Establishing a Role for Minority Source Language in Multilingual Facilitation

Jack Rueter, Niko Partanen, Khalid Alnajjar and Mika Hämäläinen

University of Helsinki (Rueter, Partanen, Alnajjar, Hämäläinen), University of Turku (Rueter)

Abstract

This document is dedicated to a young man, who, despite the number of times he has traveled around the Sun, is always open to new thoughts on ways to include languages, especially the smaller ones, and the people who speak them in far-reaching and sustainable open-source development. Since Trond Trosterud in Tromsø is attributed a terrific track record in transnational and circum-polar linguistics, we try to attract his attention further afield, to languages and phenomena he has only touched. The language phenomena addressed here come from Erzya and the Zyrian variety of Komi; Erzya has issues presented but not discussed in his dissertation, whereas Komi brings in issues of adnominal and predicate number marking in conjunction with case homonymy that have been resolved thanks to the flexibility of the infrastructure. These source languages, like others, have documented new dimensions and added shape to the ever-growing infrastructure.

Keywords: Erzya, Komi-Zyrian, morphology, multiargument marking

1. Introduction

The idea of establishing an infrastructure that allows work on individual languages of great diversity requires more than a minute for conception. In fact, the formulation of the concept is incomplete without the sleepless nights, nurturing, coaxing, feverish training and jovial interplay with it during its adolescence. It is especially important to note that endangered languages often deal with very different problems than majority languages (Hämäläinen 2021). From a linguistic point of view, infrastructure building requires not only dedication but involvement in the actual study of languages to be addressed on that platform of study as well as an awareness of language needs such as can be observed in earlier categorizations of copula and negation verbs (Trosterud 1994). Another and far reaching language feature subsequently considered in morphological complexity is homonymy that is addressed at an early point (Trosterud 2006) for enhanced application to comprehensive use of the infrastructure.

In 2006, Trond Trosterud published a dissertation on Homonymy in the Uralic Two-Argument Agreement Paradigms, where he addressed morphological phenomena beyond those of the Germanic Faroese Trosterud (2009), Saami Antonsen and Trosterud (2011) and Kven Trosterud et al. (2017) languages of Norway. This work was symptomatic of what is required for multilingual facilitation. On the one hand, this work lead to lexical research development, such as discussed in Antonsen et al. (2009b). It also engendered and strengthened concepts of re-usability Antonsen et al. (2010), shared development Gerstenberger et al. (2016); Rueter et al. (2021b). On the other hand, it opened doors to different dimensions of collaboration Trosterud and Moshagen (2021), and even new ones beyond the original infrastructure itself, e.g., Snoek et al. (2014), Simonenko, Alexandra (2020), Alnajja et al. (2020). This also went beyond the principals foreseen by the originators Hämäläinen and Wiechetek (2020), Khanna et al. (2021).

2. Peripheral notes on Erzya

In Trond’s dissertation, he addresses several languages, including Erzya and Moksha, whose description we endeavor to expand upon here. Trond’s description of four-dimensional paradigms in Mordvin (Trosterud 2006:246-303) gives insight into the phenomena of number and person marking of subject and object on the verb. He provides a plethora of erudite information on the two literary languages and their speakers. Here, we will attempt to put the notes on morphology into perspective, so that in his leisure he might return to this work to update and expand it.
ROLE FOR MINORITY SOURCE LANGUAGE

<table>
<thead>
<tr>
<th>PERS PRON</th>
<th>ABE</th>
<th>ABL</th>
<th>CMPR</th>
<th>COM</th>
<th>DAT</th>
<th>ELA</th>
<th>GEN</th>
<th>ILL</th>
<th>INE</th>
<th>LAT</th>
<th>LOC</th>
<th>NOM</th>
<th>PRL</th>
<th>TEMP</th>
<th>TRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>N SG DEF</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>N PL DEF</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>N PFXSG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>ADP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: cases and locus in Erzya

