Creating a corpus for Kven, a minority language in Norway

Pia Lane, Kristin Hagen, Anders Nøklestad and Joel Priestley
University of Oslo

Abstract

Language documentation, including the development and use of corpora, is frequently linked to revitalisation. This is also the case for the Kven language, a Finnic minoritised language, traditionally spoken in the two northernmost counties of Norway. Kven is a recognised minority language in Norway, protected by the European Charter for Regional or Minority Languages. This status led to increased efforts to document Kven, including the development of the Ruija Corpus, consisting of recordings of interviews in Kven. The corpus was an important tool for the standardisation of Kven. In this article we describe how the corpus was developed and account for search functions, including a discussion of the limitations of the corpus. We also discuss the role of corpora and other online tools for language revitalisation, with a particular focus on the standardisation of Kven and conclude by reflecting on how expertise also resides with the speakers of an endangered language and that they have a right to be involved in efforts of language documentation and revitalisation.

Keywords: Corpus linguistics, revitalisation, minority language, Kven

1. Introduction

Across the world, a large number of languages are in the process of being standardised, following a longstanding tradition within anthropological linguistics, linguistic typology and language documentation. Scholars describe and document languages, and we have sophisticated means for data analysis, for developing grammars and dictionaries, and recently also for building large electronic corpora. One such corpus is the Ruija Corpus, a speech corpus from Kven and Finnish-speaking areas in Northern Norway. In this article, we first provide some information on the Kven people and their language and outline the development of the Ruija Corpus and describe search functions, including a discussion of the limitations of the corpus, due to the lack of grammatical annotation. We then discuss the role of corpora and other online tools for language revitalisation, with a particular focus on the standardisation of Kven and conclude by reflecting on how expertise also resides with the speakers of minoritised language and that they have a right to be involved in efforts of language documentation and revitalisation.

2. Background – Kven language

The Kven language is a Finnic minoritised language, traditionally spoken in Troms and Finnmark, the northernmost counties of Norway, see the map in figure 1. The Arctic region has been multilingual for centuries, and from the beginning of the 18th century, people from Finnish-speaking areas, in what today are the northern parts of Sweden and Finland, settled along the coast of Northern Norway; some of them settled before the current national frontiers were drawn (Sundelin 1998). This group of people and their descendants are called Kven or Norwegian Finns, and particularly in the coastal areas there is a long tradition of trade and intermarriage with the Sámi population. When the idea of Norway as a nation state got foothold in the 19th century, a monolingual and homogenous nation came to be one of the cornerstones of the idea of the Norwegian nation state, and assimilatory and even oppressive policies were directed towards the Kven and Sámi populations. These policies are referred to as the Norwegianisation process, and many of these were directed towards language, and regulations limited the use of Kven and Sámi in schools and sale of land in the northern areas to people with knowledge of Norwegian (Pietikäinen et al. 2010). The Kven were seen as a 'national problem' because of their position as a border minority, "Russia's

 Nordlyd 46.1: 159–170, *Morfologi, målstrev og maskinar*: *Trond Trosterud {fyller | täyttää | deavdá | turns} 60!*, edited by Lene Antonsen, Sjur Nørstebø Moshagen and Øystein A. Vangsnes. Published at UiT The Arctic University of Norway.
http://septentrio.uit.no/index.php/nordlyd https://doi.org/10.7557/12.6345

foothold in Western Europe" (Niemi 1995: 196); hence, the assimilation policies were particularly prominent in the exposed border areas (Niemi 2003). In tandem with general processes of modernisation, these policies contributed to an extensive language shift from Kven to Norwegian, and by the 1960s most Kven children spoke Norwegian only. Norway's ratification of the European Charter for Regional or Minority Languages (under the auspices of the Council of Europe) in 1993 subsequently led to the recognition of Kven as a language in 2005, and not primarily a dialect of Finnish (Lane 2011).

In Finland, Finnish was given official status in 1863 and was developed into a language of education and administration. Vocabulary and grammatical structures from the Eastern Finnish dialects were included, and neologisms were created in order to replace Swedish loanwords (Latomaa and Nuolijärvi 2005), processes which influenced the Finnish dialects in Finland, whereas the Kven had left before this standardisation took place. Because the Kven were not in Finland during the standardisation process, their language developed differently, particularly in terms of vocabulary. Kven and Finnish are agglutinating languages with rich inflectional morphology, but the largest difference between Kven and standard Finnish is lexical due to old Swedish borrowings that have been retained in Kven and newer borrowings from Norwegian. Kven and Finnish are mutually intelligible, though Finnish speakers understand Kven better than Kven speakers understand Finnish because Finns learn Swedish in school and therefore understand older Swedish and Norwegian borrowings in Kven, whereas Kven speakers in most cases do not understand the Finnish equivalents of these borrowings (Lane 2016). Self-identification is also one of the criteria for the language – dialect distinction: the Kven speakers identify themselves as Norwegian and not Finnish (Hyltenstam and Milani 2003).

Figure 1. Map Troms and Finnmark County. Source: Kartverket – Norwegian Mapping Authority (Attribution 4.0 International (CC BY 4.0) © Kartverket https://www.kartverket.no/en).

