

# Combining Concepts and Their Translations from Structured Dictionaries of Uralic Minority Languages

Mika Hämäläinen, Liisa Lotta Tarvainen, Jack Rueter

Department of Modern Languages, University of Helsinki  
Fabianinkatu 33, 00014 University of Helsinki  
mika.hamalainen@helsinki.fi, liisa.l.tarvainen@helsinki.fi, jack.rueter@helsinki.fi

## Abstract

Building language resources for endangered languages, especially in the case of dictionaries, requires a substantial amount of manual work. This, however, is a time-consuming undertaking, and it is also why we propose an automated method for expanding the knowledge in the existing dictionaries. In this paper, we present an approach to automatically combine conceptually divided translations from multilingual dictionaries for small Uralic languages. This is done for the noun dictionaries of Skolt Sami, Erzya, Moksha and Komi-Zyrian in such a way that the combined translations are included in the dictionaries of each language and then evaluated by professional linguists fluent in these languages. Inclusion of the method as a part of the new crowdsourced MediaWiki based pipeline for editing the dictionaries is discussed. The method can be used there not only to expand the existing dictionaries but also to provide the editors with translations when they are adding a new lexical entry to the system.

**Keywords:** combining dictionaries, low-resource languages, semantic dictionaries

## 1. Introduction

Words in a language are often considered to be the basic units of a dictionary. Like the language itself, dictionaries also contain elements specific to many different fields. Languages contain words composed of different phonemes, morphemes and syntactic structures. One can use a language by merely employing nothing but words. To understand a language fully, however, one also needs to understand the culture. A word is defined in the Concise Oxford English dictionary first as "[a] single distinct meaningful element of speech or writing, used to form sentences with others". Dictionaries contain words which are referred to as lexemes, lexical items, head words or lemmas in order to specify their specific use in dictionaries and to distinguish them from their use in the spoken language. In general, the head word of a dictionary entry is given as nominative singular for adjectives and nouns, while infinitive (or third person singular) forms are used for verbs cf. (Kastovsky, 1992). Depending on the language, a dictionary provides different inflectional forms essential to the language, and examples to illustrate usage.

The problem many minority languages are facing is that their language resources are very limited. Small Uralic languages, which are the focus group of this paper, are no exception to this rule. Having good language resources in the form of a dictionary would benefit these languages, because they can be used in various tasks such as machine translation (Brown, 1997) and pedagogy (Antonsen; Huhmarniemi & Trosterud, 2009) when teaching the language to non-native speakers. Thus, their impact on the revitalization of endangered languages would be high.

Furthermore, the building of a multi-language dictionary of this variety is essential for supporting minority languages. Due to the pressure of major languages, the amount of speakers of these minority languages has diminished radically. A multi-lingual dictionary that concentrates on

these minority languages is one possibility to maintain the vocabulary of these languages and present them to larger audiences and users.

Building quality resources for low resource languages requires a good amount of manual work. Such work has been on-going for Uralic languages in the Giellatekno infrastructure<sup>1</sup> for the past decade in the form of dictionaries, morphological analyzers, constraint grammars and rule-based machine translation tools among others.

This paper builds upon the manually crafted bilingual dictionaries in the Giellatekno infrastructure and proposes an automatized way of combining semantic and translation knowledge from these dictionaries. This method will expand the existing language resources by providing completely new translations and language pairs. Later on, this work will be incorporated in the Sanat<sup>2</sup> MediaWiki platform (Rueter & Hämäläinen, 2017), which is used to crowdsource the future development of the Giellatekno language resources.

## 2. Related Work

Standardizing bilingual or multilingual dictionaries through combining entries from different corpora and sources, such as WordNet (Miller, 1995), in order to build well-structured dictionaries for machine translation and other uses has become a central source of interest over the past decades. A great many past studies, however have focused on majority languages.

The aim of (Klavans & Tzoukermann, 1990) is to focus on the methods that combine structured but incomplete or only partially complete information from the dictionaries in order to create a bilingual lexical database. The aim of their BICORD system is to create a bilingual corpus based dictionary which is able to combine lexical information both from the bilingual corpora and from the machine-

<sup>1</sup> See <http://giellatekno.uit.no/index.eng.html>

<sup>2</sup> Accessible on <http://sanat.csc.fi/>

readable dictionaries. The goal is to indicate lexical equivalence with associated translations. The basis is in linguistic research combined with the data collected from machine-readable dictionaries. Statistical techniques and machine-readable dictionaries combined with the methods from standard linguistic has been used as the central guideline. The study concentrates on the action verbs in French and in English with the approach to analyze their use and behavior in a bilingual corpus.

