



Mapping areal variation and majority language influence in North Sámi using hierarchical prosodic analysis

Katri Hiovain, Antti Suni, Martti Vainio, Juraj Šimko

University of Helsinki, Finland

firstname.secondname@helsinki.fi

Abstract

We present results of a statistical hierarchical analysis of areal variation in prosody of spoken North Sámi languages. The hierarchical analysis method compares unigram models using cross-entropy measure. The models depict distributions of Δ -features of f_0 and energy signals decomposed using Continuous Wavelet Transform. These signals are obtained from speech recordings of five areal North Sámi varieties recorded in sites in northern Finland and Norway.

We evaluate three potential sources of areal variation in prosodic characteristics of these five areal varieties: (1) traditional dialectal analysis of North Sámi, (2) influence of the relevant majority languages, and (3) geographical distance. Our results show a significant positive correlation between cross-entropy distances between models and geographical distances between recording sites, demonstrating a viability of the method for typological analysis. Prosodic characteristics of the areal varieties are also influenced by majority languages, and, to a smaller degree, by differences between the North Sámi dialectal varieties.

Index Terms: language comparison, prosodic typology, wavelet transform, statistical modelling, minority languages, North Sámi language, language contacts, language change

1. Introduction

North Sámi is one of the nine currently spoken Sámi languages of the Uralic language family. It belongs to the Western group of the Sámi languages and is spoken in the Northern parts of Finland, Sweden and Norway [1]. The language has the highest number of speakers (approx. 25,000 [2]) among Sámi languages and is an official language in six northernmost counties in Norway and legally recognized in Sweden and Finland [3]. The North Sámi language is divided into three main dialect groups: Torne Sámi, Finnmark Sámi and Sea Sámi [4]. All language material used in this paper comes from the Finnmark dialect group that is traditionally further subdivided to Western and Eastern dialect groups.

The main focus of this work is the areal variation of the Finnmark North Sámi language and the influence of majority languages – namely Finnish and Norwegian – on spoken Sámi and its prosody. This influence is continually reinforced by bilinguality of speakers (country’s majority language in addition to Sámi), migration of Sámi people to other regions of their respective countries and the growing effects of digitalization and mass-media. As a result, a rapid change in multiple language characteristics impedes mutual intelligibility between North Sámi speakers and presents new challenges to the traditional dialectological analysis [2].

In terms of linguistic typology, there are several important differences between Norwegian, Finnish and North Sámi [5, 6, 7]. Finnish and North Sámi are Finno-Ugric languages of the

Uralic family known for the agglutinative features and rich case systems. Norwegian is a North Germanic language belonging to the Indo-European family whose languages are generally more fusional with less complex case marking.

The languages also differ in prosodic characteristics such as prevailing rhythmic and intonation patterns. For example, in most Norwegian varieties, there are two lexically distinctive pitch accents, which gives Norwegian a distinguishable “singing” quality [5]. Also, in contrast to both Finnish and North Sámi, the word intonation in Norwegian is generally rising [6]. The Finnish intonation patterns, on the other hand, are relatively stable and more uniform, with falling word intonation in general [7]. All three languages have predominantly word-initial lexical stress pattern, with some exceptions in Norwegian due to unstressed prefixes and foreign loans.

The prosodic and typological features of the majority languages are influencing the North Sámi areal varieties. Native North Sámi speakers can easily recognize whether their interlocutor is a Norwegian or Finnish Sámi. In addition to prosody, the traditional dialectal differences between Western and Eastern varieties of the Finnmark North Sámi are manifested through variation in phonological features and phonemic variation also recognizable in speech [4].

Comparing prosodic characteristics of languages or language varieties needs to combine intonational and rhythmic properties, stress patterns, and suprasegmental features (presence and realization of quantity contrast, tonality, etc.). The lack of a prosodic transcription system applicable across multiple languages and language families (as in our case) further complicates this task [8].

In this paper we present results obtained using purely data-driven approach introduced in our earlier work [9]. We assess mutual similarity among North Sámi varieties by comparing statistical distributions of prosodic patterns – unigram language models – obtained from hierarchically decomposed f_0 and energy envelope signals.

We analyze speech material from DigiSámi read speech corpus [10, 11] containing recording of a number of North Sámi speakers from five sites in Finland and Norway, representing both Finnmark North Sámi dialectal groups (with a different distribution compared to majority language influence).

The same corpus was analyzed in an earlier work using similarly motivated established language recognition techniques [12]. This previous work used an i -vector representation of speech signal, that combines prosodic (energy) and spectral (MFCC) characteristics of speech signal [13]. In contrast, in our work, the distances among different areal varieties (in terms of their similarities and differences) arise solely from “prosodic” features of energy and f_0 ; our method completely disregards other information such as spectral properties of signal.

