Spicing up the Game for Underresourced Language Learning: Preliminary Experiments with Ainu Language-speaking Pepper Robot

Karol Nowakowski[*] , **Michal Ptaszynski** , **Fumito Masui**
Department of Computer Science, Kitami Institute of Technology, Koencho 165, Kitami, 090-8507, Hokkaido, Japan
nowakowski.karol.p@gmail.com, ptaszynski@cs.kitami-it.ac.jp, f-masui@mail.kitami-it.ac.jp

Abstract

We present a preliminary work towards building a conversational robot intended for use in robot-assisted learning of Ainu, a critically endangered language spoken by the native inhabitants of northern parts of the Japanese archipelago. The proposed robot can hold simple conversations, teach new words and play interactive games using the Ainu language. In a group of Ainu language experts and experienced learners whom we asked for feedback, the majority supported the idea of developing an Ainu-speaking robot and using it in language teaching. Furthermore, we investigated the performance of Japanese models for Speech Synthesis and Speech Recognition in generating and detecting speech in the Ainu language. We performed human evaluation of the robot's speech in terms of intelligibility and pronunciation, as well as automatic evaluation of Speech Recognition. Experiment results suggest that, due to similarities between phonological systems of both languages, cross-lingual knowledge transfer from Japanese can facilitate the development of speech technologies for Ainu, especially in the case of Speech Recognition. We also discuss main areas for improvement.

1 Introduction

The Ainu language is a language isolate native to northern parts of the Japanese archipelago, which – as a result of a language shift triggered by the Japanese and Russian colonization of the area, and assimilation policies – is currently recognized as nearly extinct (e.g., by Lewis *et al.* [2016]).

While multiple initiatives are being undertaken by the members of the Ainu community to preserve their mother tongue and promote it among the young generations, the number of speakers who possess the level of proficiency necessary to teach the language, is extremely small. As a consequence, access to Ainu language education is severely limited.

In recent years, Computer-Assisted Language Learning (CALL) and Robot-Assisted Language Learning (RALL) have been proposed as a way to support both native and foreign language acquisition [Randall, 2019]. We believe that they could also be helpful in addressing the challenges facing Ainu language teaching.

A major obstacle for the application of such ideas to minority languages, including Ainu, is the lack of high-volume linguistic resources (such as text and speech corpora and in particular, annotated corpora) necessary for the development of dedicated text and speech processing technologies. Advances in cross-lingual learning indicate that this problem can be, to a certain extent, alleviated by transferring knowledge from resource-rich languages. Unsurprisingly, however, such techniques tend to yield the best results for closely related languages (see, e.g., [He *et al.*, 2019]), which is not a feasible scenario for the Ainu language, as it has no known cognates. Nevertheless, given the similarity of phonological systems and some (presumably, contact-induced) grammatical constructions between Ainu and Japanese, we anticipate that it may be beneficial to use the existing Japanese resources as a starting point in the development of language processing technologies for Ainu.

In this paper, we describe a preliminary Ainu language conversational program for the Pepper robot[1], which serves as a proof of concept of how robots could support Ainu language education. Because dedicated speech technologies for Ainu are not available, and to test our assumptions about potential benefits of Japanese-Ainu cross-lingual transfer, we use Text-to-Speech and Speech Recognition models for Japanese. Finally, we conduct a survey among a group of Ainu language experts and experienced learners to evaluate the robot's speech in terms of intelligibility and pronunciation, and perform automatic evaluation of Speech Recognition performance.

2 The Ainu Language

Ainu is an agglutinating, polysynthetic language with SOV (subject-object-verb) word order. Until now, none of the numerous hypotheses about genetic relationship between Ainu and other languages or language families (see [Shibatani,

[*] Contact Author

[1] https://www.softbankrobotics.com/emea/en/pepper

1990]) have gained wider acceptance. Thus, it is usually classified as a language isolate. On the other hand, a number of grammatical constructions and phonological phenomena are believed to have developed under the influence of Japanese (see [Bugaeva, 2012]).

There are multiple regional varieties of the Ainu language. Some of them – in particular the dialects of Sakhalin – exhibit properties not found in other regions, the dissimilarities being large enough for many experts (e.g., [Refsing, 1986; Murasaki, 2009]) to describe the Hokkaido and Sakhalin dialects as mutually unintelligible. In this research, we focus on Hokkaido Ainu.

2.1 Phonology

Phonemic inventory of the Ainu language consists of five vowel phonemes: /i, e, a, o, u/, and twelve consonant phonemes: /p, t, k, c, s, h, r, m, n, y, w, ʔ/. For detailed analyses, please refer to Refsing [1986], Shibatani [1990] and Bugaeva [2012]. At the level of phonemes, phonological system of Ainu exhibits significant overlap with that of the Japanese language. There are, however, substantial differences between the two languages in phonotactics. In Japanese the basic syllable structure is CV (C = consonant, V = vowel), with only two types of consonants that may close a syllable: a geminate (doubled) consonant, and syllable-final nasal /N/. On the contrary, in Hokkaido Ainu all consonants except /c/, /h/ and /ʔ/ may occur in syllable coda position [Shibatani, 1990]. Furthermore, certain combinations of phonemes, such as /ye/ and /we/ are not permitted in modern Japanese or can only be found in foreign loan words (or have an irregular phonetic realization, as in the case of /tu/, pronounced as [tsɯ]), while the corresponding Ainu phonemes are not subject to such restrictions.