Ji's et al. (2016) paper concentrates on matching lexical entries in order to reduce the quantity of the term data and raise the quality of the lexica. The quality is enhanced by matching and combining lexical entries and resources from multiple dictionaries. The quality criteria are based on linguistic and terminological research work such as Levenshtein distance string metric. In Ji's et al. research this distance measure is tested on WordNet and BabelNet in order to change the quality criteria into language-independent frequency-based measures. The idea of the quality criteria is to detect well-constructed entries from the model dictionary, which in this research is the Princeton WordNet. The aim is to find duplications and other errors in the linked lexical entries and resources and see if reducing the term data can be compensated by more exact structure and content.

Hovy's approach (Hovy, 1998) is focused on describing the use and results of semi-automated cross-ontology concept. The goal is to create ontologies which are based in large-scale machine translations. The focus is in combining the ontologies in order to create a practical large-scale ontology for free use in the web. The primary stress lies in combining and standardizing these ontologies. The main use for these ontologies is in machine translation but they can be used in other purposes as well.

Navigli and Ponzetto (2010) are presenting a multilingual semantic network, BabelNet, with the aim to produce a lexical resource with high accuracy. By the automatic mapping between BabelNet's two resources, the English Wikipedia and WordNet, they have provided an automatic construction in order to create a large multilingual lexicon. Their aim is to present a new methodology for the automatic construction of a multilingual lexical knowledge resource of this art. The project is based on combining the lexicographic and encyclopedic knowledge. Navigli and Ponzetto have unified the word senses as concepts and semantic pointers between synsets as relations from the WordNet with all the encyclopedic entries, such as individual pages, as concepts from Wikipedia. The semantically unspecified relations in Wikipedia are collected from hyperlinked texts. These two resources give lexical knowledge of different type, from which one is concentrated on the named entities and the other on concepts. Linking the two knowledge resources and the use of two different resources has displayed that it will provide large-scale lexical resources which work as the basis for BabelNet.

Wehrli et al. (2009) present the MulTra (Multilingual Translation) project. The aim of the MulTra project is to develop an efficient grammar based translation model which is able to cover several different languages. The project is also based on object-oriented software design. The basis for this grammatically oriented approach is

abstract and generic linguistic, which is based for example on Noam Chomsky's generic grammar. The aim of the research is to grow the amount of language pairs and develop them but also to reduce the development costs. First, the project is concentrated to five large European languages (such as German and French) but also later on on languages with different writing systems (such as Russian and Greek). The approach which is based on abstract and generic grammar is seen as worth developing while there is a possibility to benefit from the abstract levels of a language. In order to reduce the complexity of the major part of bilingual lexical databases, they can be automatically derived by transitivity.

The aim of Ji's et al. (2014) paper is to present a workflow of merging and matching the anonymous special language terms from Web terminology sites. The term candidates collected from the Web terminology sites with the help of TermFactory are unified and merged with the resources that are already found in the TermFactory, which is a semantic Web framework designed for professional multilingual terminological use. The goal is to produce more high-quality term candidates for professional terminology use. One of the central themes in the paper is not to lose the provenance of the term candidates when merging them in TermFactory. The long-term goal is to produce clear terminology with high standards in order to serve high-quality translations.

### 3. Combining Dictionaries

In this part, we present our proposed approach to combine multilingual dictionaries for small Uralic languages. The dictionary of each language contains translations for one or more languages for each lemma. This means that a word in Moksha might be translated into Finnish and English while a word meaning the same thing in Erzya might have translations in Russian and Finnish. By combining these translations, we can improve the translation coverage of both dictionaries.

As a result of the combination, we will first of all have, more translations for lemmas in the four languages, and new target languages that haven't existed before in the original dictionaries. This means that Skolt Sami will get translations into a new language French through Erzya, and Erzya will get Norwegian translations through Skolt Sami, for example.

Second of all, the meaning groups, once combined, will have IDs that are shared across all of the four dictionaries, which will turn the strict tree structure of the dictionaries into a graph-like structure which is essentially a step towards having something similar to WordNet for Uralic languages.

