

## BRAHMANISM, BRAHMINS AND BRAHMIN TAMIL IN THE CONTEXT OF SPEECH TO TEXT TECHNOLOGY

R S Vignesh Raj<sup>a</sup>, Ashik Ali<sup>b</sup> Dr. BabakKhazaei<sup>c</sup>

<sup>a</sup>Research degree student, <sup>c</sup>Senior Lecturer

Sheffield Hallam University, Sheffield, United Kingdom

<sup>b</sup>Software Developer Fireflyapps Limited, Sheffield, United Kingdom

Corresponding email: classicalraj@gmail.com

### Abstract

Brahmanism in today's world is largely viewed as a 'caste' within Hinduism than its predecessor religion. This research paper explores the impact and influence religious affiliation could have on using certain advanced technologies such as the voice-to-text. This research paper attempts to reintroduce Brahmanism as a distinct religion and justifies through an empirical study and other literature as to why it is important, more specifically in language-related technology.

**Keywords:** Tamil; speech to text; technology; Brahmanism; Brahmins

### 1. Introduction

Speech-to-text is a fascinating area of research. The speech-to-text system exists for popular languages such as English. Whilst dealing with the user acceptance of technology, previous experiments suggest that it is vital to consider the religious affiliation as it could influence 'pronunciation' which is perceived to be very important for the users to be able to use voice-to-text technology in syllabic languages such as Tamil (Rama et al., 2002). Some of the previous experiments have identified the effect of code mixing, mispronunciation or in some cases the inability to correctly pronounce a syllable by the native Tamil speakers (Raj, Ali&Khazaei, 2015). This research paper aims to look into the code mixing and pronunciation aspects of 'Brahmin Tamil' - a dialect of Tamil spoken by the Brahmins who speak Tamil as their mother tongue.

### 2. Brahmanism in the context of religion

The term 'Brahmanism' has been in popular usage by several research scholars. (McGovern, 2012) in his research on Brahma: An early and ultimately doomed attempt at a Brahmanical synthesis argues about the late- to post- Vedic "Brahmanical synthesis' centered around the conception of Brahma as both supreme creator God and ultimate goal for transcending samsara. The Brahmins follow Brahmanism- which is not merely a religious sect but a system of ideas and practices definitive of an idealized social order. (Bronkhorst, 2015) in the review essay on Greater Magadha and the New Brahmanism, seemed to have used the term 'Brahmanism' to suggest the possibility of Brahmanism as a distinct religion with Sanskrit as its language. In Droixhe's review article in the Journal of the American Oriental Society (1980), the author suggests that Sanskrit was a language invented by the Brahmins and they considered it to be their sacred language. The author further mentions interesting similarities between Greek and Sanskrit. Anecdotal references suggest that in the present world, Sanskrit influence and usage could be found in the Brahmin dialect of most of the Indian languages and not just in Tamil. Therefore, it was felt that Brahmanism needed to be classed as a distinct 'religion' in order to study the Brahmin dialect as it is significantly different in the case of Tamil. It is strongly opined

that by classifying Brahmanism as a separate religion, it could enable the society and researchers to approach from a different perspective.

### 3. Related work

There is very little literature available in relation to experiments and tests on Brahmin Tamil within the context of technology and more specifically from a user 'ability' and 'acceptance' perspective. Therefore this could be considered as one of the first efforts in highlighting the necessity, to classify Brahmanism as a distinct religion and the reason why the Brahmin Tamil dialect should be considered in any Tamil-related language-based technology experiments and researches.

(McDonough & Johnson, 1997) through their investigation into the basis of the contrast among five liquids in a dialect of Tamil, the Brahmin dialect of Tamil is said to have unusual inventory of five 'distinctive' sounds: plain and retroflex rhotics, and plain and retroflex laterals, and a fifth which has been previously described as rhotic.