3. How the corpus came about

The fourth International Polar Year (IPY 2007–2008) was the largest global research initiative to be carried out for 50 years, and approximately 50 000 researchers and language technologists from 60 countries participated. The Research Council of Norway issued a Call for Proposals, which resulted in 29 funded projects. One of these was the Linguistic and Cultural Heritage Electronic Network (LICHEN) at the Department of Linguistics and Scandinavian studies (University of Oslo), directed by Pia Lane. The project was a part of a larger international cluster, whose aim was to create an electronic framework for the collection, management, online display, and exploitation of existing corpora of the languages of the northern circumpolar region. The Norwegian part of the LICHEN project involved carrying out a pilot project on the Kven language. The aim of the Norwegian LICHEN project was to digitise and transcribe old recordings of Kven, collect new data in order to test the tools being developed by the University of Oulu, and to make the recordings and transcriptions available for researchers. The main electronic system was not completed by the time the Norwegian project was concluded, so a separate speech corpus, the Ruija Corpus, was developed by Pia Lane in cooperation with the Text Laboratory at the Department of Linguistics and Scandinavian Studies, University of Oslo. Ruija (both in Kven and Finnish) refers to

Northern Norway, and the name was chosen for the corpus not to alienate speakers who prefer to refer to their language as Finnish (often modified as 'our Finnish' to distinguish it from standard Finnish).

As is often the case for minority language projects, there was limited funding and thus lack of resources to make an annotated corpus with linguistic information (such as lemmas and morphosyntactic tags). The PI of the Norwegian LICHEN project, Pia Lane, did not have a permanent position, which further limited the time available for corpus development. The corpus consists of recordings and transcriptions. When the corpus was initiated, there was no written standard for Kven, so there was no standard that could be used as a basis for the transcriptions. The recordings were transcribed by students at the University of Oulu and Mikael Voronov at the Kven Institute. The transcribers followed the same guidelines, but they were based at different institutions and worked with different supervisors who implemented the guidelines slightly differently. Consequently, there is some variation in the transcription conventions.

The corpus was launched in April 2010 with 76 hours of recordings, and data from two other projects added. These projects are *Identities in transition – a longitudinal study of language shift* and *Standardising minority languages – STANDARDS*, both financed by the Research Council of Norway and directed by Pia Lane. *Identities in transition,* a study of language shift in Bugøynes-Pykejä (a Kven community in Northern Norway) compared and contrasted interview data from the same individuals from 1975 and 2008, supplemented by interviews in Norwegian with younger speakers from the generation who had shifted to Norwegian. The STANDARDS project explores how intended users of a written standard of Kven relate to the standard and consists of interviews and video recordings of Kven speakers reading a text in Kven for the first time. The project source codes in the corpus reflect these projects: LICHEN = LI and KI, Identities in transition = id and sks,[1] SMS = STANDARDS. From 2023 the corpus will also contain interviews from the project *Voices of revitalisation,* which focuses on how new speakers experience the process of starting to speak Kven or Sámi. New speakers are individuals who have learned an indigenous or minoritised language in an educational setting, often as a part of revitalisation efforts, and who reclaim (start speaking) the language later, often at important transitional moments in life, such as when needing the language for work purposes or becoming a parent (O'Rourke and Pujolar 2015). The corpus therefore has data suited for analyses of grammar or phonology, including change over time as the main bulk of the corpus spans a period of more than 30 years. This timespan also allows for sociolinguistic studies exploring changes in language attitudes and identity construction over time.

4. Developing the corpus

The first version of the Ruija Corpus was presented in an older version of the search and postprocessing tool Glossa, developed at the Text Laboratory (Johannessen et al. 2008). The main idea behind Glossa was – and still is – to give researchers a user-friendly tool where they can concentrate on their research and do not need to learn advanced query languages or attend courses to use the tool.

This first version of Ruija contained nearly 430 000 tokens, 379 000 Kven and 50 300 Norwegian ones. There were 85 speakers in the corpus from 12 places. Some of the speakers were interviewed twice (in 1975 and 2009, as a part of the project *Identities in transition*).

The present version of the Ruija Corpus is converted to a new and even more user-friendly version of Glossa (Nøklestad et al. 2017, Søfteland et al. 2020).[2] The corpus is enriched with more speech data and in 2022 it contains almost 522 000 tokens from 12 places, with 109 speakers in total.

The main search page is very simple as figure 2 shows. You can search for one or more words in the Google-like search box in the middle of the page. Metadata categories are located on the left-hand side. The results are given as concordances. Figure 3 shows a search for "kveeni"'. 'Kveeni' is a Kven word that refers to both the Kven people and their language and may be used both as a noun and as an adjective. Above the metadata menu you can see how many speakers and words are included in the chosen selection. A click on "Show speakers" gives you a list of all speakers in the selection.

[1] sks refers to Suomalaisen Kirjallisuuden Seura (The Finnish Litterature Society) which provided the recordings from 1975.

[2] Glossa uses the IMS Open Corpus Workbench system for text search and MySQL as the metadata database. The server code is written in Clojure, a modern dialect of Lisp that runs on the Java Virtual Machine.

