Numerals and what counts

Jack Rueter
Department of Digital Humanities
University of Helsinki
jack.rueter@helsinki.fi

Niko Partanen
Department of Finnish, Finno-Ugrian and Scandinavian Studies
University of Helsinki
niko.partanen@helsinki.fi

Flammie A. Pirinen
Divvun
Uit Norgga árktalaš universitehta
Tromsø, Norway
tommi.pirinen@uit.no

Abstract

This study discusses the way different numerals and related expressions are currently annotated in the Universal Dependencies project, with a specific focus on the Uralic language family and only occasional references to the other language groups. We analyse different annotation conventions between individual treebanks, and aim to highlight some areas where further development work and systematization could prove beneficial. At the same time, the Universal Dependencies project already offers a wide range of conventions to mark nuanced variation in numerals and counting expressions, and the harmonization of conventions between different languages could be the next step to take. The discussion here makes specific reference to Universal Dependencies version 2.8, and some differences found may already have been harmonized in version 2.9. Regardless of whether this takes place or not, we believe that the study still forms an important documentation of this period in the project.

1 Introduction

Numerous treebanks in the Uralic languages have become available within the Universal Dependencies (UD) project (Zeman et al., 2021). In recent years, at least within the Uralic language family, we have seen new treebanks emerging in languages with closely related siblings that already have an existing treebank. Examples of such languages are Skolt Saami, in relation to Northern Saami (Tyers and Sheyanova, 2017), Komi-Permyak, in relation to Komi-Zyrian (Partanen et al., 2018), or Moksha in relation to Erzya (Rueter and Tyers, 2018). Although the entirety of Uralic languages is still not fully represented within the Universal Dependencies project, the situation has improved in many ways since the last survey on the state of this language family in UD was conducted (Partanen and Rueter, 2019). While more extensive surveys are useful, we think there are situations where individual nuanced features should be compared between the languages, so that consistency could be maintained and improved upon. At the same time, this may provide a thoughtful point of departure for new discussions around such features, as we believe the questions discussed here are relevant beyond the realm of Uralic languages. Even in other treatment of UD on different language groups, such as Slavic, numerals have been recognized as one category that demands special attention (Zeman, 2015). Recently Schneider and Zeldes have also discussed inconsistent nominal constructions in the English treebanks (2021), and even the issues we describe in the Uralic treebanks here can well be described in a similar vein. These are not dramatic issues, but small points of divergence that we could pay attention to, but if we decide to do so, we would also need to devise strategies to operationalize the edits in numerous languages with a long history of treebank work.

We can additionally point to recent discussions within the Universal Dependencies project where the various ways to annotate English numerical expressions have been discussed.1 Conversations such as

---

1This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/.

1https://github.com/UniversalDependencies/docs/issues/654
these are relevant for Universal Dependencies developers more widely, and for the sake of consistency such decisions should be at least considered for the other languages in the project. Our study also discusses some numeral types in the Uralic languages that are known, but not yet attested in the treebanks. Thereby, their description provides an important starting point for future work on these languages, during which these forms will inevitably be encountered.

2 Numerals in Universal Dependencies

In this paper, we discuss numerals in the Uralic languages. Probably the simplest approach would be to gather all numeral-type words on the basis of their Universal part-of-speech numeral (UPOS NUM) value or features making reference to numerals in different Uralic languages. Among the features at least NumType is one that would be presumed to be present with all numerals, although it also occurs widely with other parts of speech. The possible, currently documented numeral types are cardinal numerals, ordinal numerals, multiplicatives, fractions, distributives, sets or collective numerals and ranges. These concepts provide a good base for a relatively elaborate and nuanced system, but at this phase the UD system appears slightly asymmetric.

Potential asymmetry might be dealt with by adding a binary for the split between numerals and counted nouns versus nouns with sequential deixis-like marking. In the Erzya, Moksha languages, sequential deixis is readily attested in combination with multiplicatives and sets, but due to the fact that ordinals only comprise three combinatorial instances in Erzya, it may strike us as fruitless to introduce a plus/minus binary for ordinal. The Erzya examples below illustrate this.

