

# 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 led 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.

©2022 Jack Rueter, Niko Partanen, Khalid Alnajjar and Mika Hämäläinen. *Nordlyd* 46.1: 231-240, *Morfologi, målstrev og maskinar: Trond Trosterud {fyller | täyttää | deavdá | turns} 60!*, edited by Lene Antonsen, Sjur Nørstebø Moshagen and Øystein A. Vangsnes. Published at UiT The Arctic University of Norway. <http://septentrio.uit.no/index.php/nordlyd>  
<https://doi.org/10.7557/12.6370>



## ROLE FOR MINORITY SOURCE LANGUAGE

	ABE	ABL	CMPR	COM	DAT	ELA	GEN	ILL	INE	LAT	LOC	NOM	PRL	TEMP	TRL
PERS PRON	+	+	+	–	+	+	+	+	+	–	–	+	+	–	+
N INDEF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
N SG DEF	+	+	+	–	+	+	+	–	+	–	–	+	+	–	?
N PL DEF	+	+	+	DIAL	+	+	+	+	+	+	–	+	+	–	+
N PXSG3	+	+	+	–	+	+	+	+	+	–	–	+	+	–	?
ADP	–	+	–	–	–	–	–	–	–	+	+	–	+	–	–

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 distinguishes the person of the object. This resembles singularly object marking structures attested in languages outside the Uralic context, for example Apurinã (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 phenomenon 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, prolativ, comparative and translative may 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 ellipsis<sup>1</sup>. Perhaps, this will provide for further dimensions to syntax analysis research in Saami language research (cf. Antonsen et al. 2010, Sheyanova and Tyers 2017, Rueter and Hämäläinen 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 Gabelentz (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 3<sup>rd</sup> person singular possessive suffix’, and ADP ‘adposition or relational noun’<sup>2</sup>. 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 (Nad’kin 1968:27–28; Rueter 2010:97–100). Other case forms, such as the translative in *ks* have not been documented in the possessive

<sup>1</sup>Let’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 hyllyllä 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., *tonavtrüčas saji lavša langsoñšerñ* (lit. ‘the.student takes shelf one.that.is.on’) ‘the student takes the one that is on the shelf’.

<sup>2</sup>The relational noun *kudikele* ‘in the vestibule’ (aka postposition or adpostion), although represented by a defective paradigm without a nominative form, can be found in Erzya literature in the definite singular declension, e.g., *kudikelganñ* ‘vestibule.PRL-DEF.SG’ *Kudikelganñ tago mažaževš lšiča*. ‘Once again someone could be heard going out through the vestibule’.

	INDEF	PX1SG	PX2SG	PX3SG	PX1PL	PX2PL	PX3PL
NOM.SG	<i>jalga</i>	<i>jalgam</i>	<i>jalgat</i>	<i>jalgazo</i>	<i>jalganok</i>	<i>jalgank</i>	<i>jalgast</i>
NOM.PL	<i>jalgat</i>	<i>jalgan ~ jalgam</i>	<i>jalgat</i>	<i>jalganzo</i>	<i>jalganok</i>	<i>jalgank</i>	<i>jalgast</i>
GEN	<i>jalgañ</i>	<i>jalgan ~ jalgam</i>	<i>jalgat</i>	<i>jalganzo</i>	<i>jalganok</i>	<i>jalgank</i>	<i>jalgast</i>
DAT	<i>jalgañeñ</i>	<i>jalgañeñ ~ jalgañ turtov ~ jalgam turtov</i>	<i>jalgañeñ ~ jalgañ turtov ~ jalgat turtov</i>	<i>jalgansteñ ~ jalganzo turtov</i>	<i>jalganok turtov</i>	<i>jalgank turtov</i>	<i>jalgansteñ ~ jalgast turtov</i>
ABL	<i>jalgado</i>	<i>jalgadon ~ jalgadam</i>	<i>jalgadot</i>	<i>jalgadonzo</i>	<i>jalgadonok</i>	<i>jalgadonk</i>	<i>jalgadost</i>
ELA	<i>jalgasto</i>	<i>jalgaston ~ jalgastom</i>	<i>jalgastot</i>	<i>jalgastonzo</i>	<i>jalgastonok</i>	<i>jalgastonk</i>	<i>jalgastost</i>

 Table 2: the noun *jalga* ‘friend’

or definite singular declensions. In a similar vein, the definite plural lative seems to be missing from the enumeration of cases in the Erzya Morphology Grebneva (2000), although it can be attested in the literary works of the prolific Erzya authors Abramov and Shcheglov.