Although, in contrast to Japanese and Sakhalin Ainu, the opposition between short and long vowels is not distinctive in Hokkaido Ainu [Shibatani, 1990], vowels are often prolonged in certain forms (e.g., interjections – see [Nakagawa, 2013]) or at the end of a sentence (e.g., in imperative sentences – see [Satō, 2008]).

Both Japanese and the Ainu language have a pitch accent system, but with different characteristics: (Tokyo) Japanese exhibits a so-called "falling kernel" accent (i.e. the change from high to low pitch is distinctive), whereas in Ainu the position of the rise in pitch is important [Bugaeva, 2012]. The place of the accent in an accented word in Tokyo Japanese is not predictable without prior lexical knowledge [Shibatani, 1990]. In Ainu, apart from a few words with irregular accent, the accent falls on the first syllable if it is closed, and on the second syllable otherwise. The accent of a word may also be affected by the attachment of certain affixes [Bugaeva, 2012].

Intonation in the Ainu language is falling in declarative sentences and rising in questions [Refsing, 1986]. In Japanese, on the other hand, not all types of questions are pronounced with a rising intonation [Fujii, 1979].

2.2 Transcription

Most written texts in Ainu are transcribed using Latin alphabet and/or the Japanese *katakana* syllabary. *Katakana*, in its official version, has no means to represent closed syllables not occurring in the Japanese language. As a result, it is not possible to produce phonemically accurate transcriptions of many Ainu words containing syllable-final consonants. In older documents, such syllables were usually expressed as a combination of two characters representing open (CV) syllables. For instance, in one of the oldest Ainu language dictionaries, *Ezo hōgen moshiogusa* [Uehara and Abe, 1804], the word *apkas* (/apkas/, "to walk") was trascribed as アプカシ (/apukasi/). In contemporary transcription conventions, this problem is solved by using an extended version of the syllabary, with small-sized (*sutegana*) variants of *katakana* characters to denote syllable-final consonants (e.g., ⟨ㇰ⟩ /k/ derived from ⟨ク⟩ /ku/, and ⟨ㇷ゚⟩ /p/, from ⟨プ⟩ /pu/).

3 How Can Robots Support Ainu Language Learning?

Since inter-generational transmission of the Ainu language in the home environment has been disrupted, the only way to reverse the language shift is through promoting the study of Ainu. However, due to logistic and human resource constraints (i.e., low number of skilled instructors), Ainu language classes are only held in a limited number of larger cities, sporadically (once a week or even less frequently), and often only seasonally[2]. This means that the access to Ainu language education is severely limited. A further consequence of this situation is that the few existing courses are often held in groups of learners of varying age and/or exhibiting different levels of Ainu language skills, making it more difficult to adapt the tutoring strategy to individual needs of each participant. Another problem is short attention span of smaller children, which affects productivity of the already limited time spent in the classroom [Watanabe, 2018].

While there is a steadily growing number of publicly available learning aids (including a radio course, "Ainugo Rajio Kōza", broadcast weekly by the STV Radio in Sapporo, a collection of textbooks for multiple dialects of Ainu and educational games published by the Foundation for Ainu Culture[3], a YouTube channel[4] and an Ainu language course on a mobile language learning platform, Drops), self-learning from such materials does not include the element of interaction, which is believed to play an important role in the process of language acquisition [Long, 1996; Mackey, 1999].

In this research, we propose robot-assisted language learning as a means to increase the opportunities for learning through interaction, and potentially also to improve the efficiency of human-instructed learning sessions. In recent years, a growing body of research indicates that social robots can support language learning in both children and adults [van den Berghe *et al.*, 2019; Randall, 2019]. Promising results have been obtained in experiments exploring the

[2] As an example, the 2019 edition of the beginner level Ainu language course run by the Foundation for Ainu Culture (https://www.ff-ainu.or.jp/) was held in three cities in Hokkaidō (Sapporo, Kushiro and Shiraoi), with up to 20 lessons over the course of 6 or 8 months.

[3] https://www.ff-ainu.or.jp/web/learn/language/dialect.html

[4] www.youtube.com/channel/UCsvS5QjLwvlVhWpK48L57Cg

use of robots for teaching vocabulary [Alemi *et al.*, 2014; Kory Westlund *et al.*, 2017], speaking skills [Lee *et al.*, 2011], reading skills [Eun-ja Hyun *et al.*, 2008; Hong *et al.*, 2016], and grammar skills [Kennedy *et al.*, 2016]. In addition to that, there is strong evidence for positive effects of robots on the motivational aspect of language learning. Previous research found students working with a robot to be more satisfied with their learning experience [Eimler *et al.*, 2010; Shin and Shin, 2015], show higher motivation, engagement and confidence [Lee *et al.*, 2011; Wang *et al.*, 2013], and sustain interest and concentration for a longer time [Han *et al.*, 2008], when compared to a traditional classroom and other technologies, such as computers and tablets.

