

# Can code-switching improve the user experience with a dialogue system app for recording endangered languages?

Jacqueline Brixey\*

University of Wisconsin-Madison  
brixey@wisc.edu

David Traum

USC Institute for Creative Technologies  
traum@ict.usc.edu

## Abstract

This paper investigates whether a multilingual spoken dialogue system can be used to help collect and preserve endangered language data. In this work, we extend DAPEL (Dialogue APP for Endangered Languages), which is designed to help preserve any language. Our focus, for testing purposes, is on the American Indigenous language Choctaw. The system uses English as a common language, and we test whether incorporating code-switching—the act of alternating between languages—enhances the user experience and/or increases the amount of recorded language data. Our results indicate that users have a positive response to interacting in both languages with the system, that the system plays a meaningful role in language documentation, and, notably, that participants who speak Choctaw as their first language are more receptive to a code-switching system than to a monolingual English-based system.

## 1 Introduction

There are roughly 7,000 languages spoken in the world today. Seifart et al. (Seifart et al., 2018) found that ‘around 3,660—that is, more than half of now living languages—are currently threatened, endangered, moribund, or nearly extinct’. Many languages have not been documented or minimally documented (Gauthier et al., 2016). With many languages facing decreasing populations of speakers, it is imperative to document languages while there are still fluent speakers. However, there are numerous bottlenecks in the process of preserving a language, such as the labor, cost, and time involved in recruitment and conducting recordings of fluent speakers. Technology could be a means to efficiently and economically document a language. A dialogue system for language preservation would fulfill a similar role as a language preservation practitioner conducting an interview with a speaker and

would reduce the challenges of documentation. Additionally, speakers could use the system when their schedules permit, making capturing data from additional speakers easier. An added benefit is that users have been shown to disclose more information to a dialogue system than with a human (Lucas et al., 2014), which could indicate that people would be willing to speak more and be recorded saying more in the endangered language than with a human interviewer. A challenge for this kind of system, is that it may be difficult to build a monolingual system in the target language, because the main reason for deploying the system is because there is a lack of sufficient data in the language. (Alavi et al., 2019) showed that bilingual speakers would be willing to speak in two languages with a system that only spoke one. The system, called DAPEL (Dialogue APP for Endangered Languages), was designed for language preservation of Endangered Languages, but had not been extensively tested with these languages or their speaker populations.

In this paper, we report on testing of (a modified version of) the DAPEL system with speakers of an endangered language, Choctaw, showing that it can succeed in collecting new data from an endangered language. We modified the user interface of the system to be presented as a web page, and other enhancements for usability, as described in Section 3.2. Additionally, we investigate the impact on the user of the system employing code-switching—the use of two languages in conversation—in its utterances. We hypothesize that since the users are bilinguals, as they must understand and speak English and an endangered language, it is possible that the system using more of the endangered language could lead to a better user experience and/or more recorded audio language data.

In the next section, we review prior work on language preservation including DAPEL and other technology to assist this task. In section 3, we

\*Work completed while at USC Institute for Creative Technologies

lay out the methods for our study on Dapel’s efficacy in preserving conversational Choctaw and the impact of system code-switching, including specific research questions, modifications to the Dapel system, codeswitching framework, evaluation metrics, and experiment participants and protocols. In Section 4.1, we examine the results of the study, including survey responses from participants, comparing monolingual vs code-switching versions and L1 (native) vs L2 speakers. We also analysed the contents of Choctaw speech, including duration, number of unique and new words encountered. We conclude in section 5, noting that the best system design may depend on the type of user, and prospects for incorporating LLMs for endangered language generation.

## 2 Related literature

The field of linguistics has a long history of endangered language documentation. In the early tradition in America, linguists and anthropologists created written and oral records of Indigenous languages, typically producing three items intended and structured for an academic audience: a dictionary, a grammar, and a set of texts of the language. The language use documented in this tradition often prioritized literary and ceremonial domains and rarely documented conversational or everyday language (Rouvier, 2017).