As noted by Trosterud (2006), Mordvin language morphology deviates from that of the Finnic languages. Whereas verbs in Erzya, Moksha and Finnic take subject marking with distinctions for the categories of person and number, the Mordvin languages also distinguish these two categories in specific subject-object marking on the verb for many instances with singular subjects. A downplayed or third person plural subject in Erzya only distinguished the person of the object. This resembles singularly object marking structures attested in languages outside the Uralic context, for example Apurinâ (cf. Facundes 2000, Rueter et al. 2021a; see also Plains Cree Harrigan et al. 2017). In the nominal system, as in Finnic, the NP head may bear declensional marking specific to the categories of case, number and possessor, but Mordvin includes yet another category – definiteness. All of these categories singularly or in combination might also be followed by a so-called second declension (used for coping with contextual ellipsis), copula person and number marking, which could then be followed by an additive clitic. Hence, in addition to subject-object marking strategies with person and number combinations observed in verbal conjugation, a less frequent phonemon of four-dimensional marking in possessor-index with subject conjugation is also attested in nominals and adpositions, see (1a) below. Furthermore, the NP head in the inessive, abessive, prolate, comparative and translativa also function as adnominal attributes (Rueter 2010:19–22), Erzya and Moksha exhibit so-called secondary declension when the new NP head is lost through contextual/retrievable ellipsis¹. Perhaps, this will provide for further dimensions to syntax analysis research in Saami language research (cf. Antonsen et al. 2010, Sheynova and Tyers 2017, Rueter and Hämaaläinen 2020a) and beyond (cf. Rueter and Tyers 2018).

Whereas (Trosterud 2006:248) presents a breakdown of the Erzya case systems for number of core, local and other semantic cases with reference to Moksha as well, a rather difficult undertaking, since the two languages have at least slightly divergent polifunctionality in their case systems. For this reason, it might be more to the point of indicating three core cases (NOM, GEN, DAT) and an additional two (ABL, INE), which are also used as object markers (cf. Bartens 1999:91–94, 96; Grebneva 2000:76; Grünthal 2008:229-230, example 6). Likewise, when we speak of basic local cases we must mention a semioverlapping set (ABL, INE, ILL, ELA, LAT, LOC, PRL), which are found in the nominals, postpositions and nouns of deficient declension in the language. The number of cases varies from ten in Gabalenzi (1839), and nine in Wiedemann (1865) to fifteen in Rueter (2010). Each enumeration of case is based on different criteria. On the one hand, personal pronouns, which might not exceed nine cases, can be pitted against complex NP heads, which may be found in at least fifteen distinguishable forms, on the other.

In Table (1), NP cases have been dealt with, where possible, using binary + vs – notation. The NP heads serving as locus can be enumerated as follows: PERS PRON ‘personal pronouns’ (e.g., SG3), N INDEF ‘indefinite noun’, N SG DEF ‘definite singular noun’, N PL DEF ‘definite plural noun’, N PX3SG ‘noun with a 3rd person singular possessive suffix’, and ADP ‘adposition or relational noun’². The pronoun, definite singular and possessive paradigms are smaller than those of the indefinite and definite plural. Some forms, such as the definite plural comitative are only attested in individual dialects (cf. Bartens 1999:91–94, 96; Grebneva 2000:76; Grünthal 2008:229-230, example 6). Likewise, when we speak of basic local cases we must mention a semioverlapping set (ABL, INE, ILL, ELA, LAT, LOC, PRL), which are found in the nominals, postpositions and nouns of deficient declension in the language. The number of cases varies from ten in Gabalenzi (1839), and nine in Wiedemann (1865) to fifteen in Rueter (2010). Each enumeration of case is based on different criteria. On the one hand, personal pronouns, which might not exceed nine cases, can be pitted against complex NP heads, which may be found in at least fifteen distinguishable forms, on the other.