Our aim is thus to evaluate and compare three potential sources of areal variation in purely prosodic characteristics of

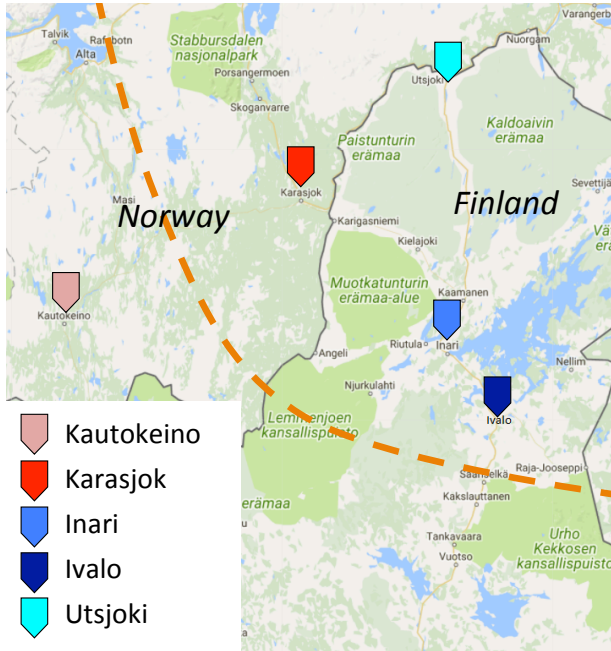


Figure 1: *The North Sámi varieties. Map compiled using Google Maps. The dashed curve separates the Westerns and Eastern dialectal variety areas (see [6]).*

the North Sámi speech material. The first is the traditional divide into two dialectal varieties, Eastern and Western North Sámi. The second is the influence of majority languages, namely Finnish and Norwegian. Finally, the third potential source of differences among varieties is the geographical distance and associated intensity of mutual language contact.

2. Methodology

2.1. Material

Table 1: *Number of speakers and the overall duration of recordings for each variety.*

Locations / varieties	Speakers (female)	Minutes
Kautokeino (skt)	4 (2)	62:07
Karasjok (skr)	6 (5)	43:02
Ivalo (siv)	6 (5)	43:29
Utsjoki (suf)	5 (1)	64:29
Inari (sin)	4 (3)	43:54

We have analyzed material from the DigiSami read speech corpus [10, 11] containing recordings from five locations in Northern Finland and Norway, representing five potential regional varieties. The locations are Kautokeino and Karasjok in Norway, and Utsjoki, Ivalo and Inari in Finland (see the map in Fig. 1). Altogether, the corpus contains material from 25 North Sámi speakers, with ages range between 16 and 65 years. Table 1 provides additional information about the corpus material.

All speakers were bilingual: speakers from Kautokeino and Karasjok were native also in Norwegian, while those from Utsjoki, Ivalo and Inari were bilingual in Finnish and North Sámi. Although the North Sámi does not have a spoken standard, it has a standard written form. This allowed to present all partic-

ipants with the same reading material, Wikipedia articles about the Sámi languages and dialects, traditional Sámi costumes and snowmobiles. The DigiSami data was previously manually annotated and segmented on sentence level. These annotations were used to divide the recordings to individual sentences.

2.1.1. Dialectal varieties and majority languages

As mentioned above, our recordings contain material from two countries. The Kautokeino and Karasjoki varieties are spoken in Norway and have presumably been under Norwegian influence while Ivalo, Inari and Utsjoki varieties are spoken in Finland and can be expected to manifest Finnish influence in their prosodic characteristics.

The DigiSami corpus also allows to evaluate additional influence traditional dialectal variety on prosody. Only the Kautokeino (**skt**) variety of these areas belong to the Western dialect area according to the traditional North Sámi dialect analysis while the rest of the varieties belong to the Eastern dialect group [1]. The Western and Eastern North Sámi dialects differ in terms of phonemic and phonetic variation as well as suprasegmental features, especially relevant for this study. For example, in the Eastern dialects (all except **skt** in the data used for this experiment), the first syllable short vowel is lengthened: *amas* 'strange' : [a:mas]. In contrast, in the Western dialect, the second syllable short vowel is lengthened: *amas* : [ama:s] (the vowel length distinction is not marked orthographically in North Sámi) [4].

2.2. Prosodic analysis and language model comparison

The methodology used for hierarchical prosodic analysis has been introduced in [9].