In the 1980s, as language communities began to use documentation records to support language revitalization efforts, two key changes in methodologies occurred. First, language documentation became an important resource for combating language shift (Dobrin et al., 2009). Second, communities have come to play a more central role in determining documentation practices and the resulting documents, with the expectation that language documentation should support community priorities for revitalization, and that funding opportunities should contribute to these efforts, even when other research goals are present within the preservation activity (Rouvier, 2017).

Despite the long tradition of language documentation, there is “a minimally adequate quantity of data for less than 1% of the world’s 7000 languages” (Gauthier et al., 2016). Technology can play an important role in scaling up documentation efforts, allowing users to be recorded simultaneously and as their schedule allows while reducing prohibitive cost aspects such as travel. Additionally, technol-

ogy has fostered greater audio and video documentation. Finally, technology has encouraged better sharing of primary language data with researchers and community members (Rouvier, 2017).

Previous systems that aimed to document a language include AIKUMA (Bird et al., 2014) and LIG-AIKUMA (Gauthier et al., 2016). The AIKUMA app (Bird et al., 2014) was designed to record parallel translation data in English and a second language. The AIKUMA app is deployed on smartphones and shows the viability of deploying mobile apps to document endangered languages. Members of the speech community can upload written translations. LIG-AIKUMA (Gauthier et al., 2016) added functionality to the AIKUMA app. Users could again translate speech and record themselves in spontaneous speech alone or with others, introducing novel texts and reading them aloud or re-speaking a previously recorded audio clip from another speaker. LIG-AIKUMA was deployed to record three African languages (Adda et al., 2016). Unlike DAPEL, LIG-AIKUMA does not directly engage with the speaker in dialogue or offer spoken conversational prompts.

The general format of DAPEL is an interview-type system. One previous interview system that informed the design of DAPEL interviewed participants to detect the presence of psychological distress indicators, such as those for PTSD and depression (DeVault et al., 2013). The virtual agent in this system, named Ellie, asked open-ended questions, such as “What would you say are some of your best qualities?” and, “What are some things that make you really mad?” A second system (Johnston et al., 2013) created an automated spoken dialogue system to communicate over the phone that asked questions drawn from government and social science surveys. Their motivation was to standardize the interview experience across participants by lowering the error and bias that human interviewers can introduce in survey results data. The experiments in this study were primarily focused on strategies for confirmation but did find that participants were satisfied with over-the-phone interviews with the dialogue system. Additionally, previous dialogue systems found that including small talk can lead to positive user impressions of dialogue systems (Kobori et al., 2016; Cassell et al., 1999). A final study (Nakamura et al., 2019) found that users strongly preferred the system with small talk to the system without.

