Hinglish 2 Hindi Converter: A Review

Rubana Khan[1], Rajatkumar Singh[2], Tathagat Khanorkar[3], Vinit Girde[4], Divyesh Ramteke[5], Payal Atey[6]
Assistant Professor[1], Student[2, 3, 4, 5, 6]
Department of Computer Technology
Priyadarshini College of Engineering, Nagpur, India

Abstract:
Hinglish is a language which is a combination of English and Hindi, especially a type of English that involve many Hindi words. The language, which is nearly English with a Hindi-ascent, serves as common language spoken by common man of India. The mother tongue of the middle-class and the so-called up-to-date people through which millions of Indians connect. It mixes Hindi and English words together, it does not follow English sentence format, it creates its very own variant of English called as Hinglish.

I. INTRODUCTION

Hinglish is a molten mixture of the English and Hindi languages that you confront very usually in India. English has been very widespread in India ever since the British *Raj*, of course, and remains a very common language in that area of the world. Most educated Indians speak English very well, and it has remained the language of the business, education, and technology sectors in that country. Hindi is one of the official languages of the Republic of India, spoken by an approximate 260 million people. It's not shocking that in a place such as India – where English is a crucial skill for success and Hindi is a common local language – that people would speak both languages, but the fascinating thing is that it's now almost universal in the area for people to speak a mixture of both languages relate to as *Hinglish*. The language is mostly a subconscious code-switching endeavor, with speakers openly mixing English words and sentences with Hindi (and even some other local languages). Most Hinglish speakers don't even realize that they are swapping back and forth – they are naturally using the best words and phrases from each language to communicate their thoughts. In the modern communication-based society, no natural language suggests to have been left untouched by the craze of code-mixing. For different communicative objective, a language uses linguistic codes from another language. This allow rise to a mixed language which is neither totally the host language nor the adventitious language. The mixed language poses a contemporary threat to the complication of machine translation. It is unavoidable to identify the "adventitious" elements in the source language and process them appropriately. The adventitious elements may not appear in their original form and may get morphologically mutated as per the host language. Further, in a convoluted sentence, a section/assertion may be in the host language while another section/assertion may be in the adventitious language. Code-mixing of Hindi and English where Hindi is the host language, is a common phenomenon in day-to-day language usage in Indian metropolis. The scenario is so regular that people have started seeing this a different variation thoroughly and calling it by the name Hinglish. In this paper, we exhibit a mechanism for machine translation of Hinglish to pure (standard) Hindi and pure English forms.

IMPACT AREA OF HINGLISH

Hindi and English languages both are individually monolinguals but the use of both this language makes them a hinglish speaker. The user feels it is continent to use hinglish, a mixture of both English and Hindi language and it also increases their fluency in a language some time. In an English Job market many people have lost their job due to problems in English language fluency but this does not mean by any means the people are entirely shifting towards Hindi language, now a day many individuals are now shifting their focus towards hinglish which has gained high value in recent times. It is very useful when the user has to use limited English and have to make a bold impression to get the opportunity. The use of hinglish can prove very beneficial, especially in a local market for the advertisement purpose. Over time, our model suggests that all three speech communities will survive, and Hinglish is not likely to replace English or Hindi, as some reports are suggested.

OBJECTIVE OF THE PROJECT

Hinglish 2 hindi converter allows the user to represent the hindi text using hinglish words. Hinglish language has really started to influence into various aspects of today's Indian youth in both home and work environments. Nowadays people are not trying to just mingle both English and Hindi words, but they are using it in such a innovatory and dynamic way which can be easily seen in today's advertising era. In this competitive advertising world English alone can't prove beneficial to sell a product in the market. Hinglish is the language of advertising and it is going to get more and more popular and will determinate the barricade of this industry in our country.

