Finite-state description, developing mental awareness

Rueter, Jack

Northern European Association for Language Technology
2023-04

Rueter , J 2023 , Finite-state description, developing mental awareness . in Rule-Based Language Technology . NEALT Monograph Series , vol. 2[1] , Northern European Association for Language Technology , Tartu , pp. 217-227 .

http://hdl.handle.net/10138/357046

Rule-Based Language Technology
2023
NEALT Monograph Series, 2

Finite-state description, developing mental awareness[1]

Jack Rueter
Department of Digital Humanities
FIN-00014 University of Helsinki, Finland
jack.rueter@helsinki.fi

Abstract

In this article, we approach finite-state description practices that must be instilled in the developer. Thoughts are presented accompanied by reference to concrete experiences with different languages and their description.

We contend that finite-state description of languages leads to development in the describer-developer. This presupposes regular interaction with developers of upstream and downstream technologies. And as more languages are described, the developer learns what to choose as a starting point, hopefully with the help of a researcher, research documentation or native speaker well versed in the workings of the language.

We maintain that finite-state work should serve more than one purpose or audience, and that, as linguists, we should be raising the bar by applying the knowledge of research to description, so that our understanding of the linguistic phenomena can be attested by others or proven false. We are providing a methodology for repeatable experimentation and rule making.

We see that each language provides something unique, while sharing some recognizable features with other languages. We stress the necessity to avoid generating characters from epsilons and offer examples where it is possible to write rules that reduce characters to epsilons instead.

We also stress the need to describe the predictable infinite set of all native phenomena, whereas the unknown and random qualities introduced through language contact cannot form a foundation for our descriptions.

Finally, we call for a playful approach to phenomena in a language, because that might bring us closer to how a child would learn the language – through repetition, mistakes and self-correction.

Key Words: *finite-state morphology, regular morphology*

---

[1] **Ref**: Rueter, Jack. 2023. Finite-state description, developing mental awareness. In: Arvi Hurskainen, Kimmo Koskenniemi, and Tommi Pirinen (eds.), Rule-Based Language Technology. NEALT Monograph Series, 2:217-227. https://dspace.ut.ee/handle/10062/89595

1 Introduction

Over the period of 1996–2022, I have worked on finite-state descriptions of languages with complex morphologies, mainly Uralic but also a few from the Americas. My work has involved collaboration with other linguists, and I have benefited greatly from input from others working with finite-state description at all levels. Finding myself on the lower tier, my intention is to share practical experience, which I have gradually accumulated and adapted to subsequent language description projects I have contributed to in what is now the GiellaLT infrastructure[2].

My endeavour is to build language descriptions that might be used and reused at many levels. This includes building for the linguists such analysers whose lexical base can be part of a dictionary project and which itself can be readily adapted for use as a spell checker, on the one hand, and a computer assisted intelligent language learning tool, on the other.

In the years of my development, I have grown to understand the need for an analysis base that supports contextual disambiguation of analyses but also contributes to the function and dependency tagging of parsed sentences. I would also like to look forward to shallow-transfer machine translation.

2 Deciding where to start

When designing a new finite-state description of a language, I ask myself many questions. The answers to these questions help me to find a starting point and, perhaps, even an ending point. First, we need to know what kinds of text the analyser is to be used for. The answer to this will provide us with information on looking for alternatives in both the two-level-model rules and the concatenation paths. Second, we need to access previous work on the morphophonology of the language and assess just how much we might actually know about the language. Even knowledge of closely related languages can prove helpful, but individual languages always seem to exhibit exciting features of their own.

The languages I have been collaborating on do not all have extensive literary corpora, nor are the texts written adhering strictly to a single orthography. In fact, most texts contain erroneous word forms that may be traced back to the author, optical character recognition, or the absence of sufficient language documentation, i.e., the word is missing from the lexicon, or the specific forms are missing from the language descriptions. In many cases, the language community represented by the corpora is hetero-genic. There are hopefully fluent native speakers and perhaps writers, but there may also be language learners and researchers with varied needs.

Most important, start from what actually appears in the language as your base, i.e., where possible use existing word forms as the stems you build upon. Any deviation from this ideal will involve more work and documentation. But if you have to use hypothetical stem forms, make them regular and something that others can implement.

