

Can we use a spoken dialogue system to document endangered languages?

Seyed Hossein Alavi and Jacqueline Brixey* and David Traum

Institute for Creative Technologies

University of Southern California

{seyedhoa, brixey, traum @ict.usc.edu }

Abstract

This paper investigates using a dialogue system to preserve endangered languages, and the viability of a multilingual dialogue system to generate a general use corpus of audio responses in . We introduce DAPEL (Dialogue APP for Endangered Languages). DAPEL elicits responses from speakers of endangered languages by having a conversation with them. We conducted a pilot user study to examine the efficacy of using an automated system like DAPEL versus a human interviewer. We also studied the effects of engaging in small-talk in a different language in between recording prompts for the target language.

1 Introduction

Humanity is facing the threat of many languages becoming extinct. With every language that dies, we lose cultural heritage and history. Documentation of endangered languages provides a means to preserve endangered languages, provide revitalization materials to prevent further loss, and serve as a resource for research.

There can be many obstacles encountered when documenting an endangered language. For some of the world's languages that are primarily oral-only languages, there are few if any written records or publications in the target language. There are also inherently few speakers of an endangered language, making it difficult to find speakers of whom to record. Additional obstacles are the human labor and financial expense involved in traveling to the communities of language speakers to collect data.

In this paper we introduce *DAPEL*, a spoken dialogue mobile app for documenting endangered languages, as a possible means for collecting audio recordings of endangered languages. Responses recorded by DAPEL will form a general use corpus that could benefit language revitalization efforts, as well as research in computer science, linguistics, and anthropology.

Our proposed system is designed to collect data from speakers from multiple languages, however, speakers must be English-bilinguals in order to understand the dialogue. We recognize that a multilingual system could be confusing, however it is data intensive to build dialogue systems that recognize, understand, and generate language, and such data does not already exist for most endangered languages. As a result, such systems currently exist in only a few majority languages. In order to capture as many endangered languages as possible, the system interacts with all users in English.¹ Using English as the language for communication in DAPEL means that individuals of many diverse linguistic backgrounds will be able to use it. One of our motivations for using English as the common languages for DAPEL is because at least one in four of the world's population is capable of communicating in English (Seidhofer, 2013). A second reason for using English is that loss of endangered languages in English majority speaking countries is occurring at a higher rate than other places globally. Research indicates that endangered languages in the United States, Canada, Australia have experienced loss rates of 75% since 1950 to present times, while regions such as sub-Saharan Africa have had loss rates of

¹It would be straightforward to change English to another large-resource language, like Spanish, Chinese, or Arabic, which might be preferable to address endangered languages where bilingual speakers speak a something other than English.

*equal contribution

less than 10% (Simons and Lewis, 2013).

To our knowledge, no dialogue systems exist to collect audio data of endangered languages. Nor do we know of any already existing dialogue systems in which the user is expected to speak two languages during a session but only interact with the system in one. We present a pilot study to investigate some of the questions related to viability of the proposed system.

2 Related work

To design DAPEL, we were informed by research on social dialogue, interviewing systems, and linguistic methods for collecting spoken language data.

The research and documentation of endangered languages is vast. The United Nations declared 2019 International Year of Indigenous Languages, aiming to raise awareness of the endangerment of indigenous languages in the world. Conferences, such as the International Conference on Language Documentation & Conservation (ICLDC), have been formed to promote research and connect researchers in the field. Seifart et al. (2018) gives a survey of the considerable body of work conducted over the past twenty five years in the field. Nevertheless, around 35-42% of the worlds languages remain substantially undocumented (Seifart et al., 2018).

The use of apps to collect data remotely has been implemented in previous works such as Aikuma (Bird et al., 2014), a mobile app in which users re-speak previous recordings and give oral translations. An app deployment for a recording system could alleviate some challenges and costs, as many people own smart phones, and remote speakers could then make recordings as their schedule permits. However, the most natural use of language is to communicate with others, including answering questions and sharing similar experiences, in a dialogue.

Dialogue systems could provide a way for speakers of endangered languages to produce audio data. Interviewing systems, such as the Sensesi Kiosk system (DeVault et al., 2014), have been shown to build rapport with interviewees, and encourage users to talk as much as possible per turn and throughout the interaction. Dialogue systems that engage in “small talk” have also been shown to increase user rapport (Bickmore and Cassell, 1999). Small talk is defined as

neutral topics that are not related to a task. Small talk in dialogue functions as a means for interlocutors to establish an interaction style (Bickmore and Cassell, 1999), and develop “communal common ground”(Clark, 1996).

3 The DAPEL system

DAPEL is proposed to be deployed as a smartphone app interface. As many people have smartphones nowadays, deploying our system as an app will make it easier to collect data because remote speakers will be able to make recordings as their schedule permits on their own phones. Our proposed system will gather audio data from bilingual people who speak in English and an unspecified endangered language.

