Variation in Spoken North Sami Language

Kristiina Jokinen[1,3], Trung Ngo Trong[1,2] and Ville Hautamäki[2]

[1]Institute of Behavioural Sciences, University of Helsinki, Finland
[2]School of Computing, University of Eastern Finland, Finland
[3]Institute of Computer Science, University of Tartu, Estonia
Kristiina.Jokinen@helsinki.fi, trung@uef.fi, villeh@cs.joensuu.fi

Abstract

The paper sets to investigate the amount of variation between the North Sami speakers living in two different majority language contexts: Finnish, spoken in Finland, and Norwegian Bokmål, spoken in Norway. We hypothesize that the majority language is a significant factor in recognizing variation of the North Sami language. Although North Sami is the biggest of the nine currently spoken Sami languages and it has become a lingua franca among the Sami speakers, there are clear differences in the pronunciation of the North Sami spoken in Finland and Norway, so that the difference can be used to recognize which majority language region the speaker comes from. Using a corpus of spoken North Sami collected in locations in Finland and Norway, we experimented in classifying the speech samples into categories based on the two majority languages. We used the i-vector methodology to model both intra- and between-dialect variations, and achieved the average recognition of about 17.31% EER for classifying the Sami speech samples. The results support our hypothesis that the variation is due to the majority language, i.e. Finnish or Norwegian, spoken in the given context, rather than individual variation.

Index Terms: Dialect recognition, accent recognition, spoken interaction, under-resourced languages, i-vector system

1. Introduction

The North Sami language (*davvisámegiella*) has about 20,000 speakers living in the area that spans over the northern parts of three countries: Norway, Sweden and Finland. The speakers are at least bilingual and can also speak the majority language of the country they live in (i.e. Norwegian, Swedish, and Finnish). This has impact on the spoken North Sami language, and the speaker's country of origin can be fairly easily distinguished based on their speaking manner [1].

North Sami enjoys official recognition: it is an official language in six northernmost counties in Norway and legally recognized in Finland and Sweden so that the Sami speakers have the right to use their tongue in all official contexts. In Finland, the Sami language can be studied as the mother tongue through the educational system up to the college level, and the first students graduated from college in 1994. The Sami languages are studied at the University of Oulu (Giellakas Institute) and at the University of Helsinki, and in recent years about half a dozen PhD dissertations have also been written in the North Sami language e.g. [2, 3]. There is also a Sami language technology research centre Giellatekno at the Arctic University of Norway in Tromso.

In this paper, we study the North Sami spoken variation in Finland and in the Finnmark area in Norway. It is clear that the speakers of North Sami in Finland and in Norway can be distinguished from each other, and we assume that there is more variation among the speakers of North Sami who live in the different majority language locations in Norway and in Finland, than among the speakers who live in different location within the same majority language context. This variation cuts across the traditional dialect boundaries according to which North Sami as a whole belongs to the Western dialects of the Sami languages.

In terms of methodology, we approach the North Sami dialect characterization task as an automatic closed set *language identification* (LID) task [4]. The methods in language recognition are categorized in two main categories: *acoustic* [5, 6] and *phonotactic* [7, 8, 9] approaches. Acoustic approach uses segmental features computed from the utterance and modeled by statistical techniques, whereas the phonotactic approach models the phoneme occurrences. Recently, an acoustic modeling technique, called i-vector modeling of speech utterances [10] has been shown to either outperform or have similar performance than the phonotactic approaches [11]. In this paper we will use i-vectors as the statistical speech utterance models.

We recognize each dialect based on the trained statistical model, and the closest matching dialect model is selected for each test utterance. The research in the dialect characterization has focused on larger dialect groups, such as automatically classifying American and British English [12] or Arabic dialects [13]. In addition, recent NIST *language recognition evaluation* (LRE) campaigns have focused on characterizing dialects. However, recognition of regional dialects has received less attention. Some inital recognition experiments were performed by DeMarco and Cox [14] on fourteen regional British English dialects using the i-vector system. They reported the identification error rate of 32%. Later in [15], they improved the identification error to 19% using a very large-scale fusion system. Similar performance was obtained using i-vectors on characterizing regional Swiss French dialects [16]. Also Finnish regional dialects of the Satakunta region were studied using i-vector based techniques, with similar results [17].

Our goal is to verify the impact of the majority language on North Sami, and thus find evidence on the issues related to assimilation, and dialect variation. We use the i-vector based automatic dialect recognition system to validate our hypothesis that the variation in North Sami dialects is due to the majority language, i.e. Finnish or Norwegian, spoken in the given context, rather than individual variation.