---

2 https://giellalt.github.io/

2.1 Information sources and what exists

Using grammars as sources is sometimes a questionable practice. They are often used for presenting prescriptive language norms. If a grammar or a schoolbook claims that some form or usage does not exist or should not be used, view this as a statement of existence. Saying, for example, that *ain't* is not a word, and that one should use *am not* instead, should clearly elicit adding the form to a description, at least a list of exceptions – these are not random combinations of letters *xqztvy* or other gibberish but actually word forms the reader is expected to associate with. What is more, the language learner may use non-standard forms learned in the language environment, whereas the researcher can benefit from alternative morphological readings – here the finite-state description may be aware of all readings available, including hypothetical ones.

2.2 Phonetic, pedagogical and practical

The language learner will make spelling mistakes. Spelling errors may derive from any number of sources, e.g., dialect variation, orthographic practices unknown to other languages of fluency, complex morphology. Skolt Saami, for example, has a standard orthography containing 38 characters: <aâbcčʒǯdđefgǧghijkǩlmnŋoõprsštuvzžää'>. An additional two,<ę'>, have been included in the description, to enhance tools addressing pedagogic needs of the community, as the two additional characters provide for a balanced description of word morphology. More work must be done in the description of spelling mistakes, so that the analyser will recognize them as well, and provide links to proper readings with writing tools[3].

2.3 In stride with or one step ahead of the other linguists

The language researcher may have a limiting prescriptive grammar that presents an incomplete paradigm. This is where Moksha Mordvin literary corpora (cf. ERME, PaBiVUS) have actually evidenced extensions needed in the transducer for nominative singular and plural forms of possessive pronouns (Alâmkin et al.). Subsequent work, however, will require that the new cells, not attested in the grammar, be tagged as standard or substandard. The idea is to become aware enough of the language that you can, with the help of the computer, suggest **regular** extensions to the paradigms. This can be seen in the description of a whole new dimension of derived diminutive forms in the Skolt Saami language ICALL program[4].

2.4 Keep within the domains of your language and do regular things

It is imperative that the finite-state description of a given language addresses both the facts and the myths of the morphology. By following in the steps of previous descriptions

3 https://divvun.org/

4 https://oahpa.no/nuorti/morfas/der/

with meticulous accuracy, we can generate regular paradigms, which in turn help to spot gaps in research and documentation. This work can and should also contribute to the work of others. Leave space in your description for semi-automated extension, but refrain from contributing to ambiguity, which should be avoided where possible. Interlink your data sets with those of others; this is intriguing and invites them to collaborate (cf. Latvian in Southern Balto-Finnic dictionary[5]; Apurinã; Lushootseed).

3 What I learned from Skolt Saami

New languages present exciting features and challenges. Greet them with enthusiasm, and your work will be rewarding. Description of the northern and eastern Saami languages presupposes gradation and vowel change. There is no vowel harmony triggered by the stem with a reflex in suffixes. In a way, these languages are simple, as their stem variation permeates entire languages.

Finnish gradation is nearly limited to plosives ‹*p, t, d, k*› and ‹*v*›, but it is accompanied by vowel harmony. In Skolt Saami, however, we are only dealing with sound variation in the stem. Gradation affects virtually all Skolt Saami consonants with predictable regularity (cf. *loaf-loaves*, *wife-wives* in English, but consider them regular), and simultaneously the vowels in the stem show patterns of height, length and palatal coloring (cf. *sing-sang-sung*, *write-wrote-written* and *mouse-mice*, but regular). Vowel length exhibits regular but challenging alignment with consonant gradation. Many of these morphophonological features correlate to what happens in the larger Northern Saami language, but there are certain phenomena that cannot be directly transferred in finite-state descriptions. Hence, they require separate consideration. Even the alignment of consonant and vowel phenomena in the same language can merit separate treatment.

3.1 Splitting gradation description into individual phenomena

The description of Skolt Saami gradation was initially begun on the Giellatekno infrastructure with the idea of transferring rules from the Northern Saami to Skolt Saami. When I took up work with the Skolt Saami description in 2013, my point of departure was slightly different. Previous work had set up a three-grade system with simultaneous rules to produce three grades, which simultaneously affected both consonants and vowels, which is how gradation works in Northern Saami: weak grade indicates a long coda vowel and short consonant; strong grade a short coda vowel and long consonant, while extra-strong grade indicates an extra-short vowel and extra-long consonant. My description, in contrast, described the vowel and consonant phenomena separately. I also began introducing triggers instead of relying on morphophonological constituents available in the stems and affixes for activating two-level rules. Two extra, pedagogic characters were introduced to the description, which was intended for dictionary and paradigm generation but not for the standard language. Half a year down the road, this

5 https://sonad.oahpa.no/lav/liv

decision proved to be a lucky one. As it turned out, there was need for a fourth, allegro, grade with a short coda vowel and a short consonant. The two-level rules had already been written, so the triggers merely had to be put in the right place. Allegro had not been sufficiently documented previously.

