

2020

Improving the language model for low-resource ASR with online text corpora

Hjortnaes, Nils

European Language Resources Association (ELRA)

Artikkelit ja abstraktit tieteellisissä konferenssijulkaisuissa

© European Language Resources Association (ELRA)

CC BY-NC <http://creativecommons.org/licenses/by-nc/4.0/>

<https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/SLTUCCURLbook.pdf>

<https://erepo.uef.fi/handle/123456789/8144>

Downloaded from University of Eastern Finland's eRepository

Improving the Language Model for Low-Resource ASR with Online Text Corpora

Nils Hjortnaes¹, Timofey Arkhangelskiy², Niko Partanen³,
Michael Rießler⁴, Francis Tyers¹

¹Indiana University, ²University of Hamburg, ³University of Helsinki, ⁴University of Eastern Finland
nhjortn@iu.edu, timarkh@gmail.com, niko.partanen@helsinki.fi,
michael.riessler@uef.fi, ftyers@iu.edu

Abstract

In this paper, we expand on previous work on automatic speech recognition in a low-resource scenario typical of data collected by field linguists. We train DeepSpeech models on 35 hours of dialectal Komi speech recordings and correct the output using language models constructed from various sources. Previous experiments showed that transfer learning using DeepSpeech can improve the accuracy of a speech recognizer for Komi, though the error rate remained very high. In this paper we present further experiments with language models created using KenLM from text materials available online. These are constructed from two corpora, one containing literary texts, one for social media content, and another combining the two. We then trained the model using each language model to explore the impact of the language model data source on the speech recognition model. Our results show significant improvements of over 25% in character error rate and nearly 20% in word error rate. This offers important methodological insight into how ASR results can be improved under low-resource conditions: transfer learning can be used to compensate the lack of training data in the target language, and online texts are a very useful resource when developing language models in this context.

Keywords: Zyrian Komi (kpv), automatic speech-recognition, documentary linguistics

1. Introduction

Speech recognition has lots of potential to be a highly useful technology while working with the world’s various endangered languages. In recent years numerous studies have been conducted on this topic, and especially the recent work on phoneme recognition is reaching very promising results on endangered language corpora (Wisniewski et al., 2020; Michaud et al., 2019). Persephone (Adams et al., 2018) and Elpis (Foley et al., 2018) have been the most widely used systems in the language documentation context, but as the field is rapidly evolving, various methods are available. One of them is Mozilla’s DeepSpeech (Hannun et al., 2014). By finding new, more effective ways to use these methods, we can open up their usage to low resource languages. Language documentation is one area in which these methods can be applied. Levow et al. (2017) propose a number of tasks demonstrating the usefulness of natural language processing to language documentation, including speech processing. In this paper we report our latest results using DeepSpeech.

In a recent study we investigated the usefulness of automatic speech-recognition (ASR) in a low-resource scenario, which is generalizable for the fieldwork-based documentation of a medium-size endangered language of Russia (Hjortnaes et al., 2020). We ran various experiments with 35 hours of spoken dialectal Zyrian Komi (Permic < Uralic, henceforth Komi) to optimize the training parameters for DeepSpeech¹ and explored the impact of transfer learning on our corpus. Although the work was a promising start, further research was acutely needed as the reached accuracy

was very low.

In this study, we continue our previous work by exploring new potential methods to improve our results. Specifically, we are looking at how to improve the language model (LM) in order to reach higher accuracy in the ASR. In our earlier experiments, tuning the language model was able to produce slightly better results, though these were very small improvements.

We presume that a domain mismatch may have been involved in our previous low results with the language model despite its relatively large size. The language model we used previously was developed from more formal varieties of written language (literature and Wikipedia), it did not match the domain of the speech data to be recognized, i.e. more informal spoken language recordings. Two obvious differences are the frequent use of discourse particles and code switching to the majority language, Russian, which are atypical of written language. Other differences between written and spoken language are the insertion of dialectal words or word forms and the preference for shorter syntactic units in the latter.