2. The North Sami Language

The nine Sami languages belong to the Finno-Ugric language family, and they are spoken in the northern parts of Finland, Norway, and Sweden, and in the Kola Peninsula in Russia. Cur-

Figure 1: Map of Northern Europe, with the inset showing the data collection sites for this study.

rently there are about 30.000-40.000 Sami speakers, but some Sami languages have already disappeared and some have only a few speakers left [18, 19]. North Sami is the biggest and most widely spoken of the Sami languages, with about 20,000 Sami speakers. It has become a kind of lingua franca among the Sami speakers, used e.g. in news broadcasting.

In practice, however, all Sami speakers are at least bilingual: they also speak the majority language (Finnish, Norwegian, Swedish). Due to active and strong revitalization efforts, especially using the so called "language nests" and legislation, the languages have gained more interest and an official status in the Nordic countries. For example, the Inari Sami language has been systematically supported and the number of speakers has now grown from some tens of elderly people to more than 300 mother tongue speakers in about 20 years [20]. It should be noted, however, that there are more ethnic Sami people than there are speakers of the Sami languages, and the definition of an "ethnic Sami" has been under intense debate [**?**].

The Sami languages form a continuum without any sharp dialect and language boundaries, but the different Sami languages are not necessarily mutually intelligible. Traditionally the Sami languages have been divided into Eastern and Western dialects: e.g. North Sami belongs to the Western group, and the geographically close Inari Sami to the Eastern group. North Sami is itself divided further into three main dialects: Torne Sami, Finnmark Sami, and Sea Sami.

In this study, we focus only on North Sami variation. Our data is collected within the DigiSami project [21, 22] in the main Finnmark region where North Sami is spoken. Figure 1 shows the North Sami speaking area and the data collection sites used in this study.

Spoken language shows variation related to individual speakers. For instance, according to [1], there are dialect variations within Karasjok area in Finnmark (one of our data collection locations), which means that within one dialect area, such as the Finnmark Sami, one can find differences in the pronunciation, especially in diphthongs.

However, the Finnmark area cuts across the current state boarders of Finland and Norway, and we hypothesize that the differences in spoken North Sami language in Finnmark are bigger between the languages spoken in the two majority language areas (Finland and Norway) than between the various data collection sites in the Finnmark area general. In other words, the variation can be related to the different majority languages spoken in the area, and it shows the influence of Finnish and Norwegian on the spoken North Sami. It is interesting that although there exists a written standard language for North Sami (and for the other Sami languages as well), such variation can also cause uncertainty in the written form.

The differences in the Sami languages mainly concern morphophonetic variation, while syntactic changes are fairly small [23]. For instance, in the Eastern dialects the vowel system contains three low vowels /ä/ : /a/ : /â/ whereas Western dialects have only two /ä/ : /a/:

gálgat /kalkah/ : galgat /kâlkah/ 'must'
gápmagiid /kaapm(a)ij/ : cápmii /cääpmij/

Also diphthongs in the Eastern dialects become open:

miessi /miessii/ 'a fawn': geassi /kiässii/: /kiessii/ 'summer'
luohti /luohtii/ 'a yoik':goahti /kuähtii/: /kuohtii/ 'teepee, trad. dwelling'

In the Eastern dialects, /k/ in between vowels has disappeared, and /p/ has changed into /v/:

johka /joohka/ : joga /joo.a/ 'river'
lohpi /lohpii/ : lobi /loovii/ 'permission'

Diphtongs are one of the most complex issues in the North Sami phoneme system. There are four diphtongs (/eä/ as in leat 'to be', /ie/ as in giella 'language', /oa/ as in boahtit 'to come', /uo/ as in vuodjat 'to swim', but they vary between the individual speakers. E.g. in our data, one of the speakers in the Utsjoki sample pronounces the diphthong "uo" as /ye'/, although in general it would be pronounced /ue/. Another interesting observation is that the phoneme /ʉ/, which does not belong to the original set of Sami phonemes, now appears in new loanwords such as "universitehta" 'university' in the Norwegian side. In North Sami spoken in Finland, the word is pronounced with an ordinary /u/.