To the left of each search result there are two or three icons: one for playing the search result in a media player (figure 4) and one for showing the search result as a waveform (figure 5). Some recordings also have a button for video presentation. Clicking on the speaker's identifier shows all the metadata available for the speaker.

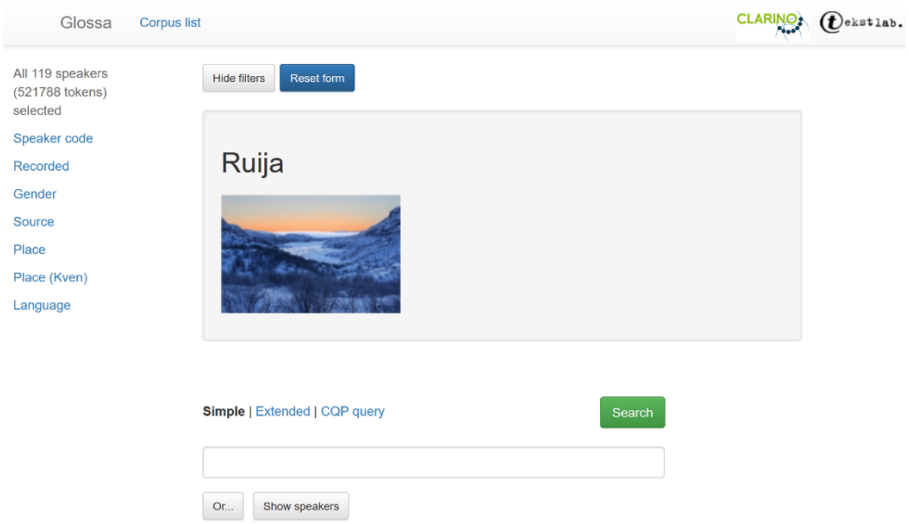

Figure 2. The main search page of the Ruija Corpus.

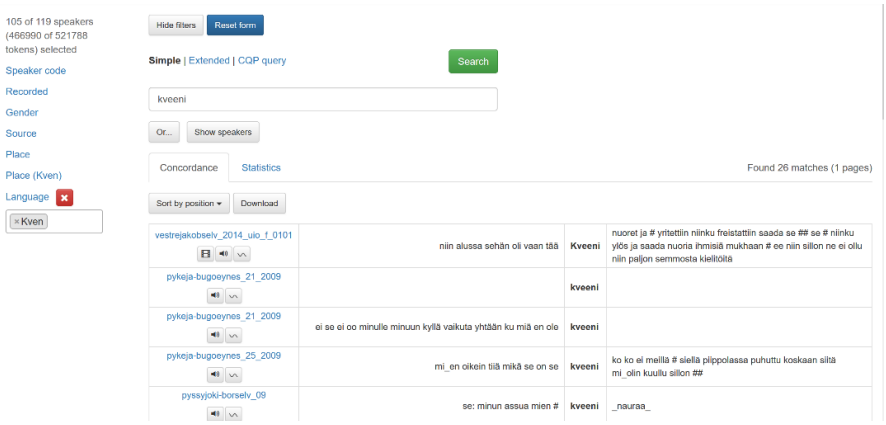

Figure 3. Example of simple search result: kveeni – 'Kven'

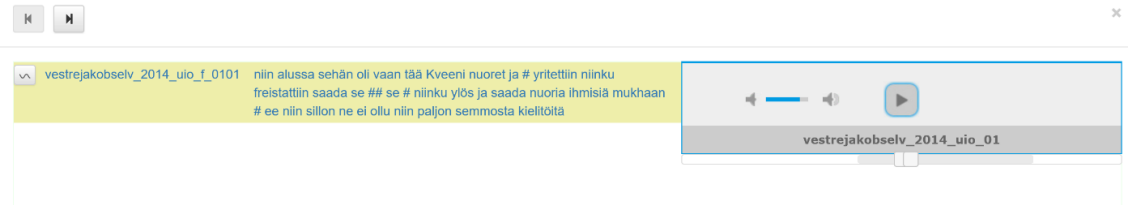

Figure 4. The search result in a media player. If you move the squares under the box left and/or right you get more context.

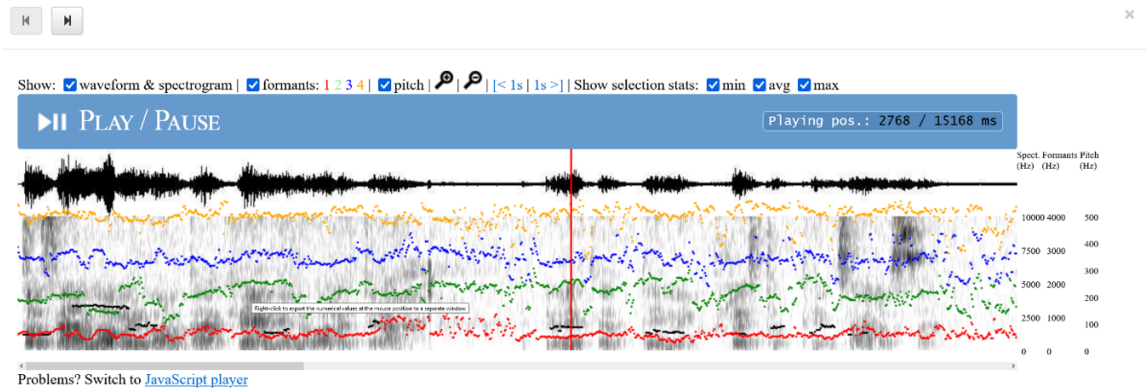

Figure 5. Example of search result as waveform.