- nummod [-Ord] vejke ‘one’
- nummod [-Ord][+Approx] kavtoška ‘couple’
- nummod [-Ord][+Sets] kavonst ‘two pairs/sets’
- nummod [-Ord][+Dist] kavtoñ-kavtoñ ‘two-by-two’
- advmod [-Ord][+Mult] kavšít ‘twice, two iterations of the verb’
- advmod [-Ord][+Mult] kavroñkirda ‘twofold, double the amount’
- advmod [-Ord][+Mult][+Approx] kavkšt’eška ‘a couple of times’
- advmod [+Ord][+Mult] omboc’ed’e ‘for the/a second time’
- advmod [+Ord][+Mult] ombonoñkirda ‘a second time’
- amod [+Ord][+Sets] omonst ‘a second set’
- amod [+Ord] omboc’e ‘second’
- det [-Ord][+Tot] kavoriñe’ek ‘the both of us’
- det [-Ord][+Approx][+Tot] kevet’eješkañest ‘the approximately 15 of them’
- det [+Ord] ombot’ks ‘the second’

Above, we can observe that the approximatives and distributives including universal quantifiers are not associated with sequential deixis in Erzya. Whereas, sequence and range might readily be combined. In counting iterations of a predicate, Erzya shows a clear distinction between it and quantification of mass (‘twice’ and ‘twofold’ cannot be equated), but this distinction becomes less obvious when applied to a sequential deixis system. A glimpse at Komi-Permyak and Komi-Zyrian will remind us that multiplicatives may also be used in a distributive context (Rueter et al., 2020, 22). Multiplicatives, sets,
distributives, etc. should not be distinguished from ordinals any more than they are from cardinals, since the term cardinal might readily be treated as a zero like nominative singular. The last three items within the list above are also exceptional as they would demand syntactic dependency ‘det’, which according to the guidelines is not allowed. Analogically, chosen conventions could possibly also be extended to the annotations of items such as English ‘both’ and Swedish ‘båge’.

Conceivably, numerals might be divided into various categories according to their semantic use. The most predominant numeral types might therefore be associated with quantification, sequence, and entity naming. Quantification articulates distinctions in the mechanisms of counting. Singular entity counting is typified by the use of cardinals (such as in Finnish yksi ‘one’, kaksi ‘two’, kolme ‘three’, etc.), and there may be different marking patterns for the counted noun.

In many languages, there are standards by which the head noun of a nummod dependency takes special marking. In Komi-Zyrian, Komi-Permyak and Hungarian, for example, the counted noun shows no deviance from its regular nominative singular marking strategies when qualified by any cardinal numeral. In Balto-Finnic, Finnish, Estonian, Livvi and Karelian, the partitive singular marks the counted nouns when they are qualified by numerals two and above, even though their syntactic position would otherwise call for a nominative singular—for other cases a fitting semantic or syntactic case is used, i.e. phrase agrees in case.

(1) a. kolmešukupolvie
three.NOM.SG generation.PAR.SG
‘two generations’ (krl: vepkar-1652.40)

b. kuutta kertua enemmän
six.PAR.SG time.PAR.SG more
‘six times more’ (krl: vepkar-1740.21)

c. šuašša muašša
hundred.INE.SG land.INE.SG
‘in a hundred lands’ (krl: vepkar-1740.6)

Contrastively, the Mordvin languages, Erzya and Moksha, exhibit a variation that has yet to be researched in depth, i.e. counted nouns do not obligatorily take special marking when qualified by cardinal numerals two and upward, see Markov (1961, 42) and Rueter (2013, 107), but perhaps also in dialect studies (Ryabov, 2016; Rueter, 2016; Levina, 2021; Agafonova and Ryabov, 2021). A similar phenomenon can be observed in Moksha (Rueter, forthcoming 2022). The Saami languages attest to two different strategies: Northern Saami takes genitive singular marking of its counted nouns when qualified by numerals two and above, whereas Skolt Saami makes a three-way split, a genitive singular marking the numeral range 2–6, and the partitive marking seven and upward (with the decline in language proficiency the use of the partitive has become less certain).

