

Caste as a Facet of Dialectal Variation in Spoken Telugu

Brad B. Miller
Brigham Young University

1 Introduction

The sociological underpinnings of dialectal variation are many. Social and geographic isolation are often the most recognizable reasons, though there are other factors that may go into any type of dialectal change. Where social isolation is a factor, we find that it is not a matter of distance between social groups, but rather a matter of lack of contact for one or another sociological reasons. As such, where we see social isolation, we would also expect to see some form of dialectal variation, be it phonological or lexical, syntactic or even metalinguistic (Coulmas, 1997; Labov, 1970).

The caste system³⁰ grew from the Hindu religion as a method of social hierarchy that came to encompass most of Indian culture. Idealistic in nature, it prescribes a person's duties within their own caste as parts of a much larger community. Unfortunately due to its description of personal purity, the caste system came to breed a culture of social isolation, where cross-caste interactions were rare and often discouraged.

John J. Gumperz was among the first to show that dialectology could be used to show certain social distinctions. Though Gumperz showed that linguistic variations could be found in many different types of social groups, he is perhaps most famous for his work on the caste groupings found in a north Indian village (Gumperz 1958). He successfully was able to show that where caste contact was more frequent, dialectal variation was decreased. As such where caste contact was infrequent, stronger differences between dialects were more notable and even identifiable by speakers.

In more recent years with the advent of independence for India as well as increased influence from western countries through mass media, Internet and social media, more voices have arisen calling for a removal of the caste system. In the government, efforts have been made to reduce certain cases of caste backwardness by creating a reservation system within positions of employment as well as systems of higher education. This reservation system allows people from castes and tribes that are considered backwards and economically deprived to be given opportunities for higher education and better paid positions of employment. In spite of these efforts, though, it seems as though the

³⁰ It ought to be noted that the current synopsis of the caste system is rather crude, especially for such a complex topic. Opinions may vary, even among Hindus as to the specifics of the caste system, however the purpose of this paper is not to consider the caste system itself, but rather to discuss its role in language variation.

government is itself sending contradictory messages. In the most recent census, India's government included caste as an identifier for its citizens (Foy, 2010) as if to say that, while this method of discrimination is frowned upon by democratic systems, it still exists and is not something that will be removed entirely in the near future.

With this in mind, the question arises as to the amount of social isolation caused by certain social factors, including the caste system. If governmental reform and other influences have truly created a more socially integrating society, we would expect to see a decreased ability in speaker's ability to identify caste through speech. As such, the current study is an attempt to gain a more accurate view of the social stigma of caste while also looking at other sociological factors that might be identifiable through speech. Using an online survey as well as additional interviews with native speakers, we will look at the dialectal variations found in spoken Telugu, a Dravidian language spoken in southeastern India. It will be shown that while there are sociological factors that are identifiable through speech, including age, birthplace, education level and socioeconomic status, caste is one that is decreasingly found in speech, thus proving that the caste system is becoming less of a socially isolating force in Indian society.

This paper will begin by outlining the structure of the online survey as well as the methodology behind the experiment. We will then examine the qualitative opinions of various native speakers and their own experience with caste-specific dialectal variations. We will then look at the results from the experiment and implications from the study, followed by unanswered questions and suggestions for future research.

2 Survey Methodology

In order to better understand native intuition on certain sociological factors that are interpretable from speech, an online survey was created and administered to native Telugu speakers throughout the world. The survey consisted of two sections: the first was a set of qualitative opinion questions that allowed native speakers to simply voice personal opinions on caste and language ability, the second section consisted of a series of stimuli and corresponding questions developed with the intention of identifying which social factors can be uncovered through listening to a short speech segment. See Appendix A for the specific questions used for both sections. One question, Section 2 Question 2, asked where a person was from. The question offered three areas, each a large region of the state with distinct histories. The three regions, Telangana, Rayalaseema and Coastal Andhra each represent dialect areas that were previously noted through personal interviews. We see a map of these three regions in (Figure 1).

