Towards an open-source universal-dependency treebank for Erzya

Rueter, Jack Michael

2018

Towards an open-source universal-dependency treebank for Erzya

Jack Rueter
University of Helsinki
Department of Modern Languages
Helsinki
`jack.rueter@helsinki.fi`

Francis Tyers
НИУ ВШЭ
School of Linguistics
Moscow
`ftyers@hse.ru`

Abstract

This article describes the first steps towards a open-source dependency treebank for Erzya based on universal dependency (UD) annotation standards. The treebank contains 610 sentences with 6661 tokens and is based on texts from a range of open-source and public domain original Erzya sources. This ensures its free availability and extensibility. Texts in the treebank are first morphologically analyzed and disambiguated after which they are annotated manually for dependency structure. In the article we present some issues in dependency syntax for Erzya and how they are analyzed in the universal-dependency framework. Preliminary statistics are given for dependency parsing of Erzya, along with points of interest for future research.

Tiivistelmä

Tässä artikkelissa kerrotaan ersän kielen avoimen puupankin ensimmäisistä askeleista, joissa sovelletaan universaaliriippuvuus-annotaatiota (UD). Puupankki sisältää 610 virkettä joissa on yhteensä 6661 tokenia ja se perustuu avoimeen ersänkieliseen originaalikirjoituksiin. Tällä tavalla varmistetaan puupankin saatavuutta ja laajennettavuutta. Puupankin tekstit on ensin analysoitu morfologisella jäsentimellä ja disambiguoitu, minkä jälkeen suoritetaan loppuyksiselitteistäminen käsin ja lisätään riippuvuussuhteet. Artikkelissa esitetään joitakin kysymyksiä, jotka esiintyvät ersän lauseoppia sovellettaessa universaaliriippuvuuskehyksiin. Annetaan alkutilastoja ersän jäsennyksestä sekä ajatuksia tulevan tutkimuksen näkemyksistä.

Abstract

Те статиясонть сёрмадтано эрзянь келень од ресурсадо, конась весеменень панжадо, чувтокс валрисьмень пурнавксто, чувтонь банкто, ды юртонзо путомадо. Валрисьмень анализэнь теемстэ нолдави тевс масторлангонь вейсэнь аннотация, конаньсэ невтеви валрисьме пелькстнэнь вейкест-вейкест эйстэ чувтокс аштема лувост (Universal Dependency UD). Статиянть сёрмадомсто чувтонь банкось ашти 610 валрисьмеде, косо весемезэ 6661 токент (валт-лотксема тешкст), материалось ашти весеменень панжадо эрзякс сёрмадозь литературанть эйстэ. Истя чувтонь банкось саеви-келейгавтови кинень мелезэ – ресурсась ванстсы оляксчинзэ. Васня пурнавксонь валрисьметненень тееви морфологиянь анализ, конась мейле седе вадрялгавтови синтаксисэнь анализсэ.

Мейле келень ванкшныцясь сонсь невти кона пелькстнэ конатнень эйстэ аштить. Статиясонть макстано зярыя кевкстемат, конат чачить эрзянь кель UD марто вастневемстэ. Макстано эрзянь келень анализдэ васнянь статистика ды арсемат-мельть келень ванкшномань сыця ёнкстнэде-тевтнеде.

1 Introduction

This article describes work towards the development of a Universal Dependencies-based dependency treebank for Erzya, a Uralic language traditionally spoken in the Volga Region. Little if any computational-linguistic research has been published on syntactic parsing for Erzya. A valuable resource in the study and development of syntactic parsing is a treebank—a corpus of parsed texts containing gold-standard syntactic annotation.

Freely available treebanks exist for many languages, one particularly interesting set is the group of over 60 languages represented in Universal Dependencies (UD), where Erzya is now one of the smaller "upcoming languages"[1]. This mutual presentation makes it possible to understand and utilize language-independent dependency tagging with direct analogy from other Uralic languages, such as Finnish, Estonian, North Sami and Hungarian, as well as languages sharing other morphosyntactic characteristics with Erzya. The UD environment also makes direct reference to terminology definition resources, such as those offered by SIL[2], and research in the World Atlas of Language Structures (WALS)[3].