In order to collect data from multiple languages and fulfil our goal of equally working to preserve all endangered languages, DAPEL was designed to request information from users in a lingua franca (initially English). To our knowledge, there are no existing multilingual dialogue systems, nor systems in which the agent asks the user to speak in two languages but only interact with the agent in one. Our work will answer if users find this a comfortable and viable conversational paradigm. DAPEL aims to collect lengthy and linguistically rich recordings from users in their respective languages. We included small talk to not only foster engagement and comfort with the system, but to also encourage longer responses.

In DAPEL (visualized in Figure 1), the system first provides an interview prompt, which the user responds to in their endangered language. The interview prompts are on a range of topics and are designed to illicit a diversity of syntactic and lexical responses. The system then asks for a summary of what the user said in response to the interview prompt. The summary act serves as a translation, and as a means to select relevant topics for small talk. The small talk topics were scripted ahead of time. Following the summary, the system engages in at maximum two small talk turns in English. An example dialogue with a recording topic would be as follows:

System [recording topic]: What sports do you like to play or watch?

User: Pues, cuando era niño, me gustaba jugar fútbol. Pero ahora de adulto, prefiero ver fútbol americano en la televisión.

System: Could you give a summary in English

of what you discussed in your recording?

User: I said that I used to like to play football, but nowadays I only watch it on the tv.

System: Oh football! My brother plays that! He isn't very good but he seems to have fun. I only like to watch football.

System: Who is your favorite team?

User: New England Patriots.

System: I hope they have a great season!

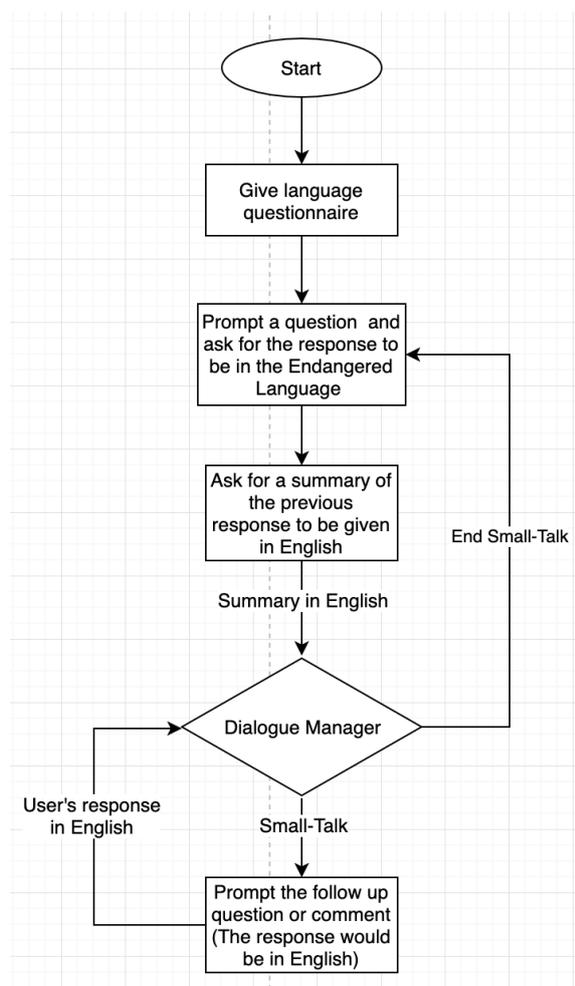


Figure 1: Flow chart of DAPEL.

4 Experiment

DAPEL would be the first multilingual system designed to collect endangered language audio recordings. It is unclear, however, if its proposed dialogue design will produce recordings as good as those with a human interviewer, if the multilingual design is comfortable or confusing, and if small talk will contribute to or detract from the quality of the audio recordings. We designed a pi-

lot experiment with three conditions to determine if our proposed system could be effective at collecting responses. Our research questions are:

1. Do people respond with lengthy and meaningful responses when prompted by a computer to record answers in an endangered language?
2. Is a summarizing act natural for the user and the dialogue flow?
3. Does social chat help users feel comfortable with the interaction? Are the small talk responses adequate?

Choosing English as the main interaction language might be a mental burden to switch back and forth between two languages for the user. In order to clarify if this design would work, we defined the first condition in the experiment in which a human interviewer was paired with a participant. The interviewer introduced the prompt, asked for a summary, and then gave at least two small talk responses from a prepared script based on the summary given.

In the second and third conditions, we conducted a Wizard of Oz (WOZ) study to evaluate how users would respond to an equivalent interaction when paired with a computer interviewer. In the second condition, prerecorded small talk questions from the same script as condition one were played to the user based upon the summaries. In the third condition, the small talk portion of the dialogue was omitted.