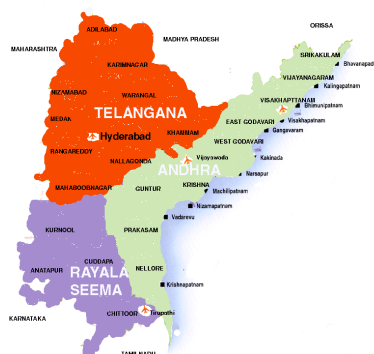


Figure 1: Three Regions of Andhra Pradesh

For the second section of the survey, stimuli were recorded using 8 male native Telugu speakers, some in the United States and others in India. Each subject was asked to reprimand a fictional person for being late to a meeting. In using a friendly scold each stimulus had at least one imperative, one stative, and one question. None of the stimuli were identical, but rather were ad-libbed by the subject in the moment with no preparation. In so doing each subject's personality was able to come through in the short audio segment.

Each audio clip was under 30 seconds in length. Stimuli subjects varied in age from 20 to 65 years old. Unfortunately only one subject was from Telangana and none were from Rayalaseema regions. At least one person was from each of the education options as well as economic options. The order in which the stimuli were presented was randomized for each subject. Only one stimulus was presented at a time, along with each of the questions on a single web page. The answers were given to each subject in the form of a drop-down option box. Due to the sensitive nature of the question of caste an option of "No Response" was included.

It should be noted now that the unique aspects of each audio clip have not been analyzed for this study. Variations in phonetics, syntax, morphology, intonation, stress, as well as a myriad of other suprasegmentals could be the causes of these perceived differences in the clips. Because the audio clips are instances of natural speech, these things are not accounted for. While in the future these things will be investigated, for purposes of this study it was simply relevant to discover if subjects could successfully determine the given factors.

Survey subjects were solicited through social media websites as well as contact lists of the Telugu Association of Utah and the Telugu Association of North America. Subjects came from nearly every major city within Andhra Pradesh, though several currently lived in the United States, the United Kingdom, or New Zealand. Each was a native speaker of Telugu, though the majority also spoke two other languages (most often English and Hindi). There were 78 test subjects, though only 12 were female. Many subjects were younger (born in the 1980's and 1990's) though several were older as well. All subjects had at least a secondary level education.

The surveys were presented using Qualtrics Online Survey software. Qualitative questions were organized in such a way that certain questions only appeared given positive responses to previous questions (see Appendix A).

Individual Interviews

Opinions on caste are quite unanimous throughout Andhra Pradesh. Caste is not something that is discernable through speech alone. Especially, say most, when that person is in school. However, it is interesting to note just how widely opinions vary, even when talking to the same informant. One young man, in particular, was quite adamant that caste was not something people even talked about any more, and was especially something that hadn't been an issue for many years. However when asked further if there were some people that spoke differently than him, he noted that his household maid spoke differently, and that most people from her community spoke in the same way that was quite difficult for him to understand.

It seems as though the question of caste causes many conflicting views and opinions. Every person interviewed was adamant that caste was no longer an issue of discrimination in India, citing governmental and educational reform. However, once those issues were brought up, it was consistently mentioned that caste was still an issue, as people's caste came into play when being considered for the reserved spots in employment or education. One interviewee even suggested that she knew of people who had implied that they belonged to a lower caste to be considered for above-mentioned reserved spots. All in all

caste is a touchy subject, and one of much confusion and debate among Indians and non-Indians alike. However when considering dialectology and caste together, most people will acknowledge that there are certain castes that are still very distinguishable from their speech, while others simply are not.

3 Data Collection and Results

In the following section we will examine each of the questions offered for each audio stimulus. With each of the questions we will show the results for each in the form of a table. Each question was run through a chi-square analysis to prove statistical significance. Each stimulus (A1-A8) can be seen along the vertical axis along with the statistics proving significance. Each asterisk indicates a p-value that is less than 0.0001. A double asterisk (**) indicates a p-value of less than 0.05. Each correct answer is found in bold. It should be noted that an ANOVA should be done in the future to eliminate any influence from individual social factors, however for purposes of this study a chi-square was sufficient.