To our knowledge, however, no previous treebank exists for either of the Mordvinic languages, although there are closed annotated corpora, such as MORMULA at the University of Turku[4], quantlang-uhlcs[5] in Helsinki, and the semi-limited ERME[6].

In building our treebank we take advantage of previous work done by Rueter in Helsinki Finite-State Transducer Technology (HFST) morphological analysis and part-of-speech tagging for Erzya on the Giellatekno infrastructure, as well as ongoing disambiguation work with Constraint Grammar (VISLCG).

The remainder of the paper is organized as follows. Section 2 gives some background linguistic information on Erzya, and outlines some special challenges in parsing Erzya. In Section 3 we describe the corpus that we annotated and the methodology used in annotating it. Section 4 gives a sketch of some decisions we have made with respect to annotation guidelines, referring back to the discussion in Section 2. For reasons of space and time, these guidelines are by no means complete, but they do present a subset of guidelines which are of particular interest.

2 Background

2.1 Erzya

Erzya is one of the two Mordvinic languages traditionally spoken in scattered villages throughout the Volga Region and former Russian Empire by well over a million in the

---

[1] `http://universaldependencies.org/`

[2] `http://www.glossary.sil.org/`

[3] `http://wals.info/`

[4] `http://www.helsinki.fi/~kopotev/finnish_corpora_eng.pdf`

[5] Quantifiers and Quantification in Finnish and Languages Spoken in the Central Volgakama Region – UHLCS `http://urn.fi/urn:nbn:fi:lb-2016012202`

[6] Erme – Erzya and Moksha Extended Corpora `http://urn.fi/urn:nbn:fi:lb-201407306`

Case	Definite	Form	Function
Nom	Def	/ś/ /t́ńe/	def subject, predicative
			topic marker
Nom	Indef	-	ind subject, predicative
			ind attribute, object
			ind Adp complement
Nom	PxSg3	/Ozo/	subject, predicative
Gen	Indef	/Oń/	Ind genitive attribute
			Ind object, Adp complement
			embedded subject, object
Gen	Def	/Ońt́/	def object
			def adp complement
Ine	Indef	/sO/	locative, instrumental
			object

Table 1: Some cases and functions. *Note that with the exception of the third person singular possessive suffix, there is generally no distinctions made for number or genitive/nominative marking in the possessive declension.*

beginning of the 20th century and down to approximately half a million in the 2010 census[7]. For some, however, Erzya is only a part of the conglomerate Mordvin index, a population with the status of most numerous among the Uralic languages in Russia. Since there is no Mordvin language, as it were, but rather the closely related (adjacent yet not contiguous) Erzya and Moksha languages with their literary representation, research in syntax has often attempted to encompass the two.

Erzya, like many Uralic languages, is agglutinative with extensive morphology, agreement and constituent ordering phenomena that present a challenge to any syntactic description of the language. The most prominent of these challenges apparent from the start are case marking, definiteness, ellipsis, numerals, and copula variation between dependent and independent morphology. An open-source finite-state morphological analyzer constructed for Erzya[8] provides ample tagging for the annotation, but there is still plenty of work to be done with disambiguation. Erzya, much like other Ural-Altaic languages (Tyers and Washington, 2015), assigns more than one function to its cases. It also attests to intricate constituent ordering and minimal conjunction/subjunction marking, which will be one topic future research.
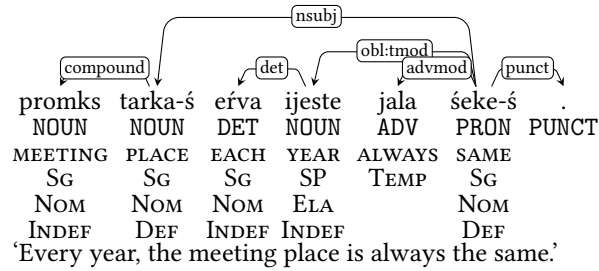
As indicated in Table 1, the definite nominative singular might be attested with the dependency relation, *nsubj, root* (in certain equative predications[9], example (1)), and to indicate a postposed topic (see example (5.b)).

