DEVELOPMENT OF SPEECH SYNTHESES FOR LOWER SORBIAN AND UPPER SORBIAN USING MARYTTS

Astrid Schmiedel, Ingmar Steiner

Serbski institut/Sorbian Institute, Bautzen, Germany
Fraunhofer IAIS, Sankt Augustin, Germany
astrid.schmiedel@serbski-institut.de, ingmar.steiner@iais.fraunhofer.de

ABSTRACT

Lower and Upper Sorbian are Slavic languages spoken in eastern Germany. As minority languages they are phonetically and phonologically less well documented than major languages and so far only benefit from limited digital resources. Thus, they are usually not considered for technologies such as text-to-speech (TTS) synthesis. This paper describes our approach, methods, and resources for creating two TTS synthesis systems for the Sorbian languages using the open-source MaryTTS platform. All data and resources for the unit-selection synthesis voices will be published under an open-source license and will be made freely available.

Keywords: Upper Sorbian, Lower Sorbian, text-to-speech synthesis, MaryTTS, unit selection speech synthesis

1. INTRODUCTION

The Sorbian people are one of four recognised autochthonous national minorities living in Germany. Their languages, Lower and Upper Sorbian, are protected in Germany under the European Charter for Regional or Minority Languages [1]. Not everyone who identifies as Sorbian is also proficient in the language; especially with regard to Lower Sorbian, there are only about 200 native speakers left (almost) all elderly. In addition, there are semi-speakers and, thanks to language revitalisation measures, also new speakers. Exact numbers of speakers are difficult to give; ten-year-old estimates speak of up to 5,000 speakers [2, p. 279] of Lower Sorbian and about 20,000 speakers [2, p. 291] for Upper Sorbian, both with a clear downward trend.

Still, as with many other minority and endangered languages, more and more Sorbian-language content and information in text form can be found on the internet. However, so far there are no technical aids such as auditory speech output or read aloud programmes, which would help achieve full accessibility for the Sorbian languages. The development of Sorbian text-to-speech (TTS) synthesis systems has both technical and linguistic prerequisites. The lack of a complete and scientifically supported standard pronunciation of Upper and Lower Sorbian is problematic. Though there are studies on Lower Sorbian and Upper Sorbian phonetics and phonology (e.g., [3, 4, 5, 6]), relevant questions, i.e., phoneme onset, number of phonemes and typical allophones, distribution, exact realisation individually and in phonetic compounds, relationship between spelling and pronunciation, need fundamental research.

The development of speech syntheses for Lower Sorbian and Upper Sorbian is a research project which will not be complete for several years, but the present aim is to fill as many of these gaps in the fundamental research as possible. Fundamental research on the phonetics of the Sorbian languages and work on the speech syntheses run in parallel, and the current focus of the fundamental research is on the connection between orthography and pronunciation in order to be able to work with preliminary orthoepic proposals.

The focus of this paper is the technical development using the open-source MaryTTS platform [7] and the voicebuilding process which it provides.

2. MATERIAL AND METHODS

Due to its modular architecture, the MaryTTS platform allows developers to access and modify the entire processing system from text input to speech output [8] and therefore enables the building of custom synthetic voices in new languages. It also supports unit-selection speech synthesis, which we choose for the benefit of its more natural synthetic sound. The process requires the creation of natural language processing (NLP) components for Lower and Upper Sorbian, and also requires appropriate speech data, i.e., natural language recordings of a speaker in the target languages in order to combine them for building the synthesised voices.

2.1. NLP components

Pronunciation prediction requires at least information on the pronunciation of the individual sounds as well as a word list as a lexical resource.