Sets of entities, i.e. sets with more than single members, are counted synthetically across the languages with various strategies. In Finnish, for example, pairs of scissors are counted by using plural forms of the cardinal numerals and the NP head noun alike, e.g. yhdet sakset ‘one pair of scissors’ (here both the numeral and the noun it qualifies are in the plural, and unlike Russian the distinction is retained for numerals five and above, too). In contrast, Erzya has its own numeral forms typically derived in -Onst, hence kavonst vasońpejel’ ‘two pairs of scissors’ with the counted noun in the plural. Although numerals of the sets type are typically introduced for counting pairs, they are, in fact, often used with larger sets, such as sets of six cups and saucers.

Iterations of predications are often counted with adverb derivations of cardinal numerals, but the productivity of these derivations still requires assessment from language to language. While Finnish only minimally utilizes the word forms in -sti: kahdesti ‘twice’, kolmesti ‘thrice’ and tuhannesti ‘a thousand times’, the Hungarian, Komi-Zyrian, Komi-Permyak, Erzya and Moksha languages use regular derivations for indicating ‘X times’, -szer/-ször/-szor, -iš, -iš, -kśt and -kśt’, respectively. Needless to say,
matters become confusing when these iterative numerals are categorized as multiplicatives in UD. The result, at least in Erzya, is that ‘being paid kavkśit’ = twice’ and ‘being paid kavonkirida = double or twofold’ are registered as the same thing, which is by no means always the state of affairs semantically, but from a syntactic perspective it is plausible.

Distributive numerals are not a simple class. They can be further categorized into subclasses, as immediately becomes apparent in the two Hungarian strategies: két-két ‘two each’ with a noun head, and kettesével ‘two at a time’ with a verb head. Whereas the former may be used as a definite numeral in the context Berta és Rudi két-két csomagot hozott ‘Berta and Rudi brought two suitcases each’, implying that a total of four suitcases were brought, the latter expression is indefinite. The indefinite distributive numeral kettesével ‘two at a time’ in nearly the same context Berta és Rudi kettesével hozta a csomagokat ‘Berta and Rudi brought the suitcases two at a time’3 would indicate that each iteration of the predication involves two suitcases, but there is no indication regarding the number of iterations – it could be any number of times. In this context, definiteness is lent by the object, i.e. ‘the suitcases’.

Approximative numerals are numerals with values slightly less or more than the number given. Finnish, for example, attests parikymmentä ‘about twenty’ from the words pari ‘couple’ and kymmentä ‘ten (partitive)’. In addition to constructions with the element pari, there are fairly regular derivations formed from other basic numerals as well: kolmisen + kymmentä ‘approximately thirty’.

In Erzya, as in Moksha, approximative forms in -ška are found for counting entities vet’eška loman ‘about five people’ and iterations kološšakšt ‘about three times’. With the use of an approximative numeral, the likelihood rises that no plural marking is indicated on the counted noun. The predominance of nominative singular marking of the NP head also holds when the approximative is marked with an N–(N + 1) strategy, i.e. vet’e-koto loman ‘five-or-six people’. The use of adjacent numerals to indicate approximate values is also found in Komi-Zyrian, i.e. vit-kvajt and vit-ō-kvajt both translate to five or six.

In Finnish, the expression of range with numerals follows the same pattern as is observed in point of departure to end destination, i.e. the elative case marks the starting point, and the illative marks the end point. In the range 5–7 kilometers, the Finnish involves viidessei seitsemän kilometriä five+elative, seven+illative and kilometer+partitive, which is the same counted noun strategy observed in basic numerals.

Fractions in Finnish can be expressed in at least two different ways. One way is to join the ordinal nominative singular with the noun osa ‘part’, hence viides + osa = viidesosa, where only the end is declined and as such is distinguished from ‘the fifth part’ of something, where we would actually be talking of sequences. Syntactically, neljä viidesosa ‘four fifths’ functions in the same manner as any noun with a cardinal qualifier, i.e. the NP head is marked with the partitive singular when in an otherwise nominative-singular position nummod(viidesosa, neljä). The second derivational expression for ‘fifth’ is viidenennes, it too is treated syntactically as a counted noun, as appears to be the case in other Uralic languages.

