standard news articles, interviews, letters from readers and folklore texts. Altogether 12 articles or other texts are included from the 2014, 2017 and 2018 volumes of the newspaper, a total of 676 entries. The 12 articles were chosen so that different genres, variable topics and different authors are represented, but also so that the length of an individual article is between 20 and 150 clausal entries.

    This source was chosen mainly due to its contemporary nature and sufficient variety. The sources for written Mansi are limited, because the language is used only in restricted circumstances. The easiest way to include folklore texts was to gather them from the same source as all other data, because the same Cyrillic orthography is applied. The data naturally includes various sentence types, such as declaratives, questions and imperatives, which vary according to the genre of the text. When something in the analysis is connected to the sentence type, this is clearly stated. Most often the same model can be applied to any sentence type, as the next section will show.

4.2. Implementation of the analysis

My template analysis was implemented in several phases, where the results of one phase led to decisions concerning the next one. The data have been first divided into clauses, and one clause provides one entry. The final result was a 9+1 slot model, which I found suitable for describing the variation of syntactic order. This analysis was not implemented in pre-planned steps, rather finishing one step always led to decisions concerning the next one. Afterwards, I named the phases as follows:

1. Providing desmemes (chains of constituents) out of clauses.
2. Creating a table with desmemes on horizontal lines.
3. Adjusting the components (constituents) and the slots (vertical columns).
4. Deciding on the number of slots and naming them. Adding a pragmatic level to the analysis.
5. Turning the template to a vertical position.
6. Analysing variation in placement of single syntactic functions and deepening the pragmatic level of the analysis.

In the fifth phase, the template was turned to a vertical position, as presented in Table 2. Beside the syntactic analysis presented above, also a preliminary pragmatic analysis was implemented: the individual referents were analysed according to their appearance in the whole text, in the same paragraph, in previous sentences and in the immediately previous sentence. In this way they were defined as primary topics, secondary topics, scene-setters and foci or pragmatically neutral arguments. Also their possible inherent topicality was examined. This analysis resulted in preliminary remarks on how the topicality or focality of an individual argument influences its placement inside the linear order. Still, the real analysis was finished in the sixth phase, when the results and conclusions were combined in the form of a pragmatic template (see Table 4 in Section 6).

The preliminary template (see Table 2) has 9+1 slots and includes alternative slots: for example, subject varies for pragmatic reasons between two slots (1 and 7), and both are marked in the table. Already in this phase, some rearrangement was carried out to combine the pragmatic analysis with the syntactic one. For example, focal DOs were moved from Slot 6 to Slot 8, in case Slots 7 and 8 were not occupied yet. The first slot of this model is marked 0, because it is occupied quite rarely and only by elements that do not affect the order of the other constituents.

Table 2: Ten slots filled by syntactic or pragmatic functions. The template is refined in Section 6.

0	Scene-setting adverbials	
1	Subj-A	topical subject
2	Temp-1	temporal adverbials
3	Loc-1	locational adverbials
4	Manner	manner adverbials
5	Loc/Temp-2	in cases of several temporal or locational adverbials in one sentence
6	D-slot	neutral DOs, IOs and directionals[a]
7	Subj-B	in cases when a scene-setting adverbial occupies Slot 2, 3 or 4
8	Focus slot	the most focal element: often represented by a syntactic DO, IO, predicative, negation particle or infinitive
9	Verb slot	verb or nominal predicate

a. For the definition of directionals or directional adverbials, see Section 5.3.

Before proceeding, note that the template is refined in the sixth phase of the analysis (see Section 6). The interdimensional nature of the analysis makes the description complicated. Because of the pragmatic variation described in Section 2, one syntactic function often correlates with particular pragmatic functions. There are syntactic functions that always occupy the same linear positions, with minor exceptions, but also pragmatic functions that occupy a certain position independently of the syntactic functions they represent: mapping out these occupancies is one of the main aims of this article. There are very clear and visible constituent-order hierarchies between syntactic functions. Some functions occupy their own slots alone; some slots are shared by several functions. If there is a function missing from an individual sentence structure, the slot in question remains unoccupied.

In the following Section 5, this approach is presented function by function, not in the numerical order of the slots but starting with the syntactic core arguments and their placement, and proceeding towards the less prominent constituents. Finally, an elaborated template model (Phase 6) is presented in Section 6.

5. Linear placement of syntactic functions in my data

In this section, I present the core findings of my analysis is detail, taking into account both the syntactic and pragmatic level simultaneously. Subsection 5.1 is devoted to the clause-final placement of verbs. In Section 5.2, the placement of subject and its variation are discussed in detail. Indirect and direct objects are discussed in Section 5.3, and the adverbials of locations, time and manner in Section 5.4. Finally, the placement of agent adverbials, nominal predicates and infinitives is presented in Section 5.5, and some questions concerning interrogatives in Section 5.6.

In the following subsections, the data examples are preceded by a template description: above each original data sample, the syntactic functions are listed in the right linear order, accompanied by their slot numbers. For example, a manner adverbial in Slot 4 is labelled "Man:4", a predicate verb in Slot 9 is labelled "V:9", etc.

5.1. On verb-finality

According to Rombandeeva (1984: 58–60), the strongest or most stable feature of Northern Mansi word order is that the verb always occupies the absolute clause-final position. My data support Rombandeeva's statement: 624 examples out of 676 are verb-final. This means that 92% of the clauses are verb-final, but not all include a predicate verb; of those clauses which do include a predicate verb, 99% are verb-final. 99% of all clauses without a predicate verb include a nominal predicate placed at slot 9 (see Section 5.5). A verb can be only followed by a conjunction or – due to recent Russian influence – an infinitive.

Examples (12) and (13) are typical verb-final sentences that represent the strong majority in my data. Independently of the pragmatic context, the verb occupies the sentence-final position in both the active (12) and the passive (13).

Man:4–V:9

(12) Кӣсыӈ хōтпа ӣр-нэ тӣла-т урыл потырт-ас.

Kāsəŋ χōtpa ēr-ne tēla-t urəl potərt-as.
every   person concern-PRS.PTCP thing-PL about speak-PST.3SG
'She talked about things that concern all of us.' (LS 1/2018: 6)

SP:1–Adv:8–V:9

(13) Йильпи   кол   тай   са̄в   ӯнтт-аве.

Jiľpi   kol   taj   sāw   ūntt-awe.
new   house   PTCL many build-PASS.3SG
'A lot of new houses are being built.' (LS 1/2018: 5)

Example (14) illustrates another default situation. The predicate is followed by a conjunction but is still regarded as sentence-final. In Table 2 the corresponding slots are Slot 9, and a conjunction following it.

SA:1–V:9–Conj

(14) Округ-т   о̄л-нэ   мир,   касащ-ēг-ыт те […]

Okrug-t   ōl-ne   mir,   kasaś-ēγ-ət te […]
district-LOC live-PRS.PTCP people want-PRS-3PL if
'If the people living in our district want […]' (LS 1/2018: 6)

To sum up, the verb or a nominal predicate occupies the verb-final position and, in any clause type (for nominal predicates, see Section 5.5), it may only be followed by conjunctions or infinitives. Complex predicates are briefly discussed in Section 5.5. There are only minor exceptions to this rule in my data, and these exceptions are clearly due to Russian influence and not discussed here.

5.2. Placement of the subject

One of the most regular alternation types in my data is the variation in the placement of the subject in active and passive clauses (SA and SP) between two stable slots. Rombandeeva (1979; 1984) mentions that the subject precedes the predicate, and my data support this statement with minor exceptions.[7] In my analysis, I reserved two slots (1 and 7) for subjects. These two slots are symmetrically occupied by subjects – not simultaneously but in either one or the other. Slot 1 is called *Subj-A*, and Slot 7, *Subj-B* in my analysis. Analysing corpus data, however, is not so simple. For example, if we have a clause with only a subject and a verb, or with a subject, a DO

---

7. In some special cases the syntactic Subject is placed just after the Verb. These cases are not relevant from the point of view of this study, and they are not discussed here.

following it, and a verb, do we have a Subj-A or a Subj-B? In other words, if the subject is both clause-initial and in the second to last slot before the predicate, which should be regarded as the primary choice? Cases like this need some further elaboration. In this study they are regarded as sentence-initial (Subj-A), because – as shown in the following – Subj-B is connected to particular pragmatic circumstances, which do not occur with sentences including only a subject, DO and predicate.

As demonstrated in Section 2, the primary topic of the utterance occupies the syntactic function of subject, and this causes variation between different syntactic structures. At the same time, the most topical elements are placed in the sentence-initial position. As a result of these two tendencies together, the syntactic subject is often in the sentence-initial position. If the SA/SP is not following a scene-setting local, directional or temporal adverbial, it occupies the sentence-initial position. In my data, 63% of the sample entries include an explicitly marked syntactic SA or SP. In 71% of the sentences with a subject, the SA/SP argument occupies the sentence-initial slot. This number includes also the cases where the SA/SP argument is in Slot 1 and all slots from slots 2 to 7 are empty. In 29% of these cases, Slot 7 is occupied by the SA/SP. (See Appendix.)

Variation between Subj-A and Subj-B in both active and passive is briefly demonstrated in examples (15–18). In the first two, the SA or SP is placed in sentence-initial position. In both (15) and (16) the syntactic subject is the primary topic of the sentence. Example (15) is from a folklore tale about a Mansi man called Zakhar and what happened to him one winter. The third-person singular subject is not only the primary topic of this sentence, but also the narrative topic of the story.

SA:1–Dir:6–V:9

(15) Сāхарка   вōр-н   мина-с.
      Sāχarka   *wōr-n*   *mina-s.*
      Zakhar   forest-LAT go-PST.3SG
      'Zakhar went to the forest.' (LS 15/2017)

Example (16) is from a news article about children who were taken to the forest to participate in traditional activities. The children are the narrative topic and also the primary topic of this sentence. In the sentence it is stated that before doing hunting exercises they were given bows and arrows. Arrows and bows are the focal elements of the clause. The children, i.e. the

semantic recipient of the clause, represent the primary topic and for that reason it occupies the syntactic function of subject. A passive structure is chosen:

SP:1–IO:8–V:9

(16) Тāн ё̄вт ос ло̄мт-ыл май-вē-с-ыт.

     Tān jōwt is lōmt-əl maj-wē-s-ət.
     3PL bow and arrow-INS give-PASS-PST-3PL

     'They were given bows and arrows.' (LS 21/2018: 4)

Subjects preceded by a conjunction are regarded here as SA/SP-initial clauses. A conjunction can be placed in several positions, but it does not influence the order of the other constituents. In example (17) we have a subordinate clause beginning with a conjunction (Slot 0), followed by the SA/SP (Slot 1).

Conj:0–SA:1–Inf+Neg:8[8] –V:9

(17) Хоты хо̄тпа-т ёхтал-аӈкве ат вēрм-ыс-ыт […]

     Xoti χōtpa-t joχtal-aŋkwe at wērm-əs-ət […]
     if person-PL arrive-INF NEG be.able-PST-3PL

     'In case people are not able to come [to the health centre]…'
     (LS 20/2018: 4)

In examples (18–20) a temporal or locational expression occupies the sentence-initial position, a locational expression follows it, and the syntactic subject comes just after these (Slot 7). In other words, the order of other constituents does not change, but the subject is postposed at Slot 7, just after the adverbials. The subject is included in an utterance describing something that happened during the mentioned period of time or at a particular location.

This kind of adverbial expression is called a scene-setting adverbial (see Section 1). The nature of scene-setting adverbials in different languages has been discussed in the literature. An important observation from the

---

8. In this template, the placement of several components of a complex predicate may overlap in Slot 8. In this case, both an infinitive and a negation particle are placed in the same slot. Consequently, both are calculated in the total number of occupancies. This exceptional arrangement concerns only parts of complex predicates – for more details see Section 5.5.

point of view of this study is that scene-setting adverbials are placed within the topic field (see e.g. Benincà & Poletto 2004). Further, for example Rizzi (2004) has stated that they are not part of the topic, as they do not correspond to the definition of topic, but rather they occupy an independent position. As proven by my data, scene-setting adverbials share the sentence-initial position with the topic, but there is no need to call them topics. The subject does not lose its (inherent) topicality as in (18), but as a scene-setting element, an adverbial occupies the sentence-initial position, and the subject occupies Slot 7.

Temp:2–Loc:3–SA:7–V:9

(18) Тӑл   сыс   округ   янытыл   ам   сӑв   мӑ-т
     Tāl   sis   okrug   janitəl   am   sāw   mā-t
     year   during   district around   1SG   a.lot.of land-PL

   яласа-с-ум […]
   jalasa-s-um […]
   visit-PST-1SG

   'During this year I have visited a lot of places around the whole district […]' (LS 1/2018: 6)

In example (19), the temporal adverbial appears as a scene-setting adverbial. The clause is from an article about a communal house-building project. In the previous sentences it has been mentioned that new houses are being built, and also the Sartyn'ya village has been mentioned among the involved areas. However, the temporal expression "this year", which is actually new information, is placed at sentence-initial position, because it sets a temporal framework within which the main predication holds.

Temp:2–Loc:3–SP:7–V:9

(19) Ты   тӑл   Сортыҥъя-т   кит   йильпи   кол-ыг
     Ti   tāl   Sortiŋja-t   kit   jilʹpi   kol-əγ
     this   year   Sartyn'ya-LOC two   new   house-DU

   ӯнтту-ве-с-ӣг […]
   ūnttu-we-s-əγ […]
   build-PASS-PST-3DU

   'This year, two brand-new houses were built in Sartyn'ya […]'
   (LS 1/2018)

Further, as stated above, in some rare cases the scene-setting adverbial is placed at Slot 0 in my template. In those cases the subject is placed at Slot 1, other adverbials follow it, but the scene-setting adverbial precedes it. In example (20) the subject follows the temporal adverbial, and other constituents occupy their normal slots.

Temp:0–S:1–Dir:6–Inf:8–V:9

(20) Аквта   тāл   тав   Āс+угорский   театр-ын   рӯпит-аӈкве
      Akwta   *tāl*   *taw*   *Ās+ugorskij*   *teatr-ən*   *rūpit-aŋkwe*
      same   year   3SG   Ob+Ugrian   theatre-LAT work-INF

патхата-с.
patχata-s.
start-PST.3SG

'During the same year, he started to work for the Ob-Ugrian theatre.' (LS 21/2018: 2)

Example (21) shows a typical polar question without a question word, provided only with intonation. The locational expression 'at home' is not a topical element, while the second-person singular pronoun and 'mother tongue' are in this context, but as a scene-setting adverbial it occupies the sentence-initial position:

Loc:0–SA:1–Man:4–V:9

(21) Юн   нāн   щāнь   лāтӈ-ыл   потырт-эг-ӣн?
      Jun   *nān*   *śāń*   *lātŋ-əl*   *potərt-eɣ-ān?*
      at.home   2SG   own   language-INS speak-PRS-2SG
      'Do you speak your mother tongue at home?' (LS 22/2019: 9)

Finally, it is relevant to ask whether SAs and SPs behave in the same way. In other words, is there any difference between active and passive clauses? My data include 332 active entries with an explicitly expressed subject, and 89 passive entries with an explicitly expressed subject. Of these, 74% of the active subjects are placed in Slot 1, while the corresponding figure for the passive is 63%. The difference of eleven percentage points is perhaps not remarkable in this context, but possible differences between subjects of active and passive clauses remain a question for further studies.

5.3. Placement of DOs, IOs and directional adverbials

In this subsection I will discuss the linear placement of direct and indirect objects and directional adverbials. As described in Sections 1 and 4.2, indirect objects include two morphosyntactic forms: semantic recipients of ditransitive clauses marked with the lative case, and semantic themes of ditransitive clauses marked with the instrumental. The category of directional adverbials includes both lative-marked (semantic Goal) and ablative-marked (semantic Source) expressions of direction. These functions share many common functional features: DOs and instrumental-marked indirect objects (IOs) represent the same semantic role, the patient or theme. Directional adverbials – ablative- and lative-inflected directional nouns – are very close to lative-marked IOs, also because half of them are marked with the same morphological case, the lative. The ablative-marked directionals include also possessive adverbials, e.g. human referents appearing in the semantic role of source. My data prove that the syntactic behaviour of directional adverbials is closer to that of IOs than, for example, locational adverbials.