(1)

promks	tarka-ś	eŕva	ijeste	jala	śeke-ś	.
NOUN	NOUN	DET	NOUN	ADV	PRON	PUNCT
MEETING	PLACE	EACH	YEAR	ALWAYS	SAME	
Sg	Sg	Sg	SP	Temp	Sg	
Nom	Nom	Nom	Ela		Nom	
Indef	Def	Indef	Indef		Def	

'Every year, the meeting place is always the same.'

3 Methodology

3.1 Corpus

To form a corpus, we were able to utilize materials by Erzya authors previously secured for language research purposes while in the Republic of Mordovia. The number of sources utilized is extremely limited, due to the elementary state of the developing treebank.

Document	Description	Sentences	Tokens	Av. length
Valskeń gudok	short story by Anoshkin, V.	4	36	9
Kirdažt	Novel manuscript by Bryzhinski, M.I.	270	3487	12.9
Veĺeń vajgeĺt	Miniatures by Chetvergov, E.	105	957	9.1
Pićipalakst	Foreword; Dunyashin, A.	75	633	8.4
Lažnića Sura	Novel by Kutorkin, A.D.	155	1534	9.9
Separate	Individual phrases	1	11	11
		610	6661	10.9

Table 2: Statistics on the composition of the corpus

The initial materials are representitive of original Erzya-language materials from the late 1920s to the turn of the new millennium. The Separate Individual phrases file will serve for documenting cited materials from scientific publications, such as the most recent Erzya syntax Агафонова et al. (2011).

The figures in Table 3 are incomplete, but do provide an initial ball park figure. It was noted that subsequent work will need to be done with extended dependency relations, as a small number of the cases returned included the values *appos* and *conj*. Also, the high number of genitive occurrences with the *obl* relation would indicate the presence of adpositions. Although the inessive had been indicated earlier as an object case, there was not a single instance where it occurred as a dependent case. The distinction of the dependency relations *nmod* versus *nmod:poss* in the statistics has been taken in following with usage in the Finnish UD projects; in subsequent contemplation, this may, in fact, be unnecessary since the genitive case already indicates a possessive relation in contrast to the inessive case with a spatial meaning as a noun modifier.

4 Annotation guidelines

When entering a group, such as UD, it is nearly imperative that all tagging conform with practices attested in the group. As such, it has been easy to compromise and

	obl	nmod	nmod:poss	obj	root	nsubj
`Case=Nom|Definite=Def`	1	_	_	_	6	170
`Case=Nom|Definite=Ind`	37	10		99	52	154
`Case=Nom|Number[psor]=Sing|Person[psor]=3`	_	_	_	_	1	41
`Case=Gen|Definite=Ind`	23	6	106	53	2	3
`Case=Gen|Definite=Def`	65	2	24	80	_	_
`Case=Dat|Definite=Ind`	10	_	_	_	1	_
`Case=Dat|Definite=Def`	33	_	_	_	_	_
`Case=Ine|Definite=Ind`	45	_	_	_	1	_

Table 3: Syntactic functions associated with different combinations of case, possession and definiteness.

relabel/learn values for various features in order to arrive at the group norm, on the one hand, and present new features, on the other. Work centered around indication for features (number, person), relations (aux, compound) as well as the handling of ellipsis and numerals. Certain phenomena clearly require additional time for thought and development. Work has been facilitated by the use of UD Annotatrix (Tyers et al., 2018), a tool for annotating treebanks in UD.