Universal quantifiers, such as the Finnish molemmat ‘both’, have more complex counterparts in Hungarian mindkettő (literally ‘all’ + ‘two’), which may also take associative marking for first, second and third persons plural in mindkettünk, mindkettettek, mindkettük, respectively. The Hungarian mindhárom ‘all three’, ‘tous les trois’ then comes as no surprise, and one begins to expect subsequent mindnégy ‘all four’. Komi-Zyrian and Erzya attest to yet another aspect: the associative personal reference can also be in the singular, allowing for access. If we are speaking of a singular ‘person’ and mention that ‘the (lit.) three of him/her are moving to town’ (Rueter, 2013), we access a definite universal quantifier pronoun with reference to this single person. This feature is not observed in Hill Mari or Udmurt (Ke’l makov and Hännikäinen, 2008, 111–112). Ordinal numerals can be associated with multiplicative, iterative and sets features. This has been observed in the presentation of some morphology for Erzya, above.

Numerals appear in entity naming, for example the Finnish viitonen ‘fiver’ may be used when making reference to money, on the one hand, but it could also be used in reference to a street car, where we would be more likely to translate it as ‘street car five’ or ‘street car number five’. Thus is fits

3cf. http://en.utdb.nullpoint.info/type/hungarian/distributive-numerals/dupldnn-sufdnv
directly into a list of problems in apposition, such as ‘the color purple’, ‘the word terrorist’ and many others including numerals discussed by Schneider and Zeldes (2021). An extension to this numeral issue is found in Finnish viitonen in reference to ‘house number five’, but the same 5 is transformed to the cardinal-form viisi if the house is 5a or 5b – viisi a or viisi b, respectively (no partitive, of course, so we are not counting letters). Here, the Erzya solution is to use the ordinal vet’eče ‘the fifth’ for 5 and vet’eče a ‘fifth a’ for 5a, which results in ambiguous homonymy.

There are differences observed across languages, where synthetic versus analytic expressions of the same numerical values might be dealt with differently. Thus, our first overview discusses the largest spread of numeral types, forms across languages. Once the collection is complete, the numeral words can be classified according to the dependencies and features. In Finnish, for example, we predict four different and regular dependencies: nummod (for cardinals and plural cardinals with plurale tantum), advmod (for counting iterations of a predication, e.g. once, twice, thrice), advcl (for distributive quantification), amod (for ordinals). Other languages, it will be noted, may have extensive det (this is not really productive in Finnish, but would be the equivalent for ‘both’ and its analogues with universal quantification of numbers three and up, probably with person marking as well, e.g. ‘the two of us’).

2.1 Numeral type

According to the Universal Dependencies documentation, some numerals can be classified as adjectives and some as adverbs.4 Thereby, in the UD guidelines both ADV and ADJ are often found as the part of speech categories for numeral expressions. At the same time, there are also situations where the NumType feature occurs with different parts of speech.

In several treebanks in the Romance languages, for example, there are pronouns such as Spanish mucho and poco which have a feature value NumType=Card. Such marking on pronouns is not common in the treebanks, although we do find English first, second, third and latter receiving POS tag PRON and feature NumType=Ord. This is also the style in Finnish, with toinen ‘second; another’ being marked similarly, and Erzya and Komi-Zyrian treebanks offer similar examples. As the combination PRON and NumType can be found only in treebanks for 10 different languages, we believe it is highly likely that similar annotations could be extended to many other languages within the project.

Nouns that are marked with NumType appear in a bit larger array of languages, all in all within 13 languages, among them, Uralic languages North Saami, Erzya and Estonian. In North Saami, these instances are collective nouns with NumType=Coll. In Erzya word pel’ ‘half’ is marked with NumType=Frac. In Estonian the only occurrences are with gene names containing numbers, such as IL-5, where NumType=Card is attested. These are all reasonable uses of NumType, as these noun types do have countable properties that are relatively well captured by the NumType feature. But again as the solutions seem language specific the annotations could be somehow harmonized or extended to more languages.