Moreover, these syntactic functions share many features concerning their placement inside the syntactic template. Slot 6 is here called the D-slot; the letter D refers to the mutual initial letter of two of the three functions, and the very adjacent phenomenon of "ditransitive". This slot is mainly occupied by lative-marked IOs, directional adverbials and DOs. Furthermore, the placement of these *D-arguments* is mainly divided between slots 6 and 8. Slot 4 can also be occupied by a D-argument, but much less frequently. The division is very clear: when the arguments in question are focal, they are placed in Slot 8, while in other cases they are placed in Slot 6. Many of the D-arguments placed in Slot 6 are also secondary topics (see Section 2) of the corresponding sentences, especially DOs.

For example, Skribnik (2001: 223) and Bíró (2015: 55) have already shown that the focal argument is always placed immediately before the predicate. In my analysis this means Slot 8, and my data strongly support their observations. Placing the most focal argument in Slot 8 explains a lot about the variation between placements of D-arguments (and of other arguments presented in Sections 5.5 and 5.6). This is demonstrated by the following examples (22–30).

Example (22) is from a story about a man and his family. His life changed when his father was called up to the army. The expression 'call up to the army' is mentioned for the first time here and is the focus of the sentence, so the directional expression occupies the focus slot:

SP:1–Temp:2–Dir:8–V:9

(22) Āщ-е   Алексей Данилович Яркин 1941 тāл-т

Āś-e   *Aleksej Danilowitš Jarkin*   *1941* *tāl-t*

farther-POSS.SG<3SG Aleksey Danilovich Yarkin 1941 year-LOC

армия-н   тот-ве-с.

armija-n *tot-we-s.*

army-LAT bring-PASS-PST.3SG

'In 1941, his father Aleksey Danilovich Yarkin was called up to the army.' (LS 9/2014: 13)

In (23), both the first-person and third-person singular referents are inherently topical elements, and the new information is that the speaker is sending greetings. The topical (secondary topic) IO is placed in Slot 6, and the focal DO in Slot 8.

SA:1–IO:6–DO:8–V:9

(23) Ам тав-ēн паща   лāтыӈ кēт-ӗг-ум.

Am *taw-ēn* *paśa*   *lātəŋ* *kēt-ēɣ-um.*

1SG 3SG-LAT greeting speech send-PRS-1SG

'I send him my greetings.' (LS 9/2014: 13)

As mentioned earlier, a nominal DO normally follows the subject and precedes the predicate, but its placement is not restricted to the slot immediately before the verb. A DO can also be followed by IOs or temporal, locational or other adverbials, if any of them is in a focus position. In my data, 50% of all DOs are placed in Slot 6 (D-slot), 45% in Slot 8 and the rest – more marginal cases that are not discussed here – in other slots. This variation can be explained by pragmatic reasons. Those DOs appearing in Slot 6 are also pragmatically classified as secondary topics (see Section 2). 99% of all DOs placed in Slot 6 are also accompanied by the objective conjugation. Correspondingly, those DOs placed in Slot 8 are pragmatically classified as foci, and they are accompanied by the subjective conjugation.

The following three examples (24–26) include a DO immediately preceding the verb. As can be seen from the context, in all of them the DO is the most focal argument of the clause. For that reason these DOs are also accompanied by the subjective conjugation (for comparison, see 27–28). In (24), the speaker has spoken about her career in the immediately preceding sentences, and she mentions the report for the first time. The focus of the sentence is what she is doing at the time:

Temp:2–DO:8–V:9

(24) Ань доклад+нэ̄пак щё̄пит-э̄г-ум.

Ań doklad+nēpak śōpit-ēγ-um.
now report+book prepare-PRS-1SG

'Currently I'm writing a report.' (LS 1/2018: 6)

Example (25) describes the situation in a village. In the previous sentences it has been explained what other things villagers do for a living, and here fishing is added to the list:

SA:1–DO:8–V:9

(25) Ма̄хум хӯл алыщл-э̄г-ыт.

Māχum χūl aliśl-ēγ-ət.
people fish catch-PRS-3PL

'People go fishing.' (LS 9/2014: 13)

Example (26) is from a story about an old man's life: first it is told that he lived in a sovkhoz and worked as a fisherman. This sentence is about the next step: later on he did many kinds of work in the same sovkhoz.

Man:2–Loc:3–DO:8–V:9

(26) Ты коныпал совхоз-т са̄всыр рӯпата ва̄р-ыс.

Ti konipal sowχoz-t sāwsər rūpata wār-əs.
this in.addition sovkhoz-LOC many.kinds work do-PST.3SG

'Besides that, he did many kinds of work at the sovkhoz.' (LS 9/2014: 13)

In both (27) and (28), the DO is placed in the sixth slot. The DO is the secondary topic of the clause (see Section 2) and accompanied by the objective conjugation. In (27) the DO is followed by a manner adverbial, and in (28) by a temporal adverbial. In (27) a speech has been discussed earlier, and it is mentioned for the first time that it was in Mansi (not in Russian):

SA:1–DO:6–Man:8–V:9

(27) Э̄ква потр-е ма̄ньщи ла̄тӈ-ыл

Ēkwa potr-e māńśi lātŋ-əl
old.lady speech-POSS.SG<3SG Mansi language-INS

ханс-ыс-тэ.

χans-əs-te.
write-PST-SG<3SG

'The lady had written her speech in the Mansi language.' (LS 9/2014: 13)

In (28), the man has already seen a reindeer watching him, and the reindeer tells the man that he has been following him for a while. The man is the secondary topic and accompanied by the objective conjugation. The temporal expression is the focal element of the clause:

SA:1–DO:6–Temp:8–V:9

(28) Ам   наӈын   хосат   тāгыл   вāг-лум   […]

Am   naŋən   χosat   tāγəl   wāγ-lum   […]
1SG   2SG.ACC long.time   completely   know-SG<1SG

'I know you very well, for a long time now […]' (LS 15/2017: 15)

Instrumental-marked IOs are placed immediately before the verb (Slot 8) in 99% of cases, because they tend to be focal. When a speaker chooses the right sentence structure for the situation (see Section 2), the focal patient is placed in the syntactic function of IO. In other words, the pragmatic variation between different three-participant constructions only allows a focal argument to occupy the syntactic function of IO.

Example (29) is from a news article about people with accommodation problems, stating that the poor family in question was given a special kind of house by the authorities.

SP:1–IO:8–V:9

(29) Тāн   тамле   кол-ыл   май-вē-с-ыт.

Tān   tamle   kol-əl   maj-wē-s-ət.
3PL   like.that   house-INS   give-PASS-PST-3PL

'They were given a house like that.' (LS 1/2018: 5)

In (30) the people in question have participated in a snow-sculpture competition, and the reader knows that they have won first prize. The new information is that they were given a diploma for their work:

SP:1–Man:4–IO:8–V:9

(30) Тāн   ты   хури   вāр-м-аныл   мāгсыл

Tān   ti   χuri   wār-m-anəl   *māγsəl*
3PL   this   sculpture   make-PTCP.PST-POSS.SG<3PL   for

янытлан   нēпак-ыл   мӯйлупта-вē-с-ыт.

janitlan   nēpak-əl   mujlupta-wē-s-ət.
honour   letter-INS   grant-PASS-PST-3PL

'They were awarded a diploma for making the [snow] sculpture.' (LS 1/2018: 16)

The following table describes the placements of DOs, directives, DOs and IOs. The table describes how the pragmatic variation between different sentence structures (see Section 2) affects the placement of individual syntactic functions. DOs and directional adverbials represent both pragmatic secondary topics and foci – and vary between the two slots correspondingly – while instrumental-marked IOs tend to be foci and occupy Slot 8. The placement of lative-marked IOs is very concentrated in Slot 6, and the most focal ones are placed at Slot 8.

Table 3: Placement of D-arguments

	6. (Neutral/Secondary Topic)	8. (Focus)	other
Directional adverbials (lative, ablative)	79%	13%	8%
DOs	50%	45%	5%
IOs (lative)	88%	12%	–
IOs (instrumental)	17%	83%	–

5.4. Placement of adverbials: time, location and manner

According to my data, the placement of adverbials is relatively regular and easy to explain. Temporal, manner and locational adverbials have their own slots, which they occupy with some exceptions due to pragmatics.[9] In short, Slot 2 is for time, Slot 3 for location and Slot 4 for manner. In the event that there is more than one adverbial of any type in one clause, the last one is placed at Slot 5. As described in Section 5.4, Slot 8 is for the most focal argument of the clause, and it can also be occupied by focal adverbials of time, location and manner.

The placement of temporal expressions is concentrated in Slot 2, which supports Rombandeeva's views (see Section 1). In 72% of cases in my data, temporal expressions are placed in Slot 2. Further, 13% of temporal expressions are placed in Slot 0 due to their scene-setting status (see Section 5.2), 11% in Slot 5, and 4% (the focal ones) placed in Slot 8.

Example (31) is from the beginning of a news text about the Russian president's visit to Khanty-Mansiysk. The president is the narrative topic and also the clausal primary topic. The temporal adverbial is in the default position.

---

9. Adverbials of location, manner and time appearing as scene-setting adverbials are discussed in Section 5.2.

SA:1–Temp:2–Man:4–Dir:6–V:9

(31) Яныг   кӯщай   э̄типāл-аг   хо̄т   ще̄с-т   товлыӈхāп-ыл
     Janiɣ   kūśai   ētipāl-aɣ   χōt   śōs-t   towləŋχāp-əl
     big   leader evening-TRANSL six   hour-LOC airplane-INS

     ӯс-ув-н   ёхт-ыс.
     ūs-uw-n   joχt-əs.
     town-POSS<1PL-LAT   arrive-PST.3SG

     'The president [of the Russian Federation] arrived in our town by
     plane at 6 o'clock in the evening.' (LS 21/2018: 2)

Example (32) is from the story about the old man, Ekur. The previous sen-
tences have explained where he studied and what he did during his studies.
The story continues here. Again, the temporal adverbial is in the default
position.

SA:1–Temp:2–Temp:5–Pred:8–V:9

(32) Екур Тобольский рыбтехникум   āстла-м-е
     Ekur Toboĺskij   ribteχnikum   āstla-m-e
     Ekur Tobolsk.ADJ college.of.fishery graduate-PST.PTCP-POSS.SG<3SG

     юипāлт   сāв   тāл   рыбак-ыг   рӯпита-с.
     juipālt   sāw   tāl   ribak-əɣ   rūpita-s.
     after   many   year   fisher-TRANSL work-PST.3SG

     'After graduating from the College of Fishery in Tobolsk, Ekur
     worked for many years as a fisherman.' (LS 9/2014: 13)

As already stated in 5.2, in some cases (22% of all entries with a scene-set-
ting adverbial) the scene-setting adverbial is placed at Slot 0 in my analysis.
In these examples, all the other arguments – including subjects – occupy
their regular slots, while the scene-setting adverbial is placed at the very
beginning of the clause. In other words, the subject is not placed at Slot
7: it appears at Slot 1 before the other adverbials, but following the scene-
setting one. In (33), the temporal expression is in a scene-setting function
and placed at Slot 0. Interestingly, the subject of the clause is not a topical
expression but belongs to an all-in-focus structure. The uncle of the man
referred to with a possessive suffix is mentioned for the first time. The fact
that he started to learn dancing and playing is new information as well.

Temp:0–S:1–Dir:6–V:9

(33) Та   пора-т   сасг-е   Вадим Важенин   ōс   та

Ta   pora-t   sasɣ-e   Wadim Waženin   ōs   ta

that time-LOC uncle-POSS.SG<3SG Vadim Vazhenin also that

ёнг-ын-кол-н   ёхт-ыс.

jong-ən-kol-n   joχt-əs.

play-PTCP.PRS-house-LAT come-PST.3SG

'By that time, also his uncle Vadim Vazhenin went to the same dancing school.' (LS 21/2018: 2)

Locational expressions[10] are placed in slots 3 (78%) and 5 (14%) – except the most focal ones, which are placed in Slot 8 (6%), and those representing scene-setting adverbials at Slot 0 (2%). In this phase of the analysis, I faced the question of how to distinguish an adverbial in Slot 3 from one in Slot 8, if there is only a subject, an adverbial and a verb in the sentence (see also Section 5.3). The question was solved in the final phase of the analysis by comparing the pragmatic analysis to the templatic order: if the adverbial is recognized as a focus in the pragmatic analysis, it is placed in Slot 8, and if not, it is placed in Slot 3.

In both (34) and (35) there is a local expression in Slot 3. Both sentences are drawn from articles about people's life stories: in (34) the local expression is a deictic pronoun, referring to a location mentioned in the previous sentence, and already familiar to the reader. The focus of the sentences is the temporal expression:

SA:1–Loc:3–Temp:5–V:9

(34) Тав   тōнт   хōт   тāл-э   ōл-ыс.

Taw   tōnt   χōt   tāl-e   ōl-əs.

3SG   there   six   year-POSS.SG<3SG live-PST.3SG

'He lived there for six years.' (LS 9/2014: 13)

In (35), the expression of location is the focus of the sentence and placed in Slot 8. The new information in the sentence is the adverbial answering the question about where the interviewed old man was born.

---

10. In this point my approach differs from Rombandeeva's: I discuss directional and locational expressions separately. Directional adverbials are discussed in Section 5.3, and they have more in common with dative adverbials.

SA:1–Loc:8–V:9

(35) Тав Хӑльӯс   район   Саранпӑвыл-т сам-ын[11]

Taw Xālŭs   rajon   Saranpāwəl-t sam-ən

3SG Beryozovo district Saranpaul-LOC eye-LAT

пат-ыс […]

pat-əs […]

become-PST.3SG

'He was born in Saranpaul in Beryozovo District […]' (LS 21/2018: 2)

In (36), there are two locational adverbials. The first was mentioned recently, but the second and more specific one is new information. The first local adverbial in Slot 3 refers to Beryozovo, where the man was said to have lived. The second is the focus of the sentence specifying the location or employer where he started to work in Beryozovo, building some new houses.

SA:1–Loc:3–Loc:8–V:9

(36) Павел Владимирович ты элы-пал Хӑльӯс   миркол-т

Pavel Wladimirovitš ti ēli-pal Xālŭs   mirkol-t

Pavel Vladimirovich this before Beryozovo municipality-LOC

кол-ыт   ӯнтт-ын   мӑ-т   рӯпита-с […]

kol-ət   ūntt-ən   mā-t   rūpita-s […]

house-PL build-PTCP.PRS land-LOC work-PST.3SG

'Before that Pavel Vladimirovich worked in Beryozovo in a communal house-building company […].' (LS 1/2018: 5)

In (37), the scene-setting locational adverbial occupies Slot 0, and the focal temporal adverbial follows it at Slot 2. There is no overt subject: as a highly topical element it is only referred to with a verb ending (zero anaphora). The name of the village (Southern Narykary) provides the framework within which the main predication holds, and the new information provided is the length of time the old man spent in the village.

Loc:0–Temp:2–Loc:8–V:9

(37) Алы   Нярихуми-т мат-нӗ-тэ   мус

Ali   Ńariχumi-t mat-nē-te   mus

Southern Narykary-LOC get.old-PRS.PTCP-POSS.SG<3SG until

---

11. *Самын патуукве* 'become visible' is a phrase meaning 'be born'. The lative-inflected noun is a part of the verb, not a separate constituent.