### 2.1. Adnominal person marking in Erzya

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 Hämäläinen (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 *jalgazo* ‘(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 archivowel, 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

	Sing	Plur	Sing	Plur
1	<i>śudan</i>	<i>śudotano</i>	<i>-An</i>	<i>-Tano</i>
2	<i>śudat</i>	<i>śudotado</i>	<i>-At</i>	<i>-Tado</i>
3	<i>śudi</i>	<i>śudít</i>	<i>-i</i>	<i>-ít</i>

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

	Sing	Plur	Sing	Plur
1	<i>viđan</i>	<i>viđtano</i>	<i>piđan</i>	<i>piđeťano</i>
2	<i>viđat</i>	<i>viđtado</i>	<i>piđat</i>	<i>piđeťado</i>
3	<i>viđi</i>	<i>viđít</i>	<i>piđi</i>	<i>piđít</i>
	<i>viđ ems</i>	‘to sow’	<i>piđe ms</i>	‘to cook’

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

harmony (Rueter 2010:62–66).

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 *viđ|ems* ‘to sow’, *viđs* ‘he/she sowed’ and *piđe|ms* ‘to cook’, *piđeś* ‘he/she cooked’ illustrating single consonants between two vowels. Here, the single consonant *đ* in *viđ|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 *jalgajan* ‘I am a friend’ from the possessum form *jalgan* ‘my friends’, the modern standard embraces either a single form *jalgan* to convey both meanings, or it adds a Southeastern dialect form *jalgam*, 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.

	Non-past	Past	Non-past (definite)	Past (definite)
1sg	<i>jalgan</i>	<i>jalgalin</i>	<i>jalgaśan</i>	<i>jalgaśelin</i>
2sg	<i>jalgat</i>	<i>jalgalit</i>	<i>jalgaśat</i>	<i>jalgaśelit</i>
3sg	<i>jalga</i>	<i>jalgal</i>	<i>jalgaś</i>	<i>jalgaśel</i>
1pl	<i>jalgatano</i>	<i>jalgalinek ~ jalgatolinek</i>	<i>jalgatrineťano</i>	<i>jalgatrineľinek</i>
2pl	<i>jalgatado</i>	<i>jalgalide ~ jalgatolide</i>	<i>jalgatrineťado</i>	<i>jalgatrineľide</i>
3pl	<i>jalgat</i>	<i>jalgat ~ jalgatolit</i>	<i>jalgatne</i>	<i>jalgatneľ</i>

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

	Non-past	Past	Non-past (definite)	Past (definite)
1SG	<i>sazoran</i>	<i>sazorolín</i>	<i>sazorošan</i>	<i>sazorošeliń</i>
2SG	<i>sazorat</i>	<i>sazorolít</i>	<i>sazorošat</i>	<i>sazorošeliť</i>
3SG	<i>sazor</i>	<i>sazoról</i>	<i>sazoroš</i>	<i>sazorošel</i>
1PL	<i>sazortano</i>	<i>sazorolínek ~ sazorotolínek</i>	<i>sazortneřano</i>	<i>sazortnelínek</i>
2PL	<i>sazortado</i>	<i>sazorolíde ~ sazortolíde</i>	<i>sazortneřado</i>	<i>sazortnelíde</i>
3PL	<i>sazort</i>	<i>sazoról ~ sazortól</i>	<i>sazortne</i>	<i>sazortnelť</i>

 Table 6: Copula person and number for the consonant-stem noun *sazor* ‘little sister’