Since written language used in social media tends to be informal (Arkhangelskiy, 2019) we hypothesized that a language model based on a social media corpus would result in significantly higher ASR accuracy. In fact, Komi is actively used in social media today and useful corpus data has recently been published by Timofey Arkhangelskiy. The *Komi-Zyrian corpora*² consist of two different sections: a standard written language corpus of 1.76 million words (called the “Main Corpus”), and the “Social Me-

¹<https://github.com/mozilla/DeepSpeech>

²<http://komi-zyrian.web-corpora.net>

Corpus	Size
Speech	35 hours
Wiki and Books	1.78M tokens
Literary	1.39M tokens
Social	1.37M tokens
Combined	2.76M tokens

Table 1: Token counts for the corpora used to create the language models. The Wiki and Books corpus also includes the Komi Republic Website and some newspaper articles.

dia Corpus” of 1.85 million words. Since these corpora are of comparable size and are a closer domain match to our speech corpus than the materials used to build the language model in our previous study, the conditions are promising to test how the larger text model influences the results.

Additionally, across these two corpora there are differences in variational sociolinguistic features, which should be taken into account during testing. The Main Corpus contains contemporary on-line press texts. Therefore it matches closely with standard written Komi. The Social Media Corpus, on the other hand, contains posts from the social media platform VKontakte³ and therefore represents the contemporary language of informal digital communication.

One significant advantage in the use of online texts is that they are available for a considerable number of minority languages and can be harvested relatively easily. Our approach for Komi is therefore generalizable to other languages, although we believe that specific conditions have to be met for endangered languages to have online materials available in sufficient quantity and quality. For online language vitality, see, e.g. Kornai (2015; Gibson (2016). First, internet access is a logical precondition as well as the basic technology for digital use of the written language, especially keyboard layouts for various platforms. Furthermore, the language needs to have a sufficiently large number of speakers and a literary standard vital enough that the speakers are familiar in writing the language (a case of another language, Kildin Saami, which does not have these conditions is described by Riebler (2013)).

Online texts also have an accumulative nature, so that the corpus grows incrementally from day to day. Therefore even a relatively limited amount of online presence can, in few decades, result in a substantially large corpus.

2. Data Acquisition

The speech data used is described more thoroughly in a previous study (Hjortnaes et al., 2020), so we only discuss it here briefly. The corpus itself will be available in the Language Bank of Finland (Blokland et al., 2020) during the spring 2020, and contains 35 hours of aligned transcriptions, primarily from northernmost Komi dialects. The transcription conventions used are close to the written standard. They use Cyrillic script, but include small adaptations to reflect dialectal differences. These adaptations are similar to texts in the recent Komi dialect dictionary by Beznosikova

³<http://vk.ru>

et al. (2012). Large portions of these materials are also available, and can be studied, via a community-oriented online portal (Fedina and Levchenko, 2017; Blokland et al., 2016 2020).⁴ This language documentation dataset is used to train the DeepSpeech model itself, but a language model is an essential part of the DeepSpeech architecture, as it is used to adjust the model’s output.

The language model used in this study is derived entirely from materials that are online. What it comes to endangered languages spoken in Russia, there is a long tradition of related work. Several corpora based on internet data have been published in in Russia in recent years, e.g. Orekhov et al. (2016) and Krylova et al. (2015), and more recent, similar work has also been conducted in Finland (Jauhiainen et al., 2019). In the future, it could be a promising avenue to combine all these sources, but for our current work we focus on one set of text corpora published last year (Arkhangelskiy, 2019), see above.

The kenlm language model (Heafield, 2011) used by DeepSpeech takes as input a plain-text file with one sentence per line. These were obtained from the annotated corpus files using `tsakorporus2kenlm script`⁵. Since it is common for social media data to be noisy and contain code switching (Baldwin et al., 2013), automatic language tagging and some text cleaning were performed when building the corpus. The latter included fixing characters with diacritics typed in one of the popular conventions, e.g. replacing *o*: with *ö* or Latin *i* with its identically looking Cyrillic counterpart. Therefore, the social media language model was based on somewhat cleaner data than the original social media posts. The conversion included two additional cleaning stages. First, only sentences with less than one-third OOV words (as determined by a rule-based Komi analyzer⁶) were included, to avoid wrongly tagged Russian sentences. Second, some numerals represented with digits were replaced with text, e.g. 2 was replaced with *кык*. All punctuation was removed. The resulting datasets used with kenlm contain 1.39M words in 153K sentences for the Main Corpus and 1.37M words in 231K sentences for the Social Media Corpus.