¹Let’s say there is a book on the table and another on the shelf. A student takes the one on the shelf. In English the word *one* covers for the contextual book. In Finnic and Mordvin, however, the noun completely disappears, but even the Finnic requires a verbal form olevan ‘the one that is’ opiskelija ottaa hyödyllä olevan (lit. ‘student takes on the shelf one that is’), whereas the NP head morphology locus in Erzya simply shifts to the attribute, i.e., tonavtńićaś saji lavśa langsońtśęńtś (lit. ‘the student takes shelf one that is’). ²The relational noun kedikele ‘in the vestibule’ (aka postposition or adposition), although represented by a defective paradigm without a nominative form, can be found in Erzya literature in the definite singular declension, e.g., kedikele-nti ‘vestibule.PRL-DEF.SG’ kedikele-nti tago mrađevis lāšča. ‘Once again someone could be heard going out through the vestibule’.

232
The Erzya language like its sibling Moksha has possessive suffixes as well as verbal and copula conjugation marking, all three of which contribute to the extensive morphologies in the Mordvin languages. Since (Trosterud 2006:264–303) provides ample analyses of verbal conjugations in Moksha and Erzya, there are only certain pieces of information that need to be addressed here, namely possessive person and copula person markers.

The shape of the nominal paradigms can be split into two types. Number is a distinguishing category of the definite declensions, whereas this distinction is only found in the nominative case of the basic or indefinite and possessive declensions. Actually, here the Saami languages documented in Giellatekno describe the abessive case without the category of number, and this idea of zero number can be readily applied to the Mordvin languages for the basic or indefinite declension cases other than the nominative. In the possessive declension, however, the Moksha language actually distinguishes number for three cases, the nominative, genitive and dative, and therefore Erzya has initially been documented with the category for number in these same three cases so as to render symmetric tagging for parallel tool development. The question of whether this should really be necessary is an issue for further development in dialect research Rueter (2020); Rueter et al. (2020a), and shallow-transfer machine translation Rueter and Hääläinen (2020b), which will also benefit from research at Giellatekno, as alluded to in Trosterud and Antonsen (2020).

In table 2, it will be noted that if there is a distinction made in the form of the possessive marker, then this distinction will be observed in the nominative singular versus other number or case categories. In Erzya, only the third person singular makes this clear distinction in the modern literary language jalga ‘(NOM.SG) his/her friend’ versus jalganzo ‘(GEN.SG; GEN.PL; NOM.PL) his/her friend’s/friends’/friends’, whereas the first person singular may become more syncretic following the Southeastern dialect m, used in all positions. This information should also be applied to (Trosterud 2006:300–303), where an enhanced understanding of the Erzya paradigm is required, i.e., of the five tables presenting Erzya possessive declension, only table 216, which presents the nominative case, does not require editing.

### 2.2. Verbal conjugation in Erzya

In an introductory to conjugation Trosterud (2006), Trosterud provides information on the non-past subject conjugation of the verb śudo|ms ‘to scold’ (see table 3), which does not contain two consecutive consonants but does, indeed, retain its stem vowel in the first and second persons plural. Here a vertical line indicates the break between stem and infinitive marker characterized as -Oms, where the upper-case O is an archi-vowel, indicative of an obligatory vowel – either the stem vowel or a middle vowel o, e as assigned by palatal-vowel
Role for minority source language

<table>
<thead>
<tr>
<th>Sing</th>
<th>Plur</th>
<th>Sing</th>
<th>Plur</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 šudan</td>
<td>šudotano</td>
<td>-An</td>
<td>-Tano</td>
</tr>
<tr>
<td>2 šudat</td>
<td>šudotado</td>
<td>-At</td>
<td>-Tado</td>
</tr>
<tr>
<td>3 ūdū</td>
<td>ūdūṭ</td>
<td>-ṭ</td>
<td>-ṭ</td>
</tr>
</tbody>
</table>

Table 3: the verb šudo|ms ‘to scold’ (not šodoms) Serebrenikov et al. 1993:635

<table>
<thead>
<tr>
<th>Sing</th>
<th>Plur</th>
<th>Sing</th>
<th>Plur</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 viḍan</td>
<td>viḍṭano</td>
<td>piḍan</td>
<td>piḍeṭano</td>
</tr>
<tr>
<td>2 viḍat</td>
<td>viḍṭado</td>
<td>piḍat</td>
<td>piḍeṭado</td>
</tr>
<tr>
<td>3 viḍi</td>
<td>viḍṭ</td>
<td>piḍi</td>
<td>piḍṭ</td>
</tr>
</tbody>
</table>