Depending on the intended role of the robot in the learning task, it can be programmed to act as an independent tutor [Lee *et al.*, 2011; Kennedy *et al.*, 2016; Kory Westlund *et al.*, 2017], or to assist a human teacher [Alemi *et al.*, 2014; Hong *et al.*, 2016]. An additional possibility investigated in previous studies is the robot taking up the role of a peer learner [Wang *et al.*, 2013; Meiirbekov *et al.*, 2016; Belpaeme *et al.*, 2018], which opens up possibilities for new forms of learning activities, such as learning by teaching [Tanaka and Matsuzoe, 2012].

Encouraged by the positive examples described above, we set out to develop a conversational robot, which could be used by students to learn the Ainu language in a more interactive manner outside of the class schedule, and at the same time would also support the work of teachers. As a proof of concept, in the remainder of this paper we describe a rule-based dialogue agent developed for the Pepper robot. It is capable of holding simple conversations, teaching new words and playing interactive games in Ainu.

An essential prerequisite that needs to be fulfilled in order to utilize the robot in Ainu language education, is to provide it with the ability to speak and receive user input in Ainu. Because there are currently no dedicated speech technologies available for the Ainu language[5], in this preliminary study we employ the Japanese Text-to-Speech and Speech Recognition models supplied with Pepper. We use this as an opportunity to examine the potential of cross-lingual transfer from Japanese for facilitating the development of speech technologies for Ainu.

As an alternative to a robot, a virtual agent could be employed, which would make the technology more accessible to individual users (e.g., it could be deployed as an application for tablet computers). However, in this study we decided to use a robot, as it has been suggested that embodied agents have advantages over virtual ones, such as the ability to manipulate physical objects and use gestures [Wit *et al.*, 2018], and are perceived in a more positive way than animated characters [van den Berghe *et al.*, 2019].

4 Materials

4.1 Pepper Robot

Pepper (shown in Figure 1) is a humanoid robot manufactured by SoftBank Robotics. It was first introduced in 2014. Below, we provide a short description of the robot's three components relevant to our experiments: Speech Synthesis (Text-to-Speech), Speech Recognition and dialogue scripting functionality.

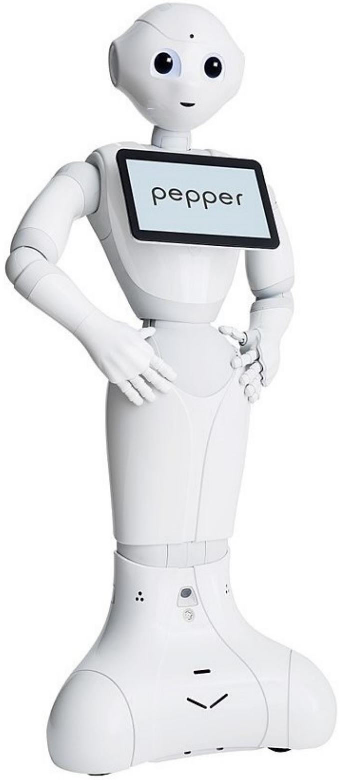

Figure 1: "Pepper the Robot", by Softbank Robotics Europe. Source: https://commons.wikimedia.org/wiki/File:Pepper_the_Robot.jpg. Licensed under the Creative Commons Attribution-ShareAlike 4.0 International license: https://creativecommons.org/licenses/by-sa/4.0/

Speech Synthesis

Conversion of Japanese text input to speech is performed by the microAITalk engine[6]. It uses a concatenative speech synthesizer based on phonemic and prosodic patterns learned from a speech corpus[7]. Speech parameters, such as pitch, speed and volume, can be locally adjusted by inserting special tags in the text.

Speech Recognition

Pepper is equipped with a closed-vocabulary Speech Recognition engine, i.e., a list of possible words/phrases must be predefined. Once speech has been detected, each phrase from

[5] In a recent work, Matsuura *et al.* [2020] reported developing an end-to-end Speech Recognition model for Ainu, but as of today, their system is not publicly available.

[6] http://doc.aldebaran.com/2-5/naoqi/audio/altexttospeech.html

[7] https://www.ai-j.jp/about/

the list is assigned a confidence level (“estimate of the probability that this phrase is indeed what has been pronounced by the human speaker”)[8].

Dialogue Scripting

A dedicated language, QiChat, is used to define “rules” for managing the flow of the conversation between the robot and humans[9]. There are two types of rules: “User rules”, triggered by human input, and “Proposal rules” generating specific robot output without any user input. Rules are grouped by “Topics”. Apart from rules, a Topic may include “Concepts” (lists of equivalent or related words/phrases, e.g., synonyms of a word), as well as user-defined functions.