3.2 Skolt Saami vowels and consonants in gradation

Stem coda vowels in Skolt Saami are subject to variation. This variation can be divided into three categories: vowel height, vowel length and palatalization. While the grammars generally speak of high vs low, there are also instances of regular conjugation involving an extra low reading, such as that found in the verb ‹*tie'tted*› ‘to know’ and the third person plural indicative present ‹*teä't'te*› ‘they know’ (see Sammallahti & Moshnikoff 1991: 169; cf. Moshnikoff et al. 2020: 324–325). The vertical line introduced for indicating extra-long geminates after diphthongs is placed between the geminate consonants, but it is also used to indicate extra-short diphthong in the fourth grade, allegro. At this time, the vertical is placed directly after the diphthong before the palatalization marker. This second usage of the vertical marker has occasionally confused language learners, who have taken it to be a palatal marker (Sammallahti 2012).

The Skolt Saami word ‹*jeä'n'n*› ‘mother’ provides an illustrative example of the vertical line, see (1a). In ‹ N+Sg+Gen › and ‹ N+Sg+Ill › the vertical line indicates an extra-long geminate, which simultaneously occurs with an extra-short geminate. In ‹ N+Sg+Loc+PxSg1›, however, there is no geminate to place the vertical line between, instead it is placed to the left of the palatal mark, immediately following the extra-short diphthong. In the standard spelling, which does not reflect pronunciation, the genitive singular and plural are aligned, as are the illative singular diminutive forms, see (1b).

(1a)

```
jeä′nn+N+Sg+Nom: jeä′n'n
      jeä′nn+N+Sg+Gen: jię′nn
      jeä′nn+N+Sg+Ill: jeän'na
      jeä′nn+N+Sg+Loc+PxSg1: jię''nstan
      jeä′nn+N+Pl+Gen: jie′nni
      jeä′nn+N+Dimin+Sg+Nom: jeännaž
```

(1b)

```
jeä′nn+N+Sg+Nom: jeä′nn
      jeä′nn+N+Sg+Loc+PxSg1: jie′nstan
      jeä′nn+N+Sg+Gen: jie′nn jeä′nn+N+Pl+Gen: jie′nni
      jeä′nn+N+Sg+Ill: jeänna jeä′nn+N+Dimin+Sg+Nom: jeännaž
```

The vowel, consonant and palatalization in the Skolt Saami word for ‘mother’ is described using a stem found in (2a). There are only two characters that map directly to each other and are retained throughout the paradigm: the word-initial letter ‹*j*› and the first ‹*n*› of the geminate.

(2a)

```
jeä′nn+N:j{ie}{eẹä}{ˈØ}{′Ø}n{ˈØ}{nØ}
```

(2b)

```
jeä′nn+N:jeä{ˈØ}{′Ø}nˈn
```

The extra-long consonant of the nominative singular retains the vertical line the geminate ‹*n*ˈ*n*› by default, and the second ‹*n*› is also retained. In this way, there are fewer paired rules to write, as the allegro loss of the second ‹n› is treated by a rule outside of the regular gradation paradigm, see (2b). The five diphthongs generated for the pedagogic paradigms are regular here and have not been split into separate sets as illustrated by ‹{ie}{eẹä}› in the stem of example (2a). Hence, here we are only providing specific character-to-epsilon place holders for two constituents in the word ‹*jeä′n*ˈ*n*› ‘mother’.

4 Describe the predictable infinite set

Komi-Zyrian is known for a regular l-v alternation. When {lv} follows a vowel and is itself followed by a morpheme with an onset vowel, the reflex is ‹l›. When it occurs in word-final position or is followed by a morpheme beginning with a consonant, however, the reflex is ‹v›, see examples (3) and (4), below. These rules hold in regularity for essentially 99% of the time.

```
(3) {lv}:l <=> Vowel _ %> Vowel ;

(4) {lv}:v <=> Vowel _ [ %> Consoant | # ] ;
```

Komi-Zyrian, like most other world languages, it should be noted, takes loanwords. The ‹l› in loanwords, in contradiction to native word practice does not undergo any alternation. Thus, we have names ‹*Mixail*› and even some common nouns ‹*portal*› ‘portal’ with ‹*l*› stems that do not participate in all aspects of regular Komi morphophonology. Hence, we are confronted with the cold fact that loanwords are unpredictable; we cannot tell what language they will come from nor is there an end to them in sight.