Table 4: the verbs viḍ|ems ‘to sow’ and pide|ms ‘to cook’ Serebrenikov et al. 1993:133, 476

Since then, it has become apparent, upon further scrutiny of the Erzya language, that there actually are near minimal pairs to be found among the verbs viḍ|ems ‘to sow’, viḍ ‘he/she sowed’ and pide|ms ‘to cook’, pide̊ ‘he/she cooked’ illustrating single consonants between two vowels. Here, the single consonant d̊ in viḍ|ems represents the stem final consonant, whereas the segment -ems is, in fact, the infinite ending (see Rueter 2016:131).

2.3. Copula complement marking in Erzya

In Trosterud 2006:251–252, the author presents extensive copula paradigms for the consonant-stem noun sazor ‘little sister’ and the mixed stem noun ovto ‘bear’, so it is important that we add the missing vowel-stem noun type jalga ‘friend’, and, instead of paradigms from an older writing tradition as might be found in (Evsev’ev 1928–29 and Bartens 1999:130–131), we can provide a modern paradigm, which, of course, still requires extensive commenting. The first comment is one to explain ordering of variants in individual paradigm cells, namely, as in our morphological development YAML testing, the first and left-most word form represents the desired form for computer generation.

Vowel-stem nouns never lose their stem-final vowel in the copula declination. The orthography that has developed since the late 1920s, however, has introduced additional syncretism, i.e., whereas the Pre-Soviet standard language (Northwestern dialect) distinguishes the copula form jalgaʃan ‘I am a friend’ from the possessum form jalgaʃ ‘my friends’, the modern standard embraces either a single form jalgaʃan to convey both meanings, or it adds a Southeastern dialect form jalgaʃel, which is also syncretic in that it means both ‘my friend’ and ‘my friends’. In the modern Central dialect standard, it must be noted the 1pl form has no final k in the non-past regardless of whether it is attached to vowel-stem, consonant-stem or mixed-stem nouns.

Table 5: Copula person and number for the vowel-stem noun jalga ‘friend’

234
For many speakers of Erzya, including many grammar writers, the presence of definite noun forms with copula marking is atypical or marginal in the Erzya language (c.f. Evsev’ev 1928-29:125–136, 149-151, 313). The infrequency of these forms in Erzya may, actually, be attributed to the fact that copula person marking occurs on the complement. Thus, both the copula subject and complement must be conceived as highly salient, and the first or second person must be established, i.e., this will not occur in introductions where the proper names are conceivably familiar, but the first and second person pronouns do not indicate a previously establish referent, and therefore are treated as the copula complements.

The mixed-stem noun type, described phonetically in (Rueter 2010:72–73), helps to explain the absence of stem middle vowels in the plural stem of ovto ‘bear’. In table (7), for example, the second person plural past tense has conceivably three valid forms. The first involves the basic nominative singular form ovto form, followed by the past tense marker and second person plural -ĺid́e. The second and third forms take the basic nominative plural as their base (ovtot and ovtt), which are also followed by the past tense marker and second person plural -ĺid́e. The distinction, therefore, is that in the mixed-stem type the final middle vowels o and e (preceded by a soft stem) tend to be dropped before NOM.PL, INE.INDEF, ILL.INDEF and ELA.INDEF markers. So far, no instances of word-final e preceded by a hard dental have been encountered in this stem type.

All of the examples given in the three tables above (5, 6, 7) address copula person marking on indefinite and definite nouns. Copula person marking can occur with other case marking as well, e.g., it can also occur with the inessive, and both of these cases can, indeed, occur in the possessive declension as well. It will be noted, of course, that nominals do not offer the spate of homonymy that verbs do, but they do present their own variety of four-dimensional marking. Let it suffice to present a nominative singular possessum, with a third person singular possessor and a second person singular copula marker in (1a) and an indefinite inessive possessum with a third person singular possessor, followed by a past tense marker and a first person plural copula marker in (1b). Further information on this can be found in (Gabelentz 1839:237, 402), (Evsev’ev 1928-29:115–125), Turunen (2010) and Rueter (2013).