4.1 Number

The Erzya language like its closely related sibling Moksha has what is often called the object or definite conjugation. Unlike Hungarian, Nenets, Khanty and Mansi, however, the object conjugation of the Mordvinic languages morphologically indicates 1st, 2nd and 3rd person as well as singular and plural for some of the object and subject referents (Keresztes (1999); Trosterud (2006) 246–303). This means that a separates set of number and person features must be present to distinguish them from the already existent Number, Person, Number[psor] and Person[psor] features. Fortunately there were already Number[obj] and Person[obj], so only Number[subj] and Person[subj] were required as new features.

(2)

   ńej-iźe
  see-Ind.Prt1.ScSg3.OcSg3
  'he/she/it saw him/her/it'

In examples (2) and (3) there is an actual non-ambiguous morphological distinction for third person singular subject in combination with third person singular object in (2) and third person plural object in (3).

(3)

   ńej-ińźe
  see-Ind.Prt1.ScSg3.OcPl3
  'he/she/it saw them'

Example (4) is also unambiguous, as both arguments are singular in value. Neither argument need be present, but both are allowed.

(4)

ńej-imik
see-Ind.Prt1.ScSg2.OcSg1
'you[sg] saw me'

Although it is possible to use personal pronouns in subject and direct object position, their presence is not the normal situation. Context awareness is required that transcends conventional sentence boundaries, hence there is person and number disambiguation present that is not discernible from the morphology but rather the larger context.

4.2   Copula and polarity

Erzya attests to varied non-verbal predication Turunen (2010) and negation Hamari (2007) strategies. The copula is divided into locative and non-locative usage, and this dichotomy can readily be observed in the morphology of the negative copulas, i.e., where the negative locative copula is represented by /*araś-*/ (which is conjugated with the help of a morphologically dependent copula), whereas the presentation of equative or class membership negation is shown with the non-flective word form /*avoĺ*/. Further negation is manifest in the converbal negation element /*apak*/ (which can also be conjugated), first preterite /*eź-*/, conjunctional /*avoĺ-*/, optative and prohibitive /*iĺa-*/, and the negative particle /*a*/. All of these can trigger the dependency relation **aux:neg**, which in the case of the polarity markers /*a*/ and /*avoĺ*/ to all parts of speech.

Proh/Opt	Ind.Prt1	Cnd	Prc.Neg	Part.Neg	Part.Neg.Emp
/iĺa-/	/eź-/	/avoĺ-/	/apak/	/a/	/avoĺ/

Table 4: Negation

The six polarity markers in Table 4. can be divided into two categories. The prohibitive/optative, indicative first preterite and conditional stems along with the polarity marker for participles and converb (elsewhere gerund) constructions are all limited to use with verbal forms.

Whereas the negative particle /*a*/ can occur with many parts of speech, it is the non-flective word form /*avoĺ*/ or emphatic negative that is used in clausal negation. Clausal negation in combination with the imperative mood evokes a contrast in the prohibitive strategy in /*iĺa-*/. The clausal negation particle /*avoĺ*/ in combination with the second person imperative produces a contrastive negative imperative, whereas the prohibitive /*iĺa-*/ (**Mood=Proh**) is combined in the modern literary norm with a connegative form. This has not been attested in WALS van der Auwera et al. (2013).