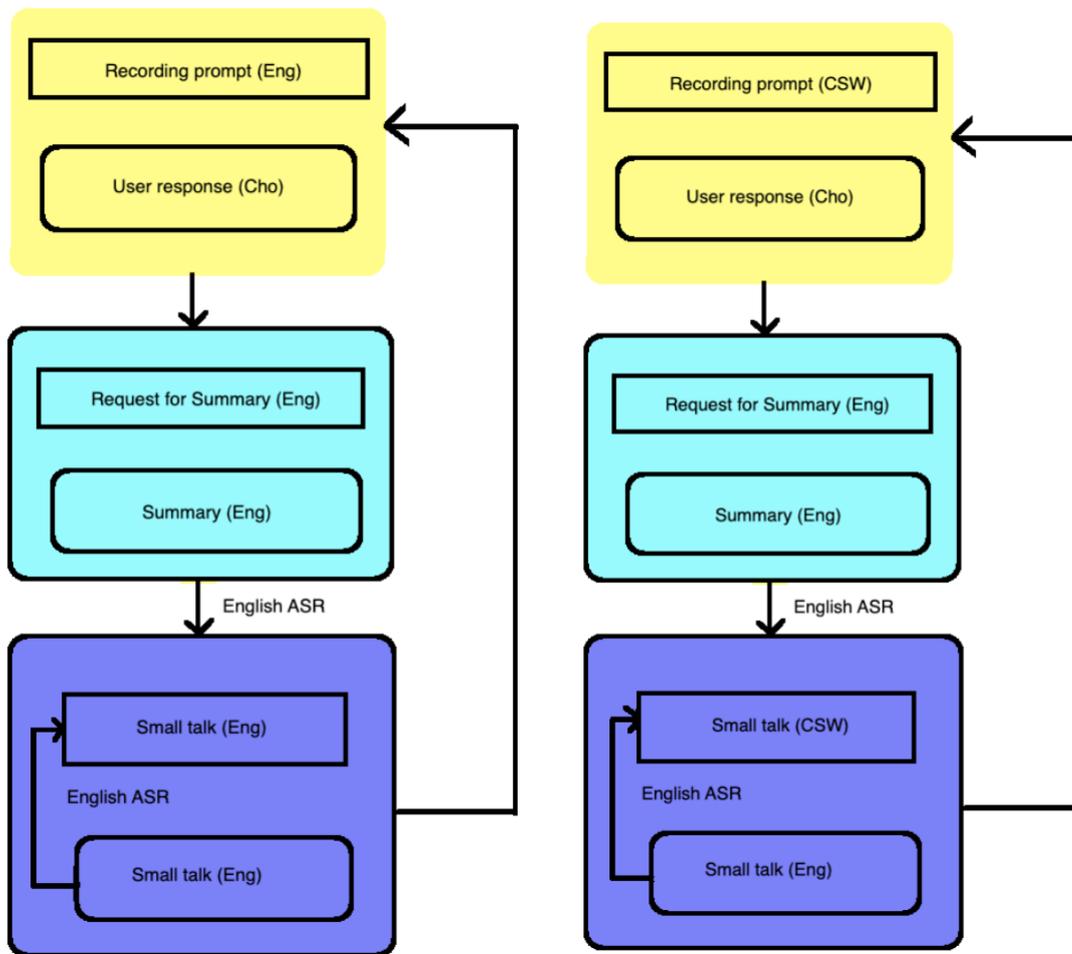


Figure 1: DAPeL system dialogue design. Colors denote parts within the conversation sequence, such as the summary part (light blue). The figure on the left is of the monolingual English system, while the figure on the right shows code-switching (in the prompt and small talk sections).

### 3 Methods

The experiment described in this work was reviewed and approved by the Choctaw Nation’s Institutional Review Board as well as the authors’ university IRB.

#### 3.1 Research questions

We address the following research questions:

1. Can DAPeL be used to collect data in endangered languages (specifically Choctaw).
2. Will users express satisfaction in using such a system?
3. Will a Code-switching version of DAPeL be more successful than a monolingual English system on either of the above?
4. Are there differences between speaker populations (Fluent first language speakers vs those

who learned Choctaw as a second language) with respect to effectiveness or user satisfaction?

#### 3.2 System design

The overall interaction of DAPeL is shown in Figure 1. All of the utterances said by DAPeL were handcrafted and recorded in advance. The dialogue sequence of DAPeL consists of a round that contains a prompt, summary, and small talk. Prompts are open-ended topics that the participant responds to in the non-English language and are intended to be broad enough for a person to be able to speak at length. The summary, typically phrased as, “Could you summarize what you just said?” is responded to in English and is where the user provides a brief overview of what they said to the prompt. The final part, small talk, is where the system can ask additional questions or engage in chit-chat in En-