(1) a. śińdře, ton avol trond-oń ċora-z-at??
   Sindre, you not Trond-GEN.INDEF son-NOM.PX3SG-COP.NONPST.SC2SG?
   ‘Sindre, aren’t you Trond’s son?’ (p.k.)
ROLE FOR MINORITY SOURCE LANGUAGE

<table>
<thead>
<tr>
<th></th>
<th>1SG OBJ</th>
<th>1PL OBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SG</td>
<td>palasamak</td>
<td>palasamiź</td>
</tr>
<tr>
<td>3SG</td>
<td>palasamam</td>
<td>palasamiź</td>
</tr>
<tr>
<td>2PL</td>
<td>palasamiź</td>
<td>palasamiź</td>
</tr>
<tr>
<td>3PL</td>
<td>palasamiź</td>
<td>palasamiź</td>
</tr>
</tbody>
</table>

Table 8: First person indicative object marking for the verb pala|ms ‘to kiss’

<table>
<thead>
<tr>
<th></th>
<th>1SG OBJ</th>
<th>1PL OBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SG</td>
<td>palamak</td>
<td>palamiśk</td>
</tr>
<tr>
<td>2PL</td>
<td>palasamiśk</td>
<td>palasamiśk</td>
</tr>
</tbody>
</table>

Table 9: First person imperative object marking for the verb pala|ms ‘to kiss’

b. išak čop kudo-so-nzo-ĺń
yesterday all.day.long home-INE-PX3SG-COP.PST.SC1SG.
‘I was at his/her house all day yesterday’ (cf. Evsev’ev 1928-29:62)

As (Trosterud 2006:253–254) continues into the discussion of the subject-object aka definite, objective and object conjugation, he applies the verbal paradigm word pala|ms ‘to kiss’\(^3\) for the description of subject-object indexing in Erzya, where the syncretism of the palasamiź form is stated to render six different readings: Thou kissest\(^4\) us (2SG>1PL), You kiss us (2PL>1PL), You kiss me (2PL>1SG), He/She kisses us (3SG>1PL), They kiss me (3PL>1SG), They kiss us (3PL>1PL), as illustrated in table (8).

The parameters for this paradigm can be defined as follows: (a) a default formative in -samiź marks the first person object, (b) the default formative is overridden when both the subject and object arguments are singular by definition. Hence, the formatives -samak ‘thou kissest me’ and -samam ‘he/she/it kisses me’ override the default form. In fact, the six readings given at (Trosterud 2006:254, example 109) might be augmented by an additional interpretation. In the spoken language, the default form is also used when the subject indicates an indefinite but specific actor. This is an aspect that would be important in the development of a rule-based translation machine. If the speaker does not want to reveal that they have been kissed by one person the default form is used. This would be treated as an analogue of the English, where a third person plural form is often used for the same purpose.

In the analysis of the abundance of paradigms afforded in Keresztes (1999) on Trosterud (2006:256–257), the author might take note of a healthy yet critical bit of information from Cygankin (1968:389), such that the formative -miśk, in this case palamiśk ‘kiss us!’, which is used in the imperative and indicates a default second person subject and a first person object. Only the cell with palamak ‘kiss me’ in it, where both arguments are singular, has an unambiguous reading 1SG>2SG. There should be more villages represented from the Southeast dialects in future research, but for back study, see also (Gabelentz 1839:237, 276), (Wiedemann 1865:74).

As indicated above, all of these notes regarding Erzya and Moksha language morphology and usage are intended to help Trond broaden his horizons in Mordvin language research. But it is definitely the morphophonological and syntactic issues forwarded for each language in this mutual GiellaLT infrastructure that have made us aware of more and more things to write about. And it is this awareness that helps us to transfer comparable questions to other languages in the infrastructure, such as Komi.