In Finnish, Icelandic and Korean treebanks we find examples of punctuation being marked with NumType=Card. No matter how the annotation is motivated, being this rare and narrowly distributed is possibly problematic for the comparability of the languages. The Estonian treebanks EDT and EWT only use NumType with two values, Card and Ord. This does not appear to rule out fractions, but they are dealt with differently, i.e. 3/4 is given the features NumForm=Digit and NumType=Card. Of course, here the value Digit indicates not written as words. A second issue in EWT is that the feature NumType=Ord is used with both UPOS NUM and ADJ. It seems that ordinal digital numerals consisting of an Arabic numeral followed by a full stop are treated as ADJ, whereas automobiles from different years have an abbreviated year digit pair followed by an apostrophe. This latter type has the UPOS value NUM, should this be the case? We will not widely compare the differences between multiple treebanks on the same language, although we do acknowledge this is an issue that needs further attention.

Having discussed the general use of NumType feature and some rarer patterns that can be found, we will next describe more in detail different numeral types and their occurrences, with references both to Uralic and other language families, as necessary.

4https://universaldependencies.org/u/pos[NUM].html
2.1.1 Cardinal numerals

The cardinal numeral type in UD is typified as an expression for counting singular items. Thus, this feature might be associated with the UD part of speech NUM (as in one, two, three, etc.). This feature value is also used with non-numerals (as in many, few, Czech kolik ‘how many’, etc.). Here, however, individual languages make a split between use of UPOS DET and NUM. The latter of which, apparently, is defended in Czech by a strong grammatical tradition, might be used for the interrogative kolik ‘how many’, which evokes cardinal numerals. Czech includes yet a third type of words as cardinals which seem to indicate the total number, e.g. čtvero (as in Čtvero ročních dob ‘The Four Season’, all four), desatero (as in ‘the Ten Commandments’, all ten). This presumably explains the definition of oba ‘both’, which in Czech is marked as UPOS NUM, whereas Talbanken deals with bågge ‘both’ as a DET. And then there is the one instance of desatero in the treebank Desatero investora ‘Lit. The ten investors’, where the word desatero has the UPOS NOUN.

This third group of cardinals, which is not observed in Swedish as a consistent counting system, appears with a nummod dependency in Czech to match the UPOS NUM. In Swedish and other languages without this counting system, words with the meaning ‘both’ are generally dealt with as DET, and they have a feature PronType=Tot.

2.1.2 Ordinal numerals

Ordinals can be seen to represent subtypes of adjectives and adverbs. In addition to the amod dependency associated with the words first, second, third, there are analogical interrogatives, etc.), there is also an advmod dependency, associated with ordinal multiplicatives, such as the Czech poprvé ‘for the first time’. By applying the feature value NumType=Ord to both UPOS ADJ and ADV, we could remove the NumType=OrdMult feature value used in Komi-Zyrian nöl’öd ‘fourth’ UPOS ADJ and nöl’ödyš ‘for the fourth time’ UPOS ADV and similarly in Erzya, Moksha and Komi-Permyak. The downside is that the parallel between cardinal and ordinal multiplicatives becomes less obvious. If we were to do so, we would be faced with the challenge of addressing numerals with three features: ordinal multiplicative and distributive.

Numerals can be classified according to what they actually enumerate or do they at all. In Erzya, the numeral type (a) vejke, kavto, kolmo, nil’e is used for counting individual entities. The pertinent dependency is nummod. (b) vejenst, kavonst, kolmonst, nil’enst is used for counting set entities from pairs of scissors to sets of cups. The pertinent dependency is nummod. NumType=Sets (c) vest’, kavkst’, kolmokst’, nil’ekst’ is used for counting iterations of a given predication. Thus this has a advmod dependency. NumType=Mult (d) Delimiting associative collectives škamost, kavonist, kolmoňest, nil’isten provide universal quantification values found in the expressions ‘alone’, ‘both’, ‘all three’ with the addition of associative reference to number and person. These numerals are used in secondary predication with reference to the subject or object. Features include PronType=Tot (e) Distributive, imperfect kavton’-kavton’, kolmońi-kolmońi, nil’ei-nil’eš NumType=Dist Aspect=Imp (f) Distributive, perfect kavton-kavto, kolmoń-kołmoń, nil’e-nil’e NumType=Dist Aspect=Perf (g) vejenkirda, kavonkirda, kolmońkirda, nil’etikirda has an advmod or amod dependency, and the feature value NumType=Mult.

