Language and Society Project

Social Stratification

Aashwin Vaish (2019114014)
Aravapalli Akhilesh (2019114016)
Chayan Kochar (2019114008)
Jerrin John Thomas (2019114012)
Padakanti Srijith (2019114002)
Suyash Vardhan Mathur (2019114006)

Introduction

The project aims to explore social stratification through a linguistic perspective. Social stratification is the division of the society into classes that can be observed in the language used by speakers. Advertisements, being targeted media, often also depict similar language of their target audience so that they can better connect with the people. We explore how viable a rigorous linguistic analysis of advertisements is to observe such stratification. Along the way some scenes in TV shows and interviews were also collected and analysed.

Background

We, humans, may be very different from one another, but have one thing in common, and that is the fact that we use language for communication. It is very much intricately embedded in all of us and this very fact makes it related to our society as a whole.

Language is central to social interaction in every society, regardless of location and time period. Language and social interaction have a reciprocal relationship: language shapes social interactions and social interactions shape language. And that is the exact reason that in this project we are aiming to conclude what type of public speak or use what type of language. By language we mean certain particular ways or certain common frameworks like hard/soft consonants, anti-rhoticity, monophthongs, code mixing, non linguistics features, etc.

For getting a better understanding and for finding a better methodology, some research paper reading was done. The first paper was “The research topic landscape in the literature of social class and inequality”. This discussed the research done in social stratification so far. The next one was “Social Class, Social Status and Stratification: Revisiting Familiar Concepts in Sociolinguistics“ by Christine Mallinson. It introduced the concepts related to social stratification and reported a case study done in Texana, North Carolina.

The third and the one which was given the most focus was “A Sociolinguistic Investigation of Social Stratification and Linguistic Variation among the Kashmiri Speech Community” by Nisar Ahmad Koka. It investigates the linguistic variation among the Kashmiri speakers at the level of phonology and lexicon of their language in accordance with certain social variables, that are, religion, education, region/socioeconomic status, age and occupation, and the reasons behind it. It looks into how the social heterogeneity of the Kashmiri speech community is reflected in the linguistic behavior of its speakers, and gives rise to variations in the use of their language at the level of phonology and lexicon. The social factors such as religion, region, education, age, occupation, socioeconomic levels/status etc. are responsible for variation in language and the resultant linguistic items are called linguistic or sociolinguistic variables.

The paper briefly touches upon informal & formal registers, diglossia, social variables and sociolinguistic variables. Then it focuses on the case study on the Kashmiri speech community. The data for the study had been collected by various methods, such as the distribution of the questionnaire, conducting interviews, and the investigator‟s direct involvement in some conversations with various members of the speech community. The analysis of variation of the phonological and lexical features due to the difference in religion, education, age and occupation.

The fourth research paper taken for reference includes the paper “Dialect Differences and Social Stratification in a North Indian Village” and its follow up paper Revisiting Khalapur I : Language Variation And Social Stratification 50 Years Later”, both of which look at Dialect Differences and Social Stratification, with focus on caste as a parameter in the village of Khalapur. This paper explores various factors, like Village/Urban differences, usage of phonetic features as a means of Social Stratification, like using /ə/ in forms like /bətau/ instead of /bUtau/, untouchables of all three castes having nasalized /I/, /ä/, /u/, vocabulary differences as a means of stratification, especially among the Hindu/Muslim community.

Data Collection & Analysis

The first task to be done was deciding the sociolinguistic variables which we would be considering. After some discussions, we narrowed it down to:

- Age
- Gender
- Region
- Language
- Religion
- Occupation
- Wealth/Income

For the dataset to be analysed, a variety of data and of different situations were required. Advertisements were manually collected and we started looking out for certain cues which might help us in our final conclusion.

We collected 49 advertisements and analyzed them separately.

But we observed certain challenges(which is described later), and to effectively overcome some of those, or rather make it clear that such challenges will always exist, we also analyzed language usage in certain TV shows, interviews and web series, which helped us in our understanding.

So, using all the above data and the analysis, we created hypotheses, certain rules which were indicating how is the society stratified by the language used by different people, whether based on wealth, age, area they live in, education, etc. The linguistic and non-linguistic features leading to social stratification were identified.

