THE ANALYSIS OF SOFTNESS IN INDIAN LANGUAGES

Aditya Raghuvanshi
Utsav Shekhar

INTRODUCTION :

We came across an idea that why are mostly soft spoken the conversations which are intended to be very polite in nature are mostly found in Urdu language. For example, Ghazal, Shayri, Qawwalis and other genres which are very famous around the world and are known for their polite nature. We then thought of going deep into this topic and we found out that there are a certain group of languages which according to general people are considered rude(for example : Mandarin, German, Tamil) and some languages are considered very sweet (for example : Bengali, Urdu, French) .

So, we thought of actually going deep into this topic and making this our project. Our initial aim was to find out why these assumptions are carried out by the people and viewing these assumptions through a linguistic perspective.

PROBLEM STATEMENT

To find Linguistic factors responsible for the assumption of the society behind some languages to be soft spoken/ sweet and some languages to be rude/ not pleasant in nature.

HYPOTHESIS

In our project, we are aiming to prove that Urdu is one of the softspoken languages or it is comparatively more soothing as compared to Hindi, and also include other foreign / Indian languges for our analysis to get a wider perspective for our problem statement.

METHODOLOGY

STAGE 1

Firstly we tested our hypothesis by analysing that the soothingness factor is lost in translation or not. We conducted a survey to collect data for our preliminary method.

Data collection

So we surveyed 50+ participants and asked them about which word according to them is more polite/beautiful/ soothing in nature.

Sr. no	Words	Percentage
1	mushkilaat	56.1%
	dikkat	43.9%
2	Alfaz	69%
	Shabd	31%
3	Ilzaam	68.3%
	Aarop	31.7%

We also gave them verses in Urdu and translated them to Hindi to check wheater the beauty/ soothingness is lost in translation or not.

Sentence	percentage
Yeh kaisa ishq hai urdu zabaan ka, maza ghulta h lavzo ka zabaan par, ki jaise paan me mehenge imam ghulta hai	82.9%
yeh kaisa pyaar hai urdu boli ka, anand milta h shabdo ka mukh par, ki jaise	17.1%

paan me keemti kimam milta hai	

We also did the opposite of translating Hindi to Urdu, and that too for very aggressive sentences just to confirm that if due to translation, the beauty. Soothingness can be lost, can it be attained back?

Sentence	Percentage
Tu janta nahi mera baap kon hai!	24.4%
Kya tumhe ilm hai ki hamare walid sahab kon hai !	75.6%

Data Analysis

Those words were Dikkat and mushkilaat , we can see that is those pairs of words the results differed from the rest. Whereas in other word pairs the differences are major. In Urdu verses translated to Hindi, data shows that people preferred Urdu are more soothing/royal to ears. And to the conclusions of our survey, we can see that even the aggressive Hindi sentences sounded more soothing to ears in Urdu.

Interpretation

The survey strongly supported our initial hypothesis that the beauty of Urdu is lost in translation. We also tweaked the survey a little bit so reduce observers paradox, we added Dikkat and mushkilaat , we can see that is those pairs of words the results differed from the rest which also strongly supported the fact that both were Urdu. Rest all of the word pairs showed that our hypothesis was correct and Urdu seemed more polite/ soothing in nature than Hindi.

In our survey we also gave them one polite and one aggressive sentence and tried finding out that is the beauty of language or the emotions they are trying to express were really lost in translation.

In our polite sentence which was originally in Urdu when translated in Hindi as per our survey shows that it clearly lost its beauty and same result was obtained when an aggressive Hindi sentence was translated to urdu, the sentence lost the feelings it was trying to convey

STAGE 2

We thought that the soothingness of the language is due to its phonetic characteristics. Se we now thought of collecting data for phonetic frequency of various languages including Urdu and Hindi and comparing them.

