MSiA 490-20 / Ling 400: Homework 4

Instructor: Klinton Bicknell

adapted from Jorge Moraleda

Turning it in: You may do this assignment alone or in a group of up to 3 people. Turn in just a write-up as a pdf (just one write-up per group).

Dialogue

This assignment involves analyzing dialogue text. Dialogue text is common on forums, chat, or phone transcripts. For this assignment, however, you'll analyze a theater play. Specifically, you'll select your favorite play for which you can get the text, analyze it in some interesting way (using whatever software you please), and write up a report. The report should describe the methods, the results, and your interpretation, as well as some visual representation of (at least some of) the results.

Grading

The grading will be based on:

- Are your methods appropriate to your questions?
- Did you implement/interpret your methods correctly?
- How interesting are your conclusions?
- How extensive is your analysis?
- How useful / appealing are the visualizations you present?
- Is there enough interpretation / conclusions, or is the report mostly just raw facts?
- Is the interpretation too speculative, and could use more facts / analysis to back it up? (Though some speculation is encouraged, so long as it's clear that it's speculative!)

Ideas

There are no restrictions on or requirements for what you might want to analyze. To help get you started thinking, a few ideas include:

- What are the temporal references? If your play contains actual dates (like in a diary) then relative times (like 'last Thursday') can be converted to absolute references (e.g., Stanford CoreNLP can do this, with the `sutime` annotator. See this online demo: http://nlp.stanford.edu:8080/sutime/process (http://nlp.stanford.edu:8080/sutime/process)).
- What are the topics of the play? You could make a break down by character. Do these topics evolve over time (e.g., by act)?
- What are the Named Entities (e.g., people, places, organizations) that appear in your play?
- Who talks (spends time) with whom? Does it change over time? Which fraction of the talk does each speaker contribute?
- Who or what does each speaker talk about? What is the sentiment of the speaker about each entity (i.e., what is the sentiment of the words that appear near them in the dialog?). You will probably

want to do coreference resolution (to identify what pronouns matches what noun) when identifying who talks about what or whom. In dialogs, participants agree about the antecedents of pronouns, so you may want to process consecutive utterances from the various dialog participants as a single unit of text for the purpose of coreference resolution.

- What is the mood of each speaker (e.g., the average sentiment of the words they utter)? Does it change over time? Does it depend on who they are talking to or who or what they are talking about?

Have fun!