How old is this person?

As can be seen in (Table 1), most subjects were able to correctly gauge the unknown speakers' ages. There were three instances where subjects assigned a different age to the speaker, though results were still statistically significant. This may have to do with voice pitch, though it is not something that will be considered for this study.

(2)

	NR	18-25	26-35	36-45	46-55	56-65	65+
A1 (30) $\chi^2(6)=107.231^*$	7	14	41	11	4	0	1
A2 (26) $\chi^2(6)=87.128^*$	4	13	32	25	2	1	1
A3 (65) $\chi^2(6)=70.795^*$	8	0	11	32	20	7	0
A4 (23) $\chi^2(6)=86.41^*$	6	11	36	19	4	1	1
A5 (55) $\chi^2(6)=119.795^*$	4	0	5	39	25	4	1
A6 (32) $\chi^2(6)=102.744^*$	4	14	39	16	5	0	0
A7 (20) $\chi^2(6)=220.487^*$	6	56	12	3	1	0	0

Where is this person from?

As was shown in the map in Figure (1), the state of Andhra Pradesh is often viewed as consisting of three smaller regions: Telangana, Rayalaseema and Coastal Andhra. Historically Telangana and Rayalaseema were never under British rule, but rather were ruled by the Muslim Nizam regime. For this reason native Telugu speakers often comment that people from that area speak using more "Persian" words, or borrowings from Urdu. Unfortunately for this study we only had one audio clip of a person from Telangana and none from Rayalaseema. However, that disparity aside, subjects were all successful in identifying a person's birthplace. The one subject that was not successfully identified is an interesting case in that he was born in Vishakhapatnam, in Coastal Andhra, but later moved to the capitol of Hyderabad in Telangana. Because he lived in both areas, it is assumed this is the reason behind this disparity.

(3)

	Telangana	Coastal Andhra	Rayalaseema	NR
A1 (T) $X^2(3)=67.949^*$	51	9	8	10
A2 (C) $X^2(3)=44.769^*$	5	44	17	12
A3 (C) $X^2(3)=49.692^*$	6	46	13	13
A4 (C) $X^2(3)=82.718^*$	4	54	11	9
A5 (C) $X^2(3)=91.641^*$	10	56	6	6
A6 (C) $X^2(3)=4.256$, $p=0.2351$	22	24	20	12
A7 (C) $X^2(3)=68.154^*$	10	51	7	10

What is this person's education level?

The original thought was that the number of English words used by each speaker would be indicative of that speaker's education level. However, upon further analysis it was found that the number varied with no indication of actual education level. In fact, as seen in the graph (4) below, the man with the highest education level used no English words, while two men of lower education levels both used many English borrowings. Disregarding the use of English words, we see that subjects were successful in identifying, for the most part, the education level of each Telugu speaker.

(4)

	NR	NE	Pri	Sec	Univ	Grad	Doc
A1 (Grad) $X^2(6)=88.026^*$ - English: 10	6	0	2	11	34	23	2
A2 (Sec) $X^2(6)=61.641^*$ - English: 0	7	6	15	33	6	11	0
A3 (Doc) $X^2(6)=42.795^*$ - English: 0	11	2	3	17	25	17	3
A4 (Pri) $X^2(6)=72.41^*$ - English: 4	7	28	20	21	1	1	0
A5 (Univ) $X^2(6)=68.641^*$ - English: 2	7	3	5	34	15	13	1
A6 (Grad) $X^2(6)=78.333^*$ - English: 1	7	0	2	10	33	22	4
A7 (Sec) $X^2(6)=81.205^*$ - English: 4	8	2	7	37	12	12	0

What is this person's economic status?