тохиӈ   тот   ōл-ыс.

toχiŋ   *tot*   *ōl-əs.*
this.way   there   live-PST.3SG

'He lived in Southern Narykary in the same way until he got old.'
(LS 9/2014: 13)

Manner adverbials are most frequently placed in Slot 4: 75% of the manner adverbials in my data are placed in Slot 4, 19% in Slot 8 due to their focus position, and 6% in Slot 0 due to their scene-setting function. Manner adverbials include adverbs (often identical with adjectives) and postpositional phrases. Example (38) is from a news text about a public event and represents a typical neutral declarative sentence. The manner adverbial is placed in its default position, before the local adverbial.

SA:1–Man:4–Loc:5–V:9

(38)   Нӣврам-ыт   щāгт-ым   тот   хӣйтыгт-ӣг-ыт   [...]

      Ńāwram-ət   *śāyt-əm*   *tot*   *χājtəyt-ēy-ət*   [...]
      child-PL   glad-ADV   there   run.around-PRS-3PL

      'Children were running around there happily [...].' (LS 1/2018: 16)

The order in (39) is exceptional in my data: the topical DO is placed in Slot 4, with the manner adverbial directly following it. This is rare from the perspective of syntactic order, but from a pragmatic point of view it is quite natural: from context we know that the horned head of the reindeer is an accessible argument, but the way he shakes it is not. In the immediately preceding sentences, we learn that the reindeer got new horns from the humans. Now the animal is proud of the horns: pride is the focal phenomenon.

SA:1–DO:4–Man:6–V:9

(39)   Сāлы   ōйка   āньт-ыӈ   поӈк-е   ёрыӈыщ   та

      Sāli   *ojka*   *āńt-əŋ*   *poŋk-e*   *jorəŋəś*   *ta*
      reindeer   male   horn-ADJ   head-POSS.SG<3SG   proudly   PTCL

      нёвумтапт-ыс-тэ.

      ńowumtapt-əs-te.
      shake-PST-SG<3SG

      'So the reindeer shook his horned head proudly.' (LS 15/2017: 15)

Finally, 17 samples in my data belong to the group "other adverbials". As mentioned above, "other adverbials" include, for example, adverbials of reason or condition, and they behave in the very same way as the adverbials of time, location or manner. Most of them function as scene-setters and are placed at Slot 0.

5.5. Placement of agent adverbials, negation particles, nominal predicates and infinitives

Quite a diverse group of functions share the same slot within the syntactic order: agent adverbials, negation particles and nominal predicates (including nouns, adjectives and participles), and the infinitives. Their syntactic functions are very systematically placed at Slot 8. However, there are two distinctive reasons for the appearance of these functions in the aforementioned slot. First, agent adverbials represent pragmatically focal arguments. This is a natural consequence of the fact that focal arguments tend to occupy syntactic non-core positions and occur inflected in oblique cases. If a semantic agent were topical, it would appear as a syntactic subject. In the event that a semantic agent is focal, it appears as a syntactic agent and is inflected in the lative case. Secondly, negation particles, infinitives and participles are closely related to the finite verb: they are parts of complex predicates[12] (see e.g. Alsina et al. 1997; Amberber et al. 2010).

The placement of infinitives is indicated in Appendix, but detailed further conclusions concerning them are not presented. The infinitive is not a syntactic function, but it is a syntactic phenomenon in many ways. In my data, 90% of the infinitive forms are placed in Slot 8: this is not because they all are focal but because they belong to complex predicates. Participle forms of verbs are not a separate group in the table in Appendix, because there are only a few of them, but they are included in the group of nominal predicates. It is still worth mentioning that those few occurrences are placed immediately before the verb.

As demonstrated in previous sections, Slot 8 is the default slot of the pragmatic function of focus. In addition, as described above, it is a default slot for elements that are parts of complex predicates. One of the most typical components in Slot 8 in my analysis is the negation particle (see

---

12. Complex predicates are predicates which are composed of more than one grammatical element (either morphemes or words), each of which contributes a nontrivial part of the information of the complex predicate (Alsina et al. 1997).

also Wagner-Nagy 2011: 80–81). Example (40) is from a newspaper article about a community-supported house-building project to help poor families. In the previous sentences, it has been told how many families and what amounts of money are involved. Now it is mentioned that the money is not paid in cash directly to the families.

SP:1–Dir:6–Neg:8–V:9

(40) Ты   олн-ыт   кāт-ын   ат   ми-ве-т.
      Ti   *oln-ət*   *kāt-ən*   *at*   *mi-we-t.*
      this   money-PL   hand-LAT   NEG   give-PASS-PRS.3PL
      'This money is not given in cash.' (LS 1/2018: 13)

Nominal predicates are predicative elements, or at least parts of complex predicates, but also inherently focal elements: their function is to give more information about the features of something which is already accessible. Mansi is a language where adjectival and some nominal predicates appear without copula. In my data, 73% of all nominal predicates are placed in Slot 8, as in (41): these ones are accompanied by a Verb. The further 15% are placed in Slot 9: these are nominal predicates without a copula (42). The remaining 12% appear in Slot 4.

SA:1–Pred:8–V:9

(41) Сāхарка   колтāгыл   мāхм-анэ   ёт   ōлупса-ныл
      Sāχarka   *koltāγəl*   *māχm-ane*   *jot*   *ōlupsa-nəl*
      Zakhar   family   people-POSS.PL<3PL   with   life-POSS.PL<3PL

      тāрвитыӈ-ыг   ēмт-ыс.
      tārvitəŋ-əγ   *jēmt-əs.*
      difficult-TRANSL   become-PST.3SG

      'Life got very hard for Zakhar and his family.' (LS 15/2017: 15)

As described above, sometimes a nominal predicate or a participle occupies Slot 9 instead of a (finite) Verb (42). That is because the nominal predicate appears without a copula and fulfils the same functions that the verb normally does. Example (42) is from a folklore tale about a man living in the Urals. In the previous sentences, the reader learns that winter has come. The focus of this sentence is on what the winter is like. The sentence is built from a shared scene-setting adverbial and two independent clauses with a subject and a nominal predicate, connected together with a conjunction.

Loc:3–SA:7–Pred:9–Conj–SA:7–Pred:9

(42) Нёр-ыт   тӯлы   ащирма-тэ   сака   нянра ос

    Ńōr-ət   *tēli*   *aśirma-te*   *saka*   *ńaŋra* *os*

    Ural-LOC   winterly   frost-POSS.SG<3SG   very   harsh and

    тӯй-тэ   осыӈ.

    tujt-te   *osəŋ.*

    snow-POSS.SG<3SG   deep

    'In the Urals the winter frost is very harsh, and the snow is deep.' (LS 15/2017: 15)

5.6. Question words and interrogative structures

Finally, some attention has to be paid to questions. In many languages, the constituent order of interrogative sentences is different from that of declarative ones. In Mansi, interrogative sentences are produced in a very similar way to declarative sentences. A more detailed picture would require a separate study based on more extensive question data, but according to the few questions appearing in my data, the constituent order proves to be not so different from that presented above. In wh-questions, the question word normally occupies the focus slot, and the other constituents are placed similarly as in declarative clauses.

We can see that both (43) and (44) are SA/SP-initial, verb-final sentences, where the wh-word occupies the focus slot:

Addr:0–SA:1–Loc:3–V:9

(43) Дмитрий Игоревич,   наӈ   хӧт   яныгма-с-ын?

    Dmitri Igorewitš   *naŋ*   *χōt*   *janiɣma-s-ən?*

    Dmitriy Igorevich   2SG   where   grow.up-PST-2SG

    'Dmitriy Igorevich, where did you grow up?' (LS 1/2020: 12)

SP:1–WH:8–V:9

(44) Аkв   тамле   ковёр   манах   хӧтал   вāр-аве?

    Akw   *tamle*   *kowjor*   *manaχ*   *χōtal*   *wār-awe?*

    one   like.this   carpet   how.many   day   make-PASS.PRS.3SG

    'How many days does it take to make a carpet like this?' (LS 1/2020: 13)

Polar questions are produced without separate interrogative structures: in speech the question is expressed with intonation, and in writing with a question mark. The constituent order follows the rules presented above, as in (45) and (46):

Loc:3–SA:7–V:9

(45) Пāвл-анын-т   ат   āстл-ым   кол-ыт   ōл-ӣг-ыт?

Pāwl-anən-t   at   āstl-əm   kol-ət   ōl-ēγ-ət?

village-2PL-LOC   NEG   prepare-PTCP.PST   house-PL be-PRS-3PL

'Are there any not ready-built houses in your village?' (LS 1/2018: 5)

SA:1–Loc:3–V:9

(46) Сӯкыръя   Саранпāвыл   ляпат   ōл-ы?

Sūkərja   Sapanpāwəl   l'apat   ōl-i?

Shchekur'ya   Saranpaul   near   be-PRS.3SG

'Is Shchekur'ya village situated near Saranpaul?' (LS 22/2019: 13)

6. Results, conclusions and further questions

In this article, I have presented a template-based analysis of Northern Mansi constituent order. Northern Mansi is a language where information structure is primarily expressed with syntactic variation between active and passive and between different clause types, but also the constituent order is connected to information structure. My analysis enables simultaneous observation of both the syntactic and pragmatic levels. The preceding section showed how syntactic functions occupy constituent order positions. Most of the functions have two or three alternative slots, and the position of a syntactic function varies between them. Within the 9+1 slots, there are very obvious alternation pairs which behave systematically in the data, and the reasons for this variation are pragmatic.

After recognizing and distinguishing the most typical types of order of syntactic functions, I connected a pragmatic level to the analysis. I added the pragmatic functions primary topic, focus, scene-setter and neutral. Table 4 shows how these pragmatic functions fit the model of ten syntactic slots. Some slots are primarily occupied by syntactic functions. When the placement of one syntactic function alternates between two slots, this can be explained by pragmatic reasons. Some slots are primarily reserved for particular pragmatic functions, and the position of the syntactic functions varies a lot. For example, the syntactic subject appears in the clause-initial position when representing the pragmatic primary topic, and closer to the clause-final verb when there is a scene-setting temporal or locational adverbial.

Table 4: Placement of syntactic and pragmatic functions in the 9+1 slot model. (An empty cell means that the pragmatic or syntactic perspective is not relevant to the slot in question.)

Slot	Pragmatic default	Syntactic default
0	Scene-setter	
1	Primary Topic	Subject
2		Time-1
3		Loc-1
4		Manner
5		Loc-2/Time-2
6	Neutral/Secondary Topic	D
7		Subject
8	Focus	
9		Predicate (verb or nominal predicate)

The default placement of the subject is in the sentence-initial position, possibly preceded by a conjunction. This tendency is connected to the topicality of the syntactic subject: in transitive structures, the pragmatic role of primary topic occupies the syntactic function of subject. The most topical arguments tend to occupy sentence-initial positions. When the subject is placed near the predicate to Slot 7, only a few particular arguments can be placed between them. According to my data, these arguments are 1) instrumental-inflected IO, 2) agent adverbial, 3) DO and 4) the infinitive or participle form of a verb, which belongs to a complex predicate. The most focal argument is placed immediately before the predicate. Because of the pragmatic variation between the different clause structures (see Section 2), this cannot be any argument: the most topical elements are placed in syntactic core roles. The most typical syntactic functions expressing the pragmatic focus function are DO, IO, directional adverbial, agent adverbial and nominal predicate.

In my data, the arrangement of constituent order is twofold. First, the basic constituent order is connected to the order of syntactic functions. Mansi is basically an SOV language. Secondly, particular syntactic functions alternate between two or more slots, depending on which pragmatic function they represent. Some slots are always occupied by a particular pragmatic function, which can be represented by several syntactic functions.

The main results of the syntactic and pragmatic analyses are:

1. The same template model can be applied to both active and passive clauses, and to different clause types.
2. The verb – or in the case of a clause without a copula, a nominal predicate – occupies the clause-final position, possibly followed only by conjunctions or other particles.
3. Topical arguments tend to occupy sentence-initial positions.
4. The most focal element is always placed in Slot 8, immediately before the verb.
5. The placement of the subject (SA/SP) varies between two slots: sentence-initial and two slots before the verb. Variation is due to pragmatic motivation.
6. DOs, IOs and directional adverbials share various common features – both semantic and syntactic – and common slots.
7. Temporal, locational and manner adverbials are placed immediately after the SA/SP, with the temporal component always coming first. When there is a clause-initial scene-setting adverbial, it is placed before the subject.

Finally it is worth mentioning that many of my results are identical or similar to the observations of Rombandeeva (1984: 58–60) on Mansi word order. My data wholly support Rombandeeva's statements about the subject and its attributes always preceding the predicate, the predicate always attending the sentence-final position, adverbials preceding the predicate, and the direct object usually being between the subject and the predicate. Also her statements that a focal expression of location is placed immediately before the predicate, and attributives always immediately precede the word they belong to, are wholly supported.

Further, the differences between Rombandeeva's observations and my results concern the placement of expression of time and location. According to Rombandeeva, the expression of time always precedes the subject and is placed in the sentence-initial position. My data show quite clearly that the expression of time precedes the subject only when it is in a scene-setting function. The default placement of temporal adverbials is in Slot 2, immediately after the subject. According to Rombandeeva, the expression of location is immediately before the predicate, or in the sentence-initial position. Also in this case my results differ from hers. According to my data, there are two stable slots for locational adverbials. The default placement is in Slot 3, and if there are several expressions of location in one sentence, the second one is in Slot 5. Further, the focal adverbials are placed at Slot 8. I also discussed the directional adverbials as a separate category, not

a part of the locational adverbials. As proven by my data, locational and directional adverbials do not behave identically: locational adverbials tend to appear before the manner and directional adverbials, while the directional ones are placed after the locational and manner adverbials.

Certain more marginal details and questions concerning Mansi constituent order have been excluded from this study and will hopefully be topics of further studies. For example, the placement of conjunctions is only briefly mentioned in the analysis, while the original template showed clear tendencies for variation among conjunctions. The slight differences between the occurrence of SAs and SPs, and a detailed analysis of the placement of complex predicates should also be discussed in more detailed studies. Moreover, different kinds of marginal chain effects, i.e. a change in one property leading to changes in others, have not been presented here, but a detailed presentation of these would be very fruitful.

Nonstandard abbreviations used in glosses

LAT	lative case	STRESS	stressed form (of a personal pronoun)
PTCL	particle	TRANSL	translative case

Other abbreviations

1-P	1-participant	IO_{INSTR}	instrumental-marked IO
2-P	2-participant	IO_{LAT}	lative-marked IO
3-P	3-participant	Loc	adverbial of location (semantic role)
Addr	addressing word		
Ag	agent (semantic role)	Man	adverbial of manner
AgA	agent adverbial (syntactic function)	Pred	nominal predicate (syntactic function)
Dir	adverbial of direction (syntactic function)	Rec	recipient (semantic role)
		SA	subject of an active clause
DO	direct object (syntactic function)	ScSA	scene-setting adverbial
DOA	Differential Object Agreement	SOC	Secondary Object Construction
DOM	Differential Object Marking		
IOC	Indirect Object Construction	SP	subject of a passive clause
IO	Indirect object (syntactic function)	Temp	adverbial of time
		WH	wh question word

References

Aissen, Judith. 2003. Differential object marking: iconicity vs. economy. *Natural Languages & Linguistic Theory* 21. 435–483.

Allan, Keith. 1987. Hierarchies and the choice of left conjuncts (with particular attention to English). *Journal of Linguistics* 23(1). 51–77. https://doi.org/10.1017/S0022226700011038

Alsina, Alex & Bresnan, Joan & Sells, Peter (eds.). 1997. *Complex predicates.* Stanford: CSLI Publications.

Amberber, Mengistu & Baker, Brett & Harvey, Mark (eds.). 2010. *Complex predicates: Cross-linguistic perspectives on event structure.* Cambridge: Cambridge University Press.

Baker, Mark. 2003. Agreement, dislocation, and partial configurationality. In Carnie, Andrew & Harley, Heidi & Willie, Mary (eds.), *Formal approaches to function in grammar*, 107–132. Amsterdam: John Benjamins Publishing Company.