4.2 Dialogue in the Ainu Language

Using Choregraphe[10] (a graphical environment for programming Pepper) and QiChat, we created a simple dialogue script in the Ainu language. The script includes 78 User rules and subrules, 20 Proposal rules, 93 Concepts and 2 function definitions. Two excerpts from the script (four Concept definitions and a single User rule with subrules) are shown in Listing 1. Possible conversation topics include: greetings, self-introduction (name, age, etc.) and asking about weather. If the human interlocutor does not speak for a specified period of time, the robot will initiate conversation by asking a question. The user can also ask to translate words (included in a predefined list) from Japanese to Ainu. Furthermore, when Pepper asks a question in the Ainu language, the user can ask him to repeat it or to explain its meaning in Japanese. In addition to conversations, the script includes two interactive games played in the Ainu language, of which one utilizes the robot’s touch sensors.

Although the bulk of the dialogues in the script are in Ainu, the parameter determining which language should be used by the robot’s Text-to-Speech engine and Speech Recognition engine, was set to Japanese. This effectively turned our experiment into an instance of cross-lingual knowledge transfer (in particular, the simplest form of it where a model trained solely on source language data is applied to the target language – see [He *et al.*, 2019]). An additional benefit of using Japanese speech technologies is the ability to communicate with the robot not only in Ainu, but also in the first language of most learners of the Ainu language.

The contents of the dialogues were based mainly on materials for the Saru dialect of Ainu (spoken in southern Hokkaido), such as [Tamura, 1996; Kayano, 1996; Nakagawa, 2013] and textbooks for the “Ainugo Rajio Kōza" [Ainu Language Radio Course][11].

In cases where the intonation of utterances generated by the Speech Synthesis engine was remarkably inconsistent with the recordings by native speakers (e.g., questions pronounced by the robot with a falling intonation, whereas it should be rising), we used the tags mentioned in Section 4.1 to modify pitch and speed. However, in this preliminary experiment we refrained from performing extensive fine-tuning[12].

Furthermore, in a number of sentences in audio materials used for reference, we observed a noticeable increase in the length of the sentence’s final vowel. While vowel length is not a distinctive feature in Hokkaido Ainu, and thus it is not reflected in written texts, we adjusted the transcriptions of such fragments in order to make the robot’s pronunciation resemble that of native speakers. Examples include the sentence-final particle *yan*, used in polite commands: in the case of *katakana* transcription it is normally expressed as ヤン (/yan/), but in one of the sentences in our dialogue script (see Listing 1) it was instead transcribed as ヤアン (/yaan/). As a result, Pepper pronounced it as [jaːn].

Since the majority of syllable-final consonants are not supported by the Japanese speech models, the corresponding full-size *katakana* characters (representing open, CV syllables) were used to denote them. For example, the word *anakne* (/anakne/, “as for”) was transcribed as アナクネ (/anakune/).

5 Evaluation Experiments

In this section, we describe the evaluation experiments conducted in order to find out the answers to the following two questions: (i) are the similarities in phonological systems of Ainu and Japanese significant enough to be leveraged in the development of speech technologies for Ainu, and (ii) what are the opinions of Ainu language experts and learners about the idea of using a robot for learning Ainu.

5.1 Survey

In the first experiment, we asked Ainu language experts and experienced learners for a judgement of the quality of speech generated by the robot, as well as for a feedback about the idea of creating an Ainu language-speaking robot and using it in language education.

Participants

The survey[13] was conducted among a group of 8 people engaged in activities related to the Ainu language, namely: 4 learners of Ainu, 1 language instructor, 1 linguist and 2 persons involved in other types of activities. 7 out of 8 participants have at least 5 years of experience with the Ainu language. All participants are also speakers of Japanese, including 7 native speakers.

Survey design

The participants were asked to watch a video demonstrating a conversation with the robot[14] and evaluate the quality of its speech (both in Japanese and Ainu) in terms of intelligibility (defined as the listener’s ability to identify the words spoken by the robot) and correctness of pronunciation. The latter was

[8] http://doc.aldebaran.com/2-5/naoqi/audio/alspeechrecognition.html

[9] http://doc.aldebaran.com/2-5/naoqi/interaction/dialog/aldialog.html

[10] http://doc.aldebaran.com/2-5/software/choregraphe/

[11] https://www.stv.jp/radio/ainugo/index.html

[12] This decision was in part motivated by the observation that insertion of multiple tags in a single word or short utterance can cause problems with the Speech Synthesis engine, leading to unnatural output and unintended pauses.

[13] https://forms.gle/bfx6VmzHeuWY1E6c8

[14] https://youtu.be/DJgVolvcees. A version with subtitles (in Ainu and Japanese) is also available: https://youtu.be/RTuGqgDNBC8.

Listing 1 Excerpts from the QiChat script

```
concept:(and_you) "\rspd=100\\vct=100\エアニ \rspd=95\\vct=120\ヘエエ?"
concept:(and_you2) ^rand["\rspd=100\\vct=100\エアニ カ エ \rspd=110\イワンケエ?" ~and_you] # Only in response to
"how are you?"
concept:(me) "\rspd=100\\vct=100\クアニ"
concept:(me_too) "~me カ"
.......................................................................................................
u:({ペッパー、} !ク エ イワンケ {"ワ エアン"} [ア ヤ]) "ク \rspd=110\\vct=100\イワンケ \rspd=100\\vct=135\ワア
ア". ~and_you2 # "How are you (Pepper)?" -> "I'm fine. And you?"
  u1:(~me_too クイワンケ {ワ} {クアン}) ~good_to_hear # "I'm fine, too" -> "Good to hear that"
  u1:({ソン ノ} ク シンキ {ワ}) "\rspd=100\\vct=100\ポン ノ シニ \rspd=100\\vct=135\ヤアン" # "I'm (really) tired" ->
  "Please get some rest"
  u1:(ク ミシム) ~in_this_case ~lets_play # "I'm bored" -> "Then let's play!"
```