- Linguistic
    - Phonetic features
    - Lexical features
    - Code switching and mixing
    - Register and jargon usage
    - Accent
- Non-Linguistic
    - Setting
    - Clothing
    - Body Language
    - Visual & Sound effects

To verify our conclusions and check it in real life we sent out surveys.

Our first survey had its shortcomings, as we just asked respondents about which ad they remember on top of their head and its relatability. Now again the respondents did not necessarily have the language aspect in mind and might have just filled it on the basis of emotional factors or requirements.

Thus to better our stance, we sent out another form. In it we asked about the many similar questions as previously like age, gender, income, education, etc etc. But this time we ourselves handpicked certain advertisements which we felt had a strong presence of certain cues and asked the respondents based on them whether they use language the same way shown in the ad.

This way we got better answers, meaningful to us, and we also observed, sometimes it did match to most of our hypothesis(keeping certain exceptions in mind), sometimes it didn't.

The data was collected in two parts (using two forms), the first data (95 responses) was to influence our approach, while the second one (195 responses) was used to confirm.

Overall, the different methods of data collection done were:

- Manual advertisement collection. A total of about 50 advertisements were collected.
- 2 Form to find social background of people and the ads they relate to. Form was prepared and distributed. 95 responses and 198 responses were recorded respectively.
- Collection of scenes from movies, TV shows, interviews.

Summary of Individual Works

Aashwin

Survey Form #1 - Data Sanitation
Total Analysed : 12 advertisements, 1 interview (3 min scene)
Language : Hindi, English
Categories of advertisements picked : Targeting Lower / Middle / Upper class, Young-Middle Age
Features Observed : Code Mixing / Switching, Accent Differences, Phonetic / Lexical Features, Various non-linguistic properties.
Survey Form #2 - Design, Sanitation / Validation

Challenges

- There is definitely a bias from the advertisers towards caring more for linguistic accuracy when dealing with "non-standard" Hindi, or targetting a minority group (e.g. Rural groups, Ethnic folk etc.)
- South India has a common culture where a lot of people from villages get their education in cities. Being the form designers, and being North Indian, we were not aware of this in time so we weren't able to design the "Region - City/Town/Village" question to accommodate this.
- Telugu speakers formed the majority of responses. It seems we weren't able to make clear in the form how much proficiency a person should have in order to say that they "knew" the language. A lot of people claimed to know a variety of Hindi that I really don't feel like they would (e.g. many urban telugu people could "Highly Relate" to Rural UP/Bihar Hindi).

Akhilesh

Survey Form #1: Framing the questions
Total Analysed: 8 Advertisements(11 characters) and 1 TV show of 2 episodes and one advertisement for English.
Language: Telugu and English
Categories of Ads picked: Targeting Lower and Middle classes of all ages.
Features Observed: Code-Mixing and Code-Switching, Accent and Jargon usage, Lexical Features and many other non-linguistic features like clothing,audio and visual effects etc.
Survey Form#2: Framing the questions and helping in developing the idea behind the form.

Challenges:

- The biggest challenge I faced is to slowly sanitise data acquired and to make it suitable for analysis without any redundancy or contradictions between responses.
- There were many cases where the conclusions made are challenged by the data we collected. In this situation, we have reached an impasse between making the decision of neglecting such data and considering it “not sufficient” for analysis OR completely changing our view and concluding our observation as wrong on the specific conclusion related to the ad.

Chayan

Total Analysis: 9 advertisements and 1 TV show
Language: Hindi, English
Features: Phonetic features, accents, code mixing, certain lexical features, non linguistic properties like income, clothing, education, region.

Survey Form 1: Helped in framing the questions
Survey Form 2: Helped in framing questions and deciding on the ads to be put out, and analysis of the Tide ad which was put in the form.

Challenges:

- Though I have said that these features means this and so on. It is not that easy. If a feature satisfies the criteria of certain relation, it is not necessary it is thus valid.
  For example, consider the case of ‘Jethalal’ in the Taarak Mehta show. His example actually exposes this error. So I observed certain phonetic features which he depicts are of lower class people, he is actually a very wealthy businessman. Hence I conclude that those features do not necessarily depict lower class. It in general may represent that he is either uneducated(which is true), from rural areas(he spent his childhood in a village), or he is a lower class person(not true in terms of wealth). This sort of helped me to be more vigilant in observing these details.
- Adding to it, like for the Tide ad, the data I collected from the surveys were not that encouraging. That was because the respondents were mainly from urban background, and thus there were not much presence of responses from rural/uneducated people. In short the data being sort of sparse and biased towards certain sections was the reason.