4.3   Dependent copula morphology

Copula morphology is dependent and independent in Erzya. While many grammarians of the past century have referred to dependent copula morphology as noun conjugation, earlier presentations, such as Wiedemann (1864) (§77), refer to it as a suffixed copula. Judging from the fact that the dependent morphology can be attached to nouns in various declensional forms, as well as adjectives, numerals, adverbs, adpositions and non-finite verb forms, the phenomenon might more readily be referred to as a clitic.
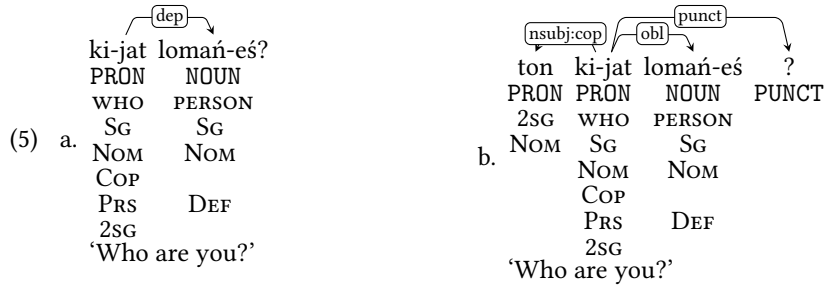
Tense	Sg1	Sg2	Sg3	Pl1	Pl2	Pl3
Nonpast	/mon odan/	/ton odat/	/son od/	/miń odtano/	/tiń odtado/	/siń odt/
Prt2	/mon odoľiń/	/ton odoľiť/	/son odoľ/	/miń odoľińek/	/tiń odoľiďe/	/siń odoľť/

Table 5: Dependent copula morphology ('young or new')

Thus we have independent copula in /uľń-/ (prt1) and /uľ-/ (prs), on the one hand, and dependent copula morphology in -/Oľ/-, etc.), on the other. Both the *nonpast* and *prt2* conjugation are virtually identical to their verbal subject conjugation counterparts; essentially the *prt2* in verbs is a combination of the short *nomen agentis* + the *prt2* used for copula function beyond the scope of finite verbs.

The dependent and independent morphologies present a challenging problem for dependency analysis in Erzya, whereas both can be in equative, class member, assertive, locative vs existential, and possessive vs belong-to predication. A dichotomy, however, is introduced where the Universal Dependency guidelines stating that the copula should be the dependent of the lexical predicate are applied only to independent copula.

The copula as dependent morphology attaches to what some scholars have considered the subject, but some scholars working with Moksha have approached this from a discourse point of view Kholodilova (2016). Instead of splitting the copula off as a separate leaf node, we have annotated the instances with dependent morphology as the head of the structure. Here constituent ordering might be an underlying factor to be considered in future work with the Erzya UD treebank.

(5) a.

ki-jat lomań-eś?
PRON NOUN
WHO PERSON
SG SG
NOM NOM
COP
PRS DEF
2SG
'Who are you?'

b.

ton ki-jat lomań-eś ?
PRON PRON NOUN PUNCT
2SG WHO PERSON
NOM SG SG
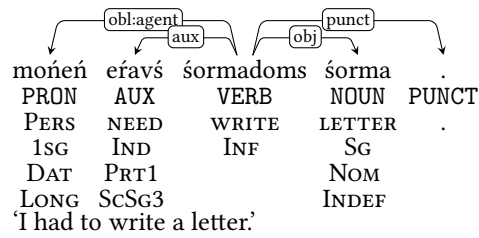NOM NOM
COP
PRS DEF
2SG
'Who are you?'

In both sentences with and without the second person singular personal pronoun, it is obvious that the word /lomań-eś/ can only be interpreted as an extra element; more than likely a topic marker. The interrogative pronoun /ki-jat/ in the predicative takes the subject second person singular, whereas the third person singular topic marker would not have triggered second person singular agreement.

4.4 Further auxiliaries

In addition to the copula and negation, the definition of additional auxiliaries in Erzya take us to necessitives. Necessitives in Erzya have parallels in Finnish, Komi-Zyrian and partially in Skolt Sami, in that a non-nominative case is used to indicate the actor, which might be construed as a subject when aligned with other European, accusative languages. In Erzya, it is the Dative that is used as in *mońeń* in example (3).
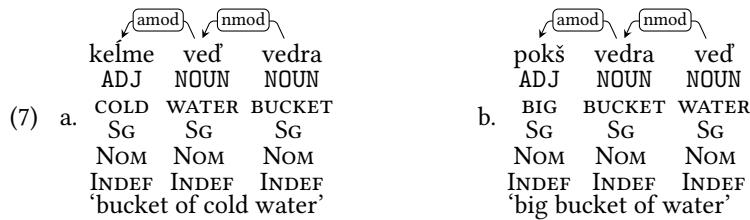
(6)

mońeń	eŕavś	śormadoms	śorma	.
PRON	AUX	VERB	NOUN	PUNCT
PERS	NEED	WRITE	LETTER	.
1SG	IND	INF	SG	
DAT	PRT1		NOM	
LONG	SCSG3		INDEF	

'I had to write a letter.'

