



Louisiana French

Joe Chu, Will Duong, Abhi Oddula, Rahul Ramki, Saharsh Shrivastava



Project Overview

Client Overview:

- Received NSF grant in 2018 to research the linguistic effects of Hurricane Katrina
- Have generated a corpus of 28 interviews of transcribed dialogue of interviewees speaking in Louisiana French
- Want to analyze verb frequencies within this corpus

Goal/Purpose Statement:

The goal of our project is to identify verb frequencies within the interview corpus. Given a verb bank, our responsibilities are to automate the process of finding and counting each conjugation of each verb within the bank.

Stakeholders:

- Dr. Katie Carmichael and Dr. Aarnes Gudmestad from the VT Department of English and Department of Modern and Classical Languages and Literatures

Objectives:

- Automate tasks for our clients to save them time from manual parsing
 - Generate verb frequencies for the verbs included in the dataset
- Assist in the overall research by helping our clients analyze the dialogue of Louisiana French speakers
- Develop a user manual for our clients if, in the future, they decide to add more interviews to their dataset

Timeline:

- February 8 - Meeting and understanding of project requirements with client
- February 26 - Presentation 1
- March 4 - Demo of current progress to client and professor
- March 6 - Completed verb 'avoir' script as a pilot to our project
- March 10 - Started implementation with TreeTagger
- March 26 - Presentation 2
- April 8 - Demo of current progress with Dr. Gudmestad
- April 11 - Presentation 3
- April 12 - Progress on final deliverable
- April 30 - Final Presentation and VTURCS

Project Design

Technology Purposes:

- Python for scripting
- TreeTagger for parts of speech tagging
- TreeTagger wrapper (Python) to use the TreeTagger commands easier
- Jupyter Notebook for a neat display of metadata for client
- Sublime Text for text editing
- Microsoft Excel for data storage

Initial Implementation

The initial implementation of our project was to simply use a python script to brute-force the entire interview corpus line-by-line and interview-by-interview. We would then use the results we got to compare it to a web-scraped verb table that we built in excel. We that this was extremely inefficient and shifted towards a more efficient way.

Final Project Design:

Our final implementation of the project is to combine our knowledge of python along with a software called TreeTagger. TreeTagger is a parts of speech tagging software that is used for multiple languages for many other research projects. For our project we used the Standard French language to be a foundation for the analysis of verb frequency in Louisiana French. We have used python scripts to feed the TreeTagger lines from the corpus and calculate us a dataset collection based off of the verb frequency.

Testing:

- TreeTagger tested with the English Parameter file and ran with English sentences
- TreeTagger tested with the Standard French Parameter file and ran with interview corpus
- Results from the data collected from testing the interview corpus verified by client.
- Cases regarding cédille, accent grave, aigu, circonflexe, tréma.