Jerrin

Advertisements in Malayalam were manually collected and 9 of them were analysed based on the chosen features, linguistic features and non-linguistic features. Most of the

advertisements targeted based on occupation. The features observed were phonetic & lexical features, code switching and mixing, register and jargon usage, accent and setting.

Some research paper reading was done. The papers were:

- "The research topic landscape in the literature of social class and inequality".
- "Social Class, Social Status and Stratification: Revisiting Familiar Concepts in Sociolinguistics" by Christine Mallinson.
- "A Sociolinguistic Investigation of Social Stratification and Linguistic Variation among the Kashmiri Speech Community" by Nisar Ahmad Koka.

In the first form, the results were too sparse to generalise. In the second form, after sanitisation, that is removing people who filled the Malayalam fields without knowing the language, there were only 12 responses. This was too less for a quantitative analysis as even the people of different categories were not there.

Challenges

The form data had a very sparse distribution. The responses were too few for a proper quantitative analysis. Also, it was noted that people filling forms usually have different perspectives to the different questions. This can lead to variation of responses from what is expected. From the research papers read, the idea of doing a case study was considered, but absence of proper dataset for carrying out was a major challenge, leading to the abandonment of the idea.

Srijith

Analysed 5 advertisements and 1 interview of 30min.
Languages: Telugu and English.
Observed Code Mixing and Code Switching, Accent Differences, Phonetic and Lexical Features and Various non-linguistic properties.

Challenges:

For the ads the advertisers have to think about the target audience and be able to convey everything they want to convey exactly within a short time. For this they focus on linguistic and non-linguistic features of the ad.

Collecting the data and the responses was difficult and the collected data was also not reasonable since many of the people didn't understand the meaning of this form as most of them are not familiar with linguistics. Collecting data in person would be a lot better and the data was also not satisfying as there were ads of 4 different languages.

Suyash

Total Analysis: 11 advertisements and 1 TV show

Language: Hindi, English
Features: Phonetic features, accents, code mixing, certain lexical features, non linguistic properties like income, clothing, education, region.

Some research paper reading was done as well:

- "Dialect Differences and Social Stratification in a North Indian Village" by Gumperz(1958)
- " REVISITING KHALAPUR i : LANGUAGE VARIATION AND SOCIAL STRATIFICATION 50 YEARS LATER" by Shailendra Mohan

Survey Form 1: Helped in framing the questions
Survey Form 2: Helped in framing questions and deciding on the ads to be put out. Did quantitative analysis for the Forevermark and Airtel advertisement along with pie charts.

Challenges:

- The data that was collected was heavily biased towards people living in the cities, and Telugu and Hindi speakers comprised a lot more among the people who filled the form than people of other languages. This led to an inherent bias in the dataset that we had for our analysis.
- We felt that there was some level of misunderstanding of the questions – like taking the narrator's voice into account for features as well apart from just the features of the speech of the characters involved in the advertisement.
- The accuracy with which people looked at the Linguistic features was rather loose, and people weren't able to accurately identify the accents which were targeted towards the rural populace.

Conclusion

We made certain assumptions for the features we came across(added in the data sheet in tabular format), and thus had a rough hypothesis for them.
For example:

- **Category**: North India, Urban English, Upper Class
  Prominent features: Hard consonants -> soft consonants, conscious effort to evade "indian" english like /r/ -> /ɹ/, /ŋg/ -> /ŋ/, /th/ -> /θ/.
- **Category**: Youth, North/Central India, Urban, Hindi, Middle-Upper Class
  Prominent features: Light to heavy English code-mixing, some features mentioned like above
- **Category**: North/Central India, Rural, Hindi, Lower class
  Prominent features: Usage of third person for self, /f/ -> /ph/, /æ/ -> /e/, fricative /z/ -> affricate /dz/

- **Category**: South India, Rural/ Suburban, Malayalam, Lower/ Middle Class
  Prominent features: Light to none code-mixing of English, some features like the before mentioned point.
- **Category**: South India, Sub-urban/Rural, Telugu, Lower-class
  Prominent features: "alana palana", "eyy", "nana", "thippalu" and similar words.

