Transforming Archived Resources with Language Technology : From Manuscripts to Language Documentation

Partanen, Niko

CEUR-WS.org
2022

Partanen , N , Blokland , R , Rießler , M & Rueter , J 2022 , Transforming Archived Resources with Language Technology : From Manuscripts to Language Documentation . in K Berglund , M La Mela & I Zwart (eds) , Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022) . CEUR Workshop Proceedings , no. 3232 , CEUR-WS.org , Aachen , pp. 370-380 , Digital Humanities in the Nordic and Baltic Countries Conference , Uppsala , Sweden , 15/03/2022 .

http://hdl.handle.net/10138/350329

Transforming Archived Resources with Language Technology: From Manuscripts to Language Documentation

Niko Partanen[1], Rogier Blokland[2], Michael Rießler[3] and Jack Rueter[1]

[1]University of Helsinki
[2]Uppsala University
[3]University of Eastern Finland

Abstract

Transcriptions in different languages are a ubiquitous data format in linguistics and in many other fields in the humanities. However, the majority of these resources remain both under-used and under-studied. This may be the case even when the materials have been published in print, but is certainly the case for the majority of unpublished transcriptions. Our paper presents a workflow adapted in the research project *Language Documentation Meets Language Technology*, which combines text recognition, automatic transliteration and forced alignment into a process which allows us to convert earlier transcribed documents to a structure that is comparable with contemporary language documentation corpora. This has complex practical and methodological considerations.

Keywords

documentary linguistics, language technology, text recognition, forced alignment, Zyrian Komi,

1. Introduction

In many fields in the humanities there is a long history of collected materials that have been stored in various archives. One type of such material is formed by hand- or typewritten linguistic transcriptions, sometimes accompanied by interlinear glosses and translations. The exact transcription and annotation conventions vary between research traditions, but vast collections of notebooks containing transcribed and translated materials in endangered languages are still ubiquitous in the field of linguistics. In linguistic research of Uralic (and other languages beyond our scope) there is a long history of publishing these transcriptions in printed volumes, but it is almost impossible to estimate how much material still remains unpublished and thus basically inaccessible for research. Furthermore, we argue that in many cases even the printed versions are not as useful as they could potentially be. The reasons are primarily

The 6th Digital Humanities in the Nordic and Baltic Countries 2022 Conference, Uppsala, Sweden, March 15-18.
niko.partanen@helsinki.fi (N. Partanen); rogier.blokland@moderna.uu.se (R. Blokland);
michael.riessler@uef.fi (M. Rießler); jack.rueter@helsinki.fi (J. Rueter)
https://researchportal.helsinki.fi/en/persons/niko-partanen (N. Partanen);
https://uefconnect.uef.fi/henkilo/michael.riessler (M. Rießler);
https://researchportal.helsinki.fi/en/persons/jack-rueter (J. Rueter)
0000-0001-8584-3880 (N. Partanen); 0000-0003-4927-7185 (R. Blokland); 0000-0002-2397-2860 (M. Rießler);
0000-0002-3076-7929 (J. Rueter)

technical: a printed page in a graphically complex linguistic transcription (typically applied in these data sources) is as non-searchable for modern computer systems as a handwritten or typed transcription. At the same time the digitization process that can be applied to these sources is remarkably similar, as we describe in this study.

There is thus a clear need and motivation to make this type of material more accessible. An excellent example is the EuroBABEL project,[1] which published transcribed, translated and annotated versions of texts in Ob-Ugric languages (Uralic) originally published in printed works written by various linguists. These texts were digitized, analysed using FLEx linguistic annotation software and made available as a database. The American Philosophical Society is also currently running a long-term project where more than 2000 pages of Tunica (a language isolate spoken in the United States) materials are processed in a comparable way.[2] This illustrates that processing a collection of this size is usually an undertaking lasting several years, involving six-figure budgets. We believe this is necessarily often the case: these old archival materials are so complex and multi-layered that publishing them in modern editions is always a large and complex task. Much of the work that would need to be done necessarily also concerns analysing the texts in detail within the context in which they were created, and connecting them to other archival sources where possible. Ideally a digital edition of a transcription is not just an online transcription, but an attempt to make the work understandable and useful for modern audiences, most importantly in ways that support the language communities from which the materials originate.