1	How did you learn to speak Choctaw, at home or at school? Tell me where and with whom you speak your language nowadays.
2	Can you share a good story your grandparents or other family members told you about their lives?
3	Who are you closest with in your family? Why is that?
4	If you could go back and talk to one ancestor, who would it be and why?
5	Do you have any siblings? How do you think being the youngest, oldest, middle, or only child affected you?
6	What's the oldest possession you currently have? Why do you still have it?
7	Can you tell me a story from your childhood?
8	What was the most interesting job you ever had?
9	What's the best piece of advice you ever received?
10	What was the strangest thing or coolest object in your childhood home? Does it have a story?
11	If money were no issue, how would you spend your time?
12	What's one thing that someone borrowed from you and never brought back, do you miss it?
13	Did you ever play any musical instruments when you were growing up? If you didn't, what would you like to learn? Tell me something about the instrument.
14	What are some of your favorite things to cook or eat? What are things that you do not like to cook or eat? Just to note, if you are sharing a cultural food, please only discuss it if it is ok for people not from your community to know about.
15	Tell me about a vacation place that you would like to visit and why.
16	What animals are you afraid of? Why?
17	What was your favorite tv show when you were a child? Tell me something about the show.
18	Do you like to read? Why or why not?
19	How important do you think science and math courses are in school, in comparison with literature, technology, arts, and history classes?
20	What is the most disgusting vegetable to eat, in your opinion? Why?
21	What piece of technology are you the most reliant on, and why?
22	What sports do you like to play or watch?

Table 1: List of all prompts

glish with the user based on what was said in the summary.

The system was presented as a webpage to the user. Each part of the dialogue was presented on a separate page: the prompt on one page, the summary on the next, and the small talk on a subsequent page. This design was intended to make the different parts of the dialogue clear to the user.

The prompt and summary recordings had an explicit "begin recording" button. This was so that participants did not feel pressured to begin talking immediately but could instead reflect on the question, make any notes, and look up words in the provided dictionaries.

The small talk portion of the dialogue was recorded during each experiment session on an external recording device. This device was left running for the duration of the experiment session to serve as a backup method if audio was not captured on the computer with which the participant interacted.

The prompts were questions selected and adapted from an online learning resource<sup>1</sup>. To ensure that the system has sufficient content, twenty-two prompts were selected, given in Table 1.

Apart from capturing a variety of vocabulary, the prompts were selected to elicit various syntactical

forms. Prompts such as 1 were intended to capture the present tense. Other prompts, such as 2 would capture the past tense. Prompts such as 15 could capture hypothetical-type tenses and conditionals. Another syntactical feature the prompts intended to potentially capture was negation forms, as in prompt 18.

To reduce user fatigue, small talk was present for only some rather than all questions. If there was a small talk option, one option was that the summary from the previous page was converted to text. The system then selected a response based on keywords in the text. For example, if the participant said they liked reading, the system asked what their favorite book was. Most small-talk interactions were one or two additional turns. For example, in prompt 15 ("Tell me about a vacation place that you would like to visit and why.") the system would first say, "That sounds like a good choice! Do you know anyone who has been there before?" Small talk included both affirmative and negative types.

### 3.3 Code-switching design

To design the code-switched utterances, we utilized a framework outlined in (Parekh et al., 2020). A code-switched utterance could be one of the three options illustrated in Table ??.

A switched phrase would typically be a noun phrase in the opposite language. To create the alter-

<sup>1</sup><https://relearnalanguage.com/language-exchange-topics/>

1. Choctaw phrase phrase in English matrix	How did you learn to speak Chahta? (How did you learn to speak Choctaw?)
2. English to Choctaw alternation	Who are you closest with in your family? Yummut katimi a? (Who are you closest with in your family? Why is that?)
3. Choctaw to English alternation	Chukka cho holisso apisa? Tell me where and with whom you speak your language nowadays. (How did you learn to speak Choctaw, at home or at at school? Tell me where and with whom you speak your language nowadays.)

Table 2: Code-switching options in DAPEL prompts and small talk. Examples are given in the column on the right.

nation utterances, switches occur at a conjunction point, a comma, before a verb, or inter-sententially.

### 3.4 Survey design

The survey utilized in this experiment was designed to evaluate the user’s sense of rapport, the naturalness of the code-switching, and the feeling of connection because of language identity. The survey consisted of thirteen 5-point Likert scale questions, and the answers were scored from 1 strongly disagree to 5 strongly agree. The final two questions were open-ended questions where participants could write sentences to respond. All survey questions were optional, and participants were informed that they could choose to skip any questions.

The full list of questions can be seen in Table 3 in the results section. Questions were selected to determine levels of rapport (1, 2, 4, 5, 6, 9) and engagement and connection (3, 8, 10, 11, 12, 13). Question 7 elicited user impressions about the system’s knowledge of Choctaw.