2.2 Numerical dependencies

Among the dependency relations assigned to the numerals, the most common is nummod. In many Slavic treebanks an additional relation of det is used, as in det:nummod. This is not used in other treebanks. In Beja treebank there is an individual occurrence of nummod:det. Another subtype of nummod, nummod:entity, appears to be used only in the Russian treebanks, especially in relation to the symbol ‘№’. Additionally nummod:flat appears only in one Polish treebank. Phenomena attested and seen necessary to annotate in the Slavic languages could also be very relevant for work with the Uralic languages, many of which have been in extensive contact with Russian.

Our analysis also indicates that the relation nummod in the Uralic languages virtually always connects to part of speech NUM. With the other languages, there is extensive variation, even though this relation is always the most common. Whether this is simply a matter of annotation conventions, linguistic description traditions or actual linguistically relevant differences, remains to be studied.
3 Discussion

As we have shown, numerals and related expressions are an area for fruitful and needed further discussion in the Universal Dependencies project. Which forms all get numeric features extends widely beyond just numerals themselves, and many lexical items that have counting properties could be annotated with NumType features, and already be annotated in different treebanks. Which of the individual solutions in different treebanks should be described better in the documentation and adapted further, and which should be harmonized in comparable uses of the treebanks, remains to be discussed, but we hope our observations help at least a bit along this path. Of course, how work on various inconsistencies should or could be coordinated across the hundreds of treebanks already in the Universal Dependencies project is not entirely clear, and remains certainly a large challenge. At the same time, new treebanks are still continuously emerging, and paying attention to various strategies used in existing treebanks should help the maintainers of these new languages to adapt their conventions. When diverse language families are included, new questions inevitably arise. For example, in Apuriná there are very few actual cardinal numbers and quantification is expressed in verbal constructions (Facundes et al., 2021; Rueter et al., 2021a).

The issue how to handle Komi-Zyrian numerals was also recently discussed in the relation to Komi morphological analyser (Rueter et al., 2021b, 67), which points to the fact that the best possible annotation scheme is often a very relevant question for uses beyond the Universal Dependencies project itself. We also believe that the classification and annotation of numerals is important from the point of view of basic linguistic research and language description. As the description of Erzya counting expressions in this study showed, the system is already very complicated and nuanced in this one language, and is just starting to be adequately described in the newest grammatical descriptions (Suikkonen and Solovyev, 2013). We presume the description of many smaller Uralic languages remains much less complete, not to even mention less studied language families of the world, which also have started to have significant presence in the Universal Dependencies project. This kind of easily accessible information about counting expression at large could be immediately beneficial, for example, in typological research, and systematic annotations and documentation in projects such as Universal Dependencies is one modern way to distribute this description.

References

Nina A. Agafonova and Ivan N. Ryabov. 2021. Ulânovskoj oblasten’ novomalyklinskoj raionon’ Őrzän’ velen’ kortvatsknèsë azorksëm’ nevtescë suffisktscë’ baška ionksost. In Mika Hämäläinen, Niko Partanen, and Khalid Alnajjar, editors, *Multilingual Facilitation*. University of Helsinki.

Sidney Da Silva Facundes, Maríia Fernanda Pereira de Freitas, and Bruna Fernanda Soares de Lima-Padovani. 2021. Number expression in apurinã (arawák). In Mika Hämäläinen, Niko Partanen, and Khalid Alnajjar, editors, *Multilingual Facilitation*. University of Helsinki.

