# Spoken Dialogue Project Report

Benny Longwill

March 16, 2019

## 1 Project approach

For this project, the Alexa Skills Kit was used to engineer a Spanish language quiz game that examines user knowledge for distinguishing the Spanish prepositions 'por' and 'para'.

### 1.1 Goal: Por and Para

The underlying goal of the quiz styled game is to master the differences between the two Spanish prepositions "por" and "para". This skill presents an obstacle to many English speakers because both words correspond to the word "for" in English in addition to several other words and idiomatic expressions (i.e., 'for', 'by', 'on', 'through', 'because of', 'in exchange for', 'in order to') [2] . Furthermore, these two Spanish prepositions cannot be exchanged for one another and have very distinct circumstances in which they can be used. That is to say that the word used in English does not determine the word used in Spanish. For this reason, non-native speakers tend to have to invest a significant amount of time learning these difference or memorizing the circumstances.[7]

### 1.2 Benefit of Quiz taking

According to a study Roediger et al (2006), there exists a testing effect in which students who take part in practice tests remember more information for longer. Through their study they compared students who took a final exam after only studying by reading, and others who did the same thing in addition to taking intermittent practice tests. They found that "on the delayed tests, prior testing produced substantially greater retention than studying, even though repeated studying increased students' confidence in their ability to remember the material."[5] As a result, if testing can be seen as a way of learning, then implementing a quiz styled game is an excellent base for the current tutoring application.

## 1.3   Voices and Accents

It is widely recognized that one of the best ways to learn a foreign language is through spoken dialogue with a native speaker. Given that the Alexa Skills Kit allows designers to create skills in multiple languages it is the ideal environment for creating a tutor that can expose the user to a variety of native speech at all times. Given this, all Spanish language in the current project (primarily in the fill-in-the-blank category questions) is produced by a randomly selected male or female native Spanish speech synthesizer. Unfortunately, there are only two voices for the Spanish language but hopefully moving forward there will be more support for the Mexican accented speech will also be supported given that it is already available through the Mexico locale.

## 1.4   Explanations

At any time during the game, quiz-takers can ask Alexa about the categories, the target prepositions, how many points they have left, or why the previous answer was correct or incorrect. She will then give her explanation and all the user has to say is "okay" in order to return to the previous state. If the user asks about categories in general without specifying one of them, Alexa will tell the user which categories are remaining.

## 1.5   Categories

In order to to make this game more entertaining, certain elements of the game "Jeopardy"[1], a televised quiz game show, have been included to better capture the attention of audiences. For example, in order to allow for more user control and more varied question types, several question categories have been implemented and can be changed at any time provided that he or she finishes the current question at hand. The categories included are:

- "What would you use": the user is given a sentence in English and has to decide the Spanish preposition translation using the context)

- "Fill in the blank": The user is told a sentence in Spanish but the target preposition is removed and replaced with a beep noise to signal to users which word is being replaced.

- "Uses": The user is given one of the contexts in English and has to respond the corresponding target preposition linked to that context.

## 1.6   Hall Of Fame

At any time during the game, a player can tell Alexa his or her name in order to be recorded in the Hall of Fame. In the *end_game()* function, code was written to write the user name to an output file that keeps track of user statistics;

however, given the server-less cloud-based nature of the Alexa framework, one can only write files within a /tmp/ folder. This folder does not persist between sessions. Therefore, the only way to keep data between sessions is to implement S3 (or Boto). Amazon Simple Storage Service (Amazon S3) is an object storage service that requires the developer to enable and configure a database 'bucket'. So although it is a possible feature given the framework, this feature was not fully implemented due to time/benefit ratio and scope of the assignment.

## 1.7 Betting System

The current application includes a betting system in which the user may request to make a point-value bet at any time for the next quiz question. Bets placed must not be greater than the amount of points that the player has. If a player gets the subsequent question correct, he or she receives the amount bet separately from the standard amount received by correctly answering the question. Furthermore, the equal amount of the bet is also removed in addition to the standard amount lost when answering a question incorrectly. Moving forward into the research domain, bets can be utilized as Judgements of learning (JOLS) similar to how they are denoted in Roediger et al, (2006).