### 3.5 Recruitment and Participants

Participants were recruited at three community centers on the Choctaw Nation reservation in Oklahoma, and at the School of Language in the Choctaw Nation headquarters in Durant, OK. Participants self-identified as being capable of carrying out a conversation in the Choctaw language.

### 3.6 Experiment sessions

All experiments were conducted in person on the Oklahoma reservation. The experiment session lasted thirty minutes, and participants could con-

tinue voluntarily. In each session, participants reviewed and signed a consent form, then interacted with the system, and finally completed a post-interaction survey.

## 4 Results

In total, 28 participants participated in the study, 14 per condition. No data was collected about the participant’s age or gender identity. To measure the experience, we analyze responses to the post-interaction survey. We also compare the average duration of the collected recordings across conditions.

### 4.1 Survey results

Results of the survey are shown in Table 3. We can see high scores for both rapport and engagement for both systems. For the negatively framed questions (2, 5, and 6), a low score indicates disagreement, thus a score of 1 for question 2 would indicate that the user did not find the system engaging. Generally, the codeswitching system received more positive scores on average. Using a two-tailed T-test, we first compared the responses to each survey question of the code-switching group against the monolingual group. P-value results are shown in the "all" column. Values with asterisks indicate statistically significant results below the 0.05 threshold, while values below 0.10 are shown in bold. Though most differences are not significant, this is expected given the small sample size, and may inform future research.

It became apparent during recruitment that there were two different cohorts within the participant population. One cohort comprises 13 second-language Choctaw speakers, all of whom were recruited in Durant. The other cohort is 15 individuals whose first language is Choctaw. We then divided the survey results based on these two cohorts. Some interesting results were observed by separating survey responses by cohort, as seen in "L1" and "L2" columns in Table 3. The L1 participants were significantly more likely to recognize the code-switching system as a Choctaw speaker (0.04) and as part of their social group (0.04). Notably, the L2 cohort did not have this same level of recognition of the code-switching system being a Choctaw speaker.

We then compared the cohorts, shown in "Mono L1 v L2" and "CSW L1 v L2" columns in Table 3. In the monolingual cohort ("Mono L1 v L2"

Question	average score		T-test p values Mono vs CSW			T-test p values L1 vs L2		
	Mono	CSW	All	L1	L2	Mono	CSW	
	1	The system understood me.	3.5	4	<b>0.09</b>	0.64	0.20	0.72
2	The system seemed unengaged.	3.1	2.6	0.47	0.20	0.53	<b>0.01*</b>	0.65
3	The system was friendly.	4.8	4.8	1	0.75	0.67	0.42	0.93
4	The system and I worked towards a common goal.	4.1	4.4	0.56	0.45	0.53	0.26	<b>0.08</b>
5	The system and I did not seem to connect.	2.5	2	0.34	0.49	0.38	0.70	0.66
6	I didn't understand the system.	2.4	1.9	0.16	0.87	0.11	0.73	0.29
7	The system knows the Choctaw language.	2.9	4.1	<b>0.06</b>	<b>0.04*</b>	0.14	0.69	<b>0.02*</b>
8	The interaction was interesting.	4.6	4.6	0.75	0.60	0.89	0.77	0.45
9	The interaction felt natural.	3.7	4.1	0.52	0.23	0.71	0.65	0.10
10	I enjoyed the small talk.	4.5	4.4	0.72	0.17	0.88	0.33	<b>0.02*</b>
11	I felt the system and I were in the same social group.	3.5	4.4	0.46	<b>0.04*</b>	0.74	0.71	<b>0.002**</b>
12	I would be willing to continue the conversation with the system for longer.	3.9	3.9	1	0.91	0.62	0.17	0.37
13	I would recommend interacting with this system to a friend.	4.2	4.1	0.87	0.37	0.35	0.19	0.55

Table 3: Survey results by system version, with significance tests comparing results based on system versions and user fluency. Text in bold indicates  $p < 0.10$ ; a single asterisk indicates  $p < 0.05$ ; two asterisks indicate  $p < 0.005$ .