Although these preprocessing steps were conducted the social media data in mind, in principle similar adjustments could possibly be useful also in other contexts where noisy text data is used in ASR.

3. Methodology

In our previous work (Hjortnaes et al., 2020), we investigated the benefit of transfer learning and found that the best results were achieved with a learning rate of 0.00001 and dropout of 10% when using transfer learning. Our model, both in the previous work and here, is the DeepSpeech⁷ architecture (Hannun et al., 2014; Ardila et al., 2020). DeepSpeech is a relatively simple five layer neural network with one bi-directional LSTM layer. It takes audio as input and outputs a stream of characters, which are then corrected by

⁴<http://videocorpora.ru/>

⁵<https://bitbucket.org/timarkh/tsakorporus2kenlm/>

⁶<https://github.com/timarkh/>

⁷[uniparser-grammar-komi-zyrian](https://github.com/mozilla/DeepSpeech)

⁷<https://github.com/mozilla/DeepSpeech>

the language model to produce the final output. The transfer learning branch⁸ allows us to reset the last n layers of the network, crucially adjusting for differences in the alphabet size between the source and target language when using transfer learning. We found that resetting the last 2 layers, which does not include the LSTM layer, was most effective. These hyper-parameters are corroborated in Meyer (2019). In this study we continue to examine what kinds of further benefits can be gained by improving the language model. For these experiments, we constructed the language models using kenlm (Heafield, 2011), as described in section 2. We then trained the speech recognition model on the same set of audio data described above, changing only the source of data used for the language model. Finally, we tuned the alpha and beta hyper-parameters of the language model which control how much we weight the LM over the output of the acoustic model and the cost of inserting spaces to separate words respectively.

4. Results

Language model	Size	CER (%)	WER (%)
None	—	70.9	100.0
Wiki	1.78M	72.1	98.1
Literary	1.39M	45.3	80.8
Social Media	1.37M	46.1	81.8
Combined	2.76M	44.7	79.8

Table 2: The best results for each source of data used to construct the language model. The CER and WER do not necessarily come from the same hyper-parameters used to integrate the language model into the speech recognition system (Hjortnaes et al., 2020).

The previous results are presented alongside our newest results in Table 2 and alone in Table 3. The best word error rate (WER) was achieved with tuned language model parameters using transfer learning (see Table 3). However, the best character error rate (CER) when using the Wikipedia corpus was achieved by disabling the LM entirely. The domain appropriate corpora, however, produced language models which significantly improved upon both the Wikipedia LM and disabling the LM altogether with a CER improvement of over 25% and a WER improvement of nearly 20%.

What is particularly interesting here is that both the literary language model and social media language model resulted in a very similar level of improvement to the performance. The combined model yielded an even greater improvement, though only by about 2%. This goes against the hypothesis that domain would be the most crucial factor here, and calls for further work on various text types.

The hyper-parameters of the language models show many similarities across difference source corpora. In all but the Wikipedia and Book corpus, the best CER was obtained when beta was set to 1, and the best WER in all cases was

with a beta of 1 as well, meaning that the language model favors inserting fewer spaces. When beta becomes large, the predictions tend towards single character words regardless of how long the gold standard is. For all LM corpus sources, as alpha increases, which favors the language model over the output of the acoustic model, the WER goes down, but the CER goes up. This implies that the LM is properly correcting words, but at the cost of other characters.

5. Discussion

It can be observed that many of the remaining errors relate to Russian code-switching within a sentence, and to dialectal forms that do not have corresponding variant in either of the text corpora used. In the following examples the incorrectly recognized words are marked with bold in the Komi sentence, and the Russian parts are marked with italics in both the Komi sentence and Russian translation. The source lines are on top and the system’s predictions are below them. Example (1) [CER: 10.0, WER: 55.5]⁹ shows an almost correctly recognized sentence, where the main problems are in words that contain dialectal morphology. Here we see that the combined model tries to suggest the comitative case form *-кӧд* from literary Komi, and in the last wordform another dialectal comitative *-кед* goes entirely unrecognized.