For each sentence in the corpus, the f_0 -contour and signal envelope (energy) contours (sampling rate 100 Hz) were extracted using a standard Praat pitch extraction routine and a wavelet-based filtering technique. The unvoiced intervals in the f_0 -contours were subsequently (linearly) interpolated and the resulting contour was smoothed (10 Hz bandwidth). The non-interpolated f_0 -contours served as voicing signals.

Unigram models were trained for each language variety using the following procedure: The f_0 and energy signals were decomposed using continuous wavelet transform (Morlet mother wavelet, $\omega_0 = 2$) to three components with pseudo-frequencies of 200 and 400 and 800 ms. A derivative (Δ -feature) was calculated for each of the resulting six components. Subsequently, these derivatives were discretized: for each speaker and each signal, the derivative values (between 5th and 95th percentile) were divided to three equally sized bins (corresponding to falling, steady and rising signal). The values below 5th and above 95th percentiles were marked as inadmissible for subsequent state calculation. These discretized signals were then combined, each admissible time point occurring within a voiced portion to speech signal was assigned one of the possible $N = 3^6 = 729$ combined states. Finally, for each language variety, a likelihood of occurrence of each particular state in all utterances of the given variety in the corpus was calculated.

The variety models were compared using cross-entropy, depicting a distance between the varieties. For two varieties represented by unigram models $(p_i)_{i=1,\dots,N}$ and $(q_i)_{i=1,\dots,N}$, where p_i , q_i are probabilities of state i in the two varieties, cross-entropy is defined as

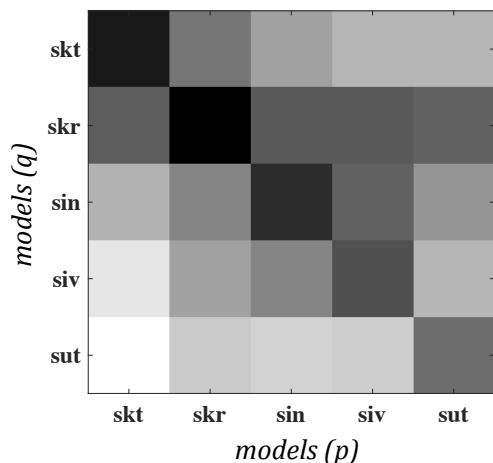


Figure 2: Confusion matrix showing cross-entropy between the unigram models (see the text). Darker cells depict lower cross-entropy (smaller distance between varieties), the brighter ones higher cross-entropy. The cross-entropy can be also interpreted as average perplexity of models (in columns) with speech material (in rows).

$$H(q, p) = - \sum_{p_i > 0} q_i \log(p_i).$$

In several cases, some of the states did not occur in the speech material for some varieties; these occurrences for (p_i) model were excluded from the sum as they would lead to an infinite cross-entropy.

This measure captures a loss of efficiency when encoding events characterized by distribution (q_i) by an encoding scheme optimized for distribution (p_i). Also, it is mathematically equivalent to an average perplexity (minus log likelihood) of the model (p_i) with a collection of states representing speech material from the variety modelled by (q_i)¹.

3. Results

The distance between the North Sámi varieties (cross-entropy) is shown in a confusion matrix in Fig. 2, with darker cells depicting lower distance/cross-entropy. The range of the cross-entropy values depicted in the matrix is rather small (5.932–6.105), reflecting the mutual similarity among the varieties and the analyzed material.

We calculated the mean cross-entropy values for each analyzed variety as an average of cross-entropy values in the row and column belonging to the variety without the entropy value on the diagonal. These “average distances from the rest” are 6.050 (**skt**), 6.015 (**skr**), 6.029 (**sin**), 6.040 (**siv**) and 6.057 (**sut**).

Cross-entropy between the two Norwegian varieties, Kautokeino (**skt**) and Karasjok (**skr**) is generally smaller than cross-entropy between these two varieties and varieties spoken in Finland, Inari (**sin**), Ivalo (**siv**) and Utsjoki (**sut**). This is particularly true for Kautokeino (**skt**) variety representing the Western North Sámi dialect; in fact, according to our measure, Karasjok model is slightly less perplexed with Inari and Ivalo material than with Kautokeino one (5.994, 5.992 vs. 5.995, second row in the confusion matrix in Fig. 2).

¹The cross-entropy is thus a generalization of the average perplexity measure used in [9].

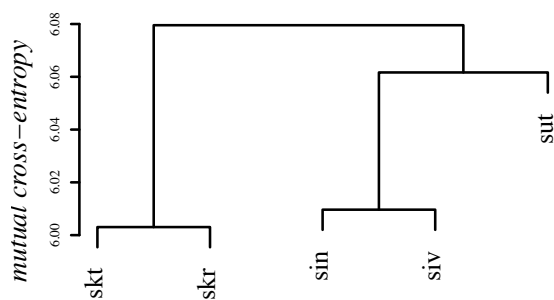


Figure 3: Clustering of the varieties using mutual cross-entropy.