Following the interaction, all participants were given a questionnaire to discuss their experience.

We recruited 18 participants, with six participants per condition. Nine females and nine males participated. We did not recruit endangered language speakers due to time constraints, but we expect the results could be generalized to most language speakers. Participants (number given in parentheses) spoke: Korean (1), Mandarin (4), Bengali (2), Spanish (2), Persian (2), French (1), Japanese (1), Dutch (1), Hebrew (1), Hindi (1), and Russian (1).

5 Results and Discussion

To answer our first research question, we measured the total duration of the time users spoke in their non-English language. Our pilot study found that participants spoke nearly as much with

#	Questionnaire results	Cond. 1	Cond. 2	Cond. 3
1	How natural did you find the interaction overall?	4.17	3.5	3.83
2	How easy was it to understand what was expected of you?	4	4	3.83
3	How natural was it to provide a summary of what was said during the recording prompt?	3.83	4	
4	How comfortable was the small talk?	4	3.83	

Table 1: Sample of questionnaire results. Participants were asked to rate from 1 to 5.

the computer system as with the human interviewer. The average duration speaking times in the non-English language was 418 seconds, 228 seconds, and 382 seconds in conditions one, two, and three, respectively. The average duration totals decreased in condition two due in part to a participant who spoke for five seconds per response. The median times were 383 seconds, 234.5 seconds, and 359 seconds, in each condition respectively. The results demonstrate that users will interact with a multilingual system, and give recorded responses in their respective languages.

For the second research question, question #3 in Table 1 shows that participants in the experiment did not find the summarizing act unnatural, and users are accepting of systems that ask them to perform this action. Although the difference might not be significant, it is interesting that users found summarizing in condition 2 more natural than condition 1. This might seem counter-intuitive at first since in condition 2 users were interacting with a WOZ system, while in condition one they were interacting with a human interviewer. This may be due to the fact that in the first part when they are speaking in their non-English language, they know that the human interviewer does not understand them at all, which might seem unnatural. People look for reactions from the other person in their normal conversations. It may be easier to interact with a system without any expectation a reaction.

For the final research question, the participants did not find the system’s small talk unnatural and all of them reported feeling neutral, positive or extremely positive about that. The small talk responses were scripted for both the human interviewer and WOZ system, however participants in condition 2 reported that the WOZ small talk responses sounded formulaic and did not always appropriately respond to the summary. This feedback might also explain why total speaking duration for condition 2 was lower than the other two conditions, as well as less natural.

A guiding research question behind our pilot study and dialogue design was “how does dialogue contribute to the user’s experience, and impact willingness to participate and disclose?” We found that users are willing to speak in two languages with a system, as well as disclose information about their language and culture. We also determined that users are willing to engage in summaries and small talk with a dialogue system, so long as the small talk responses are engaging and appropriate. The results indicate that DAPEL could be a viable system for collecting endangered language responses, even though the system engages in all dialogue acts in English.

6 Conclusion

We propose that DAPEL provides an easy and cost effective way to collect data for endangered languages. DAPEL can be implemented as a mobile app or a web application. Many people in the world have smart-phones, making it possible for languages that have not received attention for study to be recorded. We imagine that the corpus of audio recordings that the app collects could provide research opportunities in linguistics, anthropology, and computer science. Since the content collected is limited, we imagine the corpus could be used as a way for researchers to be introduced to a language. Cross-language studies would also be possible as all users would respond to the same recording prompts. The pilot study showed initial viability of the approach, but also some challenges for leveraging the impact of an automated system. As future work, we intend to improve the small-talk facility to include a wider range of feedback. We also will test whether automated NLU and dialogue management will be able to engage appropriately in dialogue and realize the potential indicated by the human and wizard conditions. We then intend to test in actual deployment capabilities with endangered languages.

References

- Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*, pages 87–92.
- Steven Bird, Florian R. Hanke, Oliver Adams, and Haejoong Lee. 2014. Aikuma: A mobile app for collaborative language documentation. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 1–5.
- Herbert H Clark. 1996. *Using language*. Cambridge university press.
- David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems.
- Barbara Seidlhofer. 2013. *Understanding English as a Lingua Franca-Oxford Applied Linguistics*. Oxford University Press, Oxford, United Kingdom.
- Frank Seifart, Nicholas Evans, Harald Hammarström, and Stephen C Levinson. 2018. Language documentation twenty-five years on. *Language*, 94(4):e324–e345.
- Gary F Simons and M Paul Lewis. 2013. The worlds languages in crisis. *Responses to language endangerment: In honor of Mickey Noonan. New directions in language documentation and language revitalization*, 3:20.