further broken up into four different aspects: pronunciation of Japanese/Ainu sounds, accents, intonation and overall impression. Intelligibility was evaluated on a three-point scale: “easy to understand", “sometimes hard to understand", “hard to understand". For pronunciation we employed a five-point scale: “perfect", “quite good", “some problems", “not good", “very bad" (when calculating average scores, presented later in this paper, we converted each grade to a numerical value, where “perfect" corresponds to 5 and “very bad", to 1).

The final question asked for an opinion as to whether an Ainu language-speaking robot such as the one developed in this research could be useful in Ainu language education. Here, the predefined options were “yes", “yes, if it's improved" and “no"; the participants were also given an option to specify their own answer.

Apart from the closed questions, the respondents were encouraged to submit any additional opinions and comments.

5.2 Speech Recognition Experiment

The goal of the second experiment was to investigate the performance of the robot's Japanese Speech Recognition engine in identifying speech uttered in the Ainu language. For that purpose, we selected a list of 30 Ainu words – of which 12 include combinations of sounds violating phonotactic constraints observed in the Japanese language – and 30 Japanese words. Each of the two word lists was then presented to the Speech Recognition engine as the list of possible phrases to detect. All words were transcribed in *kana* script (namely, *hiragana* for Japanese and *katakana* for Ainu words). As in the dialogue script (see Section 4.2), consonants occurring in syllable coda position in Ainu words were represented with the corresponding standard *katakana* characters.

Finally, a recording of each word from the respective list, uttered by a native speaker, was played to the robot 3 times and the output was recorded. Japanese voice recordings were obtained from the WWWJDIC online dictionary[15]. In the case of Ainu words, we used the audio pronunciations recorded by Shigeru Kayano for his dictionary [Kayano, 1996] and included in the online version released by the Ainu Museum[16].

Results of the experiment were evaluated in terms of Accuracy (defined as the proportion of trials where the target word yielded the highest probability) and confidence levels for words assigned the highest value in each trial.

6 Results and Discussion

As shown in Figure 2, more than half of the respondents did not have any problems with understanding Pepper's speech in Japanese. On the contrary, the speech in Ainu was at times difficult to understand for three quarters of the participants.

In their comments, multiple respondents indicated that the robot's speech was sometimes too fast, which rendered it not only unnatural, but also difficult to follow. In addition to that, several respondents pointed out problems with intonation (e.g., in questions), which in their opinion also harmed intelligibility.

The majority of the respondents rated the robot's Japanese pronunciation as either “good" or “perfect" (Figure 3). In the case of Ainu pronunciation (Figure 4), all aspects received the middle grade (“some problems") or higher from more than half of the participants.

In both cases, the lowest average scores were achieved for accents (4.00 for Japanese and 3.00 for Ainu) and intonation (4.125 and 3.00). Relatively low results for Japanese can be attributed to the characteristics of its pitch accent system (namely, unpredictability of the accent's location – see Section 2.1). This impairs the Text-to-Speech system's ability to generate correct prosodic patterns for words and phrases unseen in the training data. In Ainu, there is less uncertainty in terms of accent, but the Text-to-Speech model, trained exclusively on Japanese data, has no knowledge of the rules governing it.

The two participants with presumably the most relevant expertise (i.e., the person involved in linguistic research related to the Ainu language and the Ainu language instructor) were generally less favorable in their evaluations: on average they rated Pepper a 3.625 for Japanese speech and 2.25 for the pronunciation of Ainu, whereas the average for all respondents was 4.28 (for Japanese) and 3.16 (for Ainu). As for specific points of criticism, both of them reported hearing consonants followed by vowels in places where only a consonant (belonging to the coda of the preceding syllable) should appear.

[15] http://nihongo.monash.edu/cgi-bin/wwwjdic

[16] https://ainugo.ainu-museum.or.jp/

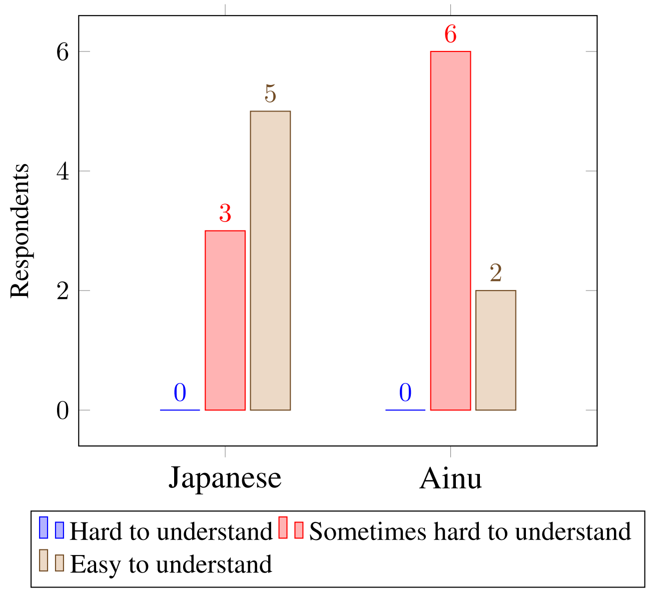

Figure 2: Evaluation of the robot's intelligibility

That, of course, is the effect of restrictions on closed syllables in the Japanese Text-to-Speech engine.