\(^3\)first introduced by (Ahlqvist 1859:23-43) in the 1860’s to replace the macabre kalma|ms ‘to bury’ verb used by Ornatov (1838) for the illustration of frequentative and habitual verbal derivations

\(^4\)The corresponding second person ending would be -st, while -eth corresponds to the third person singular -(e)s of today’s English.
3. Plural Copula Complement Marking in Komi

In the Permic languages, including Komi-Zyrian discussed here, there is a system of plural copula complement marking that differs from the plural marking in other environments. This process is well known and described, but the currently available large Komi corpora allow more nuanced investigation of possible rare features or gaps in the paradigm.

Previously, in discussions of homonymy occurring in illative versus inessive case with possessive marking, it has been noted that constraint grammar rules, such as those discussed in Trosterud (2009), Antonsen et al. (2009c), Antonsen et al. (2009a) must be found for Komi disambiguation. A good example of ambiguity is found in the Komi word for home горт and an inessive versus illative singular declension form with third person singular marking гортас ‘in his/her home’ vs ‘into his/her home’ (see ex. 2a–2b).

(2) a. А сэсся лоクトӧма горт-ас, Помӧсдінӧ.
   and then come.PAST2 home-sg.ill.plx3sg, Pomasdin.ill
   ‘And then he came home, to Pomösdin.’ Пунегова (2021)

b. Бедь кӧ суалалӧ, сідзкӧ, горт-ас некод абу.
   stick if stand.pres.ssg, if.so, home-sg.ine.plx3sg no-one is.not.
   ‘If there is a stick standing (by the door), that means nobody is at home.’ Анохин (2021)

Examples (2a–2b) illustrate distinctions made on the basis of verbal government, such that the verb локны ‘to come’ requires the illative, as seen illustrated in both гортас ‘into his/her home’ and Помӧсдінӧ ‘to Pomösdin’, while the stationary verb суалалвны ‘to stand’ takes locative government, such as the inessive in nouns. These distinctions can also be made by treating the copula complement plural marker in the same way as verbs. In (3a–3b) we can see that the copula plural marker -ӧсь is not compatible with an illative reading, thus that reading can be removed in constraint grammar disambiguation. Example (3b), however, provides us with simultaneous noun plural and copula complement plural marking. Hence, we have an analogy for one part of the independent three-dimensional paradigms treated in (Trosterud 2006:164–212), or is this actually four-dimensional, i.e., the possessive suffix introduces the person and number categories associated with the possessor, while the number category of the possessum in яс ‘pl.’ is co-located with the number category of the copulate.  

(3) a. Кирø дорын-ӧсь ӧд найӧ, горт-ас-ӧсь.
   Kirø al.ine-cop.pll you.know they, home-sg.ine.plx3sg-cop.pll
   ‘But they are ones with Kirø, they’re at his home.’ Куратова (2020)

b. Найӧ вижов либо вежов рӧма-ӧсь, перкалывэя чут-ъяс-ас-ӧсь.
   they yellowish or greenish colored-cop.pll, tan-covered fleck-pl.ine.plx3sg-cop.pll
   ‘They are yellowish or greenish in color with tan flecks.’ Ракин (2011)

The ambiguity illustrated in ex (2a–2b) has been one of the challenges in developing Komi-Zyrian Constraint Grammar, and is a classic problem in computational analysis of both Komi literary languages (cf. Rueter et al. 2020b). Examples 3a–3b show, however, that the process and involved questions go even deeper into the plurality and copular constructions. For computational description this is no issue, and the current analyser returns correctly and unambiguously ‘горт + Hom1 + N + Sg + Ine + PxSg3 + Pred + Pl’ and ‘чут + N + Pl + Ine + PxSg3 + Pred + Pl’, but from the point of linguistic description this type of structures that contain multiple plural markings have largely been outside the grammatical description tradition, and will need to be addressed in further research. Even for our computational description, the presence of repeated plural tags is a question that may need a more elegant solution. We consider this as a good example where computational and linguistic analysis can enrich one another.

4. Conclusion

Our study started with various notes on Erzya morphological paradigms. We added some important notes into the discussion about Erzya adnominal person marking, and touched briefly the verbal conjugation and
copula complement marking. To complement these observations, we also discussed plural copula complement marking in Komi, which has some behavior which we believe has not been previously discussed extensively enough, or possibly even noticed.