Discussion with Erzya and Komi native scholars has also introduced the idea of adding verbs indicating future to the list of auxiliaries, this, however, would be problematic, due to the second meaning involved, namely, inchoative/inceptive albeit an aspect marker. In this initial Erzya treebank, auxiliaries have been limited to copulas, negation and necessitives.

4.5 Compound nouns

The most recent orthographic word list of the Erzya language Бузакова et al. (2012) prescribes a mathematical strategy to compounding, i.e. if the first element is a nominative singular indefinite form (also called absolute form) with no evident derivation (sometimes a rather gray definition) two nouns are written as a single unit. This solution is not entirely related to the writing practices of the last century, and there do prove to be certain problems. The most evident problems are ensemble nouns containing mensural classifiers (for definition see Lyons (1977): p.463; cf. Rueter (2013): p.108).

(7) a.

keĺme	veď	vedra
ADJ	NOUN	NOUN
COLD	WATER	BUCKET
SG	SG	SG
NOM	NOM	NOM
INDEF	INDEF	INDEF

'bucket of cold water'

b.

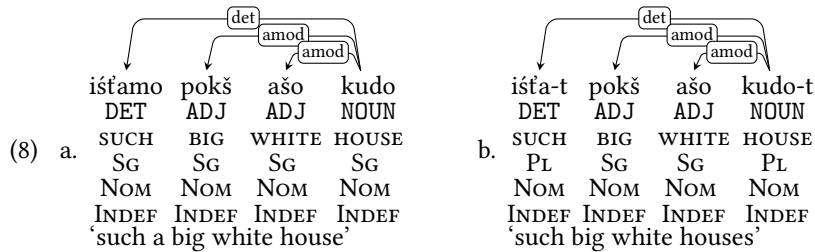| pokš	vedra	veď
ADJ	NOUN	NOUN
BIG	BUCKET	WATER
SG	SG	SG
NOM	NOM	NOM
INDEF	INDEF	INDEF

'big bucket of water'

In both instance we are talking about a measurement of water, whereas the idea of a bucket especially intended for water would be constructed in the telic noun /*veď vedra*/.
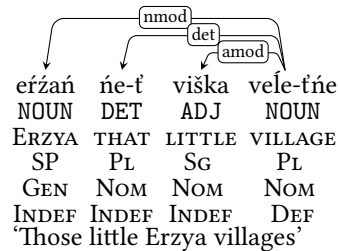
4.6 Noun head ellipsis

An analogy of symmetric negation as described in Miestamo (2013) can be applied to the description of the Erzya nominal phrase declension. In symmetric negation the structure of the negative is identical to the structure of the affirmative, except for the presence of the negative marker(s).

In a similar way, it is the final word of the Erzya noun phrase that is symmetrically declined while modifiers (with the exception of some determinatives) appear in what is termed the absolute form. Determinatives such as /*iśťamo*/ 'like this/that' can agree in number with the head noun. (NB. some descriptions Bartens (1999) maintain that regular adjectives might agree for number as well, but this type of apparent agreement, seems to be limited to parts of the northwestern dialect)

(8) a.

isťamo	pokš	ašo	kudo
DET	ADJ	ADJ	NOUN
SUCH	BIG	WHITE	HOUSE
SG	SG	SG	SG
NOM	NOM	NOM	NOM
INDEF	INDEF	INDEF	INDEF

'such a big white house'

b.