No comments were made concerning pronunciation of individual phonemes of the Ainu language. This suggests that the similarities to Japanese in this area are significant enough that a Japanese Text-to-Speech model can produce reasonable results.

Figure 5 shows the results of the survey's final question. Apart from a single respondent who was skeptical about the willingness of the Ainu people to learn their mother tongue from a machine, all participants expressed an opinion that a robot of this type would be useful in Ainu language education, either now (50%) or after improvements (37.5%).

Speech Recognition experiment results are summarized in Table 1. The robot correctly selected the target word as the most probable option in all trials for both languages. Moreover, in both cases the average confidence level exceeded 50% – the default confidence threshold in the dialogue engine, below which the speech recognition result is ignored[17]. On the other hand, in the experiment with Japanese words, only one trial yielded a confidence value below the threshold, while for words in the Ainu language, the proportion of such results was 28%. Not surprisingly, Ainu words containing syllables violating Japanese phonotactic rules perplexed the Speech Recognition engine to a greater extent than the rest of them, achieving an average confidence level of 50.36%. The value for the subset of words with two such syllables was even lower: 44.88%.

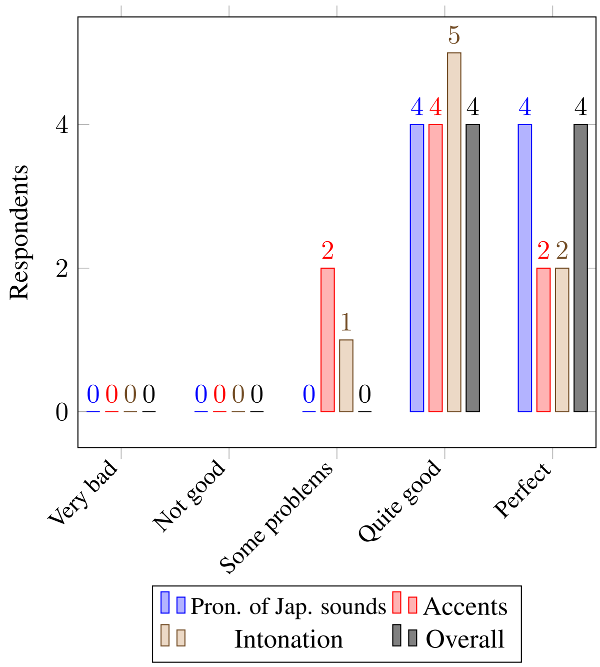

Figure 3: Evaluation of the robot's Japanese pronunciation

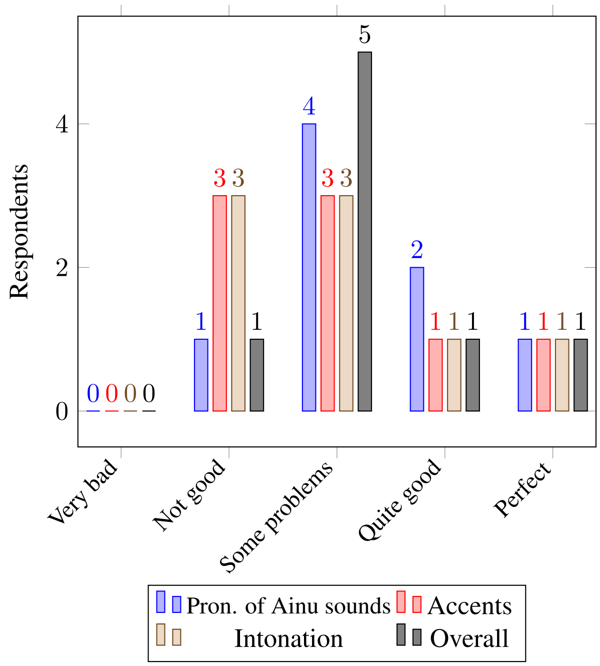

Figure 4: Evaluation of the robot's Ainu language pronunciation

[17] Although the threshold can be freely modified in Speech Recognition options, setting it to a low value may negatively affect precision of the Speech Recognition engine, causing it to recognize non-verbal sounds or background noise as speech.

