

Liana [REDACTED], Eliana [REDACTED], Anya [REDACTED], Dylan [REDACTED]
LIGN 6 - Language and Computers
Professor Will Styler
May 8th, 2022

Project Proposal

I. Introduction

Our topic for this project is to look at and compare the way natural language processing works in various languages. We are interested in this topic because we are all bilingual and interested in learning how language assistants perform in different languages. Most of the big NLP systems are developed in US English. This can be an equity issue for people with different dialects of English, as well as people who speak other languages altogether. For this reason, we are especially interested in seeing how they perform in other languages in comparison to how they perform in English. We will be using our intuitions regarding our native or heritage languages (Spanish, Brazilian Portuguese, Hindi, and Indonesian) as well as our linguistic knowledge to come up with cases in which we believe there may be some variation in performance, putting the voice assistants Siri and Google Assistant to the test.

II. Preliminary Research

We knew we wanted to compare language assistants in each of our languages, but we were unsure of what measures should be used and what linguistic factors we should target with our comparison. To guide our project, we began by doing some preliminary research into each of our languages and how they interact with voice assistants.

In Brazilian Portuguese, we found two papers that compared how Siri and Google Assistant managed different dialects caused by region and socioeconomic status (Lima et al. 2019a, 2019b). They used methods that we will be basing our research on, which will be discussed in the next section of this paper. Lima et al. (2019a, 2019b) found that Google Assistant and Siri perform significantly better with southeastern dialects, which is the accent our Portuguese speaker has. They also found that unstandard productions of words common especially common in people with a lower socioeconomic status were likely to cause issues for the voice assistant (Lima et al. 2019a), and that some common shortenings of words in colloquial speech cause issues for the voice assistants, while others the systems seem to be able to manage (Lima et al., 2019b). We also found a survey conducted by Guerino and Valentim on how Brazilians use voice assistants and what issues they have with them. More than half of the surveyed participants said they had to repeat their commands for the voice assistant to understand. Some participants also reported feeling that they felt as though the system only responded to specifically formulated commands that they had to learn before use.

So far for Hindi we haven't found any academic research relating to language assistants. However, we tested several speech to text and text to speech softwares that were made in India and found that speech to text software was remarkably accurate, whereas the text to speech software pronounced words perfectly but lacked human intonation. We will continue to look for academic research and also compare the Indian software to Siri/Google.

For Indonesian, we have found a paper describing how the Android system's Google Speech was tested for speech and text input (Areni et al. 2019). The results of the paper showed that Google's speech and text recognition is impeccable with the success rate being 100% for the trained data, and up to 96% for random data. We have also tested some text to speech softwares for the Indonesian language, with results being incredibly proper, however, the text to speech software had trouble with slang. We also tested speech to text, and surprisingly had good results despite using a variety of slang and foul language.

For Spanish, there are native Spanish-speakers all over the globe, and thus, many different varieties of Spanish. Because of this variety, some problems that Spanish virtual assistants may encounter is with speech recognition since they may have been trained on one dialect but encounter a speaker who uses a dialect unfamiliar to the system. One solution that some papers have proposed has been to use a multidialectal approach. This kind of approach utilizes a single system which contains all the different

phonetic inventories, rather than having to use different systems for each dialect. Caballero et al. 2002 found this approach to be much more successful than the mono-dialectal approach.

III. Methods and Test Examples

In order to target specific linguistic features of each language, we will create sets of sentences, one in each language, that all address a certain feature of language we believe may cause issues for voice assistants. In our research (Lima et al., 2019b) and in our lived experience, we found that voice assistants might struggle with shortening of words. We also noticed in our use of voice assistants that they don't always produce accurate responses for multi-step commands or phrases with conjunctions. In class, we have learned that novel words can be challenging for voice assistants, but also hypothesize that some languages with more phonetic-based writing systems may have an easier time with that task. In a similar vein, we want to evaluate how each language's voice assistant deals with slang words, and whether they are part of their dictionaries or not. Finally, we also want to create different forms of giving a certain command, since we found that some users complain that they need their commands to be in a specific form for an accurate response. We are also in search of other linguistic features to target, and would be interested in any advice you have on how to find appropriate features.