It is also interesting that unvoiced unaspirated stop tend to become voiced apparently due to the Norwegian influence, and also the consonant /r/ is pronounced "weaker" in Norway than in Finland. In Norway, the speakers also reduce vowels more than in Finland, especially /a/ and /ä/ in the second syllable of two-syllabic words, and in the fourth syllable in four-syllabic words.

3. Automatic Dialect Recognition

In this work, we represent an utterance using the fixed length and low-dimensional latent variable vector in the *total variability space* [10]. This is commonly called an i-vector, and it contains the variability in the utterance, such as dialect, speaker and the recording session. The *Gaussian mixture model* (GMM) supervector, $\mathbf{M}$, of an utterance is represented as [10],

$$\mathbf{M} = \mathbf{m} + \mathbf{T}\mathbf{w}, \tag{1}$$

where $\mathbf{m}$ is the utterance independent component (the universal background model or UBM supervector), $\mathbf{T}$ is a rectangular low rank matrix and $\mathbf{w}$ is an independent random vector of distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$. $\mathbf{T}$ represents the captured variabilities in the supervector space. It is estimated by the *expectation maximization* (EM) algorithm similar to estimating the $\mathbf{V}$ matrix in *joint factor analysis* (JFA) [24], with the exception that every training

utterance of a given model is treated as belonging to a different class. The extracted i-vector is then the mean of the posterior distribution of $\mathbf{w}$.

As the extracted i-vectors contain both intra- and inter-dialect variability, we use *heteroscedastic linear discriminant analysis* (HLDA) [25] to project the i-vectors onto a space where inter-dialect variability is maximized and intra-dialect variability is minimized. In standard HLDA technique, the vectors of size n are projected into subspace $p < n$, using HLDA matrix $\mathbf{A} \in \Re^{n \times n}$. *Within-class covariance normalization* (WCCN) is then used to compensate unwanted intra-class variations in the total variability space [25]. Given two i-vectors $\mathbf{w}_{\text{test}}$ and $\mathbf{w}^d_{\text{target}}$ for dialect d, cosine similarity score t is computed as follows:

$$t = \frac{\hat{\mathbf{w}}^T_{\text{test}} \hat{\mathbf{w}}^d_{\text{target}}}{\left\|\hat{\mathbf{w}}_{\text{test}}\right\| \left\|\hat{\mathbf{w}}^d_{\text{target}}\right\|} \quad (2)$$

where $\hat{\mathbf{w}}_{\text{test}}$ is,

$$\hat{\mathbf{w}}_{\text{test}} = \mathbf{A}^T \mathbf{w}_{\text{test}} \quad (3)$$

Further, $\hat{\mathbf{w}}^d_{target}$ is the average i-vector over all the training utterances in dialect d. This score a calculated for all target languages, and we identify the dialect by the one with highest degree of similarity. It is also notable that only the dialect labels involves in computing HLDA, hence, the system doesn't know which utterances belong to Norwegian or Finnish, and provides unbiased results concerned the effect of majority languages.

4. Evaluation setup

4.1. Data

Our data consists of North Sami speech samples collected in three villages in Finland: Utsjoki (Ohcejohka), Inari (Anár) and Ivalo (Avvil), and in two villages in Norway: Kautokeino (Guovdageaid) and Karasjoki (Kárášjohka). The locations were selected so that the main Sami speaking areas were represented with different North Sami dialects, and the main Sami central towns were also represented. The events were organized at high schools and in community halls and libraries, and the participants took part in three different tasks: Wikipedia article planning and writing in the Sami language, Wikipedia article reading aloud in the Sami language, and free conversation in the Sami language, see more of the DigiSami data and data collection in [21, 22]

There are 28 participants, 10 men and 18 women. Their ages range from 16 to 65 years: 17 were 16-21 years old, five 30-44 years old, and six 49-65 years old. All but one of the participants were native speakers of North Sami. One participant in Kautokeino had learnt North Sami as a second language, but he was a fluent speaker and used North Sami daily at work. The participants (or their parents if under-aged) were asked to sign a data usage agreement where they allowed the data collection and its use for research purposes.

All participants were bilingual, and spoke either Finnish (Utsjoki, Ivalo, Inari), or Norwegian (Kautokeino and Karasjoki). Most participants had lived their life in the Sami area, although not in the same town or village. Ten participants had also lived in bigger cities in the southern part of the area for a short period of time. Almost all (26) participants reported they use North Sami daily; one participant reported using North Sami weekly and one participant monthly.