Data collection

We have gathered the information in the table below by analysis of more than 5 research papers from different countries. This data has been made by using

- 46,000 Urdu words
- 2,38,604 Hindi words
- More than 9,900 phonemes from famous literary texts and speeches in Bengali
- 30,37,702 Tamil Words

IPA	Percentage Distribution of Urdu	Percentage Distribution of Bengali	Percentage Distribution of Hindi	Percentage Distribution of Italian	Percentage distribution of Tamil
ə	11.1224923	8.96	11.12	10.76	17.80666773
a:	9.76502131	11.32	8.88	11.46	5.90795608 1
r	6.60801185	7.01	5.91	6.30	4.96157507 1
n	5.44102531	4.97	3.48	7.60	1.74346513 3
i:	5.00291118	6.77	4.27	9.30	0.51739572 8
l	4.9177323	^_^	4.82	4.22	^_^
l	3.92074411	4.14	2.45	^_^	0.51739572 8
m	3.78740503	2.78	3.28	2.52	4.97744751 01
s	3.78165455	3.64	3.78	3.74	0.28742241 3
ʈ	3.62100073	3.83	0.57	^_^	^_^
k	3.04415644	^_^	7.14	3.32	10.98809416
e:	2.71781712	1.37	1.96	^_^	2.71939211 6
b	2.56866424	4.44	1.78	0.50	^_^
ʊ	2.35050568	0.09	1.09	^_^	0.83636080 6
t	2.16505294	^_^	2.89	4.62	5.22688223 5
ɖ	2.12515904	2.51	0.36	3.39	^_^
z	1.77545842	^_^	^_^	0.39	0.75014924 6
h	1.71471905	3.42	4.82	1.43	0.06371719 2

oː	1.71292203	7.82	2.31	^_^	0.023985567
p	1.70429632	2.14	2.66	2.08	6.339317172
v	1.48937241	0.58	1.62	1.48	4.280880344
ðː	1.48362194	^_^	^_^	^_^	0.023985567
j	1.42432019	1.46	1.39	1.98	0.168278089
uː	1.28702765	3.08	0.81	3.16	0.591146925
æ	1.23635161	0.98	6.85	^_^	1.425030822
ʃ	1.23024174	0.35	^_^	0.22	^_^
ʤ	1.19717652	^_^	^_^	0.38	^_^
g	1.17704986	1.59	1.31	0.48	^_^
f	1.16195487	^_^	0.35	1.06	0.735995487
d	0.99267534	^_^	1.68	3.31	^_^
tʃ	0.89527671	^_^	^_^	0.77	^_^
q	0.74756144	^_^	^_^	^_^	^_^
x	0.58978285	^_^	0.7	^_^	^_^
r	0.56138989	^_^	^_^	^_^	4.961575071
ɛ̃ː	0.49813469	0.79	^_^	2.83	3.574203358
ŋ	0.46614769	0.59	^_^	0.18	1.213103555
ãː	0.38096881	^_^	1.01	^_^	5.907956081
kʰ	0.3719837	0.88	0.7	0.20	^_^
ɔ	0.33352741	6.63	0.2	1.38	^_^
ɣ	0.28752363	^_^	1.31	^_^	^_^
tʃʰ	0.25553663	^_^	^_^	^_^	0.113836665
bʰ	0.24798913	0.47	0.67	0.25	^_^
ĩː	0.237207	^_^	4.27	^_^	0.517395728
ʤh	0.20521999	^_^	^_^	0.10	^_^
ʂʰ	0.19946952	0.75	^_^	^_^	^_^
ʈʰ	0.19084381	0.74	1.15	0.67	^_^
tʰ	0.17790525	^_^	1.15	^_^	0.522678114
ɒʰ	0.15634097	0.36	0.35	0.16	^_^
gʰ	0.17035775	0.15	0.2	0.004	^_^
æˑ	0.1347767	^_^	6.85	^_^	1.329038004
ũː	0.11932231	^_^	1.73	^_^	0.591146925
ɽʰ	0.08086602	0.64	^_^	^_^	^_^
dʰ	0.06972448	0.18	0.36	0.006	^_^
o	0.02515832	^_^	2.31	^_^	0.023985567
ʒ	0.01868904	^_^	^_^	^_^	^_^
æ̃ː	0.0154544	^_^	1.96	^_^	3.574203358
ʔ	0.01293856	^_^	^_^	^_^	^_^
eˑ	0.01186035	^_^	6.85	^_^	1.425030822
nʰ	0.00431285	^_^	^_^	^_^	1.743465133
lʰ	0.00287524	^_^	^_^	^_^	3.609840502
rʰ	0.00287524	0.02	^_^	^_^	^_^
ɔː	0.00143762	^_^	^_^	^_^	^_^