Benincà, Paola & Poletto, Cecilia. 2004. Topic, Focus and V2: Defining the CP Sublayers. In Rizzi, Luigi (ed.), *The structure of CP and IP*, 52–75. New York: Oxford University Press.

Bíró, Bernadett. 2015. Az északi és déli manysi ditranzitív szerkezetek szórendje. In É. Kiss, Katalin & Hegedűs, Attila (eds.), *Nyelvelmélet és dialektológia 3*, 43–57. Piliscsaba: Szent István Társulat, Az Apostoli Szentszék Könykiadója.

Bíró, Bernadett & Sipőcz, Katalin. 2017. The Mansi ditransitive constructions. *Finno-Ugric Languages and Linguistics* 6(1). 41–55. http://full.btk.ppke.hu/index.php/FULL/article/view/53/59

Bossong, Georg. 1985. *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen.* Tübingen: Gunter Narr Verlag.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects and topics. In Li, Charles N. (ed.), *Subject and topic*, 25–55. New York: Academic Press.

Costa, James. 1997. *Word order and constraint interaction* (Seminários de Linguística 1), 65–102. Faro: University of the Algarve.

Costa, James. 1998. *Word order variation: A constraint-based approach.* The Hague: Holland Academic Graphics.

Costa, James. 2001. The emergence of unmarked word order. In Legendre, Géraldine & Grimshaw, Jane & Vikner, Sten (eds.), *Optimality-theoretic syntax*, 171–204. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/5161.003.0010

Dalrymple, Mary & Nikolaeva, Irina. 2011. *Objects and information structure.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511993473.004

Downing, Pamela. 1995. Word order in discourse: By way of introduction. In Downing, Pamela & Noonan, Michael (eds.), *Word order in discourse* (Typological Studies on Languages 30), 1–28. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.30.02dow

É. Kiss, Katalin. 1995. Introduction. In É. Kiss, Katalin (ed.), *Discourse configurational languages*, 3–27. New York: Oxford University Press.

Flämig, Walter. 1991. *Grammatik des Deutschen: Einführung in Struktur- und Wirkungszusammenhänge. Erarbeitet auf der theoretischen Grundlage der Grundzüge einer deutschen Grammatik.* Berlin: Akademie Verlag.

Good, Jeff. 2016. *The linguistic typology of templates.* Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139057479

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of human language*, 73–113. Cambridge: MIT Press.

Hale, Ken. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1. 5–47.

Haspelmath, Martin. 2015. Ditransitive constructions. *Annual Review of Linguistics* 1. 19–41.

Hawkins, John A. 1994. *A performance theory of order and constituency.* Cambridge: Cambridge University Press. https://doi.org/10.1017/S0022226796256397

Heine, Bernd & König, Christa. 2010. On the linear order of ditransitive objects. *Language Sciences* 32(1). 87–131. https://doi.org/10.1016/j.langsci.2008.07.002

Höhle, Tilman. 1986. Der Begriff "Mittelfeld": Anmerkungen über die Theorie der topologischen Felder. In Weiss, Walter E. & Wiegand, Herbert E. & Reis, Marga (eds.), *Textlinguistik contra Stilistik. Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985* (Kontroversen, alte und neue, Bd. 3), 329–340. Tübingen: Niemeyer.

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

Kathol, Andreas. 1995. *Linearization-based German syntax.* Columbus: The Ohio State University. (Doctoral dissertation.)

Kathol, Andreas. 2000. *Linear syntax.* Oxford: Oxford University Press.

Keresztes, László. 1998. Mansi. In Abondolo, Daniel (ed.), *The Uralic languages*, 387–427. London: Routledge. https://doi.org/10.4324/9781315003283

Kulonen, Ulla-Maija. 1989. *The passive in Ob-Ugrian* (Mémoires de la Société Finno-Ougrienne 203). Helsinki: Finno-Ugrian Society.

Kulonen, Ulla-Maija. 2007. *Itämansin kielioppi ja tekstejä* (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten XV). Helsinki: Finno-Ugrian Society.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press.

LS = *Лӯимā сӭрипос* [Lūimā Sēripos]. [Newspaper published in Khanty-Mansiysk. Readable also online: http://www.khanty-yasang.ru/luima-seripos]

Malchukov, Andrej & Haspelmath, Martin & Comrie, Bernard. 2010. *Studies in ditransitive constructions: A comparative handbook.* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110220377

van Riemsdijk, Henk. 2002. The unbearable lightness of GOing: The projection parameter as a pure parameter governing the distribution of elliptic motion verbs in Germanic. *The Journal of Comparative Germanic Linguistics* 5. 143–196. https://doi.org/10.1023/A:1021251312697

Riese, Timothy. 2001. *Vogul.* München: Lincom Europa.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Haegeman, Liliane (ed.), *Elements of grammar*, 281–337. Springer: Dordrecht.
https://doi.org/10.1007/978-94-011-5420-8_7

Rizzi, Luigi. 2004. Locality and left periphery. In Belletti, Adriana (ed.), *Structures and beyond: The carthography of syntactic structures* 3, 223–251. Oxford: Oxford University Press.

Rombandeeva = Ромбандеева, Евдокия Ивановна. 1979. *Синтаксис мансийского (вогульского) языка.* Москва: Наука.

Rombandeeva, Evdokija Ivanovna. 1984. *Wogulische Syntax* (Ars Ob-Ugrica. Arbeiten auf dem Gebiet der obugrischen Sprachen, Band 4). München: Schiefer.

SHANNON, THOMAS F. 2000. On the order of (pro)nominal arguments in Dutch and German. In Shannon, Thomas F. & Snapper, Johan P. (eds.), *The Berkeley conference on Dutch linguistics 1997: Dutch linguistics at the millennium*, 145–196. Lanham: University Press of America.

SIEWIERSKA, ANNA. 1988. *Word order rules*. New York: Croom Helm.

SIPŐCZ, KATALIN. 2016. A manysi ditranzitív igék szintaxisa szemantikai megközelítésből. *Jelentés és nyelvhasználat* 3. 49–61.

SKRIBNIK, JELENA. 2001. Pragmatic structuring in Northern Mansi. In Seilenthal, Tõnu (ed.), *Congressus Nonus Internationalis Fenno-Ugristarum. Tartu 7.–13.8.2000. Pars VI*, 222–239. Tartu: Auctores.

SKRIBNIK = Скрибник, Елена Константиновна. 2004. Категории мансийского глагола и актуальное членение предложения. In Храковский, Виктор Самуилович & Володин, Александр Павлович (eds.), *Типологические обоснования в грамматике: К 70-летию проф. В.С. Храковского*, 445–458. Москва: Знак.

SONG, JAE JUNG 2012. *Word order*. Cambridge: Cambridge University Press.

SOSA, SACHIKO. 2017. *Functions of morphosyntactic alternations, and information flow in Surgut Khanty discourse*. Helsinki: University of Helsinki. (Doctoral dissertation.) https://helda.helsinki.fi/handle/10138/183285

TAYLOR, JOHN R. 1996. *Possessives in English: An exploration in cognitive grammar*. Oxford: Clarendon Press.

VILKUNA, MARIA. 1989. *Free word order in Finnish, its syntax and discourse functions* (Suomalaisen Kirjallisuuden Seuran Toimituksia 500). Helsinki: Suomalaisen Kirjallisuuden Seura.

VILKUNA, MARIA. 1995. Discourse configurationality in Finnish. In É. Kiss, Katalin 1995 (ed.), *Discourse configurational languages*, 244–268. Oxford: Oxford University Press.

VIRTANEN, SUSANNA. 2013. Informaatiorakenteen vaikutus suoran objektin merkintään mansin itämurteissa. *Journal de la Société Finno-Ougrienne* 94. 293–335. https://doi.org/10.33340/susa.82601

VIRTANEN, SUSANNA. 2014. Pragmatic direct object marking in Eastern Mansi. *Linguistics* 52(2). 391–413. https://doi.org/10.1515/ling-2013-0067

VIRTANEN, SUSANNA. 2015. *Information structure in Eastern Mansi: An information structural approach*. Helsinki: University of Helsinki. (Doctoral dissertation.) https://helda.helsinki.fi/handle/10138/152802

VIRTANEN, SUSANNA. 2016. Fokusmerkintä mansin kielen itämurteissa. *Folia Uralica Debreceniensia* 23. 293–304. http://finnugor.arts.unideb.hu/fud/fud23/17_virtanen_susanna.pdf

VAN DER WAL, GUENEVER JOHANNA. 2009. *Word order and information structure in Makhuwa-Enahara*. Utrecht: Netherlands National Graduate School of Linguistics. (Doctoral dissertation.) https://scholarlypublications.universiteitleiden.nl/handle/1887/13845

WAGNER-NAGY, BEÁTA. 2011. *On the typology of negation in Ob-Ugric and Samoyedic languages* (Mémoires de la Société Finno-Ougrienne 262). Helsinki: Finno-Ugrian Society.

ZEPTER, ALEXANDRA. 2003. *Phrase structure directionality: Having a few choices*. New Jersey: Rutgers, The State University of New Jersey. (Doctoral dissertation.) https://rucore.libraries.rutgers.edu/rutgers-lib/38429/PDF/1/play/

Appendix: Frequency of syntactic functions in my data in the 9+1 slots.

Role	0	1 S	2 Temp	3 Loc	4 Man	5 Adv2	6 D	7 S2	8 Focus	9 V
Verb	1	2	–	–	–	–	2	–	1	624
Nominal predicate	–	–	–	–	8	–	–	–	49	10
Subject, active	–	245	–	–	–	–	–	87	–	–
Subject, passive	–	56	–	–	–	–	–	33	–	–
Temporal adverbial	20	–	115	–	–	17	–	–	6	–
Locational adverbial	4	–	–	128	–	23	–	–	10	–
Manner adverbial	10	–	–	–	123	–	–	–	32	–
Directional adverbial (lative or ablative)	–	–	–	1	2	5	76	–	12	–
Other adverbials	11	–	–	2	–	–	–	–	4	–
DO	–	1	–	–	4	–	54	–	48	–
IO$_{\text{INSTR}}$	–	–	–	–	–	–	5	–	24	–
IO$_{\text{LAT}}$	–	–	–	–	–	–	22	–	3	–
Agent adverbial	–	–	–	–	5	–	–	–	9	–
Infinitive[a]	–	–	–	–	2	–	2	–	94	3
Negation	–	–	–	–	–	–	–	–	25	–

a. In this table, the placement of infinitives and some other functions may overlap. In such cases, both an infinitive and a DO, a directional or some other adverbial are placed in the same slot. Consequently, both are calculated in the total number of occurrences. This exceptional arrangement concerns only infinitives, for further details see Section 5.5.

Wotistische Überraschung versteckt sich über die Jahrtausendwende

Barbera, Manuel. 2013 (1994). *A short etymological glossary of the Votic language.* Città di Castello (PG): I libri di Emil. LXIX + 426 S.

Barbera, Manuel. 2012 (1995). *Introduzione storico-descrittiva alla lingua vota (fonologia e morfologia).* Città di Castello (PG): I libri di Emil. 448 S.

Bücher mehr als ein Vierteljahrhundert nach der Veröffentlichung zu rezensieren ist ungewöhnlich. Die Verzögerung aber beweist nur, wie lange es diesem merkwürdigen Zwillingspaar gelungen ist, sich vor der uralistischen Forschergemeinschaft zu verstecken. Manuel Barbera, heute Professor der Didaktik der Gegenwartssprachen an der Universität Turin, hat viel über mehrere Sprachfamilien publiziert; auf dem Gebiet der Uralistik s. z.B. Barbera (1993a, 1993b, 1999).

Beide hier rezensierten Werke behandeln das Westwotische. Ursprünglich beabsichtigte der Autor mit seinem etymologischen Wörterbuch zu promovieren, aber dies war dem sprachwissenschaftlichen Institut der Universität Pavia nicht recht, so dass Barbera noch die *Introduzione storico-descrittiva* verfasste.

Die Periode der Veröffentlichung zeigt sich auf lustige Weise im Layout der Werke. Die Schriften sind offenbar für den Druck mit einem Tintenstrahldrucker ausgewählt, und die meisten Diakritika hat der Autor rechts von seinen Buchstaben platzieren müssen. Auch beim Druck sind Missgeschicke passiert. Beim etymologischen Wörterbuch fehlt bedauerlicherweise Seite 107, an deren Stelle Seite 120 ein zweites Mal gedruckt wurde.

Die historisch-deskriptive Grammatik

Das Werk ist nicht ganz, was man dem Namen nach davon erwarten könnte, sondern vielmehr eine vergleichende Grammatik der ostseefinnischen Sprachen aus der Perspektive des Wotischen. Dies ist selbstverständlich eine willkommene Abwechslung zum traditionellen Fennozentrismus der Disziplin. Die historische Dimension steht stellenweise stärker, stellenweise weniger im Vordergrund, vor

allem aber bei der Darstellung der Kasusflexion, die alle finnisch-permischen Sprachen umfasst und innerhalb des Buches gewissermaßen eine separate Studie bildet.

Wertvoll ist die Behandlung der westwotischen Dialekte S. 60–71, für die Barbera aus Aristes (1968) Grammatik alle für einzelne Dörfer erwähnten lautlichen, morphologischen und lexikalischen Züge gesammelt und diese zu Verzeichnissen über Dialektzüge der jeweiligen Dörfer arrangiert hat. Diese Verzeichnisse wären ein guter Ausgangspunkt auch für eine umfangreichere Darstellung der Dialekte des Westwotischen.

Angebracht ist die indoeuropäistische Kritik an Postis Theorie zum Ursprung des Stufenwechsels sowie die Erklärung des „suffixalen Stufenwechsels" als ein vom Stufenwechsel separates Phänomen S. 146. Eine gute Innovation ist auch die Reihenfolge der Behandlung der Themen, so werden zum Beispiel bei der Darstellung des Lautsystems die neuen, „marginalen" Konsonanten nicht erst zum Schluss erwähnt, sondern gleich am Anfang behandelt.

Das Werk hätte einer Vorkontrolle durch einen erfahrenen Uralisten bedurft, denn die Menge der Fehler wirkt sich störend aus. Die meisten davon sind klein, aber einige verursachen doch Probleme im Hinblick auf das Gesamtbild. S. 91, 94 postuliert der Verfasser ohne Begründung

für das Urfinnische ein *ö* in nicht-ersten Silben, das im Wotischen geschwunden sei. Eine fehlerhafte urfi. Rekonstruktion **onne-ton* (statt *-ťton*) S. 145 führt zu einer falschen Schlussfolgerung bei der Bewertung der Entwicklung des Konsonantismus nach unbetonter Silbe.

Ein erheblicher Mangel auf Seiten der Lautlehre ist die minimale Behandlung der Diphthonge S. 97–98. Diphthonge in nicht-ersten Silben erwähnt der Verfasser überhaupt nicht. Im Bereich der Formenlehre bleibt die Rekonstruktion der wotischen Terminativendung urfi. **n-sak* S. 167 ohne Erklärung.

Der Autor ist gelegentlich etwas zu eifrig mit seiner Zeit gegangen und hat ohne Kritik Hypothesen übernommen, die zum Zeitpunkt seiner Arbeit frisch waren, wie etwa die Verwandtschaft des Jukagirischen mit den uralischen Sprachen, die vierte Längenstufe estnischer Silben und den teilweise ururalischen Ursprung der im Ostseefinnischen anzutreffenden Vergangenheitsmarkers *-si-* (S. 306).

Den Dialekt von Kukkuzi und das Krewinische fasst Barbera z.B. S. 51, 120–22 als Südwestwotisch zusammen, was offenbar ausschließlich auf Postis (1934) Auffassung zurückgeht, dass beide hinsichtlich der Entwicklung **k > *č* archaisch seien. Wenn wir jedoch dem S. 37 von Viitso übernommenen Schema

(vgl. Viitso 2008: 64) folgen, stammt der Dialekt von Kukkuzi aus demselben Zweig wie das Ingrische, was den Erhalt von urfi. *k dort besser erklärt (so auch Winkler 1997: 426).