It is interesting that most participants reported using North Sami in their main daily activities when communicating with family members (89%), teachers and co-workers (75%), but less than half (46%) used North Sami when communicating in other social situations in shops, offices, and restaurants. This may reflect the fact that the interlocutors in shops and offices are often majority language speakers. Another interesting observation is that 57% of the participants said they use North Sami with all the people they communicated with, i.e. with both family members and outsiders, while two participants reported they use North Sami when communicating with outsiders, not with family members.

Table 1: Number of speakers and utterances within each location in the Sami corpus.

Dialects	#Speakers	#utterances
Kautokeino	4	123
Karasjoki	6	82
Ivalo	6	83
Utsjoki	5	126
Inari	4	87
Total	**25**	**501**

For the experiments we selected read speech from 25 participants to the data collection. Table 1 shows the number of speakers and utterances within each dialect. For our purposes, all the audio files were partitioned into wave files of 30 seconds chunks, and downsampled to 8 kHz sampling rate.

4.2. Evaluation measurement

Results are reported in terms of average detection cost (C_{avg}) [4] and average equal error rate (EER_{avg}). EER indicates the operating point where false alarm and miss alarm probabilities are equal. C_{DET} is defined as [4],

$$\begin{aligned} C_{\text{DET}} &= C_{\text{miss}} P_{\text{tar}} P_{\text{miss}}(L_a) \\ &+ C_{\text{fa}}(1 - P_{\text{tar}}) \frac{1}{J-1} \sum_{k \neq j} P_{\text{fa}}(L_j, L_k) \end{aligned} \quad (4)$$

where P_{miss} denotes the miss probability (or false rejection rate), i.e., a test segment of dialect L_i is rejected as being in that dialect. $P_{\text{fa}}(L_i, L_k)$ is the probability when a test segment of dialect L_k is accepted as being in dialect L_i. The costs, C_{miss} and C_{fa} are both set to 1 and P_{tar}, the prior probability of a target accent, is set to 0.5 as in [4].

5. Results

To compensate for the lack of sufficient data in training and evaluation, we first trained the UBM and the T-matrix using the Finnish language PERSO corpus [26] corresponding to 30916 utterances in total. This corpus is similar to the DigiSami corpus with respect to the way the data is collected and structured. In order to reduce the factor created by speaker variation during evaluation, we used the leave-one-speaker-out (LOSO) strategy, already used in [25]. For each target accent, we randomly select a speaker, and his/her utterances are left-out which form an independent evaluation set equally combined of 5 dialects. The remaining utterances are used for training the target models. The HLDA dimensions were re-trained in each LOSO iteration from the training portion of the utterances. We repeat the experiments six times with different set of excluded speakers, so all users from each dialect are used in both training and evaluation set.

Table 2 shows the performance for each individual North Sami dialect. Generally, Norwegian dialects (i.e. Kautokeino

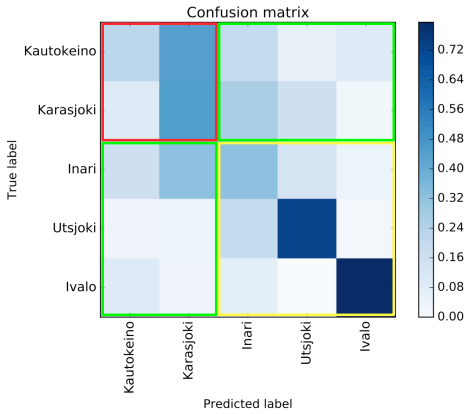

Figure 2: Confusion matrix of identified utterances.

and Karasjoki) attain lower recognition accuracy with average EER of 20.11%. Conversely, the system achieves decent performance on Finnish Sami with expected EER of 15.43%. Figure 2 shows the confusion matrix in which we treat the problem as dialect identification task. The red square region represents confusion between identifying Norwegian Sami dialects (Kautokeino and Karasjoki), while the yellow square region is the confusion among Finnish Sami dialects (Inari, Utsjoki, Ivalo), and green region indicates mis-identification rate between the two main dialects. The matrix emphasized strong confusion in identifying Kautokeino and Karasjoki, which may be caused by their close regional relation. Additionally, there is mixed confusion between Norwegian Sami and Finnish Sami represented by the amount of Karasjoki and Kautokeino utterances which are misclassified as Inari ones. This can be interpreted as a transition zone between the two major dialects. However, the mis-identification rates are significantly lower for Utsjoki and Ivalo which highlights the clear border between the spoken dialects in Norway (Kautokeino and Karasjoki) and those spoken in Finland (Ivalo, Utsjoki, Inari). Moreover, the better performance in Ivalo and Utsjoki indicates their closed relation to Finnish, since we used PERSO, Finnish corpus, for training the UBM, the results are biased toward enhancing the discrimination of Finnish North Sami. This supports the view that the majority language has an impact on the dialect among bilingual speakers. This is especially pronounced in the case of North Sami, since the majority languages Norwegian and Finnish are linguistically quite different from each other.