	Non-past	Past	Non-past (definite)	Past (definite)
1SG	<i>ovtan</i>	<i>ovtolín</i>	<i>ovtošan</i>	<i>ovtošeliń</i>
2SG	<i>ovtat</i>	<i>ovtolít</i>	<i>ovtošat</i>	<i>ovtošeliť</i>
3SG	<i>ovto</i>	<i>ovtol</i>	<i>ovtoš</i>	<i>ovtošel</i>
1PL	<i>ovtotano ~ ovttano</i>	<i>ovtolínek ~ ovtotolínek ~ ovttolínek</i>	<i>ovtoříetano ~ ovttneřano</i>	<i>ovtořínelínek ~ ovttnelínek</i>
2PL	<i>ovtotado ~ ovttado</i>	<i>ovtolíde ~ ovtotolíde ~ ovttolíde</i>	<i>ovtoříetado ~ ovttneřado</i>	<i>ovtoříelíde ~ ovttnelíde</i>
3PL	<i>ovtot ~ ovtt</i>	<i>ovtol ~ ovtotol ~ ovttol</i>	<i>ovtořie ~ ovttne</i>	<i>ovtořielť ~ ovttnelť</i>

 Table 7: Copula person and number for the mixed-stem noun *ovto* ‘bear’

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 *-líde*. 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 *-líde*. 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.INDEF, 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. *śínđre, ton avol trond-on čora-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

	1SG OBJ	1PL OBJ
2SG	<i>palasamak</i>	<i>palasamiž</i>
3SG	<i>palasamam</i>	<i>palasamiž</i>
2PL	<i>palasamiž</i>	<i>palasamiž</i>
3PL	<i>palasamiž</i>	<i>palasamiž</i>

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

	1SG OBJ	1PL OBJ
2SG	<i>palamak</i>	<i>palamišk</i>
2PL	<i>palasamišk</i>	<i>palasamišk</i>

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

- b. *išak čop kudo-so-nzo-liš*  
*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’<sup>3</sup> for the description of subject-object indexing in Erzya, where the syncretism of the *palasamiž* form is stated to render six different readings: Thou kissest<sup>4</sup> 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 *-samiž* 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 *-mišk*, in this case *palamišk* ‘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 Giella.T 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.

<sup>3</sup>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

<sup>4</sup>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 *зорм* and an inessive versus illative singular declension form with third person singular marking *зормас* ‘in his/her home’ vs ‘into his/her home’ (see ex. 2a–2b).

- (2) a. А сэсся локтöма горт-ас, Помöсдинö.  
*and then come.PAST2 home-SG.ILL.PX3SG, Pomösdin.ILL*  
 ‘And then he came home, to Pomösdin.’ Пунегова (2021)
- b. Бедь кö сулалö, сідзкö, горт-ас некод абу.  
*stick if stand.PRS.3SG, if.so, home-SG.INE.PX3SG no-one is.not*  
 ‘If there is a stick sanding (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 *зормас* ‘into his/her home’ and *Помöсдинö* ‘to Pomösdin’, while the stationary verb *сулавны* ‘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 *-öсь* 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 predicate *öсь* ‘pl.’.

- (3) a. Кирö дорын-öсь öд найö, горт-ас-öсь.  
*Kirö at.INE-COP.PL you.know they, home-SG.INE.PX3SG-COP.PL*  
 ‘But they are ones with Kirö, they’re at his home.’ Куратова (2020)
- b. Найö вижов либö вежов рöма-öсь, перкальвевья чут-ьяс-ас-öсь.  
*they yellowish or greenish colored-COP.PL, tan-covered fleck-PL-INE.PX3SG-COP.PL*  
 ‘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 Giellatekno infrastructure, Divvun and a joint GiellaLT. A few people to mention are Lene 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