Above the search box at the main search page there are two more search options: extended search and CQP search. If you know the query language of the Corpus Query Processor (CQP) search engine you can make your own advanced searches in the CQP box. If you want advanced searches by using menus and text fields, choose the *Extended* option, see figure 6. Here you can choose to search for the start of a word, the middle or the end. You can also search for the first or final word in a speech segment. Since the corpus is not lemmatised or part-of-speech tagged yet, the menu symbol on the left can only offer a box for excluding word forms.

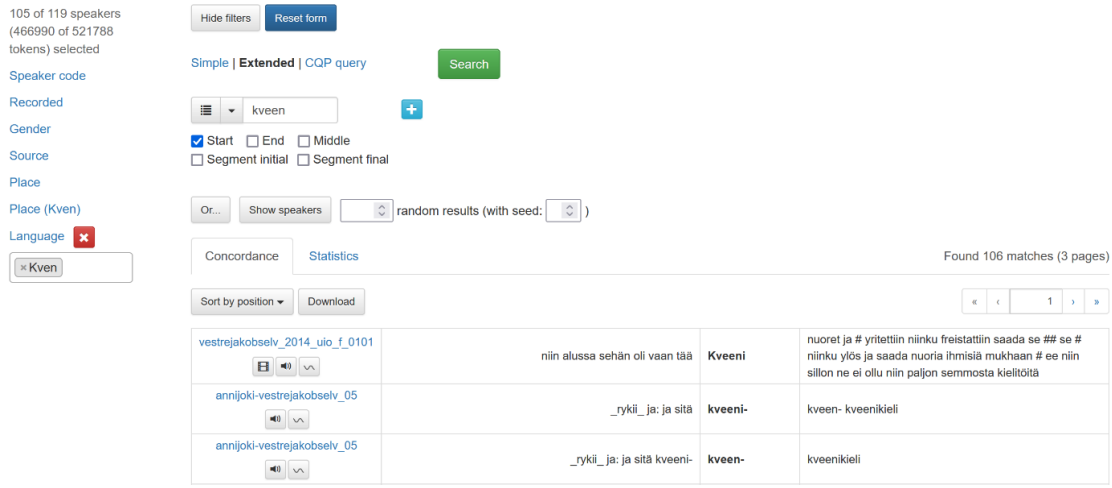

Figure 6. Example of extended search.

Figure 6 shows how to search for different word forms or inflected forms even if the corpus is not part-of-speech tagged. A search for "kveen" as the start of the word will show all words starting with "kveen" as figure 7 shows[3]. Here we have chosen to show the search result as a word list by clicking on "Statistics" below the search result. The search results are downloadable in various formats.

[3] The transcribers followed orthographic conventions, and the stem vowel of Kven was pronouced both as e or æ, transcribed as kveeni or kvääni, respectively. A search for word initial kv- would have yielded both forms.

Update stats Download: Excel Tab-separated Comma-separated

Count	Word form
27	kveenin
26	kveeni
7	kveenit
5	kveenien
4	kveeneiksi
3	kveeninkieli
3	kveenejä
3	kveen-
2	kveeniä
2	kveeniliitto
2	kveeni-
2	kveeni'asia
1	kveensk
1	kveenlan
1	kveenistä
1	kveenir_rahatki
1	kveenipuvun

Figure 7. The search result as a word list.

The Ruija Corpus is available for research, but due to the personal nature of the content in the corpus, the license is a restrictive, academic one. You have to send an application to the owner of the corpus, and after you have been granted access, you can login with Feide (the centralized identity management solution for the educational sector of Norway) or CLARIN (Common Language Resources and Technology Infrastructure).

5. Standardisation of Kven

As a consequence of the recognition of Kven, a standardisation process was initiated. The Kven Language Body, consisting of the Kven Language Council (advisory function) and the Kven Language Board (executive function) developed principles for the standardisation of Kven. They recommended that the standard should be close to Meänkieli (a closely related Finnic language in Sweden), that preference be given to forms common in several Kven dialects, that one should not aim to make the standard as removed from Finnish as possible, and that the standard should be based on standard Finnish orthography. There are some differences in grammatical structures between the Western and Eastern Kven varieties, as the settlement in the Eastern area occurred later and these areas in general had somewhat more contact with Finland (Lane 2017; Keränen 2018). An example is the interdental fricative /ð/, a phoneme that has been retained by some Kven speakers in Børselv-Pyssyjoki and is used by the writers from this village. For some it is a strong identity marker, and the Language Body therefore decided that in language regions where there is a need for additional letters to the Finnish orthography, such as *š* (alternatively sh) and *đ* / *ð*, these may be used.[4] In the Kven grammar the letter <đ> is used consistently throughout the grammar to represent /ð/ even though apart from a few Kven speakers in the Western, /ð/ has not been retained in Kven. For those who use /ð/, there is alternation between /t/ and /ð/, whereas the majority of the Kven dialects have alternation between /t/ and /Ø/,[5] see Lane (2016) for an analysis.