Compared to Kautokeino, Karasjok variety shows greater similarity to the varieties spoken in Finland. The cross-entropy between the Finnish varieties of Inari (**sin**) and Ivalo (**siv**), 6.021 and 5.998, respectively, is lower compared to cross-entropy between these two varieties and the third Finnish variety of Utsjoki (**sut**), ranging between 6.032 and 6.073.

According to our measure, the Utsjoki variety is somewhat more similar to Karasjok one than to the other two varieties spoken in Finland, Inari and Ivalo.

These rather complex patterns are summarized in the dendrogram shown in Fig. 3. For this clustering we calculated mutual cross-entropy between pairs of varieties. For each pair represented by models p and q , the mutual cross-entropy is a mean of cross-entropy $H(q, p)$ and cross-entropy $H(p, q)$. The dendrogram was then plotted using hierarchical clustering method with the mutual cross-entropy playing the role of distance (the entropies $H(p, p)$ of each model were ignored in this process).

As seen in Fig. 3, the main split occurs between the varieties spoken in Norway (**skt** and **skr**), on the one hand, and the varieties spoken in Finland (**sin**, **siv** and **sut**), on the other. The “Finnish” branch subsequently splits between Utsjoki variety and Ivalo–Inari cluster. The distance (mutual cross-entropy) between the varieties within the Ivalo–Inari cluster is considerably smaller than the distance between the cluster and the Utsjoki model; in fact this latter distance is greater than that between the two main branches in the dendrogram.

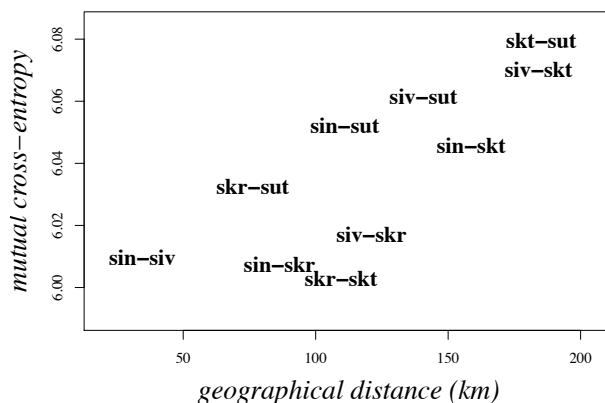


Figure 4: Clustering of the languages and varieties using mutual cross-entropy.

Finally, we also compared the similarity between the varieties as captured by cross-entropy, with geographical distances between the sites (major settlements) where these varieties are spoken. The geographical distance between sites was calculated using distance measuring tool in Google Maps. The relationship between these two distance measures is plotted in Fig. 4.

The cross-entropy positively and significantly correlates with geographical distances between the sites, the correlation equals 0.78 ($p = 0.007$).

4. Discussion

We presented here a novel method of hierarchical statistic analysis of prosody using wavelet decomposition of f_0 and intensity signals, and comparison – in terms of cross-entropy – of unigram models of Δ -features. When applied on dialectal material of North Sámi corpus, the method reveals several complementary sources of areal variation.

The mutual distances (cross-entropy) among the language models correlate positively and significantly with distances among the sites where the speech material was recorded. This finding is compatible with the standard default assumption in typology of causal links between language change, language contact and geographical distance. Perhaps even more importantly, it demonstrates that our automatic approach is suitable for investigating typological questions regarding prosodic variation, at least in closely related languages/varieties. This agreement with the “common sense” assumption also indicates that statistical distributions of hierarchically organized f_0 and energy patterns can indeed be used as a rough, and useful, representation of prosodic characteristics of speech.

The other two sources of areal variation hypothesized in Introduction is the traditional divide of our material into two dialectal varieties, Eastern and Western North Sámi, and contemporary influence of majority languages. Both of these influences are discernible in our results, albeit to a different degree.

The average distance of the sole representative of the Western dialectal variation, Kautokeino (**skt**), from all other varieties is relatively high but not as high as that of Utsjoki (**sut**). Also, the confusion matrix in Fig. 2 shows high cross-entropy values between the model of Kautokeino variety and other models except that of “fellow Norwegian” variety of Karasjok. The relationship between variety distances and geographical distances (Fig. 4), however, shows that while the cross-entropy distances for Utsjoki (**sut**) are relatively greater than would be predicted by the general trend of geographical influence (see the almost separate fit for **sut** discernible in the top part of the figure), this is not the case for Kautokeino.