- Ahlqvist, August Englebrect. 1859. *Läran om verbet in mordvinskans mokscha-dialekt*. Helsingfors. Frenckell & Son. Akademisk afhandling, som med den vidtberömda Historisk-filologiska Fackultetens vid Kejsarliga Alexanders-Universitetet i Finland samtycke till offentlig granskning framställes af August Englebrect Ahlqvist, Hist-Fil. Magister.
- Alnajja, Khalid, Mika Hämäläinen, Jack Rueter, and Niko Partanen. 2020. Ve’rdd. narrowing the gap between paper dictionaries, low-resource nlp and community involvement. In *Conference: Proceedings of the 28th International Conference on Computational Linguistics: System Demonstrations*, pp. 1–6. <https://doi.org/10.18653/v1/2020.coling-demos.1>.
- Antonsen, L., S. Huhmarniemi, and T. Trosterud. 2009a. Interactive pedagogical programs based on constraint grammar. In *Proceedings of the 17th Nordic Conference of Computational Linguistics. NEALT Proceedings Series 4. 2009*, pp. 10–17.
- Antonsen, Lena and Trond Trosterud. 2011. Next to nothing – a cheap South Saami disambiguator. In *Proceedings of the NODALIDA 2011 workshop Constraint Grammar Applications May 11, 2011 Riga, Latvia*, edited by Eckhard Bick, Kristin Hagen, Kaili Müürisepp, and Trond Trosterud, vol. 14 of *NEALT Proceedings Series*.
- Antonsen, Lene, Ciprian-Virgil Gerstenberger, Sjur Nørstebø Moshagen, and Trond Trosterud. 2009b. Ei intelligent elektronisk ordbok for samisk. In *LexicoNordica 16*, vol. 16, pp. 271–283.
- Antonsen, Lene, Saara Huhmarniemi, and Trond Trosterud. 2009c. Constraint grammar in dialogue systems. In *NEALT Proceedings Series 2009*, vol. 8, pp. 13–21.
- Antonsen, Lene, Trond Trosterud, and Linda Wiecheteck. 2010. Reusing grammatical resources for new languages. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*. European Language Resources Association (ELRA), Valletta, Malta.
- Bartens, Raija. 1999. *Mordvalaiskielten rakenne ja kehitys*. Suomalais-Ugrilaisen Seuran Toimituksia 232, Helsinki. Bartens, Raija 1999: Mordvalaiskielten rakenne ja kehitys. Suomalais-Ugrilaisen Seuran Toimituksia 232. Helsinki: Suomalais-Ugrilainen Seura.
- Cygankin, D.V. 1968. *Očerki Mordovskix Dialektov tom V*, chap. Opyt klassifikacii èrzânskix govorov mordovskogo prisur’â. Mordovskoe knižnoe izdatel’stvo, Saransk. Naučno-issledovatel’skij institut âzyka, literatury, istorii i èkonomiki pri sovete ministrov mordovskoj ASSR.
- Evsev’ev, M.E. 1928-29. *Mordovskaâ grammatika*. Central’noe izdatel’stvo narodov SSSR, Moskva. Èrzân’ grammatika.
- Facundes, Sidney da S. 2000. *The language of the Apurinã people of Brazil (Arawak)*. SUNY Buffalo. PhD Dissertation.
- Gabelentz, Herr Conon von der. 1839. Versuch einer mordwinischen grammatik. In *Zeitschrift für die Kunde des Morgenlandes*, II. 2–3, pp. 235–284, 383–419. Druck und Verlag der Dieterichschen Buchhandlung, Göttingen. Online: <https://github.com/rueter/Erzya-grammar-Gabelentz-1838-39>.
- Gerstenberger, Ciprian, Niko Partanen, Michael Rießler, and Joshua Wilbur. 2016. Utilizing language technology in the documentation of endangered Uralic languages. *Northern European Journal of Language Technology* 4: 29–47. <https://doi.org/10.3384/nejlt.2000-1533.1643>.
- Grebneva, A.M. 2000. *Èrzân’ kel’, Morfemika, valon’ teevema dy morfologiâ*, chap. Padežen’ luvos’. Re-