isťa-t	pokš	ašo	kudo-t
DET	ADJ	ADJ	NOUN
SUCH	BIG	WHITE	HOUSE
PL	SG	SG	PL
NOM	NOM	NOM	NOM
INDEF	INDEF	INDEF	INDEF

'such big white houses'

(9)

eŕźań	ńe-ť	viška	veĺe-ťńe
NOUN	DET	ADJ	NOUN
ERZYA	THAT	LITTLE	VILLAGE
SP	PL	SG	PL
GEN	NOM	NOM	NOM
INDEF	INDEF	INDEF	DEF

'Those little Erzya villages'

In an elliptical construction, where the noun head is recoverable or inferable from the context[10], the case ending is merely joined to the final modifier of the noun phrase. In practice, the final modifier can be an adjective (such as color, size, shape, etc.), a participle or converb, a determinatives (ordinal, demonstratives, collective numerals), non-core case-form nouns (including: Ine, Abe, Cmpr, Tra, Prl, etc.). If the modifiers are genitive in form, however, they generally require an additional /śe/ element, i.e. a demonstrative type element. Finally, it should be noted that this kind of construction usually takes either definite or possessive access marking, see Rueter (2010).

(10)

piže-ťńe
green-N.SP.GEN.INDEF-PL.NOM.DEF
'the green ones'

(11)

ruzoń-śe-ťńe
Russian-N.SP.GEN.INDEF-DET.DEM.SG-PL.NOM.DEF
'the Russian ones'

Less frequent, perhaps due to contextual presuppositions, comes the vowel-final modifier with the additional determinative. This more complex construction presupposes a contrastive context.

(12)

ašo-śe-ťńe
white-A.SG.NOM.INDEF-DET.DEM.SG-PL.NOM.DEF
'those white ones'

In elliptical constructions of the nominal phrase, the part-of-speech has been retained in column 4, whereas column 6 bares witness to a deluge of ordered zero-derivation, and column eight indicates the actual dependency relation of the noun phrase head.

---

[10] http://www.glossary.sil.org/term/elliptical-construction

4.7 Numerals

The Erzya language has several types of regular counting, cardinal, collective, distributive, multiplicative. There is no problem with counting nouns, such as singular concrete items, pairs or sets (such as socks, batches, broods); these numerals can be readily connected with the *nummod* dependency relation to the noun they modify. Problematic are the missing dependency relations which might better be characterized as iterative numerals; they also count entities, i.e. iterations of a predication (similar to once, twice, but not twofold, double or the second time).

Associative collectives have also been an initial problem, but by following ongoing discussions involving the use of French *tout* the relevant dependency relations would be, perhaps, *advcl* and *acl*, even *det*. Associative collectives, such as 'both of you' are regular for numerals far beyond ten.

(13)

> pand-śť kolmokśť
> pay-Ind.Prt1.ScPl3 three-Num.Iter
> 'they payed three times'

(14)

> pand-śť kolmońkirda
> pay-Ind.Prt1.ScPl3 three-Num.Mult
> 'they payed triple/threefold'

(15)

> pand-śť kolmoń-kolmoń jeuro
> pay-Ind.Prt1.ScPl3 three-Num.Distr Euro-N.Sg.Nom.Indef
> 'they payed three Euros each'

(16)

> pand-śť kolmoćeďe
> pay-Ind.Prt1.ScPl3 three-Num.Ord.Par
> 'they payed for the third time'

5 Future work

Since treebank work with Erzya is barely off the ground, there is still plenty of work to do with guidelines for further and consistent annotation. In the time since this paper was originally submitted much experience has been obtained with regard to consistent annotation. One primary undertaking in the guidelines, however, is to render them workable for the closely related Moksha language. When more treebanks have been added, there should be more opportunity to apply the Erzya model since the morphological analyzers for both languages have been designed using parallel tagging strategies where ever possible. The automation of dependency relation assignment will require further work in constituency ordering documentation, as Erzya is a so called pro-drop language attesting to low frequency for conjunctions and other syntactic structural markers.