**The places where data was unavailable is marked by the "^_^" symbol.

Data Analysis

After analysis of the above dataset. We have concluded with the following, this is the percentage of vowel phonemes used in the respective languages.

Bengali – 45.4%
Urdu – 41.73%
Hindi – 40.21%
Italian – 37.51%
Tamil – 35.92%

Interpretation

It's a known fact that a consonant cluster does not sound good and generally vowels are sounding letters in a word. So in a word if there are less consonants then the variation of sounds is lesser in it as compared to the words and us humans feel that there is a constant flow in the word pronunciation and it feels easy to pronounce it as compared to words with a lot of consonants with very less vowels.

By the above data we are trying to conclude that the languages with more number of vowels are more soothing in listening.

STAGE 3

Our Stage 2 analysis was language and linguistic-dependent analysis. Stage 3 is our analysis of languages based on society and the perspective of common people. The final stage of our methodology focused on bringing out the final conclusion from the experiment. The basic idea of the experiment revolved around the sociolinguistic fact that, the beauty and such face features and characteristics of the language are perceived by the hearer on the basis of the preconceived image of the original speakers of the language.

Data collection

The data collection for this stage majorly involved, the profiles of the native speakers of the language, during the ancient period. (The languages mainly employed for this use were Urdu, Hindi, Bengali and Tamil)

Data Analysis

We analysed, that Urdu and Bengali were mostly the languages of kinship and professionals in the king's court and were mainly associated with aristocracy. Whereas Hindi was the language and common people and working-class people. Tamil was the language of an isolated group of people for a long period of time, which were considered different from the local common speakers.

DISCUSSION

STAGE 1

So first our main motive to find out whether the beauty of words are lost in the translation from Urdu text to Hindi or vice versa. So we first tried to survey the son word level and slowly levelled it up to the sentence level

We used 3 pairs of words from which our aim was to find whether the Urdu word sounded more soothing in nature. We knew that this question may have an observers paradox so we included a pair of Urdu-Urdu words so that we may get to know if the participant is being partial in any sense or not. So most of the users responded that Urdu was more soothing or beautiful in nature as compared to its Hindi counterpart. The second part of our survey was that which of these words will be used by the speaker in formal situations, and to our prediction, the results came out to be that people tended to prefer Urdu words over Hindi words in formal attire. The last part was about the sentences in which we wanted to test that the beauty of a language is lost in translation or not, so we did two surveys on this in which on had a sentence translated from Urdu to Hindi and one part of the sentence was translated form Hindi to Urdu. Our results concluded that yes the feelings a language is originally trying to convey is lost in translation from that to some other language.

We also asked our participants the reason behind Urdu being so polite/soothing in nature, most of the responses given included that “Urdu in itself is polite” or “ some words in Urdu are polite”. We were not convinced with this hypothesis because as we know that by the phenomenon of loan words and borrowing of words, words are not permanent to any fixed language rather a single word can be a part of many languages. So rather than focusing on word to word analysis we refocused to letter to letter or phonetic analysis for getting the microscopic view of what was actually happening.

STAGE 2

After Stage 1 of our project we understood and we proved that Urdu in itself is a comparatively more soothing language.

We then thought that what might be the reason behind this soothingness, after discussions we thought that any language can be polite in its sense, and if a person who is unaware of that language is rating languages on the basis of soothingness then it has to be based on some phonetic properties of that language.