This vision of re-publishing of or corpus building with transcribed materials requires that we also look for ways to shift the workload involved away from the most manual phases, so that specialists and community members can focus on higher level tasks which arise mostly after the text is digitally readable and searchable. Although the work on text recognition in itself is very valuable, we additionally take into account the context where recordings of the transcriptions are also available. This is connected closely to the history of audio recording methods and the use of this technology in linguistics: the period when the work was conducted largely determines if there even could be a recording. The fact also remains that not everything has been archived, and, even when something has been, whether the materials are findable or still usable are different questions; see also Poa & LaPolla [1, 351]. However, on an optimistic note, we believe that in many cases the audio recordings do exist, can be found and may already be digitized.

We want to note here that digital re-publishing of this kind of material customarily involves three different steps. These are text recognition, transliteration, and forced alignment. The first is the task that extracts the searchable text from an image. We can distinguish optical character recognition (OCR) from handwritten text recognition (HTR), but the concrete technical differences are minor nowadays. Transliteration is the process where we convert digitised character strings from one to another writing system, e.g. changing the writing system from Latin to Cyrillic, or from one expert transcription system to another. In many cases this task is similar to normalization, where we would just change some of the spellings, for example.

[1] OUL and its proceeding project OUDB, see https://www.babel.gwi.uni-muenchen.de/.

[2] https://www.amphilsoc.org/blog/cnair-awarded-grant-develop-digital-linguistic-resource-tunica-biloxi-tribe-louisiana

The last step in the workflow is forced alignment. This means automatically aligning the textual representation with the audio segments where the words or sentences have been uttered. The importance of aligned transcriptions has been strongly emphasized early on in language documentation, see, e.g. discussions in [2]. In our experience the current technical solutions are satisfactory to address almost all the steps that can be identified in a pipeline that transforms these materials into a structure close to contemporary language documentation corpora. This approach was already described by Blokland et al. [3], and our work builds on that.

There has been increased discussion over the last years on using technology in endangered language contexts. For example, although there has been some success with automatic speech recognition (ASR) systems for endangered languages [4][5], it has also been pointed out that in some contexts unassisted transcription may simply be preferred [6]. Such viewpoints are very important, and it is crucial not to present technical advances automatically as real improvements, especially before some tangible long-term results can be presented and demonstrated. In our context transcribing the materials again would also mean repeating the work that has already been done decades ago, which seems hard to justify. This situation is obviously very different when no previous transcriptions have been done. To clarify our context further, the present study therefore also aims to discuss the use of a set of related technologies within the internal data management workflow of a language documentation project during the period 2017–2021. As the need for further documentation of the endangered languages continues to be a global issue and undertaking, we believe this context will remain relevant, but also acknowledge that we discuss relatively narrow working environments and goals of academic research groups at European universities.

2. Case study

In our case study we use the transcriptions the Permian Komi linguist Raisa Batalova (1931–2016) carried out in 1971 with Zyrian Komi recordings made by the Finnish linguist Erkki Itkonen (1913–1992) in the Komi Republic between 21.12.1957 and 1.1.1958 [7].

The pluricentric Komi language is a Uralic language, related to e.g. Finnish, North Saami and Hungarian. It has three main variants, Zyrian Komi, Permian Komi and Yazva Komi, all spoken in the northeast of European Russia. Zyrian Komi is spoken mostly in the Komi Republic, and has approximately 170,000 speakers. As the language is related to Finnish, Finnish Finno-Ugrists have always shown a strong interest in it, though it was not easy to visit the Komi-speaking regions during Soviet times. However, Itkonen managed to visit the Komi Republic in 1958, where he was the first Finnish linguist to do so after 1907 [7]. During this trip Itkonen was mostly in Syktyvkar, the capital of what was then the Komi Autonomous Soviet Socialist Republic, and made a number of recordings of the language, both of the standard language and of dialects, as spoken by linguists he met at the Komi section of the USSR Academy of Sciences, and by teachers and students at the Pedagogical Institute. Back in Finland Itkonen used these recordings for his own notes on Komi; in the late 1950s he transcribed and translated five of the recordings into Finnish for his private use. In 1971 the Finno-Ugric Society and the Tape Archive of the Finnish Language paid the linguist Raisa Batalova a stipend to transcribe all the recordings and translate them into Russian [8, 508]. Batalova transliterated

and translated a total of 364 pages. Both the recordings and transliterations are archived in the Tape Archive of the Finnish Language in the Institute for the Languages of Finland, and we have currently processed approximately one third of them. The present dataset therefore contains 119 pages of handwritten transcriptions and their Russian translations. The current texts are approximately 18,000 tokens of transcribed Komi, and the whole transcribed material will likely be approximately 50,000 tokens. In the Figure 2 there is a small sample from the style of transcription.