Whereas, listing or describing the instances of words that did not undergo l-v alternation seemed at first to be the smaller chore, it gradually became clear that describing documented native vocabulary and making “exception” rules for native phenomena would provide for a smaller workload over time. That is to say the infinite number of native Komi words with l-v alternation should be seen as the smaller,

predictable infinite set in comparison to the set of possible random loanwords. The unknown and unpredictable should remind us to describe the regular and known.

5 Make rules that convert to epsilon, not the other way around

Resist the urge to pull rabbits out of the hat, i.e., avoid writing rules that change epsilon into something. Whenever possible, seek out a solution that allows you to turn the rule around and render an X=>0 rule instead – there are an infinite number of epsilons between any two surface characters. It has taken a long time to learn and adhere to this practice.

In Komi-Zyrian, there were virtually two noun stems to choose from: the surface nominative singular lemma ‹*nim*› ‘name’ or the nominative singular with a third person singular possessive suffix ‹*nimys*› from which the final ‹*ys*› is removed, resulting in the same stem ‹*nim*›. For words with paragogic consonants, however, there was an actual choice: lemma, ‹*pon*› ‘dog’ versus ‹*ponj*|*ys*› : ‹*ponj*›. Since the paragogic letter is on the morphological boundary, and the same lexc lexicon is being used for a Hunspell strategy, the paragogic consonant choice has been solved using simple concatenation through continuation lexica, where one provides the paragogic ‹j› before vowels, and the other an epsilon word final or before consonants, see (5).

(5) Komi-Zyrian

```
pon+N:pon N_PARAGOGIC_J "dog" ;

LEXICON N_PARAGOGIC_J
:j VOWEL_ONSET_N_MORPHOLOGY ;
:0 CONSONANT_ONSET_N_MORPHOLOGY ;
```

The Southern Estonian Võro example places change within the stem. The stem consonant has quaternary quantity in the singular and nominative plural, whereas the oblique plural owes its complexity to binary quantity in the consonant and binary quality in the stem vowel. This can then be interpreted as six stem variants, i.e., six continuation lexica: ‹*teo*›, ‹*tego*›, ‹*teko*›, ‹*tekko*›, ‹*tekõ*›, ‹*tekkõ*›, ‹*teka*›, ‹*tekka*› required for the word ‘deed’, see (6). The stem features word-internal consonant gradation, on the one hand, and word final vowel change, on the other. In order to cope with two positions that can render epsilons, two archi-consonants have been added, one in ‹{kØ}› and the second in ‹{kgØ}› mnemonically expressing the variation.

(6) Võro

```
tego+N:te{kØ}{kgØ}o N_TEGO "deed" ;

LEXICON N_TEGO
:{back} NMN_TEGO
LEXICON NMN_TEGO
:^Pen^G2 Harm-Neutr_SG-NOM "tego-" ;
:^Pen^G1 SG-GEN_ZERO-STEM "teo-" ;
:^Pen^G3 Harm-Neutr_SG-PAR_ZERO "teko-" ;
:^Pen^G4 Harm-Neutr_SG-ILL_ZERO "tekko-" ;
```

```
:^Pen^G3^VowRM>{eõ} PL-GEN_ZERO "tekõ-" ;
:^Pen^G4^VowRM>{eõ} Harm-Neutr_PL-ILL_ZERO "tekkõ-" ;
+Dial:^Pen^G3^VowRM>{aä} PL-GEN_ZERO "teka-" ;
+Dial:^Pen^G4^VowRM>{aä} Harm-Neutr_PL-ILL_ZERO "tekka-" ;
```

Võro consonant gradation occurs both in the final coda and the penultimate coda. Since the gradation in *tego* 'deed' is directed at the penultimate coda *t*[*eg*]*o*, a double trigger strategy is applied, such that ‹^Pen› provides a mnemonic reference to the pen(ultimate) coda. The penultimate coda affected by consonant gradation may be followed by a word-final vowel *te*[*g*]*o* 'deed' or a vowel and consonant *ve*[*rr*]*ev* 'red'. Otherwise, the consonant-gradation rules can be written with nearly the same contexts. The *tego* stem type in Võro has four grades, i.e., a quaternary quantity system from ‹ ^G1 ›, the weakest, to ‹ ^G4 ›. More common, however, is the binary orthographic quantity system used for illustrating phonetic ternary quantity.

