

Towards SamiTalk: a Sami-speaking Robot linked to Sami Wikipedia

Graham Wilcock, Niklas Laxström, Juho Leinonen, Peter Smit, Mikko Kurimo and Kristiina Jokinen

Abstract We describe our work towards developing SamiTalk, a robot application for the North Sami language. With SamiTalk, users will hold spoken dialogues with a humanoid robot that speaks and recognizes North Sami. The robot will access information from the Sami Wikipedia, talk about requested topics using the Wikipedia texts, and make smooth topic shifts to related topics using the Wikipedia links. SamiTalk will be based on the existing WikiTalk system for Wikipedia-based spoken dialogues, with newly developed speech components for North Sami.

1 Introduction

In this paper we describe our work towards developing SamiTalk, a robot application for the North Sami language. This robot application is chosen because it is an interface to collaboratively edited Wikipedia information, and as a novel application, it is expected to increase the visibility of the language as well as interest in it. In particular, it is expected that young people may become more interested in using the language, which is regarded as an important and effective strategy for language revival in general. The motivation for creating robot applications to support revitalisation of endangered languages is discussed in more detail in [2].

SamiTalk will be the first robot application in the North Sami language, enabling users to hold spoken dialogues with a humanoid robot that speaks and recognizes North Sami. The robot will access information from the Sami Wikipedia, will talk about requested topics using the Wikipedia texts, and will make smooth topic shifts to related topics using the Wikipedia links. SamiTalk will be based on the existing

Graham Wilcock, Niklas Laxström, Kristiina Jokinen
University of Helsinki, Helsinki, Finland. e-mail: `firstname.lastname@helsinki.fi`

Juho Leinonen, Peter Smit, Mikko Kurimo
Aalto University, Espoo, Finland. e-mail: `firstname.lastname@aalto.fi`

WikiTalk system for Wikipedia-based spoken dialogues (see Section 2), with newly developed speech components for North Sami.

The paper is structured as follows. Section 2 summarizes the existing WikiTalk system and explains the differences between a localisation of WikiTalk for North Sami and previous localisations. Section 3 describes the development of the speech synthesizer and speech recognizer for North Sami that will be essential components of SamiTalk. Section 4 gives an example of the style of interaction for getting Wikipedia information with Samitalk. Section 5 indicates plans for future work.

2 WikiTalk and SamiTalk

The WikiTalk system [11, 3] accesses Wikipedia directly online. Using paragraphs and sentences from Wikipedia, the system can talk about whatever topics are of interest to the user. There are other applications for reading out Wikipedia articles, but simply reading out an article is a monologue rather than a dialogue. The key feature that enables WikiTalk to manage dialogues rather than monologues is its ability to handle smooth topic shifts, as described by [11]. Hyperlinks in the Wikipedia text are extracted, to be used as potential topic shifts. The system predicts that the user will often want to shift the topic to one of the extracted links.

The main challenge for dialogue modelling in WikiTalk is to present information in a way that makes the structure of the articles clear, and to distinguish between two conditions: the user shows interest and wants the system to continue on the current topic, or the user is not interested in the topic and the system should stop or find some other topic to talk about [3].

WikiTalk has been internationalised, and localised versions are now available for English, Finnish and Japanese. A number of major issues in internationalisation and localisation of spoken dialogue systems are discussed by [4], using WikiTalk as an example. In the case of WikiTalk, each new language version requires speech recognition and speech synthesis components for the language, an available Wikipedia in the language, and a localisation of the WikiTalk multimodal user interface.

Due to the internationalisation of WikiTalk, new language localisations can generally be produced relatively rapidly. However, a localisation of WikiTalk for North Sami involves more than previous localisations of WikiTalk. Up to now, the speech recognition and speech synthesis components for the languages supported by WikiTalk have been the components provided by the Nao robot. In order to use these languages, the WikiTalk system simply needs to check that the required language is installed on the robot. All Nao robots can speak English, but other languages need to be installed. Currently 19 languages (including Finnish and Japanese) are available for purchase. If the required language is installed on the robot, it can be selected and set as the current language until another language is selected.

The situation for North Sami is quite different, as it is not one of the 19 available languages. The required speech recognition and speech synthesis components are being newly developed for the DigiSami project using the new DigiSami corpus.

3 Speech technology for SamiTalk

Both a speech recognizer and a speech synthesizer are under development for North Sami. The biggest obstacle to creating these systems is the limited availability of high quality data sources. Especially for large-vocabulary continuous speech recognition, which normally requires at least 50 different speakers for a good model, the amount of available data limits both functionality and performance.

The first component that has been developed is a statistical Text-To-Speech (TTS) system. This system is built using the Ossian toolkit¹, which is suitable for building synthetic voices from small amounts of data with a minimal requirement of linguistic expertise. The toolkit utilizes vector space models [10] to predict the pronunciation, stress and other relevant factors needed to create a voice. Given the small amount of data used (approximately 3 hours), the system performs reasonably well and informal listening tests show that the results are approaching those of a commercial solution by Acapela. The main difference is that the commercial system has a hand-crafted preprocessor to transcribe numbers and abbreviations and the use of punctuation to determine prosodic structures.