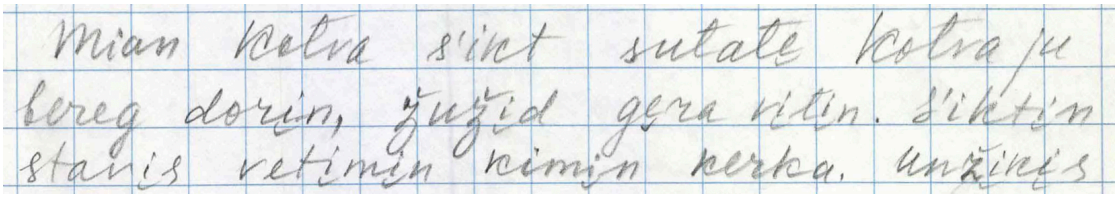


Figure 1: Example of Raisa Batalova's transcribed lines [Institute for the Languages of Finland]

The sample can be presented in Unicode characters as follows (the English translation has been added by us): *mian kołva s'ikt sułałę kołva ju bereg doṛiṇ, žužid gẹra viɬiṇ. s'iktiṇ staviṣ vetiṃiṇ kimiṇ kerka. unžikiṣ* 'Our Kolva village stands by the bank of the Kolva river, on a high hill. There are all in all about fifty houses. The most ...'. The Russian translation is handwritten on the adjacent page. Following our workflow, we can then connect this representation to the audio, as the original reel-to-reel tapes have been digitized. This is shown in Figure 2.

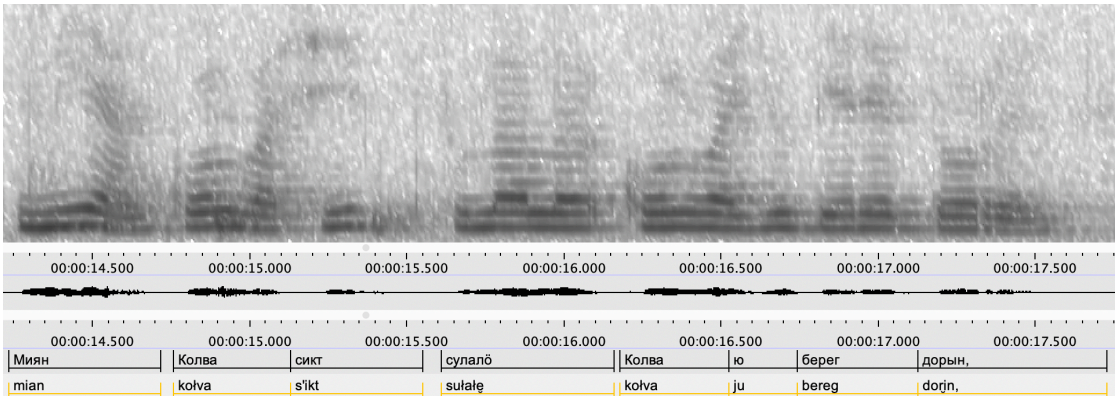


Figure 2: Word aligned data in ELAN [Institute for the Languages of Finland & IKDP-2 project]

2.1. Experiment design

To complement the dataset of 119 pages that was processed in our project, we selected 12 pages that were further annotated with word level alignment between the audio and transcription text. This includes four recordings of the Ižma dialect, which are not included in the text recognition experiments reported here. The setup is artificial, as in practice the Ižma transcriptions, i.e. those of the dialect that our working group mainly focuses on, were of course among the

most interesting for us. This also justified the more extensive work on these transcriptions, as the word level alignment was an extra step that we would not normally do. However, it was necessary in order to evaluate the forced alignment accuracy.