So we collected data sets of different languages especially those that were in limelight if people discussed about soothingess or sounds of language such as Bengali is considered the sweetest language of the world by some sources, Urdu is similar to Bengali and if we talk about the other side than Tamil and German are considered among not-so-good sounding languages. We collected their data from various sources and research papers which had the frequency of phones used which they had calculated using a large corpus and we proposed a hypothesis that a cluster of consonants which have lack of vowels generally have a multiple types of sounds which makes a word sound a bit more harsher or complex to pronounce and also sounds harsher to the hearer. This hypothesis of ours is not yet proven, and we are making this hypothesis based on the survey conducted by us with 50+ participants.

The evidences collected till now are not very apt to prove our initial hypothesis correct. So we thought of exploring this in the field of sociolinguistics which is discussed in stage 3.

STAGE 3

The main idea of stage 3 was based on the fact that the perception that the hearer carries with them is based in the perception of the native speaker of that language. So the languages which we were working on were Urdu, Hindi, Italian, Tamil and Bengali. And as we know that Urdu was spoken by the aristocracy classes of south-east Asia which includes Afghanistan, Pakistan, northern India, Iran, Iraq etc. most of the literature of Urdu is mostly composed of people having some royal background, whereas in dialects of Hindi the literature was contributed by all classes of people. So to the common people of India, Urdu sounds a bit more royal and sophisticated because of the perception of the past and lack of knowledge of Urdu among common masses in India. Urdu in today's India is mostly spoken by Royal households, poets, music composers, so common people do not consider Urdu to be a part of their daily conversation rather, they have developed a mindset that Urdu is mostly spoken by these high-status artistic societies and its not meant for daily usage. This perception is very strong and due to this people while listening to Urdu think as if it is something very elite in nature.

CONCLUSION

Our initial hypothesis that some Indian languages sound more soothing than others is disproved, accounting for the present circumstances and evidence. The collection of evidences we had are not enough to linguistically prove Urdu as a more soothing language and we don't give major weightage to the social factors which we gathered in Stage 3 of our project because they are very subjective to the type of crowd we are surveying and not universal.

ACKNOWLEDGEMENT

We would like to thank Prof. Deepti Mishra IIITH & TA Shubhankar Kamthankar for motivating us to think new and guiding us throughout our journey.

REFERENCES

https://www.livemint.com/Opinion/B9PWW3liEjY59D4Y1IGYkL/The-romance-of-Urdu-is-lost-in-translation.html

https://www.livemint.com/Opinion/B9PWW3liEjY59D4Y1IGYkL/The-romance-of-Urdu-is-lost-in-translation.html

https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0161.xml#:~:text=Linguistic%20politeness%20can%20be%20defined,what%20society%20or

https://en.wiktionary.org/wiki/Appendix:Hindi_pronunciation

https://www.sttmedia.com/characterfrequency-hindi

Haris Bin Zia[1], Agha Ali Raza[1], Awais Athar[2] PronouncUR: An Urdu Pronunciation Lexicon Generator EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

https://www.sttmedia.com/characterfrequency-italian

Estimation of the Frequency of Occurrence of Italian Phonemes in Text J. Arango*, A. DeCaprio*, S. Baik†, L. De Nardis‡, S. Shattuck-Hufnagel† and M.-G. Di Benedetto*†‡

Thennarasu Sakkan A Statistical Study of Tamil Corpus : https://www.academia.edu/8745513/A_Statistical_Study_of_Tamil_Corpus

Ferguson, Charles A., and Munier Chowdhury. "The Phonemes of Bengali." *Language*, vol. 36, no. 1, 1960, pp. 22–59. *JSTOR*, https://doi.org/10.2307/410622. Accessed 17 Nov. 2022.

https://docs.google.com/forms/d/e/1FAIpQLScK0FeaqKtx2QYr3JU6b9AEwaSfYtRVdIxZ5y2v4xPOoR9s4w/viewform?usp=sf_link