A grammar of Kven was published in Kven in 2014 and translated to Norwegian in 2017 (Söderholm 2014, 2017). The grammar was approved by the Kven Language Board, the decision-making body for the school norm of Kven. According to Evans and Dench (2006), the aim of a descriptive grammar is to capture and codify the essential structural features of a language, ideally collected as a part of a programme on language documentation, often based on a natural speech data, but sometimes supplemented by speaker acceptability judgements (p. 3). Language documentation has a broader aim than providing a grammatical

[4] In Finland, *š* is used when writing some foreign names.
[5] Ø is used to denote that there is an alternation between a consonant and zero (such as *pöytä – pöyän* 'table-NOMINATIVE - table-ACCUSATIVE).

description of a language; ideally such efforts should also document how the languages is *used*; thus, there is also an emphasis on cultural and social aspects of language (Austin 2020).

The Ruija Corpus was one of the sources for the writing of the grammar book (Söderholm 2014, 2017) in addition to other recordings and transcriptions of Kven dialect and written sources. Literature in Kven is still limited, so the author of the grammar drew primarily on novels published by Alf Nilsen-Børskog, from Børselv-Pyssyjoki in the Western dialect area, therefore the grammar is written in the Western variety of Kven while providing information about the Eastern varieties (Lane 2017). The Ruija Corpus was used both by the Kven Language Council and later by Söderholm as a base for the grammatical description of Kven as it contained recordings and transcriptions of interviews from all the core Kven areas, this allowing the council and Söderholm (who also was a member of the Kven Language Council) to map the key dialectal differences (see also Östman 2000 for a discussion on research ethics in minoritised and Indigenous contexts). Dealing with variation was one of the main challenges the Kven Language Body faced, because standardisation always entails reducing and abstracting away from diversity, as pointed out already by Milroy and Milroy (1999). Their aim was to develop a standard that could be used as a basis for developing teaching materials for learners of Kven in the educational system and for producing texts for people who spoke Kven. When considering the needs of the learners, there was a concern that a standard with too many options could impact learning negatively, whereas there also was a need to allow for enough variation for speakers of Kven to recognise 'their language' and identify with the standard. The standardisation of minority languages is an ambivalent process because it requires selecting particular forms over others — they generate and legitimise high varieties in minority languages as well as the structures to sustain their diffusion, potentially establishing linguistic standards that the language speakers themselves experience that they cannot meet (Costa, De Korne and Lane 2017). Consequently, minority language speakers are potentially faced with a double stigma (Gal, 2006; Lane 2011): their language falls short when measured against the official national language and in terms of meeting the standardised version of the minority language. The written standard may therefore be perceived by social actors as lacking both the authority and invisibility of a national language and the authenticity and legitimacy of the minority language (Woolard 2008; Lane 2015).

Currently, the most widely used version of the standard is the form closest to the one used by Söderholm (2014), though those who write Kven adapt their texts to the local context, a task for which the Ruija Corpus provides useful information. Even though the corpus is not lemmatised or part-of-speech tagged, searches for high-frequent words may still yield a substantial amount of information. One example is variation in the realisation of infinitive forms. Some Western Kven dialects have retained the proto-Finnic word final /t/ (alternation with or without /t/) whereas this is not the case for the Eastern dialect. In the absence of grammatical information, an option is to search for word forms. As almost all those who were interviewed were asked about languages they use, searches for *puhua* 'speak' would yield a lot of examples. As there is a considerable amount of phonological variation and morphophonological alternation, the best suited option is to search for the beginning of the word and then search for infinitive forms manually. Figure 8 is an example of the infinitive form of *puhua* 'speak' with word-final /t/: se oli semmonen häpy et ei pitänyt puhhut kvääanin kieltä # - it was such a shame that one should not speak Kven language # mm'[6]

Figure 8. Infinitive with word-final /t/.

The following figure shows the infinitive form of *puhua* without word-final /t/: mutta sehän oli on # eri aika nyt ## mutta sillon ei saanu puhua # suomea 'but then of course was # a different time now ## but then one wasn't allowed to speak # Finnish'.

[6] # = pause

Figure 9. Infinitive without word-final /t/.

This is of course a search method which is somewhat cumbersome and time consuming and also requires that the user has a fairly good knowledge of Kven dialects and grammar or at least a basic understanding of Finnish grammar, but this still allows us to get a basic overview of grammatical variation in Kven dialects. For researchers who wish to conduct sociolinguistic or historical studies, there are two main options for searches: one may either search for key words such as language, Kven, Finnish, Norwegian, school, war etc. in the actual corpus or search the full transcripts of each interview (there is a link to transcriptions on the Ruija Corpus main page). The latter would also allow for discourse analysis or narrative analysis.