For a text recognition experiment we needed manually corrected lines. Here we formatted the lines so that the content being evaluated consists of correctly detected lines of Komi transcription. In this way we can evaluate text recognition accuracy consistently, regardless of issues possibly caused by layout detection. It is worth noting that if the workflow would be applied without manual correction and supervision, the errors would cumulatively influence the results of each additional step. This is not addressed in the current work. For the transliteration experiment the recognized texts are aligned word by word with the orthographic variants and the audio.

3. Processing workflow

3.1. Layout detection

As the original pages essentially contained horizontal lines on a page, the layout in itself was not very complex. There were, however, issues in detecting all the lines entirely, and the built-in layout detection models customarily left parts of the lines undetected, especially at the beginning and at the end. The quality was workable, but each page took approximately 15 minutes of manual correction before text recognition could be applied.

We trained two layout detection models with the P2PaLa method [9]. The first model used 37 pages (1113 baselines), the second 61 pages (1798 baselines). Both models improved from the baseline formed by the models built-in to Transkribus. The model only detected lines within a full page-sized text area. This demonstrates that training custom line detection models is also applicable in a context where the available data is relatively small. In the later stage there was no essential improvement in line detection: the necessary manual corrections were only few, and partly based on preference. It seems likely the rest of the Batalova's collection can be handled with the current model with no adjustments.

While the recognized text was corrected we manually assigned three tags ('area-komi', 'area-russian', and 'other'), these were both text-area and line level tags, and essentially differentiated which pages had Komi text and which Russian text. In principle, for some pages, the differentiation could have been carried out differently, assigning a page-level attribute that specified the language, but some pages were partly in Komi and partly in Russian, the split usually being where the Komi text ended and the Russian translation began. These were assigned manually, as it was a very fast task to assign them and manually split the elements when needed (approx. two hours of manual work for whole collection – remaining 2/3 could be tagged in less than one workday); it also functioned as an extra quality check. The tag 'other' primarily contains notes and metadata that does not connect to the running text. It was essential for us to be able to extract from the pages only the text that has correspondence in the audio.

There are additional layout elements, such as page numbers, titles and metadata information in the beginning of the texts. These were ignored in the layout detection phase as their separation from other content was not essential for the current work, and tagging them manually if needed would be similarly simple.

3.2. Text Recognition

Following an already established example by Petzell [10], who used text-recognized Swedish dialect texts, our research group was able to build a handwritten text recognition model in Transkribus [11] with highly functional accuracy. Although the work was conducted within a larger research process, we used the materials we had created to evaluate the process in more detail. We selected an incrementally growing set of pages on which the HTR models were trained. Each model was tested against the same test set. Similar tests had been done for printed text recognition [12] and speech recognition [4]. This is a very effective design as it shows clearly where the thresholds are in the applicability of the current technology. We used the PyLaia engine [13] with 200 epochs and 20 epochs early stopping.

Experiment	Lines	Words	CER (%)	WER (%)	Training time
10 pages	289	1591	20.5	67.5	19m 41s
25 pages	662	3657	10.3	41.4	22m 48s
50 pages	1300	7315	7.6	32.7	32m 29s
75 pages	1977	10911	6.5	29.1	44m 10s
100 pages	2677	14807	6.0	26.9	51m 26s

Table 1
Results of text recognition experiments.

The results show that after 50 pages the quality increases at a significantly slower rate. Another significant observation is that even with 10 pages the character error rate is 20.5%. This means that four in five characters are correct. As a consequence it is makes sense to suggest training a HTR model at a very early stage of the transcription, as creating new proofread pages becomes increasingly faster. Thirdly, we can state that the whole training experiment is resource-efficient and fast, as the training time on the server remains very short.[3] As the whole collection is over 300 pages, it is very likely that small increases in the accuracy will continue to occur while the work expands to new pages.

The retrieved text is in the Finno-Ugric transcription, with the originally used character set represented as closely as possible. Different characters have been carefully distinguished, although there are often several suitable Unicode values that could be used. Again, as long as the choice is systematic this makes no or little difference from the point of view of subsequent tasks or use of this data.