- spublikanskoj tipografiäs' «Krasnyj Oktâbr», Saransk. Vuzon' erzân' dy finnèn'-ugran' kužotnesè student[t]nènen' tonavtnemapel'.
- Grünthal, Riho. 2008. *Transitivity in Erzya Mordvin*, pp. 235–246. Uralisztikai tanulmányok. ELTE,, International.
- Harrigan, Atticus G., Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud, and Arok Wolvengrey. 2017. Learning from the computational modelling of plains cree verbs. *Morphology* 27 4: 565–598. <https://doi.org/10.1007/s11525-017-9315-x>.
- Hämäläinen, Mika. 2021. Endangered languages are not low-resourced! In *Multilingual Facilitation*. Rootroo Ltd. <https://doi.org/10.20944/preprints202104.0113.v1>.
- Hämäläinen, Mika and Linda Wiecheteck. 2020. Morphological Disambiguation of South Sámi with FSTs and Neural Networks. In *Conference: Proceedings of the 1st Joint SLTU and CCURL Workshop*, pp. 36–40.
- Keresztes, László. 1999. *Development of Mordvin definite conjugation*. Suomalais-Ugrilaisen Seuran toimituksia, 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 Alòs i Font. 2021. Recent advances in apertium, a free/open-source rule-based machine translation platform for low-resource languages. *Machine Translation* 35 4: 475–502. <https://doi.org/10.1007/s10590-021-09260-6>.
- Nad'kin, D. T. 1968. *Očerki mordovskix dialektov*, vol. Tom V, chap. Morfologiâ nižnep'ânskogo dialekta èrzâ-mordovskogo âzyka. Mordovskoe knižnoe izdatel'stvo, Saransk.
- Ornatov, Pavel. 1838. *Mordovskaja grammatika / sostavlennaja na narechij mordvy mokshi Pavlom Ornatovym*. V Sinodalnoj tip., Moskva.
- Rueter, Jack. 2010. Adnominal Person in the Morphological System of Erzya. In *Suomalais-ugrilaisen seuran toimituksia*, 261. Suomalais-Ugrilainen Seura, Finland.
- Rueter, Jack. 2013. *Quantification in Erzya*, pp. 99–122. LINCUM Studies in Language Typology. Lincom GmbH, Germany.
- Rueter, Jack. 2020. Korpus nacional'nyx mordovskix âzykov: principy razrabotki i perspektivy funkcionirovaniâ/ dejstviâ. In *Финно-угорские народы в контексте формирования общероссийской гражданской идентичности и меняющейся окружающей среды*, 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 Hämäläinen, and Niko Partanen. 2021a. Apurinã Universal Dependencies treebank. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pp. 28–33. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.americasnlp-1.4>.
- Rueter, Jack and Mika Hämäläinen. 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 Language Resources association, Marseille, France. <https://aclanthology.org/2020.sltu-1.35>.
- Rueter, Jack and Mika Hämäläinen. 2020b. *Prerequisites For Shallow-Transfer Machine Translation Of Mordvin Languages: Language Documentation With A Purpose*, pp. 18–29. Iževsk: Institut komp'iuternyx issledovanij, Russian Federation. <https://doi.org/10.20944/preprints202104.0131.v1>.
- Rueter, Jack, Mika Hämäläinen, and Niko Partanen. 2020a. Open-source morphology for endangered mordvinic languages. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*, pp. 94–100. The Association for Computational Linguistics, United States. <https://doi.org/10.18653/v1/2020.nlposs-1.13>. Workshop for NLP Open Source Software, NLP-OSS ; Conference date: 19-11-2020 Through 19-11-2020.
- Rueter, Jack, Niko Partanen, Mika Hämäläinen, and Trond Trosterud. 2021b. Overview of open-source morphology development for the komi-zyrian language: Past and future. In *Proceedings of The Seventh International Workshop on Computational Linguistics of Uralic Languages*, pp. 62–72.
- 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 Linguistics of Uralic Languages*, pp. 15–25. <https://doi.org/10.18653/v1/2020.iwclul-1.3>.
- Rueter, Jack Michael. 2016. *Towards a systematic characterization of dialect variation in the Erzya-speaking 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-