#### 3.1 The Initial Data

The language resources in use are the dictionaries of Moksha (Rueter et al., 2018), Skolt Sami (Rueter; Rießler & Lehtinen, 2018), Erzya (Rueter et al., 2018) and Komi-Zyrian (Rueter; Kokkonen & Fedina, 2018) that are freely available in XML format in the Giellatekno infrastructure. All of these dictionaries follow the same XML syntax thus making their combination process easier. We limit the research to the noun dictionaries only. To illustrate the

syntax, an example of a simplified entry in the Skolt Sami dictionary is provided.

```

<e>
  <lg>
    <l pos="N">lää'kk</l>
  </lg>
  <mg>
    <tg xml:lang="eng">
      <t>law</t>
    </tg>
    <tg xml:lang="deu">
      <t>Gesetz</t>
    </tg>
  </mg>
  <mg>
    <tg xml:lang="eng">
      <t>right</t>
    </tg>
    <tg xml:lang="deu">
      <t>Recht</t>
    </tg>
  </mg>
</e>

```

In the example syntax, we see that the polysemic Skolt Sami word *lää'kk* is divided into two different mg elements. Mg is short for meaning group, this means that all the translations under the same meaning group refer to the same semantic concept. Therefore, when combining the translations from different dictionaries, it's important to do it based on the meaning groups. This should solve the problem of polysemy for us, provided that the meaning groups are divided in the correct way.

Combining the meaning groups poses some challenges. First of all, the meaning group division has been hand-crafted by different authors at different times, i.e. they have fuzzy boundaries. Second, they don't have ID numbers, which would link two meaning groups pointing to the same semantic concept in different dictionaries together, thus they have to be combined based on the translations they have. In the combination step, we will associate IDs to the meaning groups to solve this deficiency in the XML dictionaries.

However challenging the problem might be, our combination approach will directly benefit the limited scope of linguistic resources in these small languages. Not to mention, the Giellatekno infrastructure sports with some 20 more XML dictionaries sharing the exact same structure. Most of these dictionaries have translations in Finnish, English, Russian and Norwegian (Bokmål) but also in other Uralic and European languages depending on the authors. This means that this approach could be used in a wider set of languages in the infrastructure with little to no modifications.

In addition to meaning groups and translations, the XML dictionaries contain a variety of different kinds of information. For example, the entries can have continuation lexica for morphological analyzers, longer textual definitions, sound samples, example sentences, notes on derivation, etymology and so on. These additional pieces of information are not used in the combination approach.

### 3.2 The Combination Algorithm

The combination algorithm takes the XML dictionaries as input and produces new versions of them with more

translations added inside of the correct meaning groups. This requires the structure of the data to be changed. In the Giellatekno XMLs, meaning groups are stored under of lemmas. This means that the same meaning can be represented by multiple meaning groups inside of a dictionary in the case of synonyms. When two lemmas share a meaning, they will still have two separate meaning groups because of the structure of the XMLs.

The first step is to extract all the meaning groups from the XML dictionaries for the four languages under study and to assign them with unique IDs. This means that instead of storing the contents of a meaning group under a lemma, we simply create a pointer to the meaning group based on its ID from the lemma entry. Then, when we do the combinations and we update the ID references to the new meaning groups resulting from the combination of two or more existing meaning groups. That is, if meaning groups with IDs 1, 2 and 3 are combined, we modify the pointers from the lemma entries pointing to 2 and 3 for each language to point to the meaning group 1 which now contains the information from 1, 2 and 3.

The combination itself is done by following the formula for combining two concepts based on the similarity of their descriptions introduced by (Hovy, 1998). In their approach, the formula was used for two descriptions in English. In our approach, we treat the translations of a translation group, i.e., translations in one language, as a description. This means that, for instance, the Skolt Sami word *podd* would have one description containing the English translations *moment* and *while* and another description containing the Finnish translations *hetki* and *tovi* and so on for other languages.

$$similarity = \frac{|D1 \cap D2|}{\min(|D1|, |D2|)} * |D1 \cap D2|$$

D1 and D2 are sets of words used in the description being compared to each other. In our case, they are the translations in the same language of two different meaning groups, for example [*current, river*] and [*current, river, torrent*]. The formula gives us a similarity score between 0 and 1 that we can directly use to determine whether two meaning groups should be combined or not.