- (1) тундраын ветлі сизим во керка кари аслум
тундраын ветлі сизим во керка **карьяс лун**
вокъяскед дядьяскед
вокъяскӧд ядьяс ке
‘I worked at the tundra for seven years, built a house
(in that time) with my brothers and uncles.’

In Example (2) [CER: 15.0, WER: 54.5] an individual borrowed Russian verb *спонсируйтны* ‘to sponsor’ seems to create conditions where the model fails. It is highly unlikely that such borrowed and loosely adapted items would occur in the language model. The same example, however, displays correctly recognized the Russian noun *страховка* ‘insurance’. This illustrates how in this kind of multilingual context drawing an exact line between the languages in contact is very difficult.

- (2) никод миян оз вермы оз мӧд *спонсируйтны*
никод миян **воз веныс по оз нас пони ртны**
пока миян абу *страховка*
пока миян абу *страховка*
‘Nobody is going to *sponsor* us as long we don’t
have an *insurance*’

The Example (3) [CER: 19.0, WER: 128.5] shows how for an entirely Russian sentence, the language model is not able at all to produce the correct output, but tries to create words in standard Komi. Also here the Russian word *дело* ‘thing, issue’ is transformed into Komi *делӧ*, which would be a good approximation of how this word is often pronounced in Komi. However, this shows how finding ways to deal with Russian content is one of the major challenges with

⁸<https://github.com/mozilla/DeepSpeech/tree/transfer-learning2>

⁹Individual sentences report the number of incorrect characters for CER, not the percentage of incorrect as in WER.

CER/WER		beta				
		1	3	5	7	9
alpha	0.25	76.6/100.0	73.2/100.0	72.1 /100.0	74.0/100.0	81.8/100.0
	0.5	81.0/99.4	77.0/100.0	74.0/100.0	73.3/100.0	75.8/100.0
	0.75	85.2/ 98.1	80.1/100.0	77.2/100.0	74.8/100.0	74.7/100.0

Table 3: The impact of tuning the language model parameters on Character and Word Error Rates for the Wikipedia dump language model. (Hjortnaes et al., 2020)

CER/WER		beta				
		1	3	5	7	9
alpha	0.25	45.3 /88.1	45.9/100.0	50.0/100.0	56.8/100.0	67.3/100.0
	0.5	47.6/81.1	46.1/92.6	47.9/100.0	53.0/100.0	61.8/100.0
	0.75	51.6/ 80.8	48.6/85.2	48.3/100.0	51.6/100.0	59.0/100.0

Table 4: The impact of tuning the language model parameters on Character and Word Error Rates for the literary corpus language model.

the language documentation data we are working on. The problems are certainly similar in other highly multilingual contexts.

- (3) *это очень сложное дело не всякому идёт*
та вочис лоны делё не ся ко мый и де
‘This is a very difficult issue, it does not fit everyone...’

In Example (4) [CER: 15.0, WER: 53.8] we see a different issue. Careful listening to the original audio reveals that there truly is a segment like *мый* or *мыйке* (i.e. a pronoun which is not clearly pronounced in the recording), although it is missing from the transcriptions. In this case the model does indeed capture something which the human transcriber didn’t. On that note, detecting such mistakes in the original data would generally be a highly useful domain for speech technologies. This example also contains a Russian sequence *мало того что* ‘not only, but’, which the model, as expected, is not able to analyze.

- (4) *сыа бура сёрнитис мало того что сыа бура*
сыа бура сёрнитис малы тов то са бура
сёрнитис гашке думайтис коми кывнас
син и мыйкке думайтис коми кывнас
‘He spoke well, not only did he speak well, maybe he [even] thought in the Komi language...’

Despite the abundance of Russian confounding our results, there is a very clear difference between the accuracy of the speech recognition as a whole for different corpora. Despite being smaller, both the social media corpus and the literary corpus outperformed the larger Wikipedia corpus. This demonstrates the importance of domain in the choice of corpus for constructing the language model. Size has an impact, as can be seen from the improvement yielded by combining the literary and social media corpus, but it is far less than using a corpus of a more similar domain to the audio data. In this case, online data offers that similarity and improves our results drastically.