The following are some examples of sentence pairs we plan on using:

- a) Shortening and colloquializations of words
 - i) English - '**yall're gonna** go to the store, right? **Cuz** I need milk. (You all are going to go to the store, right? Because I need milk.)
 - ii) Portuguese - **Cês vão pro** mercado, **né?** Porque eu preciso **compra** leite. (Vocês vão para o mercado, não é? Porque eu preciso comprar leite)
 - iii) Spanish - **Vais** a ir al super, no? **Poq?** yo necesito comprar leche. (Ustedes van a ir al super, no? Porque yo necesito comprar leche.)
 - iv) Hindi : mai **jari** hoon, tu kya **kari** hai? (mai ja rahi hoon, tu kya kar rahi hai?)
 - v) Indonesian - **Gapapa** sih (nga apa apa sih)
- b) Novel names (all for placing a call)
 - i) English - Call **Goobletta Krobralapper**.
 - ii) Portuguese - Liga para a **Marbolina Vintipeça**.
 - iii) Spanish - Llama a **Marsipansita Vorelquez**.
 - iv) Hindi - **Marbooti** idhar aa!
 - v) Indonesian - Telpon **Tidiana Budiatur**
- c) Novel verbs (all in the context of how to be good at the action)
 - i) English - How to be good at **glerbing**. (verb - "glerb", clue for being a verb - "ing")
 - ii) Portuguese - Como **biquer** bem. (verb - biquer, clue for being a verb - "er")
 - iii) Spanish - cómo **bromir** bien (verb - bromir, clue for being a verb - "ir")
 - iv) Hindi - Voh khaana **faal** rahi hai. (verb - faal, clue for being a verb "aa")
 - v) Indonesian - Ajarin saya untuk bisa **menabatandi**

In order to test these sentences, we will be using methods adapted from Lima et al. in their 2019 papers. In this method, each person will read all of their English and respective languages sentences. They read a given sentence until it is transcribed and responded to correctly, or there have been 3 attempts. We will take note of how many attempts are required for each sentence, as well as what is transcribed and output for each attempt. With this, we will have data on how quickly the system is able to accurately respond to your command or question, as well as what mistakes it is making.

IV. Steps for Completion

Our project will be adhering to the following structure:

1. Academic research concerned with language assistants in our respective languages. We will be looking for any data that has already been collected regarding the challenges faced by language assistants in our respective languages. This data could include linguistic problems faced by language assistants, testimonials from individuals, comparisons between dialects etc. This will be done during week 6 and 7.
2. Next, we will design tests for how we will explore linguistic problems that language assistants run into during week 7 and 8. Test design includes the formation of sentences to challenge our language assistants with.
3. Next, we will collect data by interacting with the language assistants using our test structure during week 9. We will also transfer data into google sheets and organize it in ways we can make visualizations
4. Finally, in week 10 we will put all of our findings in our report.

V. Potential Roadblocks and Questions

As we will be working with multiple languages, we will be experiencing a few problems:

1. We are unsure of whether our sentences should always have similar meanings. Sometimes to illustrate an idea, each language has their own grammatical contractions that would lead to a complexity in interpretation.
2. We also want to experiment whether the sentences should be in the form of a command or a question, or would any arbitrary sentences work for our systems.
3. In order to obtain the most ideal results, we will aim to have each speaker carry out a number of commands in both English and their respective languages. This might lead to the confounding variable that not all of our group members have a standard English dialect, and therefore might lead to results that do not reflect the performance of a voice assistant with a general American English production.

VI. Citations

- Guerino, G. C., & Valentim, N. M. C. (2020). "Is anybody there?": exploring the use and difficulties of Brazilians with conversational systems. *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*. <https://doi.org/10.1145/3424953.3426649>
- Lima, L., Furtado, V., Furtado, E., & Almeida, V. (2019a). Empirical Analysis of Bias in Voice-based Personal Assistants. *Companion Proceedings of The 2019 World Wide Web Conference*. <https://doi.org/10.1145/3308560.3317597>
- Lima, L., Furtado, V., Furtado, E. S., Almeida, V., & da Silva, T. H. O. (2019b). Discrimination analysis of intelligent voice assistants. *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*. <https://doi.org/10.1145/3357155.3358483>
- Caballero, M., Mariño J. B., Asunción Moreno. (2002). *Multidialectal Spanish modeling for ASR*. Retrieved May 8, 2022 from <https://aclanthology.org/L02-1278/>.
- Areni, I. S., Mufidah, A. D., Amirullah, I., Wahyuni, S., Nurtanio, I., Sahali, I., & Bustamin, A. (2019). *Speech to text in Indonesian personal assistant - researchgate.net*. ResearchGate. Retrieved May 8, 2022, from https://www.researchgate.net/profile/Anugrayani-Bustamin/publication/335914888_Speech_to_Text_in_Indonesian_Personal_Assistant/links/5e25906b458515a740651f39/Speech-to-Text-in-Indonesian-Personal-Assistant.pdf