We calculate the similarity score for all the matching languages in the meaning groups. For example, if a meaning group has translations in Finnish and English, and it's compared to another one with translations in Finnish and Russian, we can only calculate the similarity based on Finnish. Furthermore, we take the minimum of the similarity scores. The hypothesis is that if one language differs in its translations, that particular language makes a semantic distinction the others don't and thus the meaning groups shouldn't be combined. Meaning groups with translations in only one language will not be considered at all in the combination. If the minimum similarity score is one, we combine the two meaning groups.

## 4. Results and Evaluation

In this part, we will discuss the results of the combination algorithm and conduct an evaluation for the combined Skolt Sami dictionary.

### 4.1 Results

The initial data set includes a total of 13095 individual meaning groups that fill the requirement of having translations to at least two different languages. After using the algorithm, the number of unique meaning groups dropped to 9914 which means that 3181 meaning groups were merged into other meaning groups. In other words, the amount of unique meaning groups was reduced by approximately 24 %.

On the average, this method added 2.6 new translations to each meaning group along with 1.5 new languages. The low number of new languages can be explained by the fact that the data is highly dominated by translations into Finnish and Russian while the other languages appear fewer times in the data.

In an effort to avoid secondary source language word reference, the four source languages, Erzya, Komi-Zyrian, Moksha and Skolt Sami, were not added to the whole number of translations. Hence the languages originally present in the Komi-Zyrian meaning groups, for example, were Russian, Finnish and English, whereas the larger number of translation sets were augmented to include the majority Norwegian, German and French languages, as well as several Sami languages (but not Erzya, Komi-Zyrian, Moksha or Skolt Sami).

### 4.2 Evaluation

Due to the fact that finding people with high enough skills in the minority languages and all the majority languages the dictionaries have translations for, we decided to limit the evaluation to the combined Skolt Sami dictionary. The combined dictionary was read through and commented by an expert linguist with high proficiency in Skolt Sami among other languages.

The annotator was asked to comment them based on two factors: are the new translations suitable for the word they have been added to and do the new translations semantically fit into the meaning group.

After the annotation was done by the annotator, we looked into the errors he had found and categorized them based on the source of the error. All in all, 1325 new translations were added to the Skolt Sami noun dictionary out of which 843 were correct translations. This means that around 63% of the new translations were correct, but as we will learn, the low accuracy wasn't entirely due to the approach.

Further analyzing the annotations, we were able to identify five different sources of error in the combined meaning groups for Skolt Sami.

#### 4.2.1 Meaning Group Division

Looking at the source of the erroneous translations, we can see that a vast majority of them, around 84% is due to a non-accurate meaning group division in one of the source language dictionaries. A 54 % of them come from the Erzya dictionary, 27 % from Komi-Zyrian, 18 % from Skolt Sami and only one percent from Moksha.

What this means, is that essentially translations that should have been in different meaning groups, were put inside of the same meaning group by the people editing the dictionaries. For example, the Erzya word *auuo* had Finnish translations *rupla* (ruble), *votka* (vodka) and *valkuainen* (egg white) inside of the same meaning group, which is against the standards set for the XML structure. As a result, the Skolt Sami word *rubbâl* which only means ruble also got the other two meanings of the Erzya word.

Inaccurate meaning group division even in the Skolt Sami dictionary itself caused wrong translations to appear with words that are synonymous only in one sense. For instance, the Skolt Sami word *puðlvvõk* had two semantically different translations in Finnish inside of the same meaning group: *sukupolvi* (generation) and *ikäryhmä* (age group), which resulted in the erroneous translation *ikäryhmä* to appear under another Skolt Sami word *puðlvv* which only means generation, not age group.

The fact that a majority of the errors came from non-properly formatted data, gives us hope that this approach would yield better results if the meaning group division was done in a more accurate fashion. This evaluation reveals that there's much work to be done in the original dictionaries as well.

#### 4.2.2 Wrong Translations

Another error source that wasn't due to the performance of the combination algorithm, was incorrect translations in the source dictionaries. This covers around 12 % of all the errors. A 33 % of the incorrect translations were found in the Erzya dictionary, a 16 % in the Komi-Zyrian dictionary 51 % in Skolt Sami and none in the Moksha dictionary.

For example, the Skolt Sami word *kruugg* had initially a wrong translation *tyre* which caused three other words with the same meaning to be added as translations: an alternative English spelling *tire*, the Norwegian *dekk* and Russian *звено*. This illustrates how a simple wrong translation can cause more erroneous translations to be included for a word.

This group of errors as well as the first one reveal more about the problems already present in the data than about the accuracy of the combination algorithm. This shows that more work is needed to do a lexicographical evaluation for all of these dictionaries.