II. LITERATURE SURVEY

Paul Taylor, Alan W Black, Richard Caley had brought up a new method for storing data related to language and its structures in a text to Speech system. Text to speech system is a method of converting normal language text into speech. The constituents of any language such as words, phones and diphones are stored as featured structures in a general object defined as linguistic item. These objects can be configured at run time and can also contain any random information in the feature structure. The words which can be pronounced similarly bt have different meaning or we can say that those words have same vocal sound are stored in the database. These kind of language structure can be shown by any graph but are commonly defined with trees or lists. Utterance or we can say a spoken word or vocal sound,

contain all types of structure of language contained in a single statement [1]. Deepa S.R., Kalika Bali, A.G. Ramakrishnan, Partha Pratim Talukdar has came up with the problem of splitting of Hindi complex words and it affects the quality of phonetizer for Hindi Speech Synthesis. As the words involved in Hindi language mainly the compound words do not have space or hypen as a connector in them, therefore there arise a problem of synthesising the Hindi compound words with those existing methods of synthesising. They proposed a new method of taking out a Hindi compound words from the compilation of whole Hindi language [2].

N. Sridhar Krishna, Partha Pratim Talukdar, Kalika Bali, A.G. Rama Krishnan gives the information about one of the smallest unit of speech which can make one word different from other word (phoneme) for Hindi.They have given us the Classification and Regression Tree (CART) based database for the supposition of words [3]. Lakshmi Sahu and Avinash Dhole described text-to-speech (TTS) system for Indian languages (Viz., Hindi, Telugu, Kannada etc.) to generate human voice or speech through a speech computer or speech synthesiser. In a text-to-speech system, synthesised speech can be generated by concatenating prices of recorded speech or from text generated on home computer. The quality of speech synthesiser is judged by its similarity to the human voice and by its ability to be understood clearly [4]. M. Habibullah Pagarkar, Lakshmi Gopala krishnan, Nimish Sheth developed a new method using davangari script to compress a complete speech. [5].

III. CONVERSION LOGIC

In this user will type the Hinghlish text which is written in English in the given textbox. For converting the given English character into a Hindi character there is requirement of a converter which convert the given English character into corresponding Hindi character. So, if the user type input as "Hum" then the output should be "हम ". For this we make two separate files, one file will store Hindi character and their equivalent English character. Second file will store the other properties of file for removing the complexity of the program. We have recorded the consonants and vowels of Hindi alphabets probably known as Barahkhadi into the equivalent English character. When user gives the input into the textbox our converter directly converts this into corresponding Hindi character without checking any semantic error. Our files map that character into Hindi character. For mapping we used type casting of character by using the phenomena of Unicode. We convert this Unicode into a string by using type casting. We make this software into java hence for storing the Hindi character and their equivalent English Character we created different classes. By using these classes, we convert their Unicode into equivalent string value.

IV. CONCLUSION

It's estimated that native Hinglish speakers will outnumber English speakers in India very soon as the language spreads. In fact, British diplomats were r e c e n t l y officially instructed to learn Hinglish in order to be able to perform their duties in the country effectively. Films and songs have been written entirely in Hinglish. This project explains a simple approach to perform transliteration (i.e. Hinglish to Hindi character mapping. This work can be extended to read a Hindi pdf file written in devnagari script.

V. REFERENCES

[1] Paultaylor, Alan W Black, Richard Caley, "THE architecture of the festival speech synthesis".

[2] Deepa S.R., KalikaBali, A.G. Ramakrishnan, Partha Pratim Talukdar, A.G. Ramakrishnan "Tools for the development of a Hindi speech synthesis system", 5th ISCA Speech Synthesis Workshop – Pittsburg.

[3] N. Sridhar Krishna, Partha Pratim Talukdar, Kalika Bali, A.G. Ramakrishnan , "Duration Modeling for Hindi Text-to-Speech Synthesis System".

[4] Lakshmi Sahu and Avinash Dhole, "Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion", International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012, ISSN 2250-3153

[5] M. Habibullah Pagarkar, Lakshmi Gopalakrishnan, Nimish Sheth, Rizwana Shaikh, Virag Shah "Language Independent Speech Compression using Devanagari Phonetics",cs.jhu.edu. 200http:// developer. android.com/ distribu te/ googleplay/p ublish/ preparing.html. (2013) .