We did not include significantly new results or experiments, but added important novel points into the discussion that has evolved for several decades, and we assume will keep on going.

Acknowledgements

Working with Trond has also meant direct or indirect collaboration with people directly connected to the Giellateknö infrastructure, Divvun and a joint GiellaLT. A few people to mention are Lena Antonsen, Chiara Argese, Tino Didriksen, Børre Gaup, Ciprian Gerstenberger, Ryan Johnson, Sjur Moshagen, Thomas Omma, Flammie A Pirinen, Francis Tyers, Heli Uibo, Kevin Unhammer plus many more.

References


JACK RUETER, NIKO PARTANEN, KHALID ALNAJJAR AND MIKA HÄMÄLÄINEN

spublikanskoj tipografiâs’ «Krasnyj Oktâbr’», Saransk. Vuzon’ erzân’ dy finnèn’-ugran’ kužotnesè stu-
dentët’nënen’ tonavtnemapel’.

tional.

Harrigan, Atticus G., Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud, and Arok Wolven-

Hämäläinen, Mika. 2021. Endangered languages are not low-resource! In Multilingual Facilitation. Rootroo

Hämäläinen, Mika and Linda Wiechetek. 2020. Morphological Disambiguation of South Sámi with FSTs and
Keresztes, László. 1999. Development of Mordvin definite conjugation. Suomalais-Ugrilaisen Seuran toimituk-
sia, 233. Suomalais-Ugrilainen Seura, Helsinki.

Khanna, Tanmai, Jonathan N. Washington, Francis M. Tyers, Sevilay Bayatlı, Daniel G. Swanson, Tommi A.
Pirinen, Irene Tang, and Hèctor Alìs i Font. 2021. Recent advances in apertium, a free/open-source

èrzâ-mordovskogo âzyka. Mordovskoe knižnoe izdatelʹstvo, Saransk.

toimituksia, 261. Suomalais-Ugrilainen Seura, Finland.

GmbH, Germany.

Rueter, Jack. 2020. Korpus nacionâl’nyh mordovskix âzykov: principy razrabotki i perspektivy
funkcionirovanâi/ dejstviâ. In Финно-угорские народы в контексте формирования обширной сокращенной
средней, pp. 118–127. Izdatel’skij centr Istoriko-
sociologičeskogo instituta, Russia. Conference date: 08-10-2020 through 09-10-2020.

Rueter, Jack, Marília Fernanda Pereira de Freitas, Sidney Da Silva Facundes, Mika Hämaläinen, and Niko
Language Processing for Indigenous Languages of the Americas, pp. 28–33. Association for Computational

Rueter, Jack and Mika Hämaläinen. 2020a. FST morphology for the endangered Skolt Sami language.
In Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages
(SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), pp. 250–257. European


Rueter, Jack, Mika Hämaläinen, and Niko Partanen. 2020a. Open-source morphology for endangered mord-
The Association for Computational Linguistics, United States. https://doi.org/10.18653/v1/2020.nlpos-
1.13. Workshop for NLP Open Source Software, NLP- OSS ; Conference date: 19-11-2020 Through 19-11-
2020.

Rueter, Jack, Niko Partanen, Mika Hämaläinen, and Trond Trosterud. 2021b. Overview of open-source

Rueter, Jack, Niko Partanen, and Larisa Ponomareva. 2020b. On the questions in developing computational
infrastructure for Komi-Permyak. In Proceedings of the Sixth International Workshop on Computational Lin-

world: Isoglosses and their Reflexes attested in and around the Dubyonki Raion, pp. 109–148. No. 10 in
Uralica Helsingiensia. University of Helsinki, Finland.

Rueter, Jack Michael and Francis M. Tyers. 2018. Towards an open-source universal-dependency tree-
Role for minority source language


Куратова, Н. Н. 2020. Марьюшка. Войыв кодзув 1 №2.


Ракин, А. 2011. Герчкан. Ви кинь 1 3.