6. Conclusion & Future work

As the combined language model was twice as large as the individual models alone, yet offered very little improvement, it remains inconclusive how large the further improvements could be with an even larger language model. We expect that simply increasing the size of the corpus will offer diminishing returns. However, we have demonstrated that creating the language model from available online materials is a very promising and effective way to improve the speech recognition in a low-resource context. By extension, this demonstrates concretely the importance of using quality data of an appropriate domain over simply using as much data as possible. Although the error rates are still relatively high, we are fast approaching a level where the ASR output starts to be sensible and useful for various purposes, primary of which would be to make transcription easier.

It is also noteworthy that the current speech dataset contains over 200 different speakers in very varying recording conditions, which is a realistic scenario for a corpus of fieldwork recordings. There is also a large amount of overlapping speech. Despite these challenges we have been able to produce relatively solid results. Therefore our study is a relevant new contribution in the line of work that attempts to eventually combine ASR systems with the fieldwork-based work of documentary linguistics.

Further experiments in this direction could include even bigger language models, which would firmly establish the role corpus size plays in the effectiveness of the LM. The National Komi Corpus¹⁰ currently contains more than 60 million tokens. This may sound unusually large for a minority language. However, as this body of texts is based on published literature, including printed books and periodicals, which have been printed in a similar, if not higher, magnitude for several other languages of the ethnic Republics of the Soviet Union and Russia, building corpora of comparable size should be possible for various minority languages of Russia as well as other minority languages in similar situations (e.g. in Western Europe). We are aware that printed books and periodicals in endangered languages are not typical of endangered languages globally. How-

¹⁰<http://komicorpora.ru/>

CER/WER		1	3	beta		
		5	7	9		
alpha	0.25	46.1 /88.6	46.8/100.0	51.2/100.0	58.5/100.0	69.9/100.0
	0.5	48.9/ 81.6	47.2/93.6	49.0/100.0	54.6/100.0	64.2/100.0
	0.75	53.5/81.8	50.1/85.9	49.5/100.0	53.1/100.0	61.3/100.0

Table 5: The impact of tuning the language model parameters on Character and Word Error Rates for the social media language model.

CER/WER		1	3	beta		
		5	7	9		
alpha	0.25	44.7 /86.9	45.5/100.0	49.7/100.0	56.7/100.0	675/100.0
	0.5	47.1/80.0	45.5/91.1	47.5/100.0	52.6/100.0	61.4/100.0
	0.75	51.2/ 79.8	48.1/83.7	47.5/100.0	50.9/100.0	58.2/100.0

Table 6: The impact of tuning the language model parameters on Character and Word Error Rates for the combined corpus language model.

ever, user-generated online communication, through websites and social media, seem to be becoming more and more available, even in contexts where standard literary materials are lacking.

The experiment here used online corpora as a source of spontaneous colloquial data that resembles the spoken transcriptions more than literary standard texts would. Although we have shown there is still some ambiguity in the kind of data we needed to improve the language model, we can experiment more in this direction in the future. For instance, there are numerous text collections available consisting of transcribed dialectal speech similar to those fieldwork-based recordings our ASR system is analysing. For these text collections there is no corresponding audio available. Potential of combining various legacy datasets systematically into language documentation corpora has been discussed before (Blokland et al., 2019), but the benefit for speech recognition may have not been previously recognized. Logically, without audio we can't use these texts in training the ASR system itself, but they could be potentially very useful as a new source of an enriched language model, matching our own speech data perfectly. Apart from simply collecting more data, finding a way to address the Russian which exists in the speech data is a potential avenue for improvement, as the current model essentially ignores it. These language models were constructed exclusively using Komi data, so any Russian which does not exactly match a Komi analogue will be a source of error.

Acknowledgements

This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. Niko Partanen and Michael Rießler collaborate within the project *Language Documentation meets Language Technology: The Next Step in the Description of Komi*, funded by Kone Foundation, Finland. Timofey Arkhangelskiy was supported by the Alexander von Humboldt foundation.

7. Bibliographical References

Adams, O., Cohn, T., Neubig, G., Cruz, H., Bird, S., and Michaud, A. (2018). Evaluating phonemic transcription

of low-resource tonal languages for language documentation. In *Proceedings of LREC 2018*.