Võro vowel variation covers both normative and substandard vowels. The word-final vowel is removed in the plural stems with ‹ ^VowRM ›, which allows for the vowel harmony trigger ‹ {back} › to select the back vowel from the ‹ {eõ} › and ‹ {aä} › archiphones. Võro morphological complexity lies in consonant gradation and vowel harmony. Whereas the vowel harmony might be directly associated with that found in Finnish with slight divergence, the gradation extends to virtually all consonants as in the Skolt and Northern Saami languages. Strong ‹ *ll* › vs weak ‹ *l* › is modeled using an archiphone ‹ *{lØ}* › preceding the constant ‹ *l* ›, i.e., 'fish' is seen in the form ‹ kala+N:ka{Øl}la ›.

In Lushootseed, a Salish language of the Pacific Northwest, language description indicates (Bates et al.; Crowgey 2019: 42) that there are seven different types of reduplication associated with the first three phonemes of the root, also observed in Lonsdale (2001). Reduplication is preposed directly to the root, and therefore three place holders in the form of ‹ {p1}› 'position 1', ‹ {p2} › 'position 2', ‹{p3}› 'position 3', immediately precede the root, see stems in example (7). The plural in these words is produced by essentially reduplicating the first phonemes, i.e., ‹*čaləs*› 'hand' becomes ‹*čalčaləs*› 'hands' and ‹*sda?*› 'name' ‹*sda?da?*› 'names'. Place holder ‹{p3}› copies the first consonant, ‹{p2}› copies the vowel, and ‹{p1}› copies the second consonant.

(7) Lushootseed

```
čaləs+N:{p3}{p2}{p1}čaləs N_RED2CVC_plus_N_RED1CVtoCVCV "hand" ;
sda?+N:s√{p3}{p2}{p1}da? N_RED2CVC "name" ;

LEXICON N_RED2CVC_plus_N_RED1CVtoCVCV
N_RED2CVC "plural" ;
N_RED1CVtoCVCV "diminutive" ;

LEXICON N_RED2CVC !! Plural
+Pl:^Red2CVC # "čal'čaləs, sda?da?" ;
+Sg: # "čaləs, sda?" ;


LEXICON N_RED1CVtoCVCV !! Diminutive
```

```
+Dimin:^Red1CVtoCVCV # "čačaləs" ;
```

A second reduplication only copies the first two letters of the root. This is used in forming the diminutive of ‹*čaləs*› ‘hand’, which results in ‹*čačaləs*› ‘little hand’. As the odd numbered place holders represent consonants and the even vowels, the two-level rules are used to deal with the values of these extensive archiphones. The consonants are relatively consistent with little deviation from a simple copy, see (8), but the vowels, which are only a few, introduce a lot of variation.

(8)

```
"{p3}:Cx in reduplication"

{p3}:Cx <=> _ {p2}: {p1}: Cx Vow: Cns [ Cns | Vow ]* [^Red2CVC: |
^Red1CVtoCVCV: ] ;

where Cx in (kʷ bʼ ǰ nʼ mʼ p pʼ b t tʼ d dᶻ c cʼ č čʼ s š l lʼ ɬ
ƛʼ y w yʼ wʼ k kʼ kʼʷ gʷ h xʷ q qʼ qʷ qʼʷ xˇ xˇʷ ʔ) ;
```

The place holders used in the description of Lushootseed are yet another way to avoid pulling rabbits out of the hat. They provide a specified position for mapping reduplication to, and they can be introduced to each root semi-automatically. Naturally, only part of the extensive reduplication system in Lushootseed can be regularly implemented, and more work is required to provide an adequate description.

6 Conclusion

Although each language has its own unique features, the same approach to resources and structuring of the lemma-and-stem pair should hold. We recommend that, where possible, minimal changes should be made to the word stems, since this will minimize the workload.

Describe the language for the language user. Remember no one writes perfectly and that the description should cater to identifying so-called incorrect and substandard language. When the grammars say “No”, make those readings available as well, especially if they represent features of regular morphology.

There should be a preference for describing the documented and documentable features of the target language, as random and unpredictable lexica and features introduced from outside of the language remain even more infinite than the nearly infinite but predictable morpholexicon of the language.

The stem should be a form from which all variants can be derived without converting from an unestablished epsilon.