2.1.1. Allophone set

In lack of a pronunciation dictionary or an orthoepy of the Sorbian languages, a list of all the allophones of Lower Sorbian and Upper Sorbian was compiled on the basis of previous research [3, 4, 5, 6] and assembled in the required form. This allophone set contains IPA transcription as well as a customised SAMPA transcription plus the according standard phonological features. A customised SAMPA transcription was necessary because some standard SAMPA transcriptions – for example [ɘ̟] and [e̞] – include a backslash or an underscore, which are reserved characters in MaryTTS. The lists contain about 90 allophones for Lower Sorbian (36 vowels, 57 consonants/compounds) and 80 allophones for Upper Sorbian (33 vowels, 47 consonants/compounds). Since there are no standard variants yet, all currently accepted pronunciations have been collected. We decided to include also non-native sounds, mostly German, e.g., long vowels. The settlement area of the Sorbs within Germany has resulted in very close language contact with German, so the additional sounds are required to correctly pronounce foreign and loan words.

2.1.2. Lexical resource

The lexical resource consists of a word list containing relevant words in both orthographic representation and phonetic transcription. To select the words, morphological generators were used, which were developed at or with participation of the Sorbian Institute as the basis for programmes for automatic spell-checking for Lower and Upper Sorbian [9, 10, 11]. Based on a registered basic word form, all correct grammatical forms can be built applying the generators. The words were therefore not selected according to frequency, but rather – due to the availability of this resource – algorithmically. The algorithm was developed to output a list of words, taking into account as many language specific sound combinations as possible (based on preliminary analysis of the accompanying research on the connection between orthography and pronunciation).

For transcription we used the Grapheme-to-Phoneme (G2P) conversion tool [12, 13] available at BAS Web Services [14]. We make use of the option to create and upload customised mappings from orthographic to phonological form to use it for Lower and Upper Sorbian as "user-defined" languages. Although this sped up the transcription process a lot, it required and still requires constant checking and manual adjustment. Currently the word lists include approximately 12,500 items for Lower Sorbian and approximately 10,400 items for Upper Sorbian; we aim at about 40,000 entries for each. Letter-to-sound rule training, required for MaryTTS, takes place automatically based on the word lists.

2.1.3. Text normalisation

As Slavic languages, Lower and Upper Sorbian are highly inflected, so this is ongoing work.

In both NLP components, we have implemented a preprocessing module for text normalisation, which handles the conversion of symbols and numbers – which would be unpronounceable by the TTS engine in their raw form – into written words. Each token in the input text is checked against a symbol mapping table, and if applicable, replaced by a normalised string. Furthermore, following the approach of [15], we detect numeric tokens, and replace them by their spell-out pronunciation, using the Unicode Consortium's ICU4J library with custom rule-sets. In this way, an input string such as "100 €" will be normalized to the words, *sto eurow*.

Further improvements, including acronym handling, are planned.

2.2. Speech Data

In addition to the NLP components, natural language material is required to generate the synthesised voices using the unit-selection method.

2.2.1. Prompt material

Fortunately, we have access to two reference corpora representing the modern Sorbian written language. They are both based on Lower and Upper Sorbian publications from the period 1990 to 2010. The Lower Sorbian Reference Corpus consists of 68,058 randomly selected but evenly distributed sentences and sentence-like units of text, the Upper Sorbian one of 70,217. The selection of sentences should be as phonetically balanced as possible. We tried to achieve this by defining suitable letter combinations in connection with markings of word boundaries as well as sentence beginnings and endings, after the sentences of the corpora had been standardised.

The Upper Sorbian version of *The North Wind and the Sun* is taken from Howson [5] whereas the

material	duration
The North Wind and the Sun	52 s
ca. 360 sentences and sentence-like units of text	90 min

(a) Lower Sorbian

material	duration
The North Wind and the Sun	49 s
ca. 380 sentences and sentence-like units of text	70 min

(b) Upper Sorbian

Table 1: Prompt material and (approx.) durations.

Lower Sorbian version was translated by an expert. Details are given in Table 1.

2.2.2. *Recordings*

The main speaker criteria are native speakers with very good pronunciation, resilience and availability.