Table 2: Recognition accuracy per each dialect in the Sami corpus. Results are reported in terms of EER (%) and C_{DET}.

Dialect	EER (%)	C_{DET} x100
Kautokeino	24.43	26.00
Karasjoki	15.79	15.74
Inari	28.26	23.95
Utsjoki	10.37	13.54
Ivalo	7.66	9.72
Average	17.31	17.79

In particular, the results from Utsjoki and Ivalo indicate that the speech of the subjects is fairly uniform and adapted to the speech of the other participants of that area. On the other hand, the confusions between Karasjoki and Inari as well as Kautokeino and Karasjoki samples show that the speakers do not form such a uniform group. In fact, the confusion between Kautokeino and Karasjoki samples may be due to the fact that both locations are in Norway, and North Sami is spoken in a more standardized manner in Norway than in Finland (cf. Utsjoki and Ivalo). The confusion between Karasjoki and Inari samples is interesting since the participants have wide age range and the locations are in different countries. It may simply reflect more active contacts between the speakers across the border, thus giving rise to assimilation of their spoken language in mutually understandable manner. It would be interesting to study assimilation of the participants' speech in interactive situations more, and compare the current data (read speech) with the speech recorded in conversational contexts.

Table 3: Confusion matrix on the Sami dialect recognition task.

		Predicted label	
		Norwegian	Finnish
True label	Norwegian	166	101
	Finnish	86	283

Furthermore, by treating the problem as a two-class identification task, Table 3 shows that the identification error of Finnish North Sami utterances is about 30%, while 60.8% of Norwegian North Sami utterances were misclassified. Cohen's kappa of the confusion table is 0.3917 with z-value of 9.8433 indicating significance. These results in Figure 2, Table 2 and Table 3 are consistent in that detecting Norwegian Sami dialects is more difficult than detecting Finnish North Sami dialects.

6. Conclusion

We studied how well the North Sami speech samples can be distinguished from each other using dialect recognition techniques. We achieved average recognition EER of 17.31%, with the main confusions taken place between North Sami spoken in Norway vs. Finland. Since the phonetic variation is observable and influenced by geographic differences and culture exchange, our results on acoustic analysis support the original hypothesis that the difference between the North Sami dialects depends on the majority language (Norwegian vs. Finnish) used by North Sami speakers.

New ways to communicate and mediate information, such as TV, radio, and recently appeared social media, have and will have a large impact on the traditional linguistic scenery by shifting language use towards "globalized" language use (see debate in [27]). It is likely that social media has a bigger impact on small language communities than on larger ones, but it is an empirical question to what extent, and which language dimensions will be preserved by new media. In the future, more data can be collected and analyzed for studying the detail effect of majority language on North Sami speakers from both acoustic or linguistic point of view.

7. Acknowledgments

This project was supported by Academy of Finland project *Fenno-Ugric Digital Citizens*(grant nro. 270082), and partially supported by Academy of Finland project *Luotettava puhujantunnistus ja puhujamuokkaus* (Reliable speaker recognition and speaker modification, grant nro. 253120) and Estonian Science Foundation project IUT 20-56. We would also like to thank all the participants in the data collection.

8. References

[1] P. Sammallahti, *The Sami languages*. Kárásjohka : Davvi girji, 1998.

[2] I. Seurujärvi, "Ale jaskkot eatnigiella : Alkuperäiskansaliikkeen ja saamen kielen merkitys saamelaisten identiteetille," Ph.D. dissertation, University of Helsinki, 2012.

[3] J. Johansen Ijäs, "Davvisámegiela finihtta vearbahámiid sojahanvuogádaga oččodeapmi vuollel golmmajahkásaš máná gielas, sámi allaskuvlla," Ph.D. dissertation, Giellagas Institute, University of Oulu, 2011.

[4] H. Li, K. A. Lee, and B. Ma, "Spoken language recognition: From fundamentals to practice," *Proceedigns of the IEEE*, vol. 101, no. 5, pp. 1136–1159, 2013.