3.3. Transliteration

After this, we applied a rule-based transliteration script that transformed the original transcription into contemporary Cyrillic orthography used for Komi (which is similar, but not identical, to Russian Cyrillic). The reason we prefer an orthographic representation in language documentation, rather than a scientific transcription, was originally argued in Gerstenberger et al. [14, 35-36]. First, this makes the work useful to the language community, whichs is already

[3] As a disclaimer, we are not aware of documentation that describes the exact setup on the Transkribus servers, but still a training time that is under one hour is clearly modest in the wider machine learning context.

familiar with the orthography. At the same time an established orthography can function as a middle stage before a more detailed transcription is made, if that is needed or desired.

The transliteration script was written in Python and uses a set of sequentially applied rules. The challenge encountered is that the Komi standard language has 40 phonemes (36 native and 4 in loanwords /t͡s/, /x/, /f/, /rʲ/) represented by 35 characters in the modern standard language. Komi features a set of consonants that can be classified according to a palatal dichotomy, similar to the palatalization dichotomy in Russian, hence the Russian use of a soft sign or fronting and non-fronting vowels has been adopted for Komi in the post-1938 era. Table 2 presents four different representations of consonants found in the transcriptions, whereas the leftmost (Batalova) and the rightmost (modern Cyrillic) indicate the source and target expression for the present project, respectively. The so-called Molodtsov alphabet (a Cyrillic-based alphabet used for Komi in the 1920s and 1930s) and IPA presentations – which are potentially relevant for other similar projects on Zyrian Komi – are also given to illustrate the underlying system. The approach presented here works in an equally effective manner also in transliteration between these writing systems, which is why we illustrate them here as well.

Batalova	**Molodtsov**	**Phonemes**	**Cyrillic**
palatal-	**neutral**		(see ‹*b*|*v*|*g*|*ž*|*k*|*m*|*p*|*r*|*f*|*x*|*c*|*š*› to ‹*б*|*в*|*г*|*ж*|*к*|*м*|*п*|*р*|*ф*|*х*|*ц*|*ш*›)
š	ш	*ʃ*	*ш (а|и|у|е|о|ы|ӧ)*
symmetric	**pairs**		(see also ‹*s*|*d*|*t*|*n*|*l*› to ‹*с*|*д*|*т*|*н*|*л*›)
z	з	*z*	*з (а|і|у|э|о|ы|ӧ)*
z’	ҙ	*ʑ*	*з (ь|я|и|ю|е|ё|ьы|ьӧ)*
asymmetric	**pairs**		
ǯ	ж	*dʒ*	*дж (а|и|у|е|о|ы|ӧ)*
ʒ’	ӂ	*dʑ*	*дз (а|и|у|е|о|ы|ӧ)*
č	щ	*tʃ*	*тш (а|и|у|е|о|ы|ӧ)*
c’/č’	ч	*tɕ*	*ч (Ø|а|и|у|е|о|ы|ӧ)*

Table 2
Phonetic equivalents in Batalova, Molodtsov, IPA, and modern Cyrillic orthography

A Python script divided the transliteration task into five sequential sets: word-initial, preletter, letter, pair-vowel and other-letters. The word-initial and preletter sets were used for dealing with multiple-character to single-letter conversion, i.e. dealing with palatal glides and and the combining UNICODE character conversion for central vowels *e̮* to *ӧ* and *i̮* to *ы*, respectively. These steps were also used for removing labialization and accent marking.

When the orthographic text was analysed with a Komi morphological analyser [15], the initial accuracy was 79.5%. This is considerably lower than the accuracy usually reached with written texts or dialect texts where the analyser is sufficiently adapted to the dialectal features, as described by Rueter et al. [15]. However, the character error rate when compared to manually verified Ižma wordforms was only 4.2%. This indicates a difference between some of the Komi dialects in the collection and the coverage of the analyser outside the already adjusted dialects.