When asked to identify a speaker's socioeconomic status, subjects were often quite close to accurate, though errors tended towards calling a person "average". In the actual test, each subject was asked to identify if a person was either "very poor", "poor", "average", "rich", or "very rich". However, it is quite apparent that the differences between calling a person "poor" and "very poor", as well as the differences between "rich" and "very rich" are extremely subjective and variant across individuals. Thus, for analysis these categories were combined. Once combined, we see that subjects did quite well.

(5)

	Poor	Average	Rich	NR
A1 (Rich) $X^2(3)=41.897^*$	1	41	20	16
A2 (Avg) $X^2(3)=49.487^*$	14	45	3	16
A3 (Rich) $X^2(3)=33.487^*$	2	37	24	15
A4 (Poor) $X^2(3)=57.282^*$	46	20	1	11
A5 (Rich) $X^2(3)=46.103^*$	0	32	36	10
A6 (Avg) $X^2(3)=50.718^*$	0	41	27	10
A7 (Avg) $X^2(3)=42.308^*$	8	44	11	15

To which caste/community does this person belong?

Researchers in the past have analyzed some form of each of these previously addressed questions at some point. We haven't seen many new trends here that make these findings unique from theirs. However, now we are coming to question of caste. This question is not only unique because it hasn't been addressed in nearly 50 years, but also because of the increased stigma associated with the topic. Asking a person to assign a caste to another person is not a common thing, and in fact most people view it as a bad thing. At the request of one of my Telugu mentors, and once it was apparent that the question was taboo to ask, it was made optional. Its taboo nature was confirmed when we saw that most subjects simply refused to answer it. Over 40 of the 78 subjects reviewed refused to answer each of the caste/community questions.

When subjects did answer, though, we saw some very thought provoking trends. Certain castes are more popular or well known across larger communities. These two castes, the Reddy and Chakali castes, were most easily identified. Other caste groups are much smaller and not as well known. One in particular, that of the Jalari community, has quite a unique way of speaking, however because it is such a small community, it is likely that subjects never had heard a Jalari speak before, and thus did not know to identify certain speech with those that caste.

The Brahmin caste, on the other hand, is the most widely known caste. It is interesting to note that the A3, who most commonly was misdiagnosed, was erroneously described as being Brahmin rather than Yadava(Golla). The Yadava caste is one of shepherders, and not typically one ascribed to men of higher education and socioeconomic power. It is likely for this reason that A3 was assumed to be a Brahmin because of the other factors. Along the same lines, the Reddy caste is a wealthy and prominent one. While this does not explain why some people heard a Vaishya man and perceived him as a Reddy, or even a Brahmin, it does explain why some leaned in that direction, seeing as it is a recognizable caste.

Though the graph shown below is noteworthy, it is far from saying that caste is 100% identifiable from speech. In personal interviews only one person stated that he was able to identify some castes from speech. All other personal informants suggested that hearing caste was something that used to happen, but no longer is something people can do. Even so, while being aware of linguistic differences unique to caste is not something people claim now, this does not prove or disprove the existence of such differences.

(6)								
	B	CH	G	J	K	R	V	Y
A1 (Reddy)	3	1	0	1	2	11	1	6
$\chi^2(7)=30.36^*$								
A2 (Gavara)	5	6	4	0	2	3	4	1
$\chi^2(7)=9.24, p=0.235$								
A3 (Golla)	21	2	1	1	5	4	3	0
$\chi^2(7)=30.36^*$								
A4 (Chakali)	0	9	3	5	0	0	2	5
$\chi^2(7)=70.459^*$								
A5 (Vaishya)	6	0	1	0	6	10	5	4
$\chi^2(7)=21.5^{**}$								
A6 (Reddy)	4	0	2	1	4	11	1	2
$\chi^2(7)=21.16^*$								
A7 (Jalari)	9	2	1	2	4	4	3	3
$\chi^2(7)=12, p=0.1$								

4 Implications and Future Research

As we have seen, there are several things that are identifiable by speech, including to some degree caste. We expected to see a decreased ability in speakers' detection of caste through speech because of the breakdown of caste barriers in society. And while this is true to some degree, some castes are still identifiable (Reddy and Chakali).