In the case of Upper Sorbian there was a wide range of potential native speakers. Test recordings were made with five eligible candidates, whose recordings were then evaluated by Sorbian linguists, including native speakers, for pronunciation clarity and possible problematic anomalies. The chosen male speaker was recorded in several sessions in 2022 in a studio environment in Bautzen, Germany. We used a Neumann TLM 103 large-diaphragm microphone mounted on a microphone arm approximately 30 cm from the speaker's mouth. The audio was sampled at 44.1 kHz (16 bit) using a Steinberg MR816 CSX audio interface along with Steinbergs Nuendo recording software. The material was presented in PDF format on a monitor and was read from screen by the speaker. At the moment, the material on Upper Sorbian comprises around 70 min of speech (see Table 1b for details).

For Lower Sorbian a male speaker with very good Lower Sorbian language skills and pronunciation was chosen. He did not learn Lower Sorbian from his parents, but through contact with his extended family, who still use it frequently. As presented in the introduction, we did have to take into account the difficult situation of Lower Sorbian. While younger speakers do learn Lower Sorbian at school or in language courses, the remaining true native speakers are not suitable for the recordings we are aiming at because of their old age and the problems resulting from it. So in the broader sense considering the context of an endangered language, the chosen speaker is to be understood as a native speaker. In several recording sessions in 2021 and 2022 the speaker was recorded in a studio environment in Cottbus, Germany. We used a AKG P420 large-diaphragm microphone mounted approximately 30 cm from the speaker's mouth. The audio was sampled at 44.1 kHz (16 bit) using a Motu 16A audio interface along with Apple Logic Studio recording software. The material was read from a printout by the speaker. At the moment, the material on Lower Sorbian comprises around 90 min of speech (see Table 1a for details).

2.3. Processing

The Lower Sorbian recordings were done in several takes and then each prompt was stored as a separate WAV file, whereas the Upper Sorbian material was stored individually directly after each prompt. In combination with the corresponding text files, the prompts were semi-automatically labelled using the MaryTTS forced-alignment plugin, a custom wrapper around the Montreal Forced Aligner [16]. The phonetic annotations were manually checked and corrected where necessary.

2.4. Voicebuilding

Another plugin handles voicebuilding, using the NLP components to assemble the processed Lower Sorbian and Upper Sorbian voice data into unit-selection voice components suitable for installation into MaryTTS. Summarizing from [8], this process comprises *(a)* extracting acoustic features from the audio, *(b)* extracting linguistic features from the text, *(c)* aligning features based on the phonetic labels, *(d)* building prosody models, and *(e)* packaging the data.

3. SUMMARY

The development of speech syntheses for Lower and Upper Sorbian is a research project which will not be complete for several years. The goal of creating a user-friendly service that enables users to have Sorbian-language texts on websites read aloud in a comprehensible way and – as far as technically feasible – with largely "good" pronunciation, requires fundamental research work on the phonetics and phonology of Lower Sorbian and Upper Sorbian. The project allows the linking of the service aspect with the closing of research desiderata in this area. It is ongoing work with constant improvements. The current state already allows an understandable synthesis of simple(st) texts in both Sorbian languages, live demo available at

https://tts-juro-matej.serbski-institut.de/. The quality of pronunciation should be improved with further data (lexical resource, audio data and especially further NLP components ideally considering different case endings). The lexical resources as well as the NLP components for Lower and Upper Sorbian are released under the Lesser GPL 3.0 license [17, 18, 19, 20]. First versions of both Sorbian synthesis voices are released free for non-commercial use [21, 22]. The Upper and Lower Sorbian Speech Databases are also publicly available under the same terms [23, 24].

4. ACKNOWLEDGEMENT

This project is supported with tax funds on the basis of the budget passed by the Saxon State Parliament (directive inclusion, no. 100664411).

The authors gratefully acknowledge one anonymous reviewer for detailed, helpful feedback.

5. REFERENCES

[1] Treaty Office of the Council of Europe, "ETS No. 148: European Charter for Regional or Minority Languages," 1998. URL: https://www.coe.int/en/web/european-charter-regional-or-minority-languages/text-of-the-charter

[2] F. Šěn, D. Scholze, and S. Hose, Eds., *Sorbisches Kulturlexikon*. Bautzen: Domowina-Verl., 2014.

[3] L. Jocz, *Wokalowy system hornjoserbskeje rěče přitomnosće [The vowel system of contemporary Upper Sorbian]*. Szczecin: volumina.pl Daniel Krzanowski, 2011.