The variety clustering derived from the confusion matrix separates Norwegian varieties from Finnish ones. This yield some support to the hypothesis about majority language influence on prosodic features of the North Sámi varieties (cf. [2]). In this respect, our results are thus compatible with those of an earlier analysis of the same corpus presented in [12] but strengthen them in two respects. First, unlike in the earlier analysis, the separation of varieties based on majority language presently emerges without any prior division of varieties into majority-language classes. Second, equally importantly, in the present work, the clustering was obtained using purely prosodic characteristics of speech.

Taken together, these findings lead us to a tentative conclusion that the majority language has a discernible influence on prosody of the analyzed North Sámi varieties. Moreover, the consequences of this influence might be manifested more strongly than those of traditional division into Western and Eastern dialectal varieties as the Kautokeino variety is not clearly separated from other varieties according to our measure.

In general, our results show that our method is sensitive enough and can be used for detecting rather small prosodic differences between closely related dialectal varieties. As it does

not require any extensive labeling or annotating the method can also be used for under-resourced languages with relatively limited speech material.

Also, the method, in principle, allows to explicitly evaluate an influence of languages in contact (e.g., the majority languages) on the investigated varieties or dialects by simply including these languages and calculating appropriate cross-entropy distances. We have actually tried this approach, but the results were difficult to interpret, most likely because the majority language speech material was somewhat stylistically different compared to the DigiSami corpus.

In our previous work where we introduced the hierarchical prosodic analysis methodology ([9]), we discussed in some detail several potential issues. Namely, rather limited size of the corpus used for analysis, and lack of ground truth that can be used to verify the results. In the present paper we mitigate these potential shortcomings to some extent. The corpus used here is considerably larger than in [9] (despite the fact that overall speech material available for the Sámi languages is rather limited). Also, we directly evaluate relatively robust and well grounded hypotheses concerning the influence of majority language, geographical distance and dialectal variety.

Another limitation with our approach is that, in current form, it allows us to see overall *consequences* of prosodic variation, e.g., in the form of variety clustering, but does not provide any information as to precisely *what* prosodic patterns might be responsible for the differences between varieties and to what extent. In principle, however, cross-entropy measure does allow for identification of particular states that contribute most to the cross-entropy value. The next natural step in developing this new method for prosodic typology will thus be a design of a semiautomatic tool for identification the prosodic patterns that are primarily responsible for the emerging variety clustering.

5. Acknowledgements

This work was partly funded by the Academy of Finland DLT project (No. 12933481). We also thank Kristiina Jokinen and the DigiSami project for the permission to use their speech data.

6. References

- [1] P. Sammallahti, *The Saami languages: an introduction*. Davvi girji, 1998.
- [2] A. Aikio, L. Arola, and N. Kunnas, "Variation in North Saami," *Globalising sociolinguistics: Challenging and expanding theory*, pp. 243–255, 2015.
- [3] I. Seurujärvi-Kari, "Ale jaskkot eatnigiella. alku-peräiskansaliikkeen ja saamen kielen merkitys saamelaisten identiteetille," *Virtittäjä*, vol. 116, no. 1, 2012.
- [4] M. Palismaa and I. M. G. Eira, "Gielas gillii, mielas millii 9-davvisámegiela suopmanat (from language to language, from mind to mind 9the dialects of North Sami)," *Davvi Girji, Kárášjohka*, 2001.
- [5] E. Sivertsen, *Fonologi: fonetikk og fonemikk for språkstudenter*. Universitetsforlaget, 1967.
- [6] K.-P. Nickel and P. Sammallahti, *Nordsamisk grammatikk*. Davvi girji, 2011.
- [7] F. Karlsson, *Finnish: An essential grammar*. Routledge, 2008.
- [8] D. Hirst and A. Di Cristo, *Intonation systems: a survey of twenty languages*. Cambridge University Press, 1998.
- [9] J. Šimko, A. Suni, K. Hiovain, and M. Vainio, "Comparing languages using hierarchical prosodic analysis," in *Proceedings of Interspeech 2017*, 2017.
- [10] K. Jokinen, "Open-domain interaction and online content in the sami language." in *LREC*, 2014, pp. 517–522.
- [11] K. Jokinen and G. Wilcock, "Community-based resource building and data collection," in *Spoken Language Technologies for Under-Resourced Languages*, 2014.
- [12] K. Jokinen, T. N. Trong, and V. Hautamäki, "Variation in spoken North Sami language." in *INTERSPEECH*, 2016, pp. 3299–3303.
- [13] N. Dehak, P. J. 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.