Acknowledgements

Working on descriptions of over twenty languages, Apurinã (apu), Erzya (myv), German (deu), Hill Mari (mrj), Hungarian (hun), Ingrian (izh), Khanty (kca), Latvian (lav), Livonian (liv), Low Saxon (nds), Lushootseed (lut), Komi-Permyak (koi), Komi-Zyrian (kpv), Mansi (mns), Meadow Mari (mrh), Moksha (mdf), Olonets-Karelian (olo), Skolt Saami (sms), Tundra Nenets (yrk), Udmurt (udm), Võro (vro), etc., meant endless hours of consultation, suggestions and chiding from my mentors, collaborators – all friends. To these people I owe my thanks: Alho Alhoniemi, Khalid Alnajjar, Lene Antonsen, Antti Arppe, Timofej Arxangelskij, Renate Blumberga, Jeremy Bradley, Andrej Čemyšev, Jölgin Cypanov, Tino Didriksen, Olga Erina, Valts Ernštreits, Sidney Facundes, Mikel Forcada, Børre Gaup, Ciprian Gerstenberger, Måns Hulden, Arvi Hurskainen, Mika Hämäläinen, Ryan Johnson, Sullõv Jüvä, Heiki-Jaan Kaalep, Neeme Kahusk, Fred Karlsson, Eino Koponen, Kimmo Koskenniemi, István Kozmács, Enye Lav, Miika Lehtinen, Krister Linden, Jorma Luutonen, Sjur Moshagen, Pēteris Paikens, Flammie Pirinen, Timo Rantakaulio, Sirkka Saarinen, Tapani Salminen, Merja Salo, Mansur Sayxunov, Dave Sienko, Janine Siewert, Miikka Silfverberg, Sven-Erik Soosaar, Pirkko Suihkonen, Trond Trosterud, Fransis Tyers, Heli Uibo, Tiit-Rein Viitso, Atro Voutilainen, Jonathan Washington, Anssi Yli-Jyrä.

References

Alâmkin et al. 2000:
Mokšen' kâl': Morfologiâ 2000. Saransk.
Bates et al. = Bates Dawn, Thom Hess, Vi Hilbert (eds.) 1994:
Lushootseed Dictionary. University of Washington Press.
Crowgey, Joshua 2019:
Braiding Language (by Computer): Lushootseed Grammar Engineering.
Facundes, Sidney da Silva 2000:
The Language of the Apurinã People of Brazil (Maipure/Arawak).
Lonsdale, Deryle 2001:
A two-level implementation of Lushootseed morphology. In *Proceedings of the 36th International Conference on Salishan and Neighbouring Languages,* 6:203–214.
Moshnikoff et al. = Satu da Jouni Moshnikoff, Eino Koponen, Miika Lehtinen 2020:
Sää'mǩiõl ǩiõllvue'ppes / Koltansaamen kielioppi. Sää′mte′ğğ – Saamelaiskäräjät
Sammallahti Pekka, Jouni Moshnikoff 1991:
Suomi–Koltansaame sanakirja / Lää'dd–sää'm sää'nnǩe'rjj. Girjegiisá.
Sammallahti Pekka 2012:
johttájedje Taas mentiin. Marko Jouste (ed.) Giellagas Institute.

Online resources

Divvun = Language tools for indigenous and minority languages:
https://divvun.org
ERME = Electronic Resource for Erzya Moksha (corpus)

http://urn.fi/urn:nbn:fi:lb-201407306
Erzya and Moksha online dictionary:
https://valks.oahpa.no/
Fu-lab = Infrastructure in Syktyvkar which has collaborated with GiellaLT
https://fu-lab.ru/laboratoriya
Komi and Udmurt online dictionary:
https://kyv.oahpa.no/
Mari online dictionary:
https://muter.oahpa.no/
Nenets and Mansi online dictionary:
https://vada.oahpa.no/
Northern Balto-Finnic online dictionary:
https://sanat.oahpa.no/
Nuorti = Skolt Saami intelligent computer assisted language learning tool:
https://oahpa.no/nuorti/morfas/der/
PaBiVUS = Parallel Bible Verses for Uralic Studies:
http://urn.fi/urn:nbn:fi:lb-2020021121
Skolt Saami online dictionary:
https://saan.oahpa.no/
Southern Balto-Finnic and Latvian online dictionary:
https://sonad.oahpa.no/
UralicNLP = Python library for using finite-state transducers from GiellaLT and Apertium.
https://github.com/mikahama/uralicNLP
Voro = Southern Estonian Võro intelligent computer assisted language learning tool:
https://oahpa.no/voro/fona/