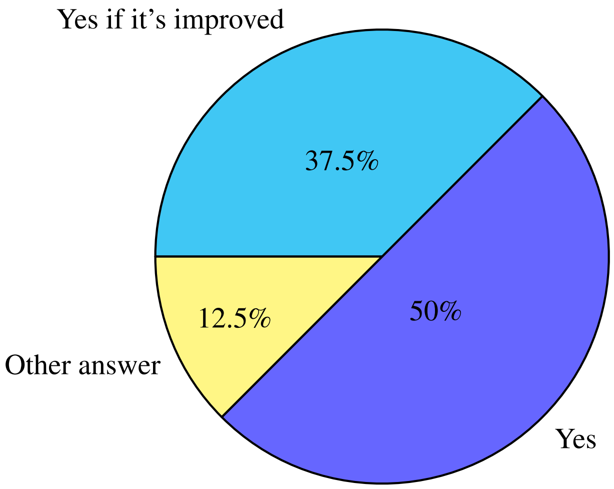

Figure 5: Answers to the question: "Do you think that an Ainu language-speaking robot such as the one in the video could be useful in Ainu language education?"

		Japanese	Ainu
	ACCURACY:	100.00%	100.00%
CONFIDENCE	**Minimum:**	45.90%	39.84%
	Maximum:	76.13%	66.47%
	Mean:	64.14%	54.29%
	Median:	64.21%	54.31%

Table 1: Speech Recognition experiment results

7 Conclusions

This research is a first step in our project to develop an Ainu language-speaking robot, intended for use in language education. In order to demonstrate the concept to potential users, we utilized existing technologies (i.e., the Pepper robot) to create a rule-based dialogue agent capable of holding simple conversations, teaching new words and playing interactive games in Ainu. After presenting the robot to a group of Ainu language learners and experts (in the form of a demonstrational video), we received positive feedback about the idea.

Existing tools for dialogue scripting, such as QiChat, provide an intuitive way to manage closed-domain conversations. We envision that, once a general framework for robot-assisted Ainu language teaching is established, language experts and instructors with basic training in computer programming could easily expand its knowledge base by designing new rules and topics.

At present, there exist no robots with Speech Synthesis and/or Speech Recognition technologies supporting the Ainu language. Leveraging similarities between phonological systems of Ainu and Japanese, in this preliminary work we utilized the Japanese models supplied with Pepper. Evaluation of the robot-generated speech by a group of experts and experienced learners revealed that – due to differences in phonotactics and suprasegmental features – employing a Japanese Text-to-Speech model alone is not sufficient to produce high-quality output (especially if we intend to use the system in language teaching). That being said, the results of both evaluation experiments seem to confirm the potential of cross-lingual transfer from Japanese for facilitating the development of dedicated speech technologies in the low-resource setting of Ainu, in particular in the context of Speech Recognition. We believe that the insights from this research will be useful in that process.

Acknowledgments

We would like to express our gratitude to all participants of our survey who took the time to share their valuable feedback.

References

[Alemi *et al.*, 2014] M. Alemi, A. Meghdari, and Maryam Ghazisaedy. Employing Humanoid Robots for Teaching English Language in Iranian Junior High-Schools. *International Journal of Humanoid Robotics*, 11, 2014.

[Belpaeme *et al.*, 2018] Tony Belpaeme, Paul Vogt, Rianne van den Berghe, Kirsten Bergmann, Tilbe Göksun, Mirjam de Haas, Junko Kanero, James Kennedy, Aylin C. Küntay, Ora Oudgenoeg-Paz, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Christopher D. Wallbridge, Bram Willemsen, Jan de Wit, Vasfiye Geçkin, Laura Hoffmann, Stefan Kopp, Emiel Krahmer, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, and Amit Kumar Pandey. Guidelines for Designing Social Robots as Second Language Tutors. *INTERNATIONAL JOURNAL OF SOCIAL ROBOTICS*, 10(3):325–341, 2018.

[Bugaeva, 2012] Anna Bugaeva. Southern Hokkaido Ainu. In Nicolas Tranter, editor, *The languages of Japan and Korea*, pages 461–509. Routledge, London, 2012.

[Eimler *et al.*, 2010] Sabrina Eimler, Astrid von der Pütten, Ulrich Schächtle, Lucas Carstens, and Nicole Krämer. Following the White Rabbit – A Robot Rabbit as Vocabulary Trainer for Beginners of English. In Gerhard Leitner, Martin Hitz, and Andreas Holzinger, editors, *HCI in Work and Learning, Life and Leisure*, pages 322–339, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[Eun-ja Hyun *et al.*, 2008] Eun-ja Hyun, So-yeon Kim, Siekyung Jang, and S. Park. Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 187–192, 2008.

[Fujii, 1979] Akihiro Fujii. Some notes on Japanese and English intonation. *Kagawa Daigaku Ippan Kyōiku Kenkyū*, 15:119–130, 1979.

[Han *et al.*, 2008] Jeonghye Han, Miheon Jo, V. Jones, and Jun H. Jo. Comparative Study on the Educational Use of Home Robots for Children. *Journal of Information Processing Systems*, 4:159–168, 2008.

[He *et al.*, 2019] Junxian He, Zhisong Zhang, Taylor Berg-Kirkpatrick, and Graham Neubig. Cross-lingual syntactic transfer through unsupervised adaptation of invertible projections. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3211—3223, Florence, Italy, 2019. Association for Computational Linguistics.