The Finno-Ugric transcription system as such has been widely used, also for Komi. Yet the published texts are not currently available as digital versions, there are only few instances of existing Cyrillic versions of the texts, and different publications differ from each other in

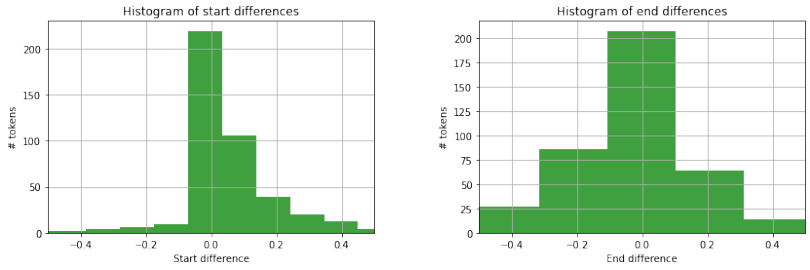
transcription details. Also Batalova's text studied here contains its own conventions that are not widely seen in other works. This creates a situation where there no exact training data that could be used exists. At the same time the approach presented above could be fairly easily extended to other publications that use their own transcription systems.

3.4. Forced alignment

Forced alignment refers to the task where a text and corresponding audio are aligned with one another. This can be done on different levels, which are usually utterances, words, phonemes or phones. In our report the texts and corresponding audio files were aligned with a forced alignment system described by Leinonen et al. [16] that is currently available in the Language Bank of Finland. We used this implementation in CSC's Puhti infrastructure as Komi was added to this system in 2021 under the macrocode **kv**. This code refers to both Permian Komi and Zyrian Komi, and indeed the current setup works for both main varieties.

The forced alignment system of Leinonen et al. [16] is an example of a cross-lingual application in this domain. The idea is that an alignment system is trained for one language and applied to another. As mentioned above, this task can be done with varying granularity. Matching longer sentences is less exact and has more margin for errors than matching words, and even more so when we discuss phonemes and phones, where the units are already very language-specific in themselves.

There are already examples of using this approach in language documentation contexts. For example, an Italian model has been used for Australian Kriol, one reason for this being that they both have similar vowel systems [17, 285]. The system we used was also based on a Finnish-language model, and the idea is that this would serve as an adequate starting point for a cross-linguistic forced alignment. The study of Leinonen et al. [16] already carried out tests in Finnish, Estonian, North Saami, and our work contributes to this work on the Uralic languages, in this case Zyrian Komi. However, it is not obvious whether there is any particular benefit in using the Finnish model for Komi as compared to any other language pair, and further testing with different languages remains important. Our results indicate that at least in the case of Komi the alignment works very well. The results are displayed in Figure 3.



(a) (b)

Figure 3: The alignment accuracy in seconds measured from the start (a) and end (b) of the word.

This shows that the majority of the aligned words are very close to their start and end, the end displaying more variation. The start difference median is 0.04 and the end difference median is 0.11, which also shows the higher fluctuation in the word end. We can complement these figures with our practical experience from manually aligning the materials. For the vast majority of the aligned words both ends had to be adjusted slightly, but it was very rare that the predicted location would be very far from the correct location. This makes the system functional when utterance level alignment is wanted, or if we want to coarsely align the text recognition result with the original audio.

4. Conclusion

We have presented a pipeline that allows us to efficiently process handwritten transcriptions, transliterate them into the modern orthography, and align them with the original audio recordings. We used approximately one third of the available dataset in these experiments, which demonstrates that, at least within these constraints, our workflow is practicable. Since the constraints under which we operate are not particularly unusual, with similar archived datasets existing for numerous endangered languages around the world, we believe that these approaches can easily be extended to new environments. The full material described here will be eventually published both in print and online, and the current study is part of the reported work in progress.

The forced aligned transcriptions can be easily converted into ELAN files [18, 19] (or similar tools used in documentary linguistics), storing the original transcription, converted orthography and later manually verified and adjusted transcription on their own tiers. At this level one can easily compare these transcriptions to modern recordings, and apply the exactly same annotation methods to the archival resources. Thus, we can bring archived transcription manuscripts and their recordings into the same unity as current language documentation endeavours based on new fieldwork. The scope of this extends much beyond the archived manuscripts and reel-to-reel tapes, but for many languages the combinations of text and unaligned audio can be found in innumerable formats and storage locations. Our workflow can be easily adjusted to most of these situations and help solving a real-world task in current documentary linguistics of under-resourced and under-researched languages.