6. Role of corpora in the revitalisation of minoritised languages

Language revitalisation may be seen as a part of Language Policy and Planning (LPP), a discipline which initially developed as a part of sociolinguistics and language-in-society studies and emerged as a field of study in the 1960s (Kaplan et al. 2000). Initially, the main efforts to plan the use and status of languages focused on national languages and nation building (Wright 2004), and language was seen as a static and delimited entity, an object which could be captured and codified. The overarching term in this period was Language Planning, with a focus on linguistic aspects and the use and status of language (Kloss 1967, Haugen 1972).[7] The structuralist period after WWII laid the foundations of what was to characterise LPP until the critical turn in the social sciences and humanities in the 1970s which brought a stronger focus on context and language use, even questioning core concepts such as language and native speaker (Lane 2015). *Language revitalisation* is commonly understood as community and individual efforts to maintain an indigenous or minoritised language or 'giving new life and vigour to a language that has been decreasing in use (or has ceased to be used altogether)' (Hinton, Huss and Roche 2018: xxi). For such efforts language documentation is important, which for Kven, the Ruija Corpus is a key part. Trond Trosterud has been a key contributor to the development of other online resources (available to the general public), such as an online dictionary (Trosterud 2019) and a morphological analyser for Kven developed by Giellatekno (Trosterud et al. 2017), the Research group for Sámi language technology, at UiT The Arctic University of Norway. The dictionary is written in the Børselv-Pyssyjoki dialect, but the analyser contains morphophonological information from different dialects. Thus, the dictionary recognises words from different dialects, such that a search for *lukea* 'read', common in the Eastern varieties, yields a Norwegian translation, but provides the Western infinitive form with a word final <t> *lukkeet*:

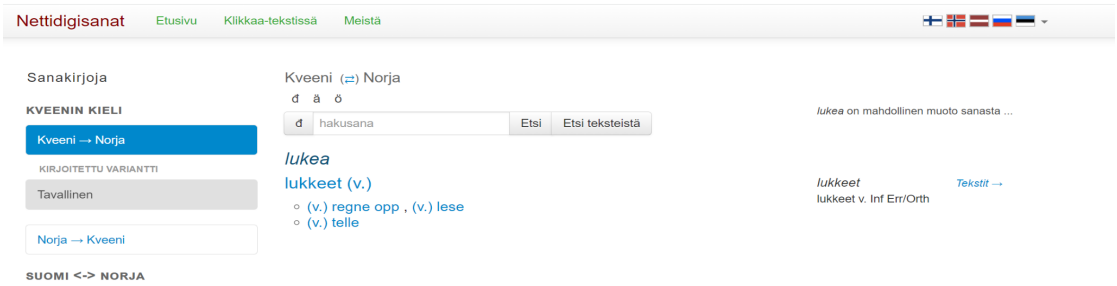

Figure 10. Example of entry from the Kven online dictionary: lukea – 'to read'.

[7] The term 'Language Planning' frequently is attributed to Haugen, but he mentions that Weinreich used the term Language Planning as a title for a seminar in 1959 (Haugen 1972: 209).

Ideally, such an online dictionary should also provide forms common in other Kven dialects, but as often is the case when documenting and standardising minoritised languages, human and monetary resources are limited; therefore, tools are built step-by-step and based on available resources. All forms of language documentation and standardisation, including making corpora, dictionaries and grammars, are carried out with a user in mind, though actors are aware of this to varying degrees, and decisions to include, and thereby exclude, some grammatical forms are not a purely linguistically based choice. The choices made by researchers and language planners may constrain future actions of intended users as these might not recognise the way they speak or lack resources or knowledge to use the developed tools. More importantly, corpora, dictionaries and grammars also prepare efforts to revitalise a minority language as such tools provide the basis for educational materials, literature and more visibility for minority languages in public space. This has indeed been the case for Kven as textbooks and a grammar for the educational system have been developed, an annual New Year speech in Kven is aired by the national Norwegian Broadcasting Corporation, street signs in Kven are introduced in the northern part of Norway giving visibility to traditional Kven place names, and the Norwegian nation state has got a Kven name (in addition to Norwegian and Sámi), namely *Norja*. There is an urgent need for more arenas for the use of Kven, both within and outside the educational system. Hinton (2018: 460) reminds us that:

> The biggest hurdle for both native speakers and language learners is to actually start using the language on a daily basis. For endangered languages, this is a major challenge. Just as elders in a community that has undergone language shift cease to use the language they grew up with because most of the community doesn't know it, so do second-language learners find themselves without interlocutors.

In spite of limited resources, dedicated researchers and language technologists have managed to develop a range of resources that have been used and will continue to be used in the process of revitalising Kven. What now remains is to continue developing the corpus, not only by including new material but also by developing the Ruija Corpus into a morphologically tagged corpus.

7. Conclusion: future prospects

Language documentation and revitalisation used to rely extensively on the role of academic experts, but there has been a shift in the field recognising that expertise also resides with the speakers of an endangered language and that they have a right to be involved in and shape these processes (Hill 2002) as a part of "a larger effort by a community to claim its right to speak a language and to set associated goals in response to community needs and perspectives" (Leonard 2017: 19). This was also the case for the documentation of Kven as the Kven Language Council had Kven members and all the members of Language Board, which held the executive function, were Kven speakers. When documenting minoritised languages, linguists work in tandem with members of local communities and participate in community efforts to sustain languages. Such participation in turn influences the balance of power and opens up space for new types of knowledge, as outlined by Eira (2007), see also Lane and Makihara (2017) for a discussion.