6   In conclusion

This has been a description of the first steps to building Erzya treebanks in accordance with Universal Dependencies. Much space has been dedicated to extensive morphological contemplation, where the matters requiring in-depth consideration are actually the minimalized set of dependency relations in tandem with morphological information, i.e., minimal use of language-dependent subfeatures. Hopefully, this work will provide a means for pivoting and sharing in what has been achieved for larger languages.

Acknowledgements

References

Raija Bartens. 1999. *Mordvalaiskielten rakenne ja kehitys*, volume 232 of *Suomalais-Ugrilaisen Seuran Toimituksia*. Suomalais-Ugrilainen Seura, Helsinki.

Arja Hamari. 2007. *The negation of stative relation clauses in the Mordvin languages*, volume 254 of *Suomalais-Ugrilaisen Seuran Toimituksia*. Suomalais-Ugrilainen Seura, Helsinki.

László Keresztes. 1999. *Development of Mordvin Definite Conjugation*, volume 233 of *Suomalais-Ugrilaisen Seuran Toimituksia*. Suomalais-Ugrilainen Seura, Helsinki.

Maria Kholodilova. 2016. *Moksha non-verbal predication*, Printon, Tallinn, pages 229–259. Uralica Helsingiensia 10.

John Lyons. 1977. *Semantics 2*. Cambridge University Press.

Matti Miestamo. 2013. *Symmetric and Asymmetric Standard Negation*, Max Planck Institute for Evolutionary Anthropology, Leipzig, chapter 113. http://wals.info/chapter/113.

Jack Rueter. 2010. *Adnominal Person in the Morphological System of Erzya*, volume 261 of *Suomalais-Ugrilaisen Seuran Toimituksia*. Suomalais-Ugrilainen Seura, Helsinki.

Jack Rueter. 2013. *On quantification in the Erzya language*, LINCOM, Muenchen, pages 99–118.

Trond Trosterud. 2006. *Homonymy in the Uralic Two-Argument Agreement Paradigms*, volume 251 of *Suomalais-Ugrilaisen Seuran Toimituksia*. Suomalais-Ugrilainen Seura, Helsinki.

Rigina Turunen. 2010. *Nonverbal Predication in Erzya*. A. S. Pakett, Tallinn.

Francis M. Tyers, Mariya Sheyanova, and Jonathan North Washington. 2018. UD Annotatrix: An annotation tool for Universal Dependencies. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*. page [to appear].

Francis M. Tyers and Jonathan N. Washington. 2015. Towards a free/open-source Universal-dependency treebank for Kazakh. In *3rd International Conference on Turkic Languages Processing, (TurkLang 2015)*. pages 276–289.

Johan van der Auwera, Ludo Lejeune, and Valentin Goussev. 2013. *The Prohibitive*, Max Planck Institute for Evolutionary Anthropology, Leipzig. http://wals.info/chapter/71.

F. J. Wiedemann. 1864. *GRAMMATIK DER ERSAMORDWINISCHEN SPRACHE*. MÉMOIRES DE L'ACADEMIE IMPERIALE DES SCIENCES DE ST.-PETERSBOURG. Buchdruckerei der Kaiserlichen Akademie der Wissenschaften., St. Petersburg–Riga–Leipzig.

Н.А. Агафонова, Р.А. Алёшкина, Г.Ф. Беспаова, Водясова Л.П., Клементьева Е.Ф., И.Н. Рябов, Г.В. Рябова, А.М. Харитонова, and В.П. Цыпкайкина, editors. 2011. *Эрзянь кель. Синтаксис: тонавтнемапель*. Издательство мордовского университета, Саранск.

Р.Н. Бузакова, Е.Н. Лисина, М.В. Мосин, and В.П. Цыпкайкина, editors. 2012. *Эрзянь келень орфографиянь валкс.* «Красный Октябрь» типографиясь, Саранск.