Through personal interviews it became apparent that while some youth are raised in households that discourage cross-caste interactions and marriage, most children do not discriminate between castes, especially in play situations. In addition school and other governmental institutions do not allow for any segregation, thus bringing everyone together to learn from the same source.

Beyond the issue of caste, it became apparent that geographic distinctions should be increased to include many smaller subregions, as people noted that speech is quite effective at placing a person from East or West Godavari, the Guntur region, and even the capitol city of Hyderabad.

Later this year I will be going to India to do a more in-depth documentation of linguistic differences between castes, education levels, men and women, age groups, as well as many other social aspects of language. In spite of it being only a pilot study, this current paper was able to show that caste distinctions are still audible to some degree. Other social and personal factors are also identifiable.

5 Acknowledgements

I would like to thank the National Science Foundation for a Graduate Research Fellowship that has made this and future research a possibility. Thanks are also due to my mentors Drs. Janis and Charles Nuckolls, as well as Dr. MV Krishnayya. Finally I would like to thank my informants and friends who provided me with such invaluable information.

Appendix A

Section 1

1. Do you speak Telugu?
2. Is there a “proper” Telugu?
3. Where is “proper” Telugu used?
 - a. In _____ region (please specify)
 - b. School
 - c. Books
 - d. Newspapers
 - e. News
 - f. Films (please specify)
4. Is there a way to speak Telugu badly?
5. Who speaks Telugu badly?
 - a. Uneducated people
 - b. People from _____ (please specify)
 - c. Lower castes
 - d. Other (please specify)
6. Can people change the way they speak to indicate they are from a specific caste or community?
7. What can people do to indicate they are from a specific caste or community?
8. Can you talk like someone from another caste-community?
9. What do you do? Which caste-community do you impersonate when you do this?

Section 2

Click play to hear part of a conversation between two people. Please answer the following questions about the man who is speaking.

1. How old is this person?
 - a. (18-25, 26-35, 36-45, 46-55, 56-65, 66-75, 75+)
2. Where is this person from?
 - a. (Telangana, Coastal Andhra, Rayalaseema)
3. What is this person’s education level?
 - a. (No education, Primary education, Secondary education, University education, Graduate education, Doctorate education)
4. What is this person’s economic status?
 - a. (Very poor, Poor, Average, Rich, Very rich)
5. To which caste/community does this person belong?
 - a. (Brahmin, Chakali, Gavara, Jalari, Kshatriya, Reddy, Vaishya, Yadava/Golla, Other)
6. Is there anything else you can tell about this person?

References

- Coulmas, Florian (Ed.). (1997). *The Handbook of Sociolinguistics*. Oxford: Blackwell Publishers.
- Foy, H. (2010, September 9). India to conduct first caste census in 80 years. Reuters. Retrieved from <http://in.reuters.com/>
- Gumperz, J. J. (1958). Dialect Differences and Social Stratification in a North Indian Village. *American Anthropologist*, 60:668-81.
- Gumperz, J. J. (1971). *Language in Social Groups*. Stanford, CA: Stanford University Press.

Texas Linguistics Forum 54:79-87

Proceedings of the Nineteenth Annual Symposium About Language and Society – Austin

April 15-17, 2011

© Miller 2011

- Iannone, Ron 2004. Higher Education's Caste System. *College Student Journal*, Vol. 38(1), p.9.
- Kachru, B. B., Kachru, Y., Sridhar, S. N. 2008. *Language in South Asia*. Cambridge University Press.
- Labov, W. (1970). The Study of Language in its Social Context. *Studium Generale*, 23: 30-87. In Fishman (1971-2), vol. 1), Giglioli (1972), and Pride and Holmes (1972).

Department of Linguistics and English Language
4064 JFSB
Provo, UT 84602

brdmiller96@gmail.com