For speech recognition two systems are currently available. The first is a large-vocabulary continuous speech recognizer that is speaker dependent [5, 8]. This system has been built from approximately 3 hours of data of a single speaker, using a variety of methods to overcome the lack of high-quality and large-quantity resources for North Sami. As this system is speaker dependent, it could only be used for the SamiTalk application by the speaker who recorded the sentences.

A second system utilizes a high-quality recognizer in a related language (Finnish). This system is used for North Sami by transcribing the North Sami words into Finnish phonemes. Although the system is speaker independent, i.e. usable by all speakers, not only those whose voices were trained upon, this system only works for a very small lexicon. This is firstly because of the lossy mapping between a North Sami word and its pronunciation in Finnish and secondly because of the manual labour involved in creating the lexicon.

Both systems are tri-state tri-phone hidden Markov models with Gaussian mixture model emission distributions, trained with the AaltoASR² toolkit [1, 6]. In the case of the open vocabulary system, Morfessor³ [9] is used to split words into segments to reduce the number of types in the lexicon. For language modelling a 20-gram model created with VariKN toolkit⁴ [7] is used. The language model is trained on the North Sami corpus provided by the University of Tromsø.

¹ Open source, available from <http://simple4all.org/product/ossian/>

² Open source, available from <https://github.com/aalto-speech/AaltoASR>

³ Open source, available from <http://www.cis.hut.fi/projects/morpho/>

⁴ Open source, available from <https://github.com/vsiivola/variKN>

4 Example interaction

This section shows an example of how users will interact with SamiTalk. The robot suggests a few topics that it can talk about using information from Sami Wikipedia. Among them, *Tiina Sanila-Aikio* is the current president of the Sami Parliament of Finland, and Japanese *Manga* shows that "local" topics such as *Snowmobile* are not the only topics in Sami Wikipedia. The user asks the robot to talk about the situation of the Sami language in Finland.

- Robot: *Man fáttá birra don háliidat gullat? Soames mu favorihta fáttát leat Sámegeiella, Roavvenjárga, Mohtorgielká, Tiina Sanila-Aikio, Manga.* (What topic do you want to hear about? Some of my favourites are: Sami language, Rovaniemi, Snowmobile, Tiina Sanila-Aikio, Manga.)
- User: *Sámegeiella* (Sami language)
- Robot: *Sámegeiela dilli Norgga, Ruoŋa, Suoma ja Ruošša bealde lea earálágan.* (The situation of Sami language is different in Norway, Sweden, Finland and Russia.)
- User: *Dilli Suomas* (The situation in Finland.)
- Robot: *Bealli Suoma 9 350 sámis máhtet sámegeiela. Suomas hállujuvvo davvisámegeiella, anárašgiella ja nuortalašgiella, main anárašgiella dušše Suomas. Buoremus dilli lea davvisámegeielain, mii ii leat seamma áitatvuloš go eará sámegeielat. Anárašgiela dilli lea vehá buoret go nuortalašgiela, dan dihte go anárašgielat giellabeassi doaibma lea joatkašuvvan guhkit áiggi. Suomas sámiid ruoktoaguovllus ássi sámiin lea leamaš riekti geavahit sámegeiela virgeolbmui jagi 1992 rájes, muhto virgeolbmui ii gáibiduvvo sámegeielmáhttu baicce geavahuvvojit jorgaleaddjit ja dulkat.* (Half of the 9,350 Sami in Finland speak Sami language. Three Sami languages are spoken in Finland: North Sami, Inari Sami and Skolt Sami. Inari Sami is spoken only in Finland. North Sami has the best situation, and it is not as endangered as other Sami languages. The situation of Inari Sami is a bit better than of Skolt Sami, because the language nest of Inari Sami has existed longer than the language nest of Skolt Sami. In Finland the Sami people who live in the Sami region have had the right to use Sami language with officials since 1992, but the language knowledge is not required of officials, they use translators and interpreters.)
- Robot: *Leage buorre, vállje čuovvovaččain: Joatkke, Ovddit, dahje daja Bisán nu ahte bisánuvvo.* (Please select from: Continue, Previous, or say Stop if you want to stop.)
- User: *Joatkke.* (Continue.)
- Robot: *Suoma sámiid ruovttoguovllus mánáin lea riekti lohkat vuodđoskuvllas ja logahagas sámegeiela eatnigiellan, válljaávnnsin dahje eaktodáhtolaččat. Dasa lassin giella oahppahuvvo maid ruovttoguovllu olggobealde Oulus, Roavvenjárggas ja Soađegili girkogilis. Sámegeiela lea vejolaš čállit studeantadutkosis eatnigiellan. Vuosttas háve dát dáhpáhuvai jagis 1994.* (In the Sami region in Finland the children have the right to study Sami as mother tongue, optional subject, or voluntarily in basic school and senior secondary school. In ad-

dition, the Sami language is taught outside the Sami region, in Oulu, Rovaniemi and the center of Sodankylä. It is possible to take the mother tongue part of the matriculation examination in Sami. This was done for the first time in 1994.)