When the Ruija Corpus was created, such concerns were less prominent in our research field, and consequently, the corpus is only available to researchers and those who are interviewed are anonymised. This is a twofold challenge: the corpus is not available to speakers and learners of Kven, and the expertise and knowledge of language and culture by those interviewed is not acknowledged because according to ethical regulations when the creating the corpus began in 2007, their anonymity had to be ensured. While the mind-set of the academic community has undergone a profound change striving to recognise the knowledge, voices, perspectives and expertise of the speakers and communities we are working with, there are still many hurdles to overcome. A future endeavour for the Ruija Corpus will be to find ways of making at least parts of the corpus available also for users outside the traditional academic community, thus recognising that knowledge and ownership to data do not reside within academia only.

References

Austin, Peter. 2020. Language documentation and revitalisation. In *Revitalizing Endangered Languages: A Practical Guide*, edited by Justyna Olko and Julia Sallabank, pp. 199–219. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108641142.014.

Costa, James, Haley De Korne and Pia Lane. 2017. Standardising minority languages: Reinventing peripheral languages in the 21st century. In *Standardizing Minority Languages: Competing Ideologies of Authority and Authenticity in the Global Periphery*, edited by Pia Lane, James Costa and Haley De Korne, pp.1–23. Routledge, New York.

Eira, Christine. 2007. Addressing the ground of language endangerment. In *Working Together for Endangered Languages: Research Challenges and Social Impacts – Proceedings of Foundation for Endangered Languages Conference XI Kuala Lumpur October 26–28* 2007, edited by Maya K. David, Nicholas Ostlerand Ceasar Dealwis, pp. 82–90. Foundation for Endangered Languages.

Evans, Nicholas and Alan Dench. 2006. Introduction: Catching language. In *Catching Language: The Standing Challenge of Grammar Writing*, edited by Felix Ameka, Alan Dench and Nicholas Evans, pp. 1–39. Mouton de Gruyter, Berlin, New York.

Gal, Susan. 2006. Contradictions of standard language in Europe: Implications for the study of publics and practices. *Social Anthropology*: 14(2), 163–181. https://doi.org/10.1111/j.1469-8676.2006.tb00032.x.

Glossa: https://www.hf.uio.no/iln/english/about/organization/text-laboratory/services/glossa/index.html

Haugen, Einar. 1972. *The ecology of language: Essays by Einar Haugen. Selected and introduced by Anwar S. Dil*. Stanford University Press, Stanford.

Hill, Jane. 2002. "Expert rhetorics" in advocacy for endangered languages: who is listening, and what do they hear? *Journal of Linguistic Anthro*pology 12(2): 119–133. https://doi.org/10.1525/jlin.2002.12.2.119

Hinton, Leanne. 2018. Approaches to and strategies for language revitalization. In *The Oxford Handbook of Endangered Languages*, edited by Kenneth Rehg and Lyle Campbell, pp. 443–465. Oxford University Press, New York. https://doi.org/10.1093/oxfordhb/9780190610029.013.22.

Hinton, Leanne, Leena Huss and Gerald Roche. 2018. Language revitalization as a growing field of study and practice. In *The Routledge Handbook of Language Revitalization,* edited by Leanne Hinton, Leena Huss and Gerald Roche, pp. xxi-xxx. Routledge. New York.

Hyltenstam, Kenneth and Tommaso M. Milani. 2003. *Kvenskans Status: Rapport for Kommunal- og regionaldepartementet og Kultur- og Kirkedepartementet i Norge*. Oslo.

IMS Open Corpus Workbench: http://cwb.sourceforge.net.

Johannessen, Janne Bondi; Nygaard, Lars; Priestley, Joel; Nøklestad, Anders. 2008. Glossa: a multilingual, multimodal, configurable user interface. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Daniel Tapias, pp. 617–621. European Language Resources Association (ELRA), Paris.

Kaplan, Robert, Richard Baldauf, Anthony Liddicoat, Pauline Bryant, Marie-Thérèse Barbaux and Martin Pütz. 2000. Current issues in language planning, *Current Issues in Language Planning*, 1(1): 1–10. https://doi.org/10.1080/14664200008668003.

Keränen, Mari. 2018. Language maintenance through corpus planning – the case of Kven. *Acta Borealia*, 35(2): 176–191. https://doi.org/10.1080/08003831.2018.1536187.

Kloss, Heinz 1967Bilingualism and nationalism. *Journal of Social Issues*, 23(2), 39–47. https://doi.org/10.1111/j.1540-4560.1967.tb00574.x

Kven online dictionary, Nettidigisanat https://sanat.oahpa.no/

Latomaa, Sirkku and Pirkko Nuolijärvi. 2005. The language situation in Finland. In *Language Planning and Policy in Europe, Vol. 1. Hungary, Finland and Sweden*, edited by Robert B. Kaplan and Richard B. Baldauf, pp. 125–232. Multilingual Matters, Clevedon.