	Mono [14]	Csw [13]	L1 [13]	L2 [14]
Average recording duration for prompts	0:05:16 (0:03:13)	0:08:56 (0:17:10)	0:08:50 (0:16:41)	0:05:05 (0:02:02)
Total recording time for prompts	1:13:42	1:56:02	1:06:10	2:03:34
Average recording duration for summaries	0:03:35 (0:02:32)	0:04:30 (0:08:24)	0:03:22 (0:01:52)	0:04:39 (0:08:14)
Total recording time for summaries	0:50:16	0:58:35	0:43:43	1:05:08

Table 4: Audio durations for prompts are given in hours:minutes:seconds. Standard deviations are given in parentheses. The number of participants is given in the column headers in square brackets; the groups are not even as L1 and L2 were not explicitly recruited, and one participant from the CSW group was excluded due to technical recording issues.

column), we compared the survey results of the L1 cohort (eight participants) against the L2 cohort (six participants) that interacted with the monolingual system. The L1 cohort was significantly more likely to rate the monolingual system as unengaged than the L2 cohort (0.02) paired with the same system. The code-switching cohort ("CSW L1 v L2" column) makes the same comparison, with five L1 speakers and nine L2 speakers. The L1 group thought the system worked with them towards a common goal (0.08), felt the system knew Choctaw (0.02), enjoyed the small talk (0.02), and significantly felt the system was in the same social group as them (0.002).

## 4.2 Audio duration

Analyzing the audio durations presented several challenges. First, not all participants completed the same prompts. Participants were encouraged to skip any they did not feel comfortable or interested in answering. There is no prompt that all participants responded to make a balanced comparison.

Second, some participants had additional free

time to continue the interaction for longer, such as Participant 23, who was the only participant to reach the final prompt, as this individual had no other time commitments. Participant 23 also gave the most prolonged responses for most prompts, which heavily skewed overall durations and averages for the code-switching group in which they participated. A second factor was that some participants could not continue longer even if they wished to because other participants were scheduled immediately after their experiment session.

Finally, the external audio recorder malfunctioned when capturing small talk for participants 1 through 12. As a result, only half of the participants' small talk was recorded and could be analyzed for duration.

An overview of the average duration for prompts and summaries is given in Table 4. One participant was excluded from the L2 group and CSW group due to technical recording issues. The average and total recording durations for both prompts and summaries were generally longer in the code-switching group, but as the standard deviations

indicate, there was an enormous variation in the durations observed between individuals. In the code-switching group and L2 cohort, the variation was led mainly by one participant who spoke more per response and responded to nearly all of the prompts.

There was limited statistical significance found comparing groups (L1 v L2) and conditions (Mono v CSW) for total recording time for a given prompt and for the durations for prompt responses from individuals. This indicates that all participants felt equally engaged with all of the prompts, regardless of system condition.

### 4.3 Review of language collected

Over 1,500 unique Choctaw words were said in the total collected audio, a substantial contribution to the documentation of an endangered language. The system captured roughly 500 new words that were not present in one Choctaw-language data set, the ChoCo corpus (Brixey and Artstein, 2021). For comparison, the ChoCo corpus contains roughly 300,000 tokens. Many of the 500 new words were inflected forms, which were not attested to in previous Choctaw language dictionaries (see (Byington, 1915) or (The Choctaw Nation of Oklahoma Dictionary Committee, 2016)), however, were verified to be valid new words by the transcriptionist.

Many participants engaged in code-switching in their responses, many of which were noun-phrase switches. There were several interesting code-switched user examples from the collected audio. One participant borrowed the word "story" from English and formed a contraction with the subject pronoun "ut", creating the inflected word "story't". The framework linguistic literature stated that morphemes and nouns would not be code-switched (Poplack, 2000), and this example shows a noun phrase with the subject marker in Choctaw while the noun is in English. Additionally, the contraction would not be considered acceptable by the literature.