- User: *Joatkke*. (Continue.)
- Robot: *Sámit leat ain juo guovtteriellagat, mii dárkkuha dan, ahte sii hállat sámegiela lassin riikkaid váldogiela. Suomas skuvllain oahpahuvo davvisámegiela lassin maid anáraš- ja nuortalašgillii, vaikke eanas oahpahas lea davvisámegillii. Sámegeilat oahpahusa birra leat Suomas sullii 500 oahppi, geain 150 ožžot oahpahusa sámegillii. Váldooassi oahpahas dáhpáhuvá 1-6-luohkái.* (The Sami are now always bilingual, which means that in addition to Sami they speak the main language of the state they live in. In Finland the schools teach in North Sami, Inari Sami and Skolt Sami, although mostly the teaching language is North Sami. Under the teaching in Sami language there are around 500 pupils in Finland, and 150 of them gain teaching in Sami. The main part of the teaching in Sami language is given in classes 1-6.)
- User: *Joatkke*. (Continue.)
- Robot: *Suomas golbma universitehta fállat sámegiela oahpahusa. Oulu universitehta Giellagas-instituhtas lea vejolaš studeret sámegiela váldoávvnnasin. Oalgeávvnnasin sámegiela lea vejolaš studeret Helssega universitehtas ja Lappi universitehtas.* (In Finland three universities offer education in Sami language. The Giellagas Institute of Oulu University offers Sami language as main subject. It is possible to study Sami language as minor subject at University of Helsinki and University of Lapland.)
- User: *Bisán*. (Stop.)
- Robot: *Háliidatgo duođat bisánit?* (Do you really want to stop?)
- User: *Juo*. (Yes.)
- Robot: *Ortnegis. Giitu. Báze dearvan.* (OK. Thank you. Goodbye.)

5 Conclusions and future work

The speech components for North Sami are under further development and have not yet been integrated and tested with the Nao robot software and the WikiTalk software. For demo purposes we hope to use a speaker independent, but closed vocabulary, speech recognizer. The dialogue has been restricted in order to show the general concept even with this limitation.

In future work, we will explore options for creating a large-vocabulary continuous speech recognizer that is speaker independent with an open vocabulary. The options include not only collecting more data, but also techniques for utilizing available North Sami resources such as news broadcasts, and resources from related Finno-Ugric languages such as Finnish.

Acknowledgements We thank our colleagues in the DigiSami project for Sami and English translations and the University of Tromsø for access to their North Sami datasets, and acknowledge the computational resources provided by the Aalto Science-IT project. The work has received funding from the EC's 7th Framework Programme (FP7/2007-2013) under grant agreement n°287678 and from the Academy of Finland under the Finnish Centre of Excellence Program 2012–2017 (grant n°251170, COIN) and through the project Fenno-Ugric Digital Citizens (grant n°270082).

References

1. Hirsimäki, T., Pyllkkönen, J., Kurimo, M.: Importance of high-order n-gram models in morph-based speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on* 17(4), 724–732 (2009)
2. Jokinen, K., Hiovain, K., Laxström, N., Rauhala, I., Wilcock, G.: DigiSami and digital natives: Interaction technology for the North Sami language. In: *Proceedings of Seventh International Workshop on Spoken Dialogue Systems (IWSDS 2016)*. Saariselkä (2016)
3. Jokinen, K., Wilcock, G.: Multimodal open-domain conversations with the Nao robot. In: J. Mariani, S. Rosset, M. Garnier-Rizet, L. Devillers (eds.) *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice*, pp. 213–224. Springer (2014)
4. Laxström, N., Wilcock, G., Jokinen, K.: Internationalisation and localisation of spoken dialogue systems. In: *Proceedings of Seventh International Workshop on Spoken Dialogue Systems (IWSDS 2016)*. Saariselkä (2016)
5. Leinonen, J.: Automatic speech recognition for human-robot interaction using an under-resourced language. Master's thesis, Aalto University, School of Electrical Engineering, Department of Signal Processing and Acoustics, Espoo (2015)
6. Pyllkkönen, J.: An efficient one-pass decoder for Finnish large vocabulary continuous speech recognition. In: *Proceedings of The 2nd Baltic Conference on Human Language Technologies*, pp. 167–172 (2005)
7. Siivola, V., Creutz, M., Kurimo, M.: Morfessor and variKN machine learning tools for speech and language technology. In: *INTERSPEECH*, pp. 1549–1552 (2007)
8. Smit, P., Leinonen, J., Jokinen, K., Kurimo, M.: Automatic speech recognition for Northern Sami with comparison to other Uralic languages. In: *Proceedings of Second International Workshop on Computational Linguistics for Uralic Languages* (2016)
9. Virpioja, S., Smit, P., Grönroos, S.A., Kurimo, M.: Morfessor 2.0: Python implementation and extensions for Morfessor baseline. Tech. rep. (2013)
10. Watts, O.: Unsupervised learning for text-to-speech synthesis. Ph.D. thesis, University of Edinburgh (2012)
11. Wilcock, G.: WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In: *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*, pp. 57–69. Mumbai (2012)