Das etymologische Wörterbuch

Dieses besondere Werk hat die bisherige komplette Nichtbeachtung nicht verdient. Es ist besser gelungen als die Grammatik. Eines seiner erklärten Hauptziele, die Darstellung des gesamten in gedruckten Quellen bekannten Wortschatzes des Westwotischen, hat zwar nach dem Erscheinen von VKS seine Bedeutung verloren, was auch der Autor selber in seinem Nachwort eingesteht. Seine Ziele und Praktiken auf dem Gebiet der wotischen Lexikographie hat Barbera auch in zwei Artikeln (2001, 2002) dargestellt.

Auch dass die Arbeit am Werk mitten in das zweite goldene Zeitalter der uralistischen Lehnwortforschung fiel, hat das Wörterbuch schnell veralten lassen, wenngleich es gerade im Bereich der Entlehnungen am stärksten ist. Das bedeutendste Verdienst des Buches ist die gründliche Besprechung der mutmaßlichen indogermanischen Ausgangsformen angenommener Lehnwörter in den Lemmata.

Einem Uralisten können die Wortartikel übertrieben lang wirken, insbesondere im Fall junger Entlehnungen aus dem Russischen, aber hierzu trägt die Tradition seiner eigenen etymologischen Wörterbücher bei, deren Hauptaugenmerk auf den alten Erbwörtern und deren Entsprechungen liegt. Barbera stellt diese Konstellation auf den Kopf. Sein logischer Ausgangspunkt ist die gleichmäßig exakte Beschreibung aller hinreichend sicheren Informationen über die Hintergründe eines Wortes, unabhängig davon, in welcher Sprachfamilie es auftritt.

Dank seiner Vertrautheit mit der indogermanistischen Forschungsliteratur gelingt die Analyse mutmaßlicher Entlehnungen – beispielsweise *kańevo*, *Narva*, *têrva* (fi. *terva*) und *terve* – Barbera besser als jedem etymologischen Wörterbuch der ostseefinnischen Sprachen, abgesehen von LägLoS. Für das Wort *irsi* (fi. *hirsi*) legt Barbera eine ganz neue, interessante baltische Erklärung vor. Mehrere Wörter unterschiedlichen Alters, die aus derselben indogermanischen Wortsippe, aber aus verschiedenen Quellen entlehnt wurden, verbindet der Autor vorbildlich mit Querverweisen untereinander.

Dementsprechend scheitert das Werk am deutlichsten an der Erklärung solcher Wörter, bei der eine Hintergrunduntersuchung der als Ausgangsformen einer Entlehnung angesetzten Wörter nicht stattfindet. Beispiele hierfür sind *mettsä* (fi. *metsä*), *ümpäri* (fi. *ympäri*) und

varsi. Bei der Bewertung vorgeschlagener Entsprechungen aus entfernt verwandten Sprachen hat sich der Autor als Nicht-Uralist vor allem auf UEW und Collinder gestützt und bietet daher kaum Neues.

Insgesamt hätte Barbera sich mit seiner Sachkenntnis seinerzeit hervorragend geeignet, die SSA-Redaktion zu vervollständigen, deren Stärken auf dem Gebiet des uralischen Lexikons lagen. Ohne Verbindungen zur uralistischen etymologischen Gemeinschaft wurde das wotische etymologische Wörterbuch jedoch zu einem Nachschlagewerk anderer Art, das auf seine eigene, interessante Weise die Standardquellen der ostseefinnischen Etymologie ergänzt und ist mit seinen umfangreichen indoeuropäistischen Literaturangaben vor allem für die Lehnwortforschung hilfreich.

Santeri Junttila

Literatur

Ariste, Paul. 1968. *A grammar of the Votic language*. Bloomington & The Hague: Indiana University & Mouton.

Barbera, Manuel. 1993a. *La gradazione baltofinnica*. Londra: Lothian Foundation Press.

Barbera, Manuel. 1993b. Problemi di ricostruzione nel consonantismo uralico. *Ponto-Baltica* V. 7–60.

Barbera, Manuel. 1999. Appunti su definitezza e partitivo nelle lingue baltofinniche. *Bollettino dell'Atlante linguistico italiano* 3a serie XXIII. 119–188.

Barbera, Manuel. 2001. Providing a future for a disappearing language: Some notes on Votic lexicography. *Linguistica Lettica* VII. 180–213.

Barbera, Manuel. 2002. Formats of etymology: the case of 'beer' and 'juniper' in Votic. In *Proceedings of the Tenth EURALEX International Congress, Copenhagen, Denmark, August 13-17 2002*. Vol. II. 563–568. Copenhagen: Center for Sprogteknologi.

Collinder, Björn. 1955. *Fenno-Ugric vocabulary: An etymological dictionary of the Uralic languages*. Stockholm: Almqvist & Wiksell.

LägLoS I–III = Kylstra, A. D. & Hahmo, Sirkka-Liisa & Hofstra, Tette & Nikkilä, Osmo. 1991–2012. *Lexikon der älteren germanischen Lehnwörter in den ostseefinnischen Sprachen* 1–3. Amsterdam & Atlanta: Rodopi.

Posti, Lauri. 1934. Vatjan kielen *k > tš* äänteenmuutoksen iästä. *Kalevalaseuran vuosikirja* 14. 139–148.

SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (Hrsg.). 1992–2000. *Suomen sanojen alkuperä. Etymologinen sanakirja* 1–3. Helsinki: Kotimaisten kielten tutkimuskeskus & Suomalaisen Kirjallisuuden Seura.

UEW = Rédei, Károly (Hrsg.). 1988. *Uralisches etymologisches Wörterbuch* I–II. Wiesbaden: Harrassowitz.

Viitso, Tiit-Rein. 2008. *Liivi keel ja läänemeresoome keelemaastikud*. Tartu & Tallinn: Eesti keele sihtasutus.

VKS = Adler, Elna & Grünberg, Silja (Hrsg.). 1990–2011. *Vadja keele sõnaraamat. Словарь водского языка* 1–7. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Winkler, Eberhard. 1997. *Krewinisch: Zur Erforschung einer ausgestorbenen ostseefinnischen Sprachform*. Wiesbaden: Harrassowitz.

Languages of hunter-gatherers through a historical-typological lens

GÜLDEMANN, TOM & MCCONVELL, PATRICK & RHODES, RICHARD A. (eds.). 2020. *The language of hunter-gatherers.* Cambridge: Cambridge University Press. xxii + 720 pp.

1. Introduction

This is a review of the collective volume *The language of hunter-gatherers*, edited by Tom Güldemann, Patrick McConvell, and Richard A. Rhodes, featuring 23 chapters, organized into 7 parts ("Introduction", "Africa", "Tropical Asia", "New Guinea and Australia", "Northeastern Eurasia", "North America", "South America"), an appendix "Preliminary worldwide survey of forager languages", a language index, and a subject index. Already from this table of contents, one can immediately see the main aim of the collection: to provide a worldwide survey of hunter-gatherer societies through the lens of their languages. The enterprise is not solely descriptive but rather serves to answer the main research question, whether languages of hunter-gatherers are in any sense different from languages of food-producers. No modern linguist seriously expects the specifics to lie at the synchronous level of linguistic structures (though see Bickel & Nichols' chapter discussed below), but the social and sociolinguistic histories of foraging communities might have particular properties when compared to those of food-producing ones, and these properties might shape their languages. Thus, the main analytic focus of the volume is on the spread and diversification of hunter-gatherer languages, on variation and contact processes observed for them, as well as on linguistic encoding of cultural knowledge specific to hunter-gatherers (note that the terms "hunter-gatherers" and "foragers" are used interchangeably in the volume and hence in this review).

2. Overview of the volume as a whole

The volume originated in the workshop "Historical linguistics and hunter-gatherer populations in a global perspective" held at the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany), and the emphasis on historical linguistics has shaped its main research question and has remained

central for most of the individual papers. Many of the papers address the dynamics of forager languages, using the reconstructed evolution of languages of an area to inform the history of subsistence, migrations, and population contact in that area (e.g. the papers by Güldemann, Blevins, Rischel, Burenhult, Reid, Harvey, and Epps). In yet other papers, the argument goes in the opposite direction: externally observed factors like the difference in subsistence of a given linguistic group from its neighbors or known population movements serve as tools which could help to understand the areal linguistic history (e.g. the contributions by Savà & Tosco, Ross, and Vidal & Braunstein). In doing so, these studies rely on an assumption that both the peculiarities of subsistence and the peculiarities of a language normally change slowly and gradually, so that atypical abrupt shifts would leave traces that can be detected and analyzed. One paper (Hill) is also devoted to verification and refinement of a particular theory of hunter-gatherer language spreads postulated earlier by one of the editors of the volume (McConvell).

At the same time, there are also papers with a purely synchronic focus interested in possible linguistic contrasts between languages of foragers and those of food producers.

Bickel & Nichols report on a wide typological survey on the topic. Brown examines a possible difference in botanic terms between foragers and food producers with a case study of oak terms in languages of the USA, but with broader conclusions in mind. Donohue checks for recurrent differences between languages of foragers and those of their food-producing neighbor(s) with three case studies from New Guinea. Harrison & Anderson are concerned with how the forager lifestyle might be reflected in the peculiarities of two closely related Turkic languages of South Siberia (Tofa and Todzhu). De Reuse refutes the hypothesis of hunter-gatherer primitivism based specifically on the much discussed Eskimo terms for 'snow', where the (wrongly) presupposed absence of a general term has been taken as evidence of weak logical ability.

Some papers of the volume report fully original studies performed specifically for the volume (e.g. those by Bickel & Nichols, Rischel, Donohue, Sutton, Anderson & Harrison, de Reuse, Hill, and Epps), even though some of them include previous findings of the authors, published or unpublished. Some are mainly summaries of the authors' expertise in an area, compiled specifically for the volume (e.g. the papers by Rhodes,

Ross, Burenhult, Soriente, Harvey, McConvell, and Vidal & Braunstein). Note however that in many cases, the border between these categories is quite fuzzy. Finally, some papers are not original altogether (e.g. Brown) or reproduce linguistic studies published elsewhere, but now with an added (extended) anthropological or ethnographic discussion (e.g. Sava & Tosco, Guldemann, and Reid).

In most papers, the focus on subsistence is omnipresent, though in some others the reference to subsistence serves merely as a background, often of only minor importance, for a purely linguistic study (e.g. the contributions by de Reuse, and in particular by Vajda and by Rhodes). These papers might be excellent linguistic work, but they do not really contribute to the central discussion of the volume, that of the possible relationship between languages and subsistence types of a given area or typologically.

The editors of the volume have evidently not strived to ensure a particular unity for the collection. There are very few cross-references between papers (even where they would seem the most natural, cf. Sutton's brief reference to a particular language spread in Australia described in detail by Harvey in the chapter immediately following Sutton's) and most areal or typological overviews undertaken in the volume's papers do not mention cases described elsewhere in the book (e.g. see the papers by Gunnarsdóttir & Stoneking, Rischel, Epps, etc.). In a sense, readers thus have an opportunity to get a retrospective view of the field, that is, what the state of linguistic knowledge on hunter-gatherer languages was up to the time when these papers were written.

While the hunter-gatherers of the Americas, Africa, Australia, and Oceania are regularly mentioned in typologically oriented linguistic literature in English, this is usually not the case for Asia. So, each paper of the section "South Asia", aside from treating its own topics of interest, serves also as a very welcome introduction to the minority languages of the area. Unfortunately, this cannot be said about the section "Northeast Asia"; it contains only two papers, none of which gives an overview of the area comparable to the South Asian overviews. In particular, the Uralic language family is not represented in the volume at all. Even though Anderson & Harrison note in their chapter, "Siberian reindeer herding peoples (e.g. Khanty, Evenki) are included both in the expanded foregoing definition [from Panter-Brick et al. (2001: 2)] and in the *Cambridge encyclopedia of hunters and*

gatherers (Lee and Daly 1999)", the editors of the volume might have had a different opinion. Indeed, in the extensive appendix "Preliminary worldwide survey of forager languages", compiled by all of the authors, only such Uralic languages as Khanty, Mansi, Nganasan, and Selkup are stated to definitely belong to forager languages, with Forest Nenets (and e.g. Evenki and Even) being classified as "uncertain cases", and Forest Enets and Tundra Enets absent altogether. To the best of my knowledge, the Forest Nenets are no more pastoralists than the Nganasans, and the Tundra Enets, and in particular the Forest Enets, are also as clear cases of foragers as the Nganasans are.

The volume took fifteen years to mature, with a workshop back in 2006 setting the first presentation of most of its papers. On the whole, many papers were updated at least a couple of years before publication, though some show signs of outdatedness. Note for example the multiple references to an obsolete Proto-Austronesian reconstruction from 1972 in Reid's paper (p. 232), updated only in a footnote, or Gunnarsdóttir & Stoneking's paper on the population genetics of hunter-gatherers, explicitly saying that it was last updated in 2009, which means that today it has a merely historical value (e.g. it mentions

methods based on autosomal DNA analysis only in passing and ignores the recent groundbreaking continent-wide studies using these methods).

The quality of the technical editing is good, even though greater perfection could have been expected from such a major academic publisher. There are occasional misprints and missing or inadequate references, maps, or tables, albeit they are not too numerous. For example, Rischel refers to a paper as "the *PLoS* paper" without any further details (p. 153); Tables 11.5–11.6 referred to in the text are missing (p. 292); Map 12.6 supposed to show languages of North-Western Melanesia is missing (p. 319), as well as Table 12.5 (p. 321); Map 13.1 does not show the areas which the text describes it as showing (p. 337); Table 13.4 does not use italics, though its meaning is explained in the text (p. 344); lists of abbreviations are missing from the papers by Harvey and McConvell; a part of footnotes in McConvell's paper are faulty, with 6 in the text referring to 7 in the list of footnotes, and so on up to 10 in the text, with 11 in the text absent from the list of references, and 6 in the list of references missing in the text; some glosses of examples 17.1, 17.2 (p. 477) and 17.8 (p. 484) are mixed up, as are translations for 'son' and 'God' in the article text and in the glosses

of example 17.14 (p. 486). I will not bother readers with a list of misprints, but they amount to no more than a couple of dozen for the 700+ pages of the volume.

3. Overview of the individual papers

The rest of this review is devoted to a concise critical description of the individual papers, completed by a brief conclusion on the relevance of the volume to the field of Uralic linguistics. The papers are presented mainly in the order of their appearance in the volume, though I have taken the liberty to group some of them together, which makes an occasional change to the default order.

The book opens with an extended **introductory paper** by the editors, where they first summarize the anthropological starting point of the debate, the "otherness" of foragers that is expected, at least by anthropologists, to be somehow reflected in their languages. The editors, and actually many authors of individual papers, openly confront the view expressed in Bellwood (2013) (and his earlier works) that puts agricultural language spreads at the forefront and marginalize foragers as participants in the world's linguistic history. After the useful overview of anthropological conceptions of foragers, the editors

turn to their languages, mainly focusing on their relationship to and contacts with languages of food-producers. The bibliography of this introductory paper is worthwhile on its own, providing any neophyte in the topic with an excellent starting point for independent discoveries.

The paper by **Ellen Dröfn Gunnarsdóttir & Mark Stoneking**, "Genetic landscape of present-day hunter-gatherer groups", is an overview of genetic research on the difference between foragers and food-producers. As has been already mentioned, it is quite outdated, though a reader can learn that (a) groups of foragers generally exhibit lower genetic diversity in their mitochondrial DNA (mtDNA) than their agricultural neighbors, and (b) differences between foragers and food-producers are greater for Y-chromosomes than for mtDNA. The former is explained by the numerical drop in forager populations conditioned by the expansion of agriculturalists, but also increased migration rates within the food-producers, while the latter is conditioned by higher migration rates for females since most human populations practice patrilocality. However, these observations are mainly based on African populations since only those were well researched 10–15 years ago.