Valentin Kel’makov and Sara Hännikäinen. 2008. *Udmurtin kielioppia ja harjoituksia*. Apuneuvoja suomalais-ugrilaisen kielen opintoja varten — Hilfsmittel für das Studium der finnisch-ugrischen Sprachen, XIV. Finno-Ugrian Society, Helsinki, Finland.

Mariä Z. Levina. 2021. Elektronnyj ázykovoj korpus kak faktor soxraneniâ mordovskix (mokšansogo i ѐrzanskogo) ázykov. In Mika Hämäläinen, Niko Partanen, and Khalid Alnajjar, editors, *Multilingual Facilitation*. University of Helsinki.

F. P. Markov. 1961. Priatytyskij dialektr Òrzâ-mordovskogo ázyka (the priiatyrsk dialect of the erzya-mordvin language). In *Očerki mordovskix dialektov*, tom V, pages 7–99, Saransk, Mordovia ASSR, USSR. Mordovskoe kniznoe izdatel’stvo.

Niko Partanen and Jack Rueter. 2019. Survey of Uralic Universal Dependencies development. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 78–86, Paris, France, August. Association for Computational Linguistics.

Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau, and Michael Rießler. 2018. The first Komi-Zyrian Universal Dependencies treebanks. In *Second Workshop on Universal Dependencies (UDW 2018), November 2018, Brussels, Belgium*, pages 126–132.
Jack Michael Rueter and Francis M Tyers. 2018. Towards an open-source universal-dependency treebank for Erzya. In International Workshop for Computational Linguistics of Uralic Languages.

Jack Rueter, Niko Partanen, and Larisa Ponomareva. 2020. On the questions in developing computational infrastructure for Komi-Permyak. In Proceedings of the Sixth International Workshop on Computational Linguistics of Uralic Languages, pages 15–25.

Jack Rueter, Marilía 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, pages 28–33, Online, June. Association for Computational Linguistics.

Jack Rueter, 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.

Jack Rueter. 2013. Quantification in Erzya. In Pirkko Suihkonen and Valery Solovyev, editors, Typology of Quantification:, LINCOM Studies in Language Typology, pages 99–122, Germany, 12. Lincom GmbH.

Jack Rueter. 2016. Towards a systematic characterization of dialect variation in the erzya-speaking world: Isoglosses and their reflexes attested in and around the dubyonki raien. In Ksenia Shagal and Heini Arjava, editors, Mordvin Languages in the Field, volume 10 of Uralica Helsingiensia, page 91–108.

Jack M. Rueter. forthcoming 2022. Mordvin. In Daniel M. Abondolo; Riitta Valijärvi, editor, The Uralic Languages. Routledge.

Ivan Ryabov. 2016. Ob issledovanii ´erzˆanskix dialektov metodami lingvisti ˇceskoj geografii [on research of the erzya dialects with linguistic-geographic methods]. In Ksenia Shagal and Heini Arjava, editors, Mordvin Languages in the Field, volume 10 of Uralica Helsingiensia, page 91–108.

Nathan Schneider and Amir Zeldes. 2021. Mischievous Nominal Constructions in Universal Dependencies. arXiv preprint arXiv:2108.12928.

Pirkko Suihkonen and Valery Solovyev, editors. 2013. Typology of Quantification: On Quantifiers and Quantification in Finnish and Languages Spoken in the Central Volga-Kama Region, volume 28 of Studies in Language Typology. LINCOM, Munich. Quantification in Erzya, Finnish, Russian, Tatar, Udmurt, with appendices in Chuvash, English, Erzya, Finnish, Russian, Tatar and Udmurt.

Francis Tyers and Mariya Sheyanova. 2017. Annotation schemes in North Sámi dependency parsing. In Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages, pages 66–75.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noêmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabriéle Aleksandraciūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arıcan, Jack Rueter, Marília Fernanda Pereira de Freitas, Sidney Da Silva Facundes, Mika Hämäläinen, and Trond Trosterud. 2020. On the questions in developing computational infrastructure for Komi-Permyak. In Proceedings of the Sixth International Workshop on Computational Linguistics of Uralic Languages, pages 15–25.