[5] N. F. Chen, W. Shen, and J. P. Campbell, "A linguistically-informative approach to dialect recognition using dialect-discriminating context-dependent phonetic models," in *ICASSP*, 2010, pp. 5014–5017.

[6] F. Biadsy, J. Hirschberg, and M. Collins, "Dialect recognition using a phone-GMM-supervector-based SVM kernel." in *Interspeech*, 2010, pp. 753–756.

[7] N. F. Chen, W. Shen, J. P. Campbell, and P. A. Torres-Carrasquillo, "Informative dialect recognition using context-dependent pronunciation modelling." in *ICASSP*, 2011, pp. 4396–4399.

[8] F. Biadsy, H. Soltau, L. Mangu, J. Navratil, and J. Hirschberg, "Discriminative phonotactics for dialect recognition using context-dependent phone classifiers," in *Speaker Odyssey*, 2010.

[9] S. Sinha, A. Jain, and S. S. Agrawal, "Speech processing for Hindi dialect recognition," *Advances in Intelligent Systems and Computing*, vol. 264, pp. 161–169, 2014.

[10] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[11] E. Singer, P. Torres-Carrasquillo, D. Reynolds, A. McCree, F. Richardson, N. Dehak, and D. Sturim, "The MITLL NIST LRE 2011 language recognition system," in *Speaker Odyssey*, Singapore, 2012, pp. 209–215.

[12] N. F. Chen, S. W. Tam, W. Shen, and J. Campbell, "Characterizing phonetic transformations and acoustic differences across english dialects," *IEEE Transactions on Audio, Speech, and Language Processing*, 2014.

[13] F. Biadsy, "Automatic dialect and accent recognition and its application to speech recognition," Ph.D. dissertation, Columbia University, 2011.

[14] A. DeMarco and S. J. Cox, "Iterative classification of regional British accents in i-vector space," in *MLSLP*, 2012, pp. 1–4.

[15] A. DeMarco and S. J. Cox, "Native accent classification via i-vectors and speaker compensation fusion," in *Interspeech*, 2013, pp. 1472–1476.

[16] A. Lazaridis, E. Khoury, J.-P. Goldman, M. Avanzi, S. Marcel, and P. N. Garner, "Swiss French regional accent identification," in *Speaker Odyssey*, Joensuu, Finland, 2014.

[17] H. Behravan, V. Hautamäki, S. M. Siniscalchi, E. Khoury, T. Kurki, T. Kinnunen, and C.-H. Lee, "Dialect levelling in Finnish: a universal speech attribute approach," in *Interspeech*, 2014, pp. 2165–2169.

[18] U.-M. Kulonen, I. Seurujärvi-Kari, and R. Pulkkinen, *The Sámi: A cultural encyclopaedia*. Suomalaisen Kirjallisuuden Seura, Helsinki, Finland, 2005.

[19] V.-P. Lehtola, *The Sami People: Traditions in transition*. University of Alaska Press, 2004.

[20] M.-L. Olthuis, "Uhanalaisen kielen elvytys: esimerkkinä Inarinsaame ("revitalisation of an endangered language: an example of the Inari Sami language")," *Virittäjä*, vol. 107, pp. 568–579, 2003.

[21] K. Jokinen, "Open-domain interaction and online content in the sami language," *Proceedings of the Language Resources and Evaluation Conference (LREC-2014)*, June 2014, Reykjavik, Iceland.

[22] K. Jokinen and G. Wilcock, "Community-based resource building and data collection," *Proceedings of the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU'14)*, pp. 201–206, May 2014, St Petersburg, Russia.

[23] M. Palismaa and I. M. G. Eira, *Gielas gillii, mielas millii 9, Davvisámegiela suopmanat*. Davvi girji os, 2001.

[24] D. Matrouf, N. Scheffer, B. G. B. Fauve, and J.-F. Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification," in *Interspeech*, 2007, pp. 1242–1245.

[25] H. Behravan, V. Hautamäki, and T. Kinnunen, "Factors affecting i-vector based foreign accent recognition: A case study in spoken Finnish," *Speech Communication*, vol. 66, pp. 118–129, 2015.

[26] Perso speech database. [Online]. Available: http://islrn.org/resources/651-108-565-673-5/

[27] P. J. Trudgill, "Diffusion, drift, and the irrelevance of media influence," *Journal of Sociolinguistics*, vol. 18, no. 2, pp. 213–222, 2014.