Acknowledgements

We want to thank the Kone Foundation (Helsinki) for their support in our research projects *Iźva Komi Documentation Project* in 2014–2016 and *Language Documentation Meets Language Technology: The Next Step in the Description of Komi* in 2017–2021. We also want to thank the Institute for the Languages of Finland for giving us access to the Komi recordings used in this study. We are also grateful to all the Komi speakers and colleagues who have collaborated with us over the years.

References

[1] D. Poa, R. J. LaPolla, Minority languages of China, in: O. Miyaoka, M. E. Krauss (Eds.), The Vanishing Languages of the Pacific, Oxford University Press, 2007, pp. 337–354.

[2] J. Gippert, U. Mosel, N. Himmelmann (Eds.), Essentials of language documentation, number 178 in Trends in Linguistics. Studies and Monographs, Mouton de Gruyter, 2006.

[3] R. Blokland, N. Partanen, M. Rießler, J. Wilbur, Using computational approaches to integrate endangered language legacy data into documentation corpora: Past experiences and challenges ahead, in: Proceedings of the Workshop on Computational Methods for Endangered Languages, volume 2, 2019.

[4] N. Partanen, M. Hämäläinen, T. Klooster, Speech recognition for endangered and extinct Samoyedic languages, in: Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation, 2020.

[5] N. Hjortnaes, N. Partanen, M. Rießler, F. M. Tyers, Towards a speech recognizer for Komi, an endangered and low-resource Uralic language, in: Proceedings of the Sixth International Workshop on Computational Linguistics of Uralic Languages, 2020.

[6] E. Prud'hommeaux, R. Jimerson, R. Hatcher, K. Michelson, Automatic speech recognition for supporting endangered language documentation, Language Documentation & Conservation 15 (2021) 491–513.

[7] E. Itkonen, Komin tasavallan kielitieteeseen tutustumassa, Virittäjä 62 (1958) 66–66.

[8] M. Korhonen, Suomalais-ugrilaisen seuran vuosikertomus v. 1971, (Journal de la Société Finno-Ougrienne 72 (1973) 505–512.

[9] L. Quirós, P2pala: Page to page layout analysis toolkit, https://github.com/lquirosd/P2PaLA, 2017. GitHub repository.

[10] E. M. Petzell, Handwritten text recognition and linguistic research, in: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, 2020.

[11] P. Kahle, S. Colutto, G. Hackl, G. Mühlberger, Transkribus-a service platform for transcription, recognition and retrieval of historical documents, in: 2017 14th IAPR International Conference on Document Analysis and Recognition, volume 4, 2017.

[12] N. Partanen, M. Rießler, An OCR system for the Unified Northern Alphabet, in: Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages, 2019.

[13] J. Puigcerver, C. Mocholí, Pylaia, https://github.com/jpuigcerver/PyLaia, 2018. GitHub repository.

[14] C. Gerstenberger, N. Partanen, M. Rießler, J. Wilbur, Utilizing language technology in the documentation of endangered Uralic languages, Northern European Journal of Language Technology 4 (2016) 29–47.

[15] J. Rueter, N. Partanen, M. Hämäläinen, T. Trosterud, et al., Overview of open-source morphology development for the Komi-Zyrian language: Past and future, in: Proceedings of the Seventh International Workshop on Computational Linguistics of Uralic Languages, 2021.

[16] J. Leinonen, S. Virpioja, M. Kurimo, et al., Grapheme-based cross-language forced alignment: Results with Uralic languages, in: Proceedings of the 23rd Nordic Conference on Computational Linguistics, 2021.

[17] C. Jones, W. Li, A. Almeida, A. German, Evaluating cross-linguistic forced alignment of conversational data in north Australian Kriol, an under-resourced language, Language Documentation and Conservation (2019) 281–299.
[18] Elan (version 6.3) [computer software], 2022. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.
[19] H. Brugman, A. Russel, Annotating multi-media/multi-modal resources with ELAN, in: LREC, 2004, pp. 2065–2068.

A. Online Resources

Our study has used the following online resources.

- Documentation of the Aalto-ASR system in the Language Bank of Finland's infrastructure at CSC's Puhti server [In Finnish].
- Forced alignment evaluation scripts in Aalto-ASR GitHub project.
- Documentation: How to Train PyLaia-Models in Transkribus