

Tanishq Chaudhary 2019114007

Group work

For our Project, we have looked at Language Varieties from a Sociolinguistic perspective, primarily focusing on Gender differences to employ data from popular Bollywood movies, that represent the population as a whole.

Our work was in three phases:

1. Collecting the data from Bollywood scripts, through various sources. After this, some manual verification had to be done, and the data was manually tagged and processed into JSON files. This had two major advantages:
 1. Character wise, scene wise, and multiple layers of disambiguation could be done easily, employing rudimentary data cleaning tools
 2. As our data was scripts, narration and directors notes needed to be removed, for both ease of reading and removal of bias
2. Read relevant research papers, and employ hypotheses on which ideas can be supported/seen in our data. This phase was done and presented in our first Presentation. We also had an involved discussion on what ideas we should focus more on, and which ideas should be discarded for various reasons.
3. Conduct a survey to prove or disprove the ideas derived from the research paper, and to also view them with respect to the movies chosen.

General Overview

Much of my work focused on looking at scripts of all the movies. We used scripts with English + romanised Hindi for our purposes, which then had to be cleansed. I did the cleaning for the movies: Veere Di Wedding and Dangal. After reading the scripts from the said movies, I discussed with Prajnaya about how meaningful would be the contextual information of each scene, and how viable and useful it would be in practice. After that, I created all the visualizations that were presented in the class, along with many more that I had to discard, for their lack of insights. Visualizations helped the rest of the teammates to get a great starting point to start creating their insights. The script was made to be used by everyone so it has a CLI, which is on Github itself. Apart from this, I also read up on numerous articles by Robin Lakoff to get a good idea of the observations from a feministic perspective.

In essence, The core component of my work revolved on a general overview about our entire project, so that everyone working on individual movies could communicate and we could relay important information, as mentioned in the first page. Along with everyone else, I was also involved with reading papers & articles in the entire process. I made functionality in the CLI to confirm and deny the hypothesis made for a few parameters as we described them in class. I will also use the ideas learned here to help devise the survey process for the next phase.

Individual Analysis

Lakoff, Robin. "Language and woman's place."

Quick Overview: This was the central source of ideas because it makes testable predictions about the differences in language between the genders. I read through the key points made, and focused on the differences in usages of slang and question tags. I did exploration of more parameters in discussions with Prajnaya.

Some Ideas in Depth:

- Lexicon differences: As explained by Prajnaya, women, according to Lakoff, use a more correct form of the language; using hyper-correct grammar. But, Lakoff claims that there happens something within the word level itself; the vocabulary used *by* and used *for* women are different from that of man.
 - There are largely 3 sets of words that one can use. One is neutral words, second is women only (restricted domain) and third is common words (used by all genders).
 - A man can only use words which are outside female-only sets. Women, on the other hand, can use all of the sets. They can also use particular sets to represent a part of their identity.
- Question Tags: Briefly said, the use of question tags reduce the *assertiveness* in the utterance, and introduce some gentleness and tenderness.
 1. Is Shivansh here? Yes/No question
 2. Shivansh is here, isn't he? Tag question
 3. It's quite hot in here. Assertive sentence.
 4. It's quite hot in here, isn't it? Small talk
 5. The NEP is quite bad, isn't it? Asking for opinion
 6. I am feeling sick, am I not?
 - 2 is interesting as it allows for a bit more leeway; "isn't he here" is easier to answer than "Is he here"
 - 5 and 4 differ as 5 is more topical, and expects something more than yes/no answer; but in a different manner than 4
 - 6 is quite redundant, but Lakoff claims this is still used by women

Ali, Syed Fahar, and Azmat Noori. "Feminism in Indian Cinema: Women Characters in Films of 1970's to 2000's"

Description: This paper provided an insight into how characters in Bollywood have generally been represented, and we used the understanding I gained from this as a parameter for the selection of films. Understanding from this film was also used to make testable predictions about movies which we discussed in our meets. The idea of looking through linguistics through a sociological perspective was expressed in this movie, where sociological description of the systematic sexism in the portrayal of females in Bollywood led to usage of concepts learned in our class to make Linguistic inferences.

Observations

All of the visual work was done by me, with Prajnaya only making the word clouds. All the good observations were pointed out in the class, and the attached slides show them well.

These observations come from instances of the following for all the movies:

- pronoun split (best case in Dangal)
- language split (seen in Pink and Queen)
- pure language split (only english/hindi sentences) (seen in Pink mostly; due to court related locations)
- frequencies of words used (general idea of words used. Names along with functional words came on the top for all the movies)
- splitting the use of language pre and post an event (mostly used in Queen, for her visits outside India)

For testing Lakoff's ideas, I combined all of the utterances of all the characters, from all of the movies.

- Slangs showed the most distinctions, but their overall quantity was quite low. Since they were feminist movies, all of them showed a greater use of slangs by females than men. The difference though, is in the ratio of approximately 1:2 for men:women. This can be only confirmed by a larger dataset. This can be explained by women becoming more and more involved in workplace related activities, along with general inclusion of women in traditionally "men only" fields. Another interesting thought was that the feminist themed movies show a lot of conversations between close female friends. This allows for talking about taboo subjects for women; which is a common theme to explore and showcase for such movies.
- Surprisingly, the number of questions asked were greater for men than women. This is very tricky to explain, and I could not find any good reasons for such a result.
- the use of hedges was not statistically significant; men and women showed an almost equal number of usage of the chosen hedge phrases. This essentially disregards the claims by Lakoff.

The rest of the tested parameters showed little to no distinctions; which we would now like to test and check with the opinions from the people themselves; in the form of a survey.

Challenges Faced

A lot of work went into the initial steps itself; working on data cleaning. The code-mixed data was not easy to deal with; especially as later on I looked into combining all the movie scripts' data into one single file; due to the differences in romanization of hindi. One such example is the use of "Hain" === "Hai" === "Hayin" in different movies.

Even after figuring out, to confirm Lakoff's ideas and observations, I needed to figure out exactly what to look for; for example using "sort of" in english is a feature in female-speak. This can translate to a lot of different phrases in Hindi, so I had to go over the scripts with multiple translations, searching if there is any match; and thus something insightful that can be drawn. Hedges, along with apology phrases and Slangs are some of the things that I focused on exploring.