# 2 Resources, software and data

This application was developed using the Alexa Skills Kit, a cloud based system that maintains support for creating a voice user interface through automated speech recognition. This kit also includes natural language understanding support in order to understand user intents, and a back-end cloud service to process intents that allow the skill designer to tell Alexa how to respond. Although no software is needed to access this skill kit, an Amazon Web Service user account is required to access and create a back-end lambda function and also an Alexa Developer account is required to access the development console and create JSON interaction model file.

The material collected to create the application quiz questions was taken from the answer key for a "Jeopardy" style online game designed for live classroom play with up to ten individuals or teams. Teams choose a question, then try to give the best answer to score points.[3]

# 3 Related and motivating work

In the field of artificial intelligence, text-based tutoring systems are currently one of the most successful endeavors and have influenced the current application to a great extent.

In a study by Seneff (2007), researchers developed several successful systems in which students engage in dialogue with the computer either over the telephone or through audio/typed input at a Web page. They also included several

different domains of conversation in their applications. For this reason, the current application was motivated to include different categories and question types and move in the direction of having a purely speech-based dialogue system.

Furthermore in Schwitter Islam (2003), a speech-based virtual tutoring system designed for walk-up-and-use situations in teaching environments employed a user interaction similar to that of a quiz game show, S-Tutor attempts to educate students in language technology.[4] From this I wanted to employ similar game show aspects but bring it forward specifically into the Spanish Language domain.

# 4 Challenges

How to improve communication success between users and spoken dialogue systems is a question still left unanswered in the field speech-to-speech dialogue systems. The current Alexa Skill tends to work well with carefully spoken speech but tends to often fail due to two key technological challenges in spontaneous natural language, (i.e., the complexity/ambiguity/informality of natural spoken language and context-based derivation of intent). In short, when people speak in a natural speech pattern it's challenging to map their words to the specific intents in the application.

Finally, the Alexa framework also presents a limited turn based dialogue. Given the nature of this system, it is difficult to implement a more rapid style game-play that does not rely on question response.

# 5 Resolution to challenges and conclusions

By itself the Alexa system's speech recognition achieves the correct transcription in the majority of utterances and keeps the application running despite adverse language input due two main additions.

First and foremost, the slots for user responses in the current application have been engineered to be very limited and predictable. The Alexa skill Kit supports several slot types that define how data in the slot is recognized and handled; however, in the case of the current project I have limited the question types to be very closed, such that the program only expects one of the two responses when answering questions: "por" or "para". Of course the program itself is much broader than this, but the same concept is used for most intents and very much limits the amount of space in which the program can become confused.

Furthermore, the program is also assisted by the incorporation of a built-in intent called "AMAZON.FallbackIntent". This intent is used to handle out-of-domain-requests that are not included as part of any of the other intents. Therefore, in the current application when an unexpected utterance is recieved, the program will continue running by calling the *get_repeat_response(intent,session)* function that allows Alexa to repeat the most recent prompt she has saved

(Which is strategically not necessarily the most recent prompt made, but rather the most recent prompt that supports the quiz as an FST).

# References

[1] CBS Press Express — Jeopardy!, www.cbspressexpress.com/cbs-television-distribution/shows/jeopardy/about.

[2] "Por vs Para." SpanishDict, www.spanishdict.com/guide/por-vs-para.

[3] Por vs. Para Jeopardy Review Game Answer Key, www.superteachertools.us/jeopardyx/answerkey.php?game=536826.

[4] R. Schwitter, Md Tawhidul Islam. S-Tutor - A Speech-based Tutoring System. In: U. Hoppe, F. Verdejo, J. Kay (eds), Artificial Intelligence in Education, Shaping the Future of Learning through Intelligent Technologies, IOS Press Ohmsha, pp. 503-505, 2003.

[5] Roediger, Henry and Jeffrey D. Karpicke. "Test-enhanced learning: taking memory tests improves long-term retention." Psychological science 17 3 (2006): 249-55 .

[6] Seneff, Stephanie. "Web-based dialogue and translation games for spoken language learning." SLaTE (2007).

[7] Waltz, Terry. "HOW SHOULD I TEACH POR AND PARA?" TPRS QA, 2018, tprsquestionsandanswers.wordpress.com/2018/01/09/how-should-i-teach-por-and-para/.