Though the data we got from the responses did not agree much with some ads and our hypothesis, that was because of the sparse and biased nature of the data. Also false responses by the respondents(unknowingly), as all these linguistics aspects being complex for them, was also majorly responsible. Most challenges we faced were regarding the forms and the data collected from it, which is described later. Still we feel that it was better than form #1, and overall though we faced many difficulties, and thus might have not got the result one would have wanted, we tried our best to get into some meaningful results.

All in all we would conclude that survey form #1 was a shot in the dark. Data collection during phase #1 was still quite aimless. After phase #1, most of our team including me was still quite at a loss at how to approach the problem. So we had some long meetings and made a plan for the next phases. Data collection b/w phase 2 and 3 was much more aimed and focused, with a goal to create a second survey form, with inputs from Prof. Radhika. We gathered up dimensions and features that we all had analysed individually and formally debated what should constitute the dimensions and features. However, we feel some points were less discussed, possibly rushed. In the end, minding all the challenges and sources of errors we could find, we would say that we might not have been able to conclude a hard hitting research, but the journey of this group project was definitely insightful, and the experience will streamline the next projects we deal with.

The form helped to conclude our Hypothesis regarding Upper Class, who use slightly accented English, and belonged to the Upper Class. The Forevermark responses helped us conclude similar observations of English relatability to high earning upper class people.

The Airtel advertisement helped us confirm the hypothesis about certain slang usage that is primarily common among teenagers and younger generations, as well as the distribution of these slangs among the Hindi speaking population. Pediasure was able to confirm the gender bias advertisers show selling child-care products towards mothers. Tide Halwai was able to confirm (in part) the relation of UP / Bihar rural Hindi with nearby languages and areas.

Challenges:

- Filling survey forms introduces a lot of errors from the correspondents' side that are uncontrollable by us. e.g. Misunderstanding a question, errors in filling the form.

This was the biggest factor as a disadvantage in form #1. We had asked people to write ads they remember on top of their head and its relatability to them. Here, the data we collected from it clearly showed that people misunderstood the question as “relatability to the product”(which is not necessarily bad) or “relatability to characters”, where emotional factors would play the role, etc.
Though in the second form we tried to rectify such issues, some general errors and exceptions were always there.

- This was necessary due to the team make-up, but conducting a survey in 4 languages definitely diluted the results obtained. Things like how people perceive dialects, how much code-mixing people are used to (to not even consider as “mixing” in layman terms) is different in different languages. The societies of North and South India are quite different (as in the above challenge).
- Also regarding the data collected, we did face the challenge that many respondents were obviously not so vigilant to focus on the details, hence obviously some data was definitely not matching. That is why I focused more on the ‘Not similar’ and ‘highly similar’ instead of the option ‘similar’ in the question regarding how much do they relate to the language used in the ad. But even then our responses were not much in support of our hypothesis mainly due to the majority of respondents being upper-middle class.

References

- Guo L, Li S, Lu R, Yin L, Gorson-Deruel A, King L (2018) The research topic landscape in the literature of social class and inequality. PLoS ONE 13(7): e0199510. https://doi.org/10.1371/journal.pone.0199510
- Mallinson, Christine (2007/10/01) Social Class, Social Status and Stratification: Revisiting Familiar Concepts in Sociolinguistics
- Koka, Nisar (2014/09/01) A Sociolinguistic Investigation of Social Stratification and Linguistic Variation among the Kashmiri Speech Community
- GUMPERZ, J.J. (1958), Dialect Differences and Social Stratification in a North Indian Village1. American Anthropologist, 60: 668-682. https://doi.org/10.1525/aa.1958.60.4.02a00050
- Hasnain, S & Mohan, Shailendra. (2013). REVISITING KHALAPUR i : LANGUAGE VARIATION AND SOCIAL STRATIFICATION 50 YEARS LATER ii. 7434. 131-140.
- Youtube Links for the advertisements are present in the data analysis sheet (in Github repository)