- bank for Erzya. In *International Workshop for Computational Linguistics of Uralic Languages, IWCLUL*. <https://doi.org/10.18653/v1/W18-0210>.
- Serebrennikov, B. A., R.N. Buzakova, and M.V. Mosin. 1993. *Èrzânsko-russkij slovar'*. Russkij Âzyk, Digora, Moskva.
- Sheyanova, Mariya and Francis M. Tyers. 2017. Annotation schemes in North Sámi dependency parsing. In *Proceedings of the 3rd International Workshop for Computational Linguistics of Uralic Languages*, pp. 66–75.
- Simonenko, Alexandra. 2020. Existential possession in Meadow Mari. In *Approaches to predicative possession: the view from Slavic and Finno-Ugric*, edited by Dalmi, Gréte and Witkos, Jacek and Ceglowski, Piotr, pp. 162–181. Bloomsbury Academic. <https://doi.org/10.5040/9781350062498.ch-008>.
- Snoek, Conor, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen, and Trond Trosterud. 2014. Modeling the noun morphology of plains cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pp. 34–42. Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-2205>.
- Trosterud, Reino Sindre, Trond Trosterud, Anna-Kaisa Räisänen, Leena Niiranen, Mervi Haavisto, and Kaisa Maliniemi. 2017. A morphological analyser for kven. In *Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages*, pp. 76–88. Association for Computational Linguistics, St. Petersburg, Russia. <https://doi.org/10.18653/v1/W17-0608>. Online: <https://aclanthology.org/W17-0608>.
- Trosterud, Trond. 1994. Auxiliaries, negative verbs and word order in the sami and finnic languages. In *Minor Uralic Languages: Structure and Development*, edited by Ago Künnap, pp. 173–181. Tartu.
- Trosterud, Trond. 2006. *Homonymy in the Uralic Two-Argument Agreement Paradigms*. Suomalais-Ugrilainen Seuran Toimituksia 251. Suomalais-Ugrilainen Seura, Helsinki.
- Trosterud, Trond. 2009. A constraint grammar for Faroese. In *Proceedings of the NODALIDA 2009 workshop Constraint Grammar and robust parsing*, edited by Eckhard Bick, Kristin Hagen, Kaili Müürisep, and Trond Trosterud, vol. 8 of *NEALT Proceedings Series*, pp. 1–7. Northern European Association for Language Technology (NEALT, <http://omilia.uio.no/nealt>). Electronically published at Tartu University Library (Estonia) <http://hdl.handle.net/10062/14180>.
- Trosterud, Trond and Lene Antonsen. 2020. Hva er viktig for forståelse? Om maskinoversetting fra nord-samisk. In *Bauta: Janne Bondi Johannessen in memoriam*, edited by K. Hagen, A. Hjelde, K. Stjernholm, and Ø. A. Vangsnes, vol. 11(2) of *Oslo Studies in Language*, pp. 489–502. Oslo: University of Oslo. <https://doi.org/10.5617/osla.8514>.
- Trosterud, Trond and Sjur Moshagen. 2021. Soft on errors? The correcting mechanism of a Skolt Sami speller. In *Multilingual Facilitation*, pp. 197–207. <https://doi.org/10.31885/9789515150257.19>.
- Turunen, Rigina. 2010. *Nonverbal predication in Erzya: Studies on morphosyntactic variation and part of speech distinctions*. Ph.D. thesis, University of Helsinki.
- Wiedemann, F.J. 1865. *Grammatik der Ersä-Mordwinischer Sprache*, vol. IX №5 of VII. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg. Online: <https://github.com/rueter/Erzya-grammar-Wiedemann-1865>.
- Анохин, Павел. 2021. Съѡд вѡр шѡрѡд да нюрѡд. *Коми му* 2021-08-26.
- Куратова, Н. Н. 2020. Марьюшка. *Войвыв кодзув* 1 №2.
- Пунегова, Надежда. 2021. Илона Артеева: Уджыд мед вѡлі съѡлѡм сертиыд. *Коми му* 2021-09-09.
- Ракин, А. 2011. Герчкан. *Би кинь* 1 3.