Lane, Pia. 2011. The birth of the Kven language in Norway: Emancipation through state recognition. *International Journal of the Sociology of Language* 209: 7–74. https://doi.org/10.1515/ijsl.2011.021.
Lane Pia. 2016. Standardising Kven: Participation and the role of users. *Sociolinguistica 30*: 105–124. https://doi.org/10.1515/soci-2016-0007.
Lane, Pia. 2015. Minority language standardisation and the role of users. *Language Policy* 14, 263–283. https://doi.org/10.1007/s10993-014-9342-y
Lane, Pia. 2017. Language standardisation as frozen mediated actions – the materiality of language standardization. In *Standardizing Minority Languages: Competing Ideologies of Authority and Authenticity in the Global Periphery*, edited by Pia Lane**,** James Costa and Haley De Korne, pp. 101–117. Routledge, New York. https://doi.org/10.4324/9781315647722.
Lane, Pia and Miki Makihara. 2017. Indigenous peoples and their languages. In *The Oxford Handbook of Language and Society*, edited by Ofelia García, Nelson Flores and Massimilano Spotti*,* pp. 299–230. Oxford University Press, New York. https://doi.org/10.1093/oxfordhb/9780190212896.013.7.
Leonard, Wesley. 2017. Producing language reclamation by decolonising 'language'. *Language Documentation and Description* 14: 15–36. http://www.elpublishing.org/PID/150.
Milroy, James and Leslie Milroy. 1999. *Authority in Language: Investigating Standard English*. Routledge, London.
Niemi, Einar. 1995. The Finns in northern Scandinavia and minority policy. In *Ethnicity and Nation Building in the Nordic World*, edited by Sven Tägil*,* pp. 145–178. Hurst and co, London.
Niemi, Einar. 2003. Regimeskifte, innvandrere og fremmede. In *Norsk innvandringshistorie. I nasjonalstatens tid 1814–1940,* edited by Einar Niemi, Jan Eivind Myhre and Knut Kjeldstadli, pp. 11–47. Pax forlag, Valdres.
Nøklestad, Anders, Kristin Hagen, Janne Bondi Johannessen, Michal Kosek and Joel Priestley. 2017. A modernised version of the Glossa corpus search system. In *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa),* edited by Jörg Tiedemann and Nina Tahmasebi, pp. 251–254. Association for Computational Linguistics, Gothenburg.
O'Rourke, Bernadette and Joan Pujolar. 2015. New speakers of minority languages: the challenging opportunity – Foreword. *International Journal of the Sociology of Language* 231: 1–20. https://doi.org/10.1515/ijsl-2014-0029.
Pietikäinen, Sari, Leena Huss, Sirkka Laihiala-Kankainen, Ulla Aikio-Puoskari and Pia Lane. 2010. Regulating multilingualism. *Acta Borealia*: 27(1): 1–23. https://doi.org/10.1080/08003831.2010.486923
Ruija Corpus: https://tekstlab.uio.no/glossa2/ruija
Sundelin, Egil. 1998. Kvenene – en nasjonal minoritet i Nord-Troms og Finnmark? In *Kvenenes historie og kultur*, edited by Helge Guttormsen, pp. 35–48. Nord-Troms historielag, Skjervøy.
Söderholm, Eira. 2014. *Kainun Kielen Grammatikki*. Suomalaisen Kirjallisuuden Seura, Helsinki.
Söderholm, Eira. 2017. *Kvensk Grammatikk*. Cappelen Damm Akademisk, Oslo.
Søfteland, Åshild, Anders Nøklestad, Joel Priestley and Kristin Hagen. 2020. Glossa som forskningsverktøy. Hva folk søker etter og hva resultatene brukes til. *Oslo Studies in Language*: 11(2): 449–464. https://doi.org/10.5617/osla.8512.
Trosterud, Sindre Reino, Trond Trosterud, Anna-Kaisa Räisänen, Leena Niiranen, Mervi Haavisto and Kaisa Maliniemi. 2017. A morphological analyser for Kven. In *Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages*, pp. 76–88, St. Petersburg, Russia. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-0608.
Trosterud, Trond. 2019. Kva bruker vi minoritetsspråksordbøker til? Ein studie av brukarloggane for tolv tospråklege ordbøker. *LexicoNordica* 26: 177–198.
Woolard, Katheryn. 2008. Language and identity choice in Catalonia: The interplay of contrasting ideologies of linguistic authority. In *Lengua, Nación e Identidad: La Regulación del Plurilingüismo en España y América Latina*, edited by Kirsten Siiselbeck, Ulrike Mühlschlegel, and Peter Masson, pp. 303–323. Vervuert, Frankfurt am Main /Iberoamericana, Madrid.

Wright, Sue. 2004. *Language Policy and Language Planning: From Nationalism to Globalisation*, Palgrave Macmillan, Basingstoke.

Östman, Jan-Ola. 2000. Ethics and appropriation – with special reference to Hwalbáy. In *Issues of Minority Peoples*, edited by Frances Karttunen and Jan-Ola Östman, pp. 37–60. Department of General Linguistics, University of Helsinki.