One interesting finding was that many people spoke informally with the system. It is common to shorten words in Choctaw in informal settings, and this was noted in many responses, for example, saying *chukma* instead of *achukma* ("good"). The Choctaw transcriptionist (a fluent speaker) also noted that many people spoke in a slang style, frequently substituting "g" in places of "k," resulting in words like *achugma* instead of *achukma*. A second general finding was the people were eager to

		Word Count	Unique Word Count
1	High total	3592 (CSW, L2)	1474 (CSW, L2)
2	Low total	32 (CSW, L2)	27 (CSW, L2)
3	Average Mono	404	233
4	Standard Deviation	27.76	13.05
5	Average CSW	537	275
6	Standard Deviation	109.66	20.32
7	Average L1	356	238
8	Standard Deviation	18.05	9.89
9	Average L2	533	249
10	Standard Deviation	60.72	22.57

Table 5: Overview of total and total unique word counts for all prompt responses in interaction.

speak Choctaw in other sections of the interaction, such as in the summaries. Three participants gave summaries in Choctaw, and two gave two summaries in Choctaw each. Several participants also spoke in Choctaw for the small talk.

Finally, we analyzed how lexically rich the prompt responses were. We first looked at the total word counts of transcriptions of each response to the prompts; the overview is given in Table 5. The unique word count measurement was measured by response. The average total word count and total unique word count were both higher for the code-switching cohort. These word counts are raw scores of all words said in the response, including any English words. The participants with the highest and lowest word counts were both in the code-switching group. No statistical significance was found between the different groups or cohorts for word count or unique word count.

## 5 Discussion

We found that the best system design as to whether to include code-switching may depend on the speaker's fluency. We observed that the code-switching system was better than the monolingual English system for L1 speakers, but did not have an impact on L2 participants. We observed that L1 speakers of Choctaw highly identified with the code-switching system over the monolingual system. L2 speakers did not show this same level of identity relation with either system. It is unclear why L2 speakers showed this difference and we leave it to future work to investigate. Neither group rated a high preference for speaking longer with the system or recommending the system to others. All open-ended survey responses were positive or neutral; no participants felt negatively about the interaction.

While the L1 group's preferences for the code-

switching system over the L2 group manifested in higher survey scores, it did not impact the duration of the audio collected. Instead, all speakers were as likely to record long responses or voluntarily continue past the required maximum experiment time regardless of the system. We leave it to future work to further explore the differences between L1 and L2 speakers when interacting with bilingual dialogue systems.

Future directions could consider additional forms of code-switching throughout the system's utterances. Additional future directions could consider the use of LLMs. LLMs were not considered for this study as current performance on low-resource languages, such as Choctaw, is frequently unreliable. An example of this language unreliability is given in Figure 2 in the annex. LLMs could potentially be deployed for English-only sections, or throughout, should DAPEL be deployed in a higher-resourced language.

## Acknowledgments

We thank the Choctaw Nation and Choctaw Nation IRB for their edits and additions to this work. We thank the many Choctaw speakers who contributed their time and knowledge, yakoke. Many thanks to Dora Wickson for her invaluable translations and transcriptions. We also thank the anonymous reviewers for their feedback.