Balthasar Bickel & Johanna Nichols in "Linguistic typology and

hunter-gatherer languages" study a hypothesis of a possible systematic linguistic differentiation between languages of foragers and those of food producers. After an introduction discussing the importance that this correlation, if discovered, would have for linguistic typology, they check more than two hundred phonetic, phonological, morphological, morphosyntactic, and syntactic variables represented in the AUTOTYP and WALS databases. With careful statistical analysis, they conclude that the languages of the two types of societies definitely belong to "the same grammatical ilk" (p. 72), and so the hypothesis is not supported by the current data on the world's languages. Since the main claim of the paper is negative, it is quite short, but its importance for the volume, and thus the field, should not be underestimated. Even though this is rather an open secret for typologists, the value of checking the linguistic equality of hunter-gatherers' languages is high for representatives of other disciplines. This could be particularly true for anthropology where the divide between the two types of societies has long been much more important than in linguistics.

The essence of **Cecil H. Brown**'s paper "Ethnobiology and the hunter-gatherer/food producer divide" is a study of the difference in botanic terms between foragers and food producers, postulating for the latter a tendency to have binomial terms. The suggested explanation for the difference is more attention to plants from the part of cultivators as opposed to foragers, but it has no further grounding except for the difference discussed in the paper and for a reference to Berlin (1992). However promising, this study is limited to terms for the oak and to languages of the United States only (excluding Alaska). This makes it an interesting start for ethnobiological research into the topic of the volume, but it is clearly insufficient for making wider generalizations.

The papers by Güldemann, Blevins, and Epps turned out to be my favorites of the whole volume, serving as perfect examples of studies where historical linguistics sheds light on the human history of the respective areas. All three are brilliantly written: fully accessible to linguists unfamiliar with these languages, they seem not to be compromised for specialists either, providing accurate reconstructions, grammatical in the case of Güldemann and mainly lexical in the case of Blevins and Epps. These papers provide exemplary embodiments of the ways in which historical linguistics can inform models of hunter-gatherer prehistory. Last but not least, they show by their own

examples that an argument from historical linguistics can be formulated in such a way that it is easy to read without previous knowledge of the history of this particular family – something not so often encountered within the subdiscipline of historical linguistics, including the long-established field of Uralic comparative linguistics.

Tom Güldemann in his paper "Changing profile when encroaching on forager territory: Toward the history of the Khoe-Kwadi family in Southern Africa" suggests a new history of South African languages. Modern Khoe-Kwadi languages are shown to be "a geographically marginalized remnant reflex of an earlier population spread supplanted by a later population spread" (p. 134), in the latter case by Bantu speakers. The earlier spread involved pastoralists speaking languages of the Khoe-Kwadi meeting indigenous foragers whose languages were related to the modern southernmost Kx'a and Tuu families (Khoe-Kwadi, Kx'a and Tuu together were earlier described as "Khoisan"). This spread coincided in time with, and most probably was conditioned by, the precipitation peak in summer rainfall 3,000–2,000 BP. When most of the area turned into desert once again, the pastoralists reverted to the foraging economy of the original population, thus today

confusing researchers who have long held foraging to be the original state of the Khoe-Kwadi speakers. The main merit of this paper is the independent analysis of purely linguistic and non-linguistic (archaeological, genetic and cultural) data, where scenarios arrived at separately are compatible with each other, making thus a strong argument for the historical reconstruction.

The paper by **Juliette Blevins** "Linguistic clues to Andamanese prehistory: Understanding the north-south divide" uncovers the linguistic and cultural history of the Andaman Islands in the Indian Ocean. These islands are inhabited by "Negritos", "a descriptive term for dark-skinned frizzy-haired people of insular and mainland Southeast Asia, usually of short stature" (p. 200–201) who are thought to represent populations predating the influx of Austronesian and Austroasiatic speakers to the area. The inhabitants of the northern part of the island speak Great Andamanese languages and those of the southern part speak Ongan languages. Despite the genetic and cultural unity of the inhabitants of the Andaman Islands, the two linguistic families are not related to each other, with the former being Austronesian and the latter being Austroasiatic, as Blevins suggests in her paper. The attribution of Great

Andamanese to Austronesian was shown by her in a previous paper, while the relatedness of Ongan to Austroasiatic is demonstrated for the first time in the current paper of the volume. These linguistic reconstructions hint either at distinct migrations to the two parts of the Andaman Islands, or to an ancient population split combined with extensive contacts with various parts of the mainland.

The paper by **Patience Epps**, "Language and subsistence patterns in the Amazonian Vaupés", is relatively short. Devoted to the Nadahup family, modern forest foragers of Amazonia, it studies the main elements of their horticultural and non-horticultural plant vocabularies, as well as those relating to cultivars. While the common wild-growing plant terminology is reconstructible for all four Nadahup languages, words for cultivars are of more recent origin, though attested in a genealogically defined subgroup of the family. Among the latter, many words are of Tukanoan and Arawakan origin, speakers of which are the Nadahups' agriculturalist neighbors. This discovery by means of classical comparative-linguistics methods allows for a relative dating of the introduction of exchange practices with the agriculturalists and of the beginning of limited horticulturalism among

the Nadahup speakers. An earlier published study of the Nadahup numeral system, also summarized in the paper, supports the same scenario.

Graziano Savà & Mauro Tosco's paper "Hunters and gatherers in East Africa and the case of Ongota (Southwest Ethiopia)" contributes to the debate on the possible historical ways in which hunter-gatherer communities developed, in particular in East Africa. It puts an earlier linguistic study by one of the authors into a larger, mainly ethnographic, context. The paper deals with Ongota, a small hunter-gatherer community in Southwest Ethiopia surrounded by food producers speaking Cushitic and Omotic languages (both Afro-Asiatic). Linguistically, the moribund Ongota language is very different from any of its neighbors largely due to its poor and isolating morphology. While the authors cannot show that a scenario of a separate language family spoken once by several hunter-gatherer communities is impossible, they insist that the current evidence is too weak to postulate this. Based on this and the ethnographic evidence connecting the Ongotas to a northern Omotic neighbor, they suggest the following cyclical development: first, a group of Omotic-speaking food-producers became marginalized

and shifted to foraging; while migrating to their current location, they experienced influences from other languages leading to a morphological profile unusual for the area; finally, today an assimilation to neighboring East Cushitic pastoralists is complete (the language is remembered only by the elders, there are no cultural distinctions, except for the lack of cattle by the Ongota). In their discussion of the Ongota case, Savà and Tosco are quite critical of other African cases where small hunter-gatherer communities with non-related languages have been claimed to represent an earlier, significantly more numerous, population. They focus on the weakness of the existing argument that can hold only if hunter-gatherers are thought to always precede food producers in a given location and no "devolution" of food-producers into hunter-gatherers is possible. If this view is not taken for granted, all these cases can equally well be analyzed as such reversals.

The paper by **Jørgen Rischel**, "Hunter-gatherers in South and Southeast Asia: The Mlabri", starts with a concise but very meaningful overview of all hunter-gatherer groups of South and Southeast Asia with their linguistic affiliation. The author then concentrates on the linguistic history of one particular group, Mlabri, which has been shown by an earlier genetic study to represent a recently (500–800 years old) isolated forager group surrounded by food-producers, with their language showing both Mon-Khmer (more) and Sino-Tibetan features (less). Based on myths and detailed linguistic analysis, Rischel suggests that the Mlabri reverted to gathering from food-production with an aim "to preserve, or even re-establish, the ethnic identity and integrity of the group" (p. 152). Their language shows traces of the encounters that this small isolated group had in the past.

Niclas Burenhult in his paper "Foraging and the history of languages in the Malay Peninsula" uncovers the complicated human and linguistic history of the Malay peninsula and forms it into a clear, easy-to-read narrative. A part of the Malay population are hunter-gatherers belonging to the "Negrito" genetic profile and speaking North Aslian Languages (Aslian < Austroasiatic); they are known as the Semang ethnographic unity (Burenhult provides a table of all Semang ethnolinguistic groups with detailed comments on their languages, locations, and current sociolinguistic situations; the data may seem peripheral to the bird's-eye view presented in the paper, but are clearly invaluable for those interested in this particular area). While two

opposing hypotheses about the origin of the Semang have been earlier formulated, the author shows that the truth lies in between, and the Semang should be viewed "as neither isolated relics, nor recent 'by-products', but as a successful blend of old and new" (p. 190). To reach this conclusion, he sums up the analysis of lexical divergence and convergence within the Aslian languages (published earlier in a co-authored paper) and compares it to the published results of recent genetic studies; confronting the two allows him to arrive at a new level of understanding of language spreads on the peninsula. Last but not least, he adds an original sociolinguistic perspective to his analysis, suggesting that the types of linguistic interaction observed by the Semang are determined by their mobile lifestyle and "patterns of their group disintegration and regrouping […] in response to changing subsistence conditions" (p. 187). High levels of individual variation and fuzziness of language boundaries are for Burenhult logical linguistic consequences of these social patterns. Personally, I was struck by this unexpected parallel to the recent past of the Northern Samoyedic groups (see Khanina (2021) for a sociolinguistic analysis of the former), with this parallel showcasing how promising cross-linguistic

comparison can be in the case of the sociolinguistics of hunter-gatherer communities.

The paper "Historical linguistics and Philippine hunter-gatherers" by **Lawrence A. Reid** is devoted to languages of Philippines foragers, who, with one exception, have been described as a "Negrito" population (see an explanation of the term above). The main focus of the paper is on historical patterns of interaction between Philippine farmers and foragers, all of whom speak Austronesian languages today. Given that phenotypically foragers are different from the rest of Austronesian speakers, a shift – or rather many shifts – by the original Philippines population to the languages of the agricultural newcomers can be assumed. What remains to be determined is the particular scenario of historical relationships and their human and linguistic outcomes, and they are discussed in the paper with references to six case studies. Reid suggests that the close contacts between incoming farmers and the local population started soon after the arrival of the farmers, and the latter were clearly socially dominating in these interactions. Eventual conflicts and other social reasons led to regular breaks in contacts and thus to linguistic diversification, supported by the Negritos' eagerness to maintain separate

linguistic identities. Original and very informative in its essence, this paper deals with evidence rather freely, and also risks losing its potential readership because of its less than superlative writing style. Without attention to the needs of linguists and anthropologists unfamiliar with languages of the Philippines (no introduction, no genealogical information for languages, insufficient maps, etc.), the author stands a chance of making his message accessible only to specialists in the area or very patient readers ready to invest their time in a certain amount of decryption.

In her paper "Hunter-gatherers of Borneo and their languages", **Antonia Soriente** presents an expert overview of forager languages of Borneo island. Unfortunate confusion and ambiguity of exonyms of the local nomadic groups (who usually do not use any autonyms) make it a difficult enterprise even for specialists to track their respective histories. Luckily, Soriente has devised a way to present the groups in quite a reader-friendly manner, with a combination of prose, informative tables, and maps. The organization of information on the genealogical subgrouping of their languages (all North Borneo Malayo-Polynesian < Austronesian) is less optimal, though this is indeed a complicated matter substantially,

with many of the affiliations being yet uncertain and some of them elucidated directly in the given paper. The vastly underdescribed state of the general linguistic situation of Borneo is exemplified by an illuminative case study of a particular language/dialect, Penan Benalui, spoken by just 450 nomadic hunter-gatherers. Most of the case study is devoted to a comparison of Penan Benalui to some other neighboring languages, all ultimately related to it, with the aim of finding its particular place among the subbranches of the North Sarawakan languages, and simultaneously to refine the number and the status of these subbranches themselves. The detailed and clearly valuable comparative description of the phonology, grammar, and lexicon of Penan Benalui is not always easy to follow for a non-specialist, e.g. only by the end of the paper could I figure out what the main logical genealogical alternatives considered in the paper actually were. With the main topic of the volume in mind, Soriente pays particular attention to knowledge systems encoded by the Penan Benalui lexicon and how it differs from the lexicons of its sedentary neighbors. Finally, she concludes her paper with a hope that future molecular anthropology studies will contribute to the discussion of the complicated past and present

relationship between various linguistic groups of Borneo island.

Malcolm Ross's paper "The linguistic situation in Near Oceania before agriculture" gives a broad overview of the history and linguistic situation of Near Oceania from its settlement by the first humans ca. 21,000 BP up to the arrival of Austronesian-speaking agriculturalists from the Asian mainland ca. 3,300 BP. He includes mainland New Guinea and Northwest Island Melanesia (= the Bismarcks, Bouganville, and the Solomon Islands) into his survey, basically delimiting himself by the territory where the so-called "Papuan" languages are spoken today (not a genealogical unity but a cover label for any non-Austronesian language of the area). The correlation between subsistence and languages in this part of the world is quite strong, with the languages of the biggest Papuan family, Trans-New-Guinea, being spoken by farmers and all the other Papuan languages, belonging to no less than 23 distinct families, being spoken by sedentary foragers living from "wild-food production" (i.e. reliance on wild sago-plants with occasional planting and minimal tending, e.g. by clearing underbrush). As can be expected, the Trans-New-Guinea farmers occupy most of the territory, with sago-dependent communities scattered around the coastal areas. Ross refers to changes in the biogeography of the past to explain the persistence of non-farmers in these particular areas, and thus claims the non-Trans-New-Guinea languages to be relics of the former linguistic diversity of Near Oceania. Interestingly, all languages of New Guinea could be genealogically connected, but since the time depth of their dispersal lies beyond the reach of the method of comparative linguistics, this continues to be only a hypothesis.

Mark Donohue's paper "Language, locality and lifestyle in New Guinea" starts with a short overview of diversity ranges in New Guinea's language communities in terms of geography and lifestyle. It then proceeds to an analysis of two case studies from North-Central New Guinea and one from the Western Highlands, in all of which the focus is on contrasting features of a language spoken by foragers and a language or languages spoken by its food-producing neighbors. These cases include both the neighborhood of related languages, as well as of unrelated ones. While the idea of the author could have been to identify some common linguistic features that would recurrently differentiate languages of foragers from those of food-producers, he concludes with a simpler claim that there is always at least one feature

by which a hunter-gatherer population's language differs from the languages of its neighbors. This generalization is indeed supported by the data Donohue presents in this paper, but it seems to be true basically for all neighboring language pairs of the world. This is in line with the conclusions of the typological paper of this volume, that by Bickel & Nichols, stating that there are no regular differences between languages based on the subsistence mode of their speakers.

Peter Sutton's paper "Small language survival and large language expansion on a hunter-gatherer continent" is an impressive original study that would be hardly realizable without the author's exceptional level of expertise in the languages and peoples of Australia. First, Sutton provides a thorough reconstruction of language group sizes in Australia at the time of colonization (1788), proving with robust numbers the abundance of very small languages unattested anywhere beyond Australia (note also the seven-page Appendix to the paper with estate/language data for ca. 150 cases of Australian linguistic varieties, in itself definitely valuable). He then proposes a theoretical explanation for the observed group sizes and their dynamics, where a state of equilibrium, ideologically based on the famous Australian

link between land and linguistic identity, is diversified by episodes of punctuation conditioned rather by external factors. Finally, he tentatively puts forward a hypothesis that the Pama-Nyungan expansion was triggered by a catastrophic population collapse outside the modern non-Pama-Nyungan zone at ca. 3,800 years BP, caused itself by epidemics brought by Asian seafarers via northern Australia. Last but not least, Sutton contributes to the main topic of the volume by suggesting that Australian evidence argues that "languages are not automatically endangered by their smallness" (p. 370), and that small hunter-gatherer languages can thrive in their multitude for millennia when not disturbed by spreads of agriculturalists.