#### 4.2.3 Homonymy

Homonymy in one of the majority language translations accounts for a 1 % of the errors. This might occur if the difference in meaning is not captured by other translations in the meaning group.

More specifically, the Finnish homonym *kuusi* which can either mean the number *six* or a *spruce* caused translations for the number *six* in other languages to occur under the Skolt Sami word *lösttkuõss* which refers to the tree. This was due to the lack of translations in different languages for the word *lösttkuõss* in the original dictionary. The word was only translated into Finnish and Latin while the word *kuudäs* referring to the number *six* had translations to Norwegian, Finnish, English, German and Russian.

The error caused by homonymy in one language could be avoided by either making sure that homonymous words have at least one synonym in the translations. This would make sure that the combination algorithm would consider the meaning groups different. Another way would be to modify the algorithm so that instead of just requiring translations in at least two languages, it would require at least two shared languages between the meaning groups it compares. This would effectively solve the issue of homonyms, but it might also result in a lower number of combinations of meaning groups.

#### 4.2.4 Polysemy

Another minor source of error, around 2 %, comes from polysemy. Polysemy can manifest much like homonymy in some cases. For example, the Skolt Sami word *nju'nnjel* refers to a *smew*, which in Finnish is *uivelo*. The Finnish word can also mean a *weak person* which caused such Russian translations *щуплый*, *хилый* and *субтильный* to be added to the Skolt Sami word. The remedy for this kind of an error from polysemy is exactly the same as for homonymy.

An interesting error caused by polysemy is when there are enough different languages to be used in comparing the meaning groups, but only one of them makes a distinction in meaning. The Skolt Sami word *njuhččâm* got translations to French *langue*, Finnish *kieli* and Russian *язык*. These are all right translations because they mean *tongue*, the problem, however, is that they also mean *language* which is why the English word *language* got also added as a translation. This is an interesting example, how many distant languages might share the same polysemy, while others don't.

#### 4.2.5 Peculiar Translations

The rest of the errors, around 1 %, came from unusual translations. An example of such a translation is a cultural translation of the Komi-Zyrian word *саралан кар*, which was translated as *capital* and *seat of the Czar*. While in the Russian context, *seat of the Czar* might very well refer to the exact same city as the word *capital*, in Skolt Sami context, that is not the case as it is also spoken in Finland and Norway. This is the only case of a cultural translation being added to Skolt Sami from another language, but if this method is applied in the context of socio-culturally very distant languages, these kinds of errors might be more frequent.

Another case of an unusual translation is the use of a non-nominative case in the translations. This was, for instance, in the case of the Erzya word *валске* which was translated into Finnish in nominative *aamu* (morning) and in adessive *aamulla* (in the morning). This resulted in the Skolt Sami word *ee'däädöž* to receive the translation *aamulla*, however the Skolt Sami word alone in nominative doesn't bare the meaning of *in the morning*. These non-nominative translations are a peculiar case, as putting them into a separate meaning group doesn't really make too much sense. This would call for a different kind of a translation field in the XML schema to separate translations from a translations that serve more as a usage example.

## 5. Discussion and Future Work

We have limited this paper to work with nouns only. In the future, we intend to apply this approach to all parts-of-speech, and eventually, to all the languages in the Giellatekno infrastructure. From the evaluation, we can say that the division into meaning groups in the Giellatekno dictionaries is not as accurate as it should be. This calls for more work in that field, as well.

Erroneous translations in the source data were found and that clearly points towards more manual work that should be done in order to make sure that the dictionaries are a reliable source for different language users and learners. The source data was also infested by missing meaning groups, meaning that the translations referring to entirely different concepts were, against the structure imposed by the Giellatekno XML standard, added inside of a single meaning group.

Whereas the wrong translations must mostly be fixed manually due to a lack of a better resource to evaluate the correctness of the translations automatically, the meaning group division, however, is something we are interested in looking at from an automated perspective in the future. Given that the translations are mainly in majority languages with a plethora of LRs available, we could probably have a look at utilizing distributional semantics approaches, such as *word2vec*, to automatically divide meaning groups that contain translations referring to entirely different concepts.

In this paper, we only discussed combining meaning groups and the translations inside of them. When we solve the meaning group division problem in the source data, we can actually link words of different minority languages together through the meaning groups they refer to, i.e., if we know that the Skolt Sami word *piännai* and the Erzya word *киска* both refer to the same meaning group for *dog*, we can include the Skolt Sami word in the Erzya dictionary as a translation and vice versa.