[Hong *et al.*, 2016] Zeng-Wei Hong, Y. Huang, Marie Hsu, and Wei-Wei Shen. Authoring Robot-Assisted Instructional Materials for Improving Learning Performance and Motivation in EFL Classrooms. *Educational Technology & Society*, 19:337–349, 2016.

[Kayano, 1996] Shigeru Kayano. *Kayano Shigeru no Ainugo jiten* [Shigeru Kayano's Ainu dictionary]. Sanseidō, Tōkyō, 1996.

[Kennedy *et al.*, 2016] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme. Social robot tutoring for child second language learning. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 231–238, 2016.

[Kory Westlund *et al.*, 2017] Jacqueline M. Kory Westlund, Sooyeon Jeong, Hae W. Park, Samuel Ronfard, Aradhana Adhikari, Paul L. Harris, David DeSteno, and Cynthia L. Breazeal. Flat vs. Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative. *Frontiers in Human Neuroscience*, 11:295, 2017.

[Lee *et al.*, 2011] Sungjin Lee, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, Gary Geunbae Lee, Seongdae Sagong, and Munsang Kim. On the effectiveness of Robot-Assisted Language Learning. *ReCALL*, 23(1):25–58, 2011.

[Lewis *et al.*, 2016] M.P. Lewis, G.F. Simons, and C.D. Fennig (Eds.). Ethnologue: Languages of the World, Nineteenth edition, 2016.

[Long, 1996] M. H. Long. The role of the linguistic environment in second language acquisition. In W. C. Ritchie and T. K. Bhatia, editors, *Handbook of second language acquisition*, pages 413–468. Academic Press, New York, 1996.

[Mackey, 1999] Alison Mackey. INPUT, INTERACTION, AND SECOND LANGUAGE DEVELOPMENT: An Empirical Study of Question Formation in ESL. *Studies in Second Language Acquisition*, 21(4):557–587, 1999.

[Matsuura *et al.*, 2020] Kohei Matsuura, Sei Ueno, Masato Mimura, Shinsuke Sakai, and Tatsuya Kawahara. Speech Corpus of Ainu Folklore and End-to-end Speech Recognition for Ainu Language. *ArXiv*, abs/2002.06675, 2020.

[Meiirbekov *et al.*, 2016] S. Meiirbekov, K. Balkibekov, Z. Jalankuzov, and A. Sandygulova. "You win, I lose": Towards adapting robot's teaching strategy. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 475–476, 2016.

[Murasaki, 2009] Kyōko Murasaki. *Karafuto Ainugo nyūmon kaiwa / First Step for the Sakhalin Ainu Language*. Ryokugeisha, Kushiro, 2009.

[Nakagawa, 2013] Hiroshi Nakagawa. *Nyū ekusupuresu Ainugo*. Hakusuisha, Tōkyō, 2013.

[Randall, 2019] Natasha Randall. A Survey of Robot-Assisted Language Learning (RALL). In *ACM Transactions on Human-Robot Interaction*, volume 9, December 2019.

[Refsing, 1986] K. Refsing. *The Ainu language. The morphology and syntax of the Shizunai dialect*. Aarhus University Press, Aarhus, 1986.

[Satō, 2008] Tomomi Satō. *Ainugo bunpō no kiso* [basics of Ainu grammar]. Daigaku Shorin, Tōkyō, 2008.

[Shibatani, 1990] M. Shibatani. *The languages of Japan*. Cambridge University Press, London, 1990.

[Shin and Shin, 2015] Jae-eun Shin and Dong-Hee Shin. Robot as a facilitator in language conversation class. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, HRI'15 Extended Abstracts, page 11–12, New York, NY, USA, 2015. Association for Computing Machinery.

[Tamura, 1996] S. Tamura. *Ainugo jiten: Saru hōgen. The Ainu-Japanese Dictionary: Saru dialect*. Sōfūkan, Tōkyō, 1996.

[Tanaka and Matsuzoe, 2012] Fumihide Tanaka and Shizuko Matsuzoe. Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, pages 78–95, 2012.

[Uehara and Abe, 1804] K. Uehara and Ch. Abe. *Ezo hōgen moshiogusa* [Ezo dialect dictionary]. 1804.

[van den Berghe *et al.*, 2019] Rianne van den Berghe, Josje Verhagen, Ora Oudgenoeg-Paz, Sanne van der Ven, and Paul Leseman. Social Robots for Language Learning: A Review. *Review of Educational Research*, 89(2):259–295, 2019.

[Wang *et al.*, 2013] Yi Hsuan Wang, S. Young, and J. Jang. Using Tangible Companions for Enhancing Learning English Conversation. *Journal of Educational Technology & Society*, 16:296–309, 2013.

[Watanabe, 2018] Kaori Watanabe. Shuto-ken ni okeru Ainu-go kyōiku no genjō [Current situation of Ainu language education in Tokyo metropolitan area]. *Journal of Chiba University Eurasian Society*, 20:229–251, 2018.

[Wit *et al.*, 2018] J. Wit, Thorsten Schodde, Bram Willemsen, K. Bergmann, M. Haas, Stefan Kopp, E. Krahmer, and P. Vogt. The Effect of a Robot's Gestures and Adaptive Tutoring on Children's Acquisition of Second Language Vocabularies. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018.