Mark Harvey's paper "Language and population shift in pre-colonial Australia: Non-Pama-Nyungan languages" discusses mainly non-Pama-Nyungan languages of Australia with a particular focus on regional and supraregional changes in the geographical locations of languages. These are meticulously reconstructed for three groups of languages based on (a) lexical and grammatical borrowings dated for protolanguage levels, (b) significant differences in age between families deduced from the level of reconstructability of each particular

protolanguage, and (c) internal structure and overall transparency of place names for the current territory of the languages involved. To these non-Pama-Nyungan spreads, Harvey adds some observations on the directions of a supraregional spread for Pama-Nyungan languages, based in turn on the geographic differences in their internal diversity and in the space they occupy. Both types of spreads are framed by the author as examples of a possibility which turns out to be not so rare after all: the spreading of language families without rapid increases in population based on the adoption of new modes of production. So, while quite specific in its descriptive part, this paper suggests rather broad methodological conclusions, and also sets an illustrative example of implications that can be drawn from a thoughtful combination of a historical-linguistics approach and attention to geographies of languages.

Patrick McConvell's paper "The spread of Pama-Nyungan in Australia" is a mix of methodological reflections on the typology of language spreads, with a focus on foragers' spreads, and of analysis of specific semantic changes in Pama-Nyungan. The paper looks rather like a summary of previous work, as most of its claims are not supported by direct evidence,

but readers are referred to various published works, many of them by the author himself. For example, McConvell insists on the validity of Pama-Nyungan as a linguistic family, arguing with Dixon and presumably other skeptics, but an outsider to Australian linguistics like myself would rather be interested in more substance of the debate than in a plain advertisement of the proposed view. The analysis of semantic changes includes case studies from the kinship domain, as well as from terms for directions, flora, and fauna. They are meant to support the main methodological idea that semantic changes in these domains occur in a language (family) during its spread to uninhabited or sparsely inhabited territory, while borrowings indicate a language (family) spreading to densely populated territory with an active presence of other languages. This is an instance of McConvell's more general typology of "upward (or skirting)" vs. "downward (or encroaching)" forager spreads (cf. Evans & McConvell 1998; McConvell 2001, 2010), discussed in detail in another paper of the volume, that by Hill. Finally, it could be mentioned that McConvell's paper is not very friendly to linguists unfamiliar with these languages or at least with Australian geography (e.g. of 10 case studies, only one is

accompanied by a map), and the descriptions of particular cases are not always detailed and precise, resembling more illustrations than evidence-based argumentation.

The only part of the volume that deals with languages somewhat familiar to me is "Northeastern Eurasia". It includes only two papers: one by Vajda, "Typological accommodation in Central Siberia", devoted to the grammatical evolution of the Yeniseian family of Central Siberia, and the other one by Anderson & Harrison, "Hunter-gatherers in South Siberia", devoted to language-encoded cultural data from the two South Siberian Turkic languages Tofa and Todzhu.

Edward J. Vajda demonstrates how Yeniseian languages, originally prefixing, accommodated through centuries to the areal suffixing profile set by neighboring Samoyedic, Tungusic, and Turkic languages. This has been achieved by (a) innovating a new root position at the leftmost edge of the phonological verb, and (b) developing nominal relational morphemes into a case-like system. The scenario is well presented and overall convincing, even though the sociolinguistic argument is a bit underdeveloped. Besides, it is not always clear from the prose whether the author has skipped an exposition of some evidence due to the lack of space or

there is no solid evidence available to him yet. Examples are the cases of borrowings in the domains of spiritual culture and family (p. 474), the Kets' unilateral bilingualism in the past (p. 475), and the prefixing nature of the Proto-Yeniseian verb (p. 488).

Gregory D. S. Anderson & K. David Harrison take it as a starting point that the foraging lifeway leaves clear traces in a language, and their paper is devoted to enumeration of such traces for Tofa and Todzhu (with significantly more data from Tofa), foragers speaking Turkic languages. They assume that these traces can primarily be found in features atypical for Turkic languages, which are well known for the pastoralism of their speakers. Crucial for their argument is historical evidence: two travelers of the 18th century described the Tofas as a group with a Samoyedic language, and one of them even recorded some Tofa words, undoubtedly Samoyedic. A possible Yeniseian substrate has also been suggested in the literature for Tofa and Todzhu. The paper itself does not focus on linguistic attribution of the substrate, but rather on its connectedness to the hunter-gatherer lifestyle. The connection is quite convincing and supported by parallels from other hunter-gatherer languages (with a reader directed

to references for more information). However, from a Uralic, or even Samoyedic perspective, one cannot help noticing that no Tofa or Todzhu feature discussed by Anderson & Harrison could be attributed specifically to Samoyedic, at least without a separate dedicated study. In other words, the Samoyedic character of the substrate is taken for granted, which is not precisely accurate given its importance for the central claim of the paper: it is from their previous language(s) that these people brought the un-Turkic features corresponding to their forager lifestyle. The paper would have gained from a more open treatment of the potential sources for the non-Turkic features, including a possible description, here missing, of the ethnolinguistic history of the area and of the current neighbors of Tofa and Todzhu.

Willem J. de Reuse's paper "Primitivism in hunter and gatherer languages: The case of Eskimo words for snow" revises the discussion in the recent literature of the Eskimo terms for 'snow'. While it has recently been successfully shown that there are two, or three at most, general terms for snow, and not only several dozen specific terms, this paper brings forward an anthropological consequence of this linguistic analysis. De Reuse not only focuses the reader's attention on the Eskimo ability for logical reasoning and generalizations evidenced by the lexicons of their languages, but also deconstructs the unfortunate myth from the linguistic point of view. The main part of the paper breaks down the details of Eskimo grammar which obscure the parallels with English, as an example of a food-producer language. Once the parallels are restored, the Eskimo system no longer looks so unusual, though the level of elaboration in the domains of snow and ice vocabulary is clearly higher given its importance for survival in the Arctic region.

The paper by **Richard A. Rhodes**, "Language shift in the Subarctic and Central Plains", explores the last 500 years of the history of Algonquian-speaking hunter-gatherer groups of the Great Lakes area in North America, with a particular focus on Cree and Ojibwe-Potawatomi. It has been designed with two methodological aims in mind, and so can be of particular interest not only for those curious about the linguistic history of North America, but much more broadly. On the one hand, a careful description of language spreads in a recent period with written historical documentation can, by the author's intention, positively contribute to building a typology of spreads, invaluable for reconstruction of prehistoric spreads. On the other hand,

which is more immediately relatable to most readers, Rhodes discusses actual Algonquian spreads of the Great Lakes as examples of logically possible types of language spreads, sometimes drawing parallels from more well-documented European ethnolinguistic history. The paper is very dense in the data discussed, though Rhodes' style of presentation is not always friendly to those unfamiliar with languages and geography of the area. For example, as many as six maps for different historic periods are provided, but the language names featured on them are inconsistent, presumably reflecting various traditions, and many of the maps lack geographic details referred to in the text (e.g. names of lakes or of modern states); the description of vowel reduction (p. 568) could have gained a lot from a table, a chart, or a map that would help a non-specialist to follow the author's logic, etc.

Jane H. Hill in her paper "Uto-Aztecan hunter-gatherers" checks the hypothesis formulated by Evans and McConvell (1998) and McConvell (2001) on the general features of hunter-gatherers' spreads against the data of the Northern Uto-Aztecan languages spoken in the USA. The hypothesis predicts that "downstream" (later also called "encroaching" by McConvell, see above) spreads of languages into

densely inhabited territories will find a linguistic reflection in more significant substratum phenomena as compared to cases of "upstream" (or "skirting") spreads into mostly empty territories. Hill discusses this hypothesis in detail and also summarizes some other ideas expressed on sociogeographical factors in hunter-gatherer language variation and spreads. I find this methodological section particularly appealing, not only for a good collection of relevant features, but also for its accessible writing style. Returning to the Northern Uto-Aztecan languages, its Takic and Numic subgroups are taken as perfect examples of the two kinds of spreads, "downstream" and "upstream", since the former migrated to Southern California, famous for its linguistic diversity, while the latter spread into the relatively unpopulated Great Basin; the two spreads largely coincide in their dating. Surprisingly, Hill discovers that neither lexical diversity, measured as the percentage of non-Uto-Aztecan words in a wordlist, nor typological diversity, calculated with the help of the WALS inventory of features (Haspelmath et al. 2005), indicate any difference in the amount of substratum within the languages of the two subgroups. As discussed at the end of the paper, this result indicates that the reality is more complicated than the

model: effects of particular events in some historical periods can be blurred by subsequent events, substratum effects are hard to measure in a uniform way through various languages, and there might be more sociolinguistic details to take into account, e.g. the linguistic effects of outside females marrying into a community, that may run parallel to any type of spread.

Alejandra Vidal & José Braunstein in their paper, somewhat misleadingly entitled "The Southern Plains and the continental tip", trace the general human and linguistic history of the last 4000 years for the Gran Chaco area in South America, which comprises ca. 50 social units speaking 18 different languages from 6 distinct genealogical groups. Due to the ecology of the region, it was impossible to carry out agricultural activities here, so that not only were the original Gran Chaco population foragers, but groups migrating to the area also reverted to foraging, at least partly. The authors describe the consecutive stages in the peopling of the area and the linguistic consequences of these events. They suggest specific patterns of inter- and intraethnic communications that are responsible for unexpectedly high rates of linguistic divergence (for related languages) and of linguistic convergence (for all languages, including the unrelated ones) observed today.

4. Conclusion

The overwhelming majority of the papers of the volume dealt with languages I knew very little about, but while reading them I could not help returning in my mind to the history of Samoyedic, my personal area of interest. That was because different scenarios attested in the history of those languages repeatedly suggested to me new questions I could apply to my own data. This is indeed one of the basic aims of linguistic typology: to supply descriptive linguists with an array of possibilities attested in the world's languages, so that each feature of every particular language could be described in reference to that pool of options. This way, more thorough descriptions can be achieved, which in their turn will successfully feed the next round of typological, and ultimately theoretical, thinking. The same logic applies here: for advancing the field of language evolution and history, in the broadest sense of these terms, it is crucial to treat the history of each family in the context of other known histories and spreads. To make such an analysis possible, every historical linguist should make use of comparable terminology and be fully explicit in one's reasoning,

setting aside the cryptic conversation of specialists only talking to each other that has been far too common in the comparative-historical field, including Uralistics.

Finishing this enthusiastic review, I find it appropriate to express my acknowledgement to the editors of the volume for putting all these papers together. Every contribution being solid on its own, jointly they picture an impressively detailed portrait of hunter-gatherers' languages and bring historical-typological linguistics to a new level. By far, this could be one of the first attempts to build a typology of language spreads bottom-up, without promotion of any specific theory, but giving equal voice to a diversity of languages, types of evidence, and ways of reasoning.

Olesya Khanina

References

Bellwood, Peter. 2013. *First migrants: Ancient migration in global perspective*. Chichester: Wiley-Blackwell.

Berlin, Brent. 1992. *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.

Evans, Nicholas & McConvell, Patrick. 1998. The enigma of Pama-Nyungan expansion in Australia. In Blench, Roger & Spriggs, Matthew (eds.), *Archaeology and language II*, 174–192. London: Routledge.

Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.). 2005. *The World atlas of linguistic structures*. Oxford: Oxford University Press.

Khanina, Olesya. 2021. Languages and ideologies at the Lower Yenisei (Siberia): Reconstructing past multilingualism. In Dobrushina, Nina & Khanina, Olesya & Pakendorf, Brigitte (eds.), special issue "Typology of small-scale multilingualism" of *International Journal of Bilingualism* 25(4). 1059–1080.

Lee, Richard B. & Daly, Richard (eds.). 1999. *Cambridge encyclopedia of hunters and gatherers*. Cambridge: Cambridge University Press.

McConvell, Patrick. 2001. Language shift and language spread among hunter-gatherers. In Panter-Brick, Catherine & Layton, Robert H. & Rowley-Conwy, Peter (eds.), *Hunter-gatherers: An interdisciplinary perspective*, 143–169. Cambridge: Cambridge University Press.

McConvell, Patrick. 2010. The archaeolinguistics of migration. In Lucassen, Jan & Lucassen, Leo & Manning, Patrick (eds.), *Migration history in world history: Multidisciplinary approaches*, 155–190. Leiden: Brill.

Panter-Brick, Catherine & Layton, Robert H. & Rowley-Conwy, Peter. 2001. Lines of enquiry. In Panter-Brick, Catherine & Layton, Robert H. & Rowley-Conwy, Peter (eds.), *Hunter-gatherers: An interdisciplinary perspective*, 1–11. Cambridge: Cambridge University Press.

Nikolaj Isanbaev 1929–2020

Professor Nikolaj Isanbaevič Isanbaev, der auf eine lange und vielseitige Laufbahn im Bereich der marischen Sprachforschung zurückblicken konnte, verstarb am 24. September 2020 im ehrwürdigen Alter von 91 Jahren. Er war am 27. Februar 1929 im Dorf Nižnij Kačmaš im Bezirk Kaltasy in Baschkortostan geboren. In Baschkortostan hat sich die marische Kultur stellenweise sehr vital erhalten, und auch die schulischen Verhältnisse waren dort im Hinblick auf die Muttersprache günstig. Daher hat das Gebiet viele marische Sprachwissenschaftler hervorgebracht, die ihre akademische Ausbildung in Joškar-Ola erhielten.

Nachdem er in Kaltasy die Schule besucht hatte, erhielt Isanbaev 1944 einen Studienplatz an der marischen Abteilung des pädagogischen Instituts in Krasnokamsk, wo er seine Ausbildung 1948 abschloss. Er setzte sein Studium in der Abteilung für marische Sprache und Literatur der historisch-philologischen Fakultät des nach N. K. Krupskaja benannten pädagogischen Instituts von Mari in Joškar-Ola fort. In den Jahren 1952–1955 absolvierte er ein weiterführendes Studium in der Aspirantur am sprachwissenschaftlichen Institut der Akademie der Wissenschaften der UdSSR in Moskau. Zu dessen Abschluss verfasste Isanbaev 1955 unter Leitung des Akademiemitglieds Boris Aleksandrovič Serebrennikov eine Kandidatur-Dissertation über die *n*-Gerundien im Mari. Sein erster Artikel, den er auf der Grundlage seiner Kandidatur-Arbeit schrieb und der sich auf die syntaktischen Funktionen des *n*-Gerundiums konzentrierte, erschien 1958 in der Zeitschrift *Nyelvtudományi Közlemények*. Die Analyse der Gerundien setzte er auch später fort: 1961 erschien seine Monographie über die Gerundien des Mari (*Деепричастия в марийском языке*).

Nach der Aspirantur arbeitete Isanbaev eine Zeitlang am sprachwissenschaftlichen Institut der Akademie der Wissenschaften, kehrte dann aber nach Mari zurück, an die Sprachabteilung des Forschungsinstituts von Mari. Von dort ging er für einige Jahre an die Abteilung für Russisch des pädagogischen Instituts von Mari, bis er 1960 an das Forschungsinstitut zurückkehrte. Er war lange als Leiter der Sprachabteilung tätig. In den Jahren 1996–2013, d.h. bis zu seiner Pensionierung, war Isanbaev

gleichzeitig auch Professor am Finnisch-ugrischen Institut der staatlichen Universität von Mari.

Isanbaev promovierte 1993 über die marisch–türkischen Sprachkontakte. Er hatte 1989 den ersten Teil seiner Monographie *Марийско-тюркские языковые контакты* veröffentlicht, in dem er den historischen Hintergrund dieser Kontakte und die Geschichte ihrer Erforschung untersuchte und auch die Kriterien erörterte, mit deren Hilfe baschkirische und tatarische Lehnwörter voneinander zu unterscheiden sind. Auch mit den westlichen Quellen, von von Strahlenberg bis zu Bereczki, war Isanbaev genau vertraut. Der zweite Teil der Monographie erschien unter demselben Titel 1994; es handelt sich um eine wörterbuchförmige Darstellung der tatarischen und baschkirischen Lehnwörter im Mari. Die Zahl der Lemmata ist erheblich größer als im einschlägigen Klassiker, der Monographie *Die tatarischen Lehnwörter im Tscheremissischen* (1923) von Martti Räsänen, denn Isanbaev kannte und erforschte die Ostdialekte des Mari, und gerade in diesen gibt es erheblich mehr tatarische Lehnwörter als in den anderen Dialekten. Im Wörterbuch bezieht er auch zu den Interpretationen früherer Forscher wie Räsänen und M. R. Fedotov systematisch Stellung.