One of the main motivations for this research was the MediaWiki based online dictionary (Rueter & Hämäläinen, 2017) we are building on the Helsinki-based Sanat server<sup>3</sup>. This dictionary makes it possible to edit Giellatekno's language resources easily in a wiki environment. In the future, we want to incorporate the approach proposed in this paper as a part of the MediaWiki workflow when editing and adding lemmas to the online dictionary.

## 6. Conclusions

In this paper, we have proposed an approach to combine dictionaries for minority languages in an automatic fashion. The majority of the errors in the results produced by the approach (around 96 %) were due to errors either in translations or meaning group division in the original dictionaries. This shows that providing that the original data has only a minimal number of errors, this method can produce reliable combinations of different dictionaries automatically.

This method helps us add beneficial information to all the minority-language LRs by deriving from manual work done separately for each of them. Because of the

<sup>3</sup> <http://sanat.csc.fi>

notoriously limited amount of resources for these minority, or even endangered, languages, this automatic method provides a notable improvement to the existing LRs.

The improved dictionaries will tangibly enhance the NLP technology resources already in place in the Giellatekno infrastructure, one such area to benefit might be machine translation.

## 7. Bibliography

- Antonsen, L., Huhmarniemi, S., & Trosterud, T. (2009). Interactive pedagogical programs based on constraint grammar. *Proceedings of the 17th Nordic Conference of Computational Linguistics. Nealt Proceedings Series 4*.
- Brown, R. D. (1997). Automated dictionary extraction for “knowledge-free” example-based translation. *7th International Conference on Theoretical and Methodological Issues in Machine Translation*, (pp. 111–118).
- Hovy, E. (1998). Combining and Standardizing Large-Scale, Practical Ontologies for Machine Translation and Other Uses. *In The First International Conference on Language Resources and Evaluation (LREC)*, (pp. 535–542).
- Ji, K., Wang, S., & Carlson, L. (2014). Matching and Merging Anonymous Terms from Web Sources. *International journal of web and semantic technology*, (pp. 53-69).
- Ji, K., Wang, S., & Carlson, L. (2016). Multilingual Dictionary Linking and Aggregation: Quality from Consistency. *Proceedings of the Fourth International Workshop on Linked Data for Information Extraction (LD4IE)*, (pp. 51-62).
- Kastovsky, D. (1992). Semantics and vocabulary. In R. M. Hogg (Ed.), *The Cambridge history of the English language I* (pp. 290-408). Cambridge University Press.
- Klavans, J., & Tzoukermann, E. (1990). The BICORD System Combining Lexical Information from Bilingual Corpora and Machine Readable Dictionaries. *COLING 1990 Volume 3: Papers presented to the 13th International Conference on Computational Linguistics*.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), pp. 39-41.
- Navigli, R., & Ponzetto, S. (2010). BabelNet: Building a Very Large multilingual Semantic Network. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (pp. 216-225).
- Rueter, J., & Hämäläinen, M. (2017). Synchronized Mediawiki Based Analyzer Dictionary Development. *The Third International Workshop on Computational Linguistics for Uralic Languages*, (pp. 56–62).
- Wehrli, E., Nerima, L., & Scherrer, Y. (2009). Deep Linguistic Multilingual Translation and Bilingual Dictionaries. *Proceedings of the Fourth Workshop on Statistical Machine Translation*, (pp. 90-94).

## 8. Language Resource References

- Rueter, Jack, Bolkin, Oleksandr, Abramov, Kuzma, & Erina, Olga. (2018). rueter/gt-xml-lexica-myv: Erzya To X nouns (Version v1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1146850>
- Rueter, Jack, Herrala, Eeva, Feoktistov, Aleksandr, & Salo, Merja. (2018, January 13). rueter/gt-xml-lexica-mdf v1.0 (Version v1.0). Zenodo. <http://doi.org/10.5281/zenodo.1146846>
- Rueter, Jack, Kokkonen, Paula, & Fedina, Marina. (2018). rueter/gt-xml-lexica-kpv: Komi-Zyrian to X nouns (Version v1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1146852>
- Rueter, Jack, Riebler, Michael, & Lehtinen, Miika. (2018). rueter/gt-xml-lexica-sms: Skolt Sami to X nouns (Version v1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.1146854>