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., and Weber, G. (2020). Common Voice. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*.

Arkhangelskiy, T. (2019). Corpora of social media in minority Uralic languages. In *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages*, pages 125–140. Tartu.

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., and Wang, L. (2013). How Noisy Social Media Text, How Diffrent Social Media Sources. In *International Joint Conference on Natural Language Processing*, pages 356–364. Nagoya, Japan.

Beznosikova, L. M., Ajbabina, E. A., Zaboieva, N. K., and Kosnyreva, R. I. (2012). *Komi sërnisikas kvyčykör*. Kola, Syktyvkar.

Blokland, R., Chuprov, V., Levchenko, D., Fedina, M., Fedina, M., Partanen, N., and Rießler, M. (2016-2020). Komi mediateka.

Blokland, R., Partanen, N., Rießler, M., and Wilbur, J. (2019). Using computational approaches to integrate endangered language legacy data into documentation corpora. In *Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 24–30.

Blokland, R., Fedina, M., Partanen, N., and Rießler, M. (2020). Spoken Komi Corpus. The Language Bank of Finland.

Fedina, M. S. and Levčenko, Dmitriy, A. (2017). Iz opyta sozdaniya komi mediateki. In Marina S. Fedina, editor, *Elektronnaja pismennost narodov Rossijskoj Federacii*, pages 220–227. Syktyvkar.

Foley, B., Arnold, J. T., Coto-Solano, R., Durantin, G., Ellison, T. M., van Esch, D., Heath, S., Kratochvil, F., Maxwell-Smith, Z., Nash, D., et al. (2018). Building speech recognition systems for language documentation. In *Proceedings of Spoken Language Technologies for Under-resourced Languages (SLTU 2018)*, pages 205–209.

- Gibson, M. (2016). Assessing digital vitality. In *Proceedings of the LREC 2016 Workshop, CCURL*, pages 46–51. Portorož.
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., and Ng, A. Y. (2014). Deep Speech.
- Heafield, K. (2011). Kenlm. In *Proceedings of the sixth workshop on statistical machine translation*, pages 187–197.
- Hjortnaes, N., Partanen, N., Rießler, M., and M. Tyers, F. (2020). 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*, pages 31–37. Association for Computational Linguistics, Vienna.
- Jauhiainen, H., Jauhiainen, T., and Lindén, K. (2019). Wanca in Korp. In *Data and humanities (RDHUM) 2019 Conference: Data, methods and tools*, page 21.
- Kornai, A. (2015). A new method of language vitality assessment. *Linguistic and Cultural Diversity in Cyberspace*, pages 132–138.
- Krylova, I., Orekhov, B., Stepanova, E., and Zaydelman, L. (2015). Languages of Russia. In *Russian Summer School in Information Retrieval*, pages 179–185.
- Levow, G.-A., Bender, E. M., Littell, P., Howell, K., Cheliah, S., Crowgey, J., Garrette, D., Good, J., Hargus, S., Inman, D., Maxwell, M., Tjalve, M., and Xia, F. (2017). STREAMLInED challenges: Aligning research interests with shared tasks. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 39–47, Honolulu, March. Association for Computational Linguistics.
- Meyer, J. (2019). *Multi-task and transfer learning in low-resource speech recognition*. Ph.D. thesis, University of Arizona.
- Michaud, A., Adams, O., Cox, C., and Guillaume, S. (2019). Phonetic lessons from automatic phonemic transcription. In *19th International Congress of Phonetic Sciences (CPhS XIX)*. Melbourne.
- Orekhov, B., Krylova, I., Popov, I., Stepanova, L., and Zaydelman, L. (2016). Russian minority languages on the web. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016*, pages 498–508.
- Rießler, M. (2013). Towards a digital infrastructure for Kildin Saami. In Erich Kasten et al., editors, *Sustaining indigenous knowledge*, pages 195–218. Kulturstiftung Sibirien.
- Wisniewski, G., Guillaume, S., and Michaud, A. (2020). Phonemic transcription of low-resource languages. In *Proceedings of Spoken Language Technologies for Under-resourced Languages (SLTU 2020)*.