## References

- Gilles Adda, Sebastian Stüker, Martine Adda-Decker, Odette Ambouroué, Laurent Besacier, David Blachon, Hélène Bonneau-Maynard, Pierre Godard, Fatima Hamlaoui, Dmitry Idiatov, et al. 2016. Breaking the unwritten language barrier: The bulb project. *Procedia Computer Science*, 81:8–14.
- Seyed Hossein Alavi, Jacqueline Brixey, and David Traum. 2019. Can we use a spoken dialogue system to document endangered languages? In *Dialog for Good (DiGo)*.
- 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.
- Jacqueline Brixey and Ron Artstein. 2021. Choco: a multimodal corpus of the choctaw language. *Language Resources and Evaluation*, 55:241–257.
- Cyrus Byington. 1915. *A Dictionary of the Choctaw Language*. US Government Printing Office. Edited by John R. Swanton and Henry S. Halbert. Smithsonian Institution Bureau of American Ethnology Bulletin 46.
- Justine Cassell, Timothy Bickmore, Mark Billingham, Lee Campbell, Kenny Chang, Hannes Vilhjálmsón, and Hao Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 520–527.
- David DeVault, Kallirroi Georgila, Ron Artstein, Fabrizio Morbini, David Traum, Stefan Scherer, Albert A Rizzo, and Louis-Philippe Morency. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proceedings of the SIGDIAL 2013 Conference*, pages 193–202.
- Lise Dobrin, Peter K Austin, and David Nathan. 2009. Dying to be counted: The commodification of endangered languages in documentary linguistics. *Language documentation and description*, 6:37–52.
- Elodie Gauthier, David Blachon, Laurent Besacier, Guy-Noel Kouarata, Martine Adda-Decker, Annie Rialland, Gilles Adda, and Grégoire Bachman. 2016. Ligaikuma: A mobile app to collect parallel speech for under-resourced language studies. In *Interspeech 2016 (short demo paper)*.
- Michael Johnston, Patrick Ehlen, Frederick G Conrad, Michael F Schober, Christopher Antoun, Stefanie Fail, Andrew Hupp, Lucas Vickers, Huiying Yan, and Chan Zhang. 2013. Spoken dialog systems for automated survey interviewing. In *Proceedings of the SIGDIAL 2013 Conference*, pages 329–333.
- Takahiro Kobori, Mikio Nakano, and Tomoaki Nakamura. 2016. Small talk improves user impressions of interview dialogue systems. In *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue*, pages 370–380.
- Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.
- Tomoaki Nakamura, Takahiro Kobori, and Mikio Nakano. 2019. Learning dialogue strategies for interview dialogue systems that can engage in small talk. In *Proceedings of the 9th International Workshop on Spoken Dialogue System Technology*, pages 307–317. Springer.
- Tanmay Parekh, Emily Ahn, Yulia Tsvetkov, and Alan W Black. 2020. Understanding linguistic accommodation in code-switched human-machine dialogues. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 565–577.
- Shana Poplack. 2000. Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *The bilingualism reader*, 18(2):221–256.

Ruth Rouvier. 2017. Language documentation, revitalization and reclamation: Supporting young learners and their communities. White paper.

Frank Seifart, Nicholas Evans, Harald Hammarström, and Stephen C Levinson. 2018. Language documentation twenty-five years on. *Language*, 94(4):e324–e345.

The Choctaw Nation of Oklahoma Dictionary Committee. 2016. *Chahta Anumpa Tosholi Himona: New Choctaw Dictionary*, 1st edition. Choctaw Print Services.

## A Annex



### Anonymous

Write a speech for someone learning Choctaw to give to their hosts when invited to dinner, put in English words instead where the speaker would not know the appropriate Choctaw words.



### ChatGPT

Ladies and gentlemen, thank you so much for welcoming me into your home tonight. I am truly honored to be here, and I want to express my sincere gratitude for your hospitality.

I am currently learning Choctaw, and while I'm not yet fluent, I would like to share a few words of appreciation in the language. Please bear with me, and feel free to correct me if I make any mistakes.

Hello you I see I. Tree path (past). Toward your good your see word. You word man.

"Halito chi pisa la li hosh. Iti anowa tok. Pila chinchukma chi pisa anumpa. Ish anumpa hattak."

(Translation: Greetings and thank you for having me. It's a pleasure to be here. I appreciate your kindness. Let's enjoy the meal.)

I want to learn more about Choctaw culture and language, and I am grateful for the opportunity to be immersed in such a welcoming environment. Thank you for this wonderful dinner, and I look forward to continuing my journey in learning and understanding Choctaw better.

He sees you your good. Chi pisa chinchukma (Thank you very much).

Figure 2: Example of Choctaw output from ChatGPT. Retrieved January 24, 2024. "Anonymous" represents input from the user. English translations are given in red next to the Choctaw text. A translation is provided by ChatGPT in parentheses, and substantial discrepancies can be observed.