Isanbaev beteiligte sich aktiv an der Abfassung mehrerer wichtiger elementarer Werke zur marischen Sprache. Er arbeitete in der Redaktion der ersten acht Bände (1990–2003) des 11-bändigen erläuternden Wörterbuchs des Mari (*Словарь марийского языка*). Auch an der Abfassung der Bände über Morphologie und Syntax (1961) der dreiteiligen Grammatik des Mari (*Современный марийский язык*) war er beteiligt.

Isanbaev befasste sich in seinen Untersuchungen mit einem breiten Spektrum von Themen. In seinen Artikeln untersuchte er die Phonetik (zum Beispiel das Lautsystem einzelner Ostdialekte), die Morphologie und die Syntax, und er war in der synchronischen Sprachforschung ebenso bewandert wie in der diachronischen. Über diese Themenbereiche hielt er auch Vorlesungen an der Universität und leitete Aufbaustudierende an.

Isanbaev erhielt sowohl in der Zeit der Sowjetunion als auch Russlands mehrere staatliche Auszeichnungen, er war u.a. Verdienter Wissenschaftler von Marij El und Veteran der Arbeit. 1995 wurde Isanbaev zum assoziierten Mitglied der Finnisch-Ugrischen Gesellschaft berufen.

Sirkka Saarinen

Ksenofont Nikanorovič Sanukov 1935–2020

Ksenofont Nikanorovič Sanukov was born into a peasant family in the village of Nosyoli (Носёли) in the Gornomariyskiy district on 5 February 1935. After beginning his studies at the teaching institute in Kozmodemyansk (1953), Sanukov transferred to the Mari State Pedagogical Institute, from which he graduated in 1958. He devoted himself to a career in research and completed his postgraduate studies in history at the History Institute of the Academy of Sciences of the Soviet Union, graduating as a Candidate of Sciences in 1970 and as a Doctor of Sciences in 1986. He was appointed Professor in 1989.

Sanukov's career was varied and multifaceted. He worked as a teacher at his alma mater, the Mari State Pedagogical Institute (1958–1960), as a journalist for the youth magazine *Молодой коммунист* ['Young Communist'] (1960–1962), as a researcher at the Mari Scientific Research Institute (1962–1966) and also as its Director (1982–1986). Sanukov is credited with reforming the teaching of history in Mari institutes of higher education. He was Director of the History Department at the Mari State Pedagogical Institute from 1986 to 1991, after which he served as Professor and Head of the Regional History Department at Mari State University in 1993–2003. The purpose of the school of thought he established was to examine historical phenomena through the widest possible lens, including the difficult and unspoken stages of history, and to approach Mari history as a part of the Mari people's identity and of general human history. (Л. П. Москвина & С. К. Свечников: "Научная школа К. Н. Санукова" ['The Scientific school of K. N. Sanukov'], *Вестник Марийского государственного университета* 3/2016.)

His students valued their teacher, and more than ten doctoral dissertations were completed under his supervision, including those of L. P. Moskvina, S. K. Svečnikov, R. I. Čuzaev and O. N. Sutyrina, to name a few. In addition to earning the appreciation of his students, Sanukov was awarded a state prize from the Mari Republic in 2005 and presented with a Second Class Medal of the Order "For Merit to the Fatherland" (Медаль ордена «За перед Отечеством») in 2006 in recognition of his academic career.

Sanukov's life's work can be viewed from three different perspectives: as an academic, as a populariser of science and as an advocate for Finno-Ugric cooperation.

The primary focus of Sanukov's research was political history. His doctoral thesis, *Рабочий класс – ведущая сила экономического сотрудничества народов СССР (60-70-е годы)* ['The working class – the leading force in the economic cooperation of the peoples of the USSR in the 1960s and 70s'] (1986), explored the recent past of Soviet economic policy. After this, in the permissive atmosphere of *glasnost* and *perestroika*, Sanukov's interests turned to the 1930s, which represented a turning point in Mari history. He focused his studies on the foundation of the Mari state in the light of Mari culture and the Mari identity. He was convinced that the Mari people had their own history within the framework of Russian and Soviet history. He familiarised himself with the works of pioneers of Mari historical research, including V. A. Muhin, A. K. Eškin and I. N. Smirnov, and with the influential historians of the 1950s, such as N. P. Kalistratov and A. V. Hlebnikov. Based on primary archival sources, Sanukov established terminology for the field of history and began studying the stages of Mari regional administration and the subsequent foundation of the Republic in an objective and source-critical manner. He shed light on previously unspoken periods of history, including the famine and Stalin's purges. Sanukov published a series of monographs on these topics: *Марийцы: Прошлое, настоящее, будущее* ['The Mari: Past, present, future'] (1992), *Из истории Марий Эл: Трагедия 30-х годов* ['On the history of Mari El: The tragedy of the 1930s'] (2000) and *Марийская автономия* ['Mari autonomy'] (2010). Following an objective investigation into these difficult stages of Mari history, he published the volume *Трагедия народа: Книга памяти жертв политических репрессий Республики Марий Эл* (1996-1997), a book in memory of the victims of political repression in the Republic of Mari El. Sanukov's written *oeuvre* is extensive. Some of his articles have been published in German and English on the interpretation of the Mari national identity ("Historische Voraussetzungen der nationalen Identifikation", Veröffentlichungen der Societas Uralo-Altaica 66, Wiesbaden 2005, 1–23) and Stalin's persecution ("Stalinist terror in the Mari Republic: the attack on 'Finno-Ugrian bourgeois nationalism'", *The Slavonic and East-European Review* 74, 1996, 658–682). His work on Mari history has been translated into Hungarian (*A cseremiszek múltja, jelene, jövője* ['The past, present and future of the Mari'], Budapesti Finnugor Füzetek 2, 1996).

Sanukov also presented the findings of his research to the broader audience. Together with historian A. G. Ivanov, he published a textbook on

Mari history, *Марий калыкын историйже: кокла ийготан да кугурак классыште тунемше-влаклан лудшаш книга* ['History of the Mari people: A textbook for secondary and high school students'], first in Mari in 1998 and then in Russian in 1999. In addition to studying the history of the Mari community, he also explored the fates of people who influenced the course of history, in works such as *Председатель исполкома: Очерк жизни и деятельности И. П. Петрова (1893–1938)* ['Chairman of the Executive Committee: A study of the life and work of I. P. Petrov'] (2000), *Наши земляки: Пути и судьбы* ['Our countrymen: Paths and destinies'] (2011) and *Из истории Марий Эл: страницы известные и неизвестные* ['On the history of Mari El: Pages known and unknown'] (2013). He was instrumental in establishing the Mari regional chapter of the Memorial Society and served as its Chair in 1989–1996 and 2000–2003.

Sanukov was widely known and respected throughout the Finno-Ugric world. He was Chair of the Mari El–Hungary Society from its inception in 1986 until its disbanding in 2006. Sanukov was one of the organisers of the first Finno-Ugric Literary Congress, which took place in Yoshkar-Ola in the summer of 1989. He participated actively in Finno-Ugric scientific and political events in Germany, Finland, Estonia and Hungary. The scientific journal *Финно-угроведение* was founded at Sanukov's initiative in 1994, and he served as its editor-in-chief until 2018. He was Secretary-General of the 10th International Finno-Ugrists' Congress held in Yoshkar-Ola in 2005.

Sanukov was a socially conscious and highly visible figure in the public forum. He encouraged and supported the activities of Mari Ushem (Марий ушем), participated in its annual meetings and presented an extensive overview of Mari history to the organisation in 2016. He was one of the few cultural influencers who dared to oppose the erection of a memorial to the Russian warlord Obolenskiy-Nogotkov in the centre of Yoshkar-Ola in May 2007.

Professor Ksenofont Sanukov passed away on 14 October 2020.

Ildikó Lehtinen

Zoja Zorina 1946–2020

Die kleine Schar der gebürtigen bergmarischen Wissenschaftler schmolz am 30. Oktober 2020 noch weiter zusammen, als das Leben der vor allem als Phonetikerin bekannten Professorin Dr. phil. Zoja Zorina endete. Zorina wurde am 1. November 1946 in dem Dorf Pernjangaš im Gebiet der Bergmari geboren. Nachdem sie eine Internatsschule in Kozmodemjansk, der Hauptstadt des Gebiets, besucht hatte, erhielt sie einen Studienplatz an der Fakultät für Fremdsprachen des nach N. K. Krupskaja benannten staatlichen pädagogischen Instituts von Mari in Joškar-Ola. Dort machte sie eine Ausbildung zur Lehrerin für Deutsch und Französisch. Nach einigen Jahren im Schuldienst wurde sie 1972 Lehrerin an der Staatlichen Universität von Mari.

In den Jahren 1979–1982 setzte Zorina ihr Studium unter Leitung von Prof. Ljudmila Verbickaja, der angesehenen Erforscherin der Phonetik des Russischen, in der Aspirantur an der Universität Leningrad fort, wo sie 1982 im Bereich der experimentellen Phonetik ihre Kandidatur-Dissertation über den regionalen Wechsel und die Variabilität der Vokale des Bergmari vorlegte. Danach setzte sie ihre Tätigkeit an der Universität von Mari mehr als 30 Jahre lang fort, zuerst als Dozentin, später als Professorin und Leiterin des Instituts für Fremdsprachen. Das Studienjahr 1985/86 verbrachte Zorina als Stipendiatin in Frankreich. 1997 verteidigte sie an der Universität St. Petersburg ihre Doktordissertation über die phonetischen Besonderheiten des Russischen in einer zweisprachigen Umgebung; die Arbeit beruhte auf dem Vergleich der russischen Sprache der in der Republik Mari lebenden Berg- und Wiesenmari und Russen.

Zorina veröffentlichte über 80 Untersuchungen, von denen viele die Phonetik behandelten – nicht nur das Lautsystem des Bergmari oder allgemeiner der marischen Sprachen, sondern auch generell das Erlernen der Laute einer Fremdsprache, zum Beispiel die Frage, wie Sprecher des Mari oder des Russischen die Vokale im gegenseitigen Sprechen wahrnehmen oder wie sie die Vokale des Französischen oder Deutschen artikulieren können. Zu den unter Finnougristen bekanntesten Monographien Zorinas zählen zweifellos die Darstellung der Lautstruktur des Bergmari, *Звуковой строй современного горномарийского языка* aus dem Jahr 2007 und die Bestandsaufnahme des Kasussystems des Bergmari *Падежная система современного горномарийского языка* aus dem Jahr 2010; beide

verfasste sie gemeinsam mit ihrer Tochter Oksana Wikström. In ihren Artikeln erörterte Zorina auch die Methodik des Fremdsprachenunterrichts sowie ethno- und psycholinguistische Fragen. Sie verfasste das erste Frequenzwörterbuch der marischen Sprachen, das unmittelbar vor dem internationalen Finnougristenkongress in Joškar-Ola erschien. Zorina nahm auch an den Finnougristenkongressen in Debrecen (1990), Jyväskylä (1995) und Piliscsaba (2010) sowie an internationalen Phonetikerkonferenzen teil und ihre Untersuchungen über das Mari erschienen auch in französischen Publikationen.

Zorina war eine tatkräftige Organisatorin, und so fielen ihr an der Universität von Mari zahlreiche administrative Aufgaben zu. Neben ihrer Tätigkeit als Institutsleiterin war sie Dekanin der Fakultät für internationale Beziehungen und für die Lehre verantwortliche Prorektorin. Sie plante die jährlich stattfindenden regionalen Konferenzen für Fremdsprachenunterricht und Kommunikation und war nicht nur an ihrer eigenen Universität, sondern auch an der Universität Nižnij Novgorod im Evaluierungskomitee für Dissertationen tätig. Ihre breit gefächerte Aktivität und Unternehmungslust bezeugt auch eine Kuriosität in ihrem Meritenverzeichnis: ein Patent für die Verwendung der Kassettentechnologie beim Kohlanbau.

Zorina erhielt neben akademischen Auszeichnungen Ehrentitel sowohl der Republik Mari als auch der Russischen Föderation, sie wurde unter anderem als Verdiente Wissenschaftlerin von Marij El und als Veteranin der Arbeit ausgezeichnet.

Sirkka Saarinen

Sananjalka on vuodesta 1959 lähtien ilmestynyt laadukasta humanistista tutkimusta julkaiseva tieteellinen vuosikirja.

Sananjalassa julkaistaan vuosittain noin kaksitoista vertaisarvioitua artikkelia suomen ja sen sukukielten, kotimaisen kirjallisuuden, folkloristiikan, kansatieteen, uskontotieteen ja arkeologian alalta. Lisäksi vuosikirja tarjoaa luettavaksi kiinnostavia katsausartikkeleita ja arvosteluja tuoreista humanistista alaa koskettavista kirjoista.

Sananjalkaa julkaisee vuonna 1929 perustettu Suomen Kielen Seura.

Sananjalka

Numerossa 63 muun muassa:

Tommi Kurki: Salienssi, subjektiivinen etäisyys ja paikannusten tarkkuus. Kansandialektologisia havaintoja lounaissuomalaisesta näytteestä

Heidi Niva: Parhaillaan mutta myöhemmin. *Olla* V-*mAssA* -rakenne ennakoinnin ja intentioiden ilmaisuissa

Lasse Hämäläinen ja Teemu Ruokolainen: Kukkaa, amfea, subua ja essoja. Huumausaineiden slanginimitykset Tor-verkon suomalaisella kauppapaikalla

Maria Takala-Roszczenko: Kirjailija Iivo Härkönen ja unelma ortodoksisesta virsilaulusta

Stina Savolainen: "Enda gången man hör eller pratar finska så är i skolan". Ruotsinkielisten opiskelijoiden kokemuksia suomen kielestä ruotsinkielisessä arjessa

Lisäksi uusimmassakin numerossa on runsaasti muita artikkeleita, katsauksia ja kirja-arvosteluja.

VIRITTÄJÄ

Vanhin ja suurin suomenkielinen kielitieteellinen lehti

Ajankohtaisia ja monipuolisia kirjoituksia suomen kielen ja sen sukukielten tutkimuksesta ja opetuksesta

Neljä numeroa vuodessa, yhteensä noin 630 sivua

Kotisivuilla laaja hakemistotietokanta julkaistuista kirjoituksista

Kirjaesittelyjä, katsauksia ja keskustelua

Aikaisempia vuosikertoja ja tuoreita lektioita verkkolehdessä

https://journal.fi/virittaja

virittajablogi.kotikielenseura.fi

Virittäjässä viime aikoina
- Santeri Junttila Osmo Nikkilän Angervojen alkuperistä (2/2021)
- Katri Priiki Verkkokeskustelijoiden käsitykset *hän-* ja *se*-pronominien norminvastaisesta käytöstä (3/2021)

Tulossa: juhlanumero 4/2021 – katsauksia 125-vuotiaan Virittäjän historiaan
- Jussi Ylikoski Fennougristiikan historiaa Virittäjän valossa
- Sampsa Holopainen, Santteri Junttila & Petri Kallio Lainasanatutkimus Virittäjässä ja muualla

Tilaushinnat 2022
- Suomeen 82 € (Kotikielen Seuran jäsenille 79 €, opiskelijoille 33 €)
- Ulkomaille 82 € (Kotikielen Seuran jäsenille 79 €, opiskelijoille 39 €)
- Verkkolehden tilaus Kotikielen Seuran jäsenille 42 €, opiskelijoille 22 €