[4] ——, *Studien zum obersorbischen Konsonantismus*. Szczecin: volumina.pl Daniel Krzanowski, 2015.

[5] P. Howson, "Upper Sorbian," *Journal of the International Phonetic Association*, vol. 47, no. 3, pp. 359–367, 2017, doi:10.1017/S0025100316000414.

[6] F. Kaulfürst, "Pšinosk k dolnoserbskej ortoepiji na zakładźe projekta awdijowych datajow za nimsko-dolnoserbski internetowy słownik [The creation and integration of sound data in the German - Lower Sorbian internet dictionary as a starting point for a discussion of Lower Sorbian orthoepy]," *Lětopis*, vol. 66, no. 1, pp. 3–41, 2019.

[7] M. Schröder and J. Trouvain, "The German text-to-speech synthesis system MARY: A tool for research, development and teaching," *International Journal of Speech Technology*, vol. 6, no. 4, pp. 365–377, 2003, doi:10.1023/A:1025708916924.

[8] I. Steiner and S. Le Maguer, "Creating new language and voice components for the updated MaryTTS text-to-speech synthesis platform," in *Proc. 11th Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan, 2018, pp. 3171–3175. URL: http://www.lrec-conf.org/proceedings/lrec2018/pdf/1045.pdf

[9] "Lower Sorbian spell-checker," https://www.niedersorbisch.de/ortografija/kontrola.

[10] "Upper Sorbian spell-checker," https://soblex.de/download/download.html.

[11] "Soblex," https://www.soblex.de/?cmd=help.

[12] U. Reichel, "PermA and Balloon: Tools for string alignment and text processing," in *Proc. Interspeech*, Portland, Oregon, 2012, pp. 1874–1877, doi:10.21437/Interspeech.2012-509.

[13] ——, "Language-independent grapheme-phoneme conversion and word stress assignment as a web service," in *Proc. 25th Conference on Electronic Speech Signal Processing (ESSV)*, Dresden, Germany, 2014, pp. 42–49. URL: https://www.essv.de/pdf/2014_42_49.pdf

[14] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017, doi:10.1016/j.csl.2017.01.005.

[15] I. Steiner, S. Le Maguer, J. Manzoni, P. Gilles, and J. Trouvain, "Developing new language tools for MaryTTS: the case of Luxembourgish," in *Proc. 28th Conference on Electronic Speech Signal Processing (ESSV)*, Saarbrücken, Germany, 2017, pp. 186–192. URL: https://www.essv.de/pdf/2017_186_192.pdf

[16] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proc. Interspeech*, 2017, pp. 498–502, doi:10.21437/Interspeech.2017-1386.

[17] "Upper Sorbian lexicon for MaryTTS | Hornjoserbski leksikon za MaryTTS," 2023. URL: https://github.com/marytts/marytts-lexicon-hsb

[18] "Lower Sorbian lexicon for MaryTTS | Dolnoserbski leksikon za MaryTTS," 2023. URL: https://github.com/marytts/marytts-lexicon-dsb

[19] "Upper Sorbian language component for MaryTTS | Hornjoserbska rěčna komponenta za MaryTTS," 2023. URL: https://github.com/marytts/marytts-lang-hsb

[20] "Lower Sorbian language component for MaryTTS | Dolnoserbska rěcna komponenta za MaryTTS," 2023. URL: https://github.com/marytts/marytts-lang-dsb

[21] "A male Upper Sorbian unit selection voice for MaryTTS," 2023. URL: https://github.com/marytts/voice-serbski-institut-hsb-matej

[22] "A male Lower Sorbian unit selection voice for MaryTTS," 2023. URL: https://github.com/marytts/voice-serbski-institut-dsb-juro

[23] "Upper Sorbian voice data for MaryTTS," 2023. URL: https://github.com/marytts/serbski-institut-hsb-data

[24] "Lower Sorbian voice data for MaryTTS," 2023. URL: https://github.com/marytts/serbski-institut-dsb-data