Some of the words that the authors have chosen for the purpose of the study were

பரம (param, Supreme being), பழம் (pazham, fruit), பரி (pari, horse), பழி (pazhi, blame)

As a part of their experiment, the female speaker of the Brahmin Tamil dialect had to read the word list at three self-selected rates (slow, normal and fast). The authors observed that the speakers who did not speak this dialect could not produce this sound or considered this sound belonging to another dialect than the one they spoke. The results of the investigations concluded that the Brahmin speakers had no difficulties in perceiving the intended sounds and the Non Brahmin speakers tended to confuse ழ ழ (zha, La)

#### A. Findings from our previous experiments

In order to understand how native Tamil speakers perceive, understand, speak and write Tamil language, an experiment that involved listening and writing components was created. The speaking component was deliberately excluded as a part of the experiment with an intention to observe the speech pattern and pronunciation of the participants while they spoke the language in a more relaxed environment. The selection of the participation was on the basis of Tamil as 'mother tongue' regardless of the geographic domicile. However, most of the Brahmin participants were 'identifiable' through certain gestures and words very unique to the Brahmin dialect. The quantitative dictation experiment was based on the anecdotal evidence that a vast majority of the native Tamil speakers have difficulty in pronouncing the syllables ழ and ழ (Raj, Ali&Khazaei, 2015). Therefore, the words for the dictation experiment was carefully chosen with more emphasis on syllables unique to Tamil language and on the basis of common pronunciation error. The participants were asked to write the words in Tamil script and in Roman script so the consistency, perception and expectation could be studied. (Raj, 2014)

Tamil being a syllabic language, the pronunciation is consistent with the written form of the language. Therefore in an ideal scenario, the expectation was that all the participants would get the words right in both scripts and as a 'native' Tamil speaker they should be able to correctly pronounce the syllables.

For the purpose of this research paper, we shall classify the participants into Brahmins and Non-Brahmins based on their 'identity as Brahmins'. The term Non-Brahmins shall refer to Hindus, Muslims, Christians, and other participants with other religious beliefs including atheism.

### **B. Preliminary observation**

- All participants regardless of their religious affiliation displayed some degree of interest in participating in the research experiment.
- Code-mixing element was apparent in almost all the participants' speech regardless of their religious belief. The Brahmin speech had some amount of Sanskrit vocabulary in addition to English.
- Participants were uncomfortable in both proper English and in proper Tamil. Most of them communicated their feelings and spoke in a code mixed version which they refer it to as 'Tanglish'.
- Most participants were educated in schools where the medium of instruction was English. From the interaction, almost all the Brahmin participants went to a school where the medium of instruction was English but studied either Tamil or Sanskrit as a second language.

It was felt that this focus group exercise was essential in order to capture the participant's code mixing element while speaking the language and perception which otherwise would not have been possible. The findings of this experiment could possibly validate the quantitative results as well.

### **C. Focus group experiment and findings**

The participants were asked to transliterate thirty Tamil words in Roman script and its equivalent Tamil scripts. In some cases, certain words were deliberately mispronounced. The participants were expected to write as they heard. The objective was to understand how participants perceived Tamil pronunciation and their expectations on such. It is opined that the experiment would hopefully answer the question:

*'Would people expect a mispronounced word to appear right in the actual system?'*  
*'Does Tamil being a syllabic language offer this flexibility?'*

Although the reasons for mispronunciation was out of scope, it is felt that understanding the issue could help determine the success rate of the user acceptance and ability to use such technologies.

It was found that the transliteration of Tamil words in Roman script by the Brahmin participants itself was significantly different from the Non Brahmin participants. The Brahmin participants used the most appropriate letters used to represent ழ in the Roman script (*zha*) while the Non Brahmin participants used ('l') to represent the same syllable. The Brahmin speech pattern was 'clear' in the sense that every syllable was pronounced as it ought to be pronounced. The key observation was that the Tamil Brahmin dialect used an extensive sound set as it included the Sanskrit sounds as well. Therefore names like 'Prabha' were pronounced as it should be pronounced by the Brahmins with ease. The syllable (pra) exists in Sanskrit as it is

while it is represented as (pira) in Tamil. Therefore there is a very subtle difference when pronouncing words that begin with such a syllable.

### **Brahmin participant:**

*"I like Tamil because that is my mother tongue but I love Sanskrit"*

The comment by the participant suggested that the participant used 'Tamil' as a linguistic identity but associated more with Sanskrit. For instance, a participant chose to study Sanskrit as a second language because the religious scriptures are in Sanskrit. Tamil as a language was more than sufficient to speak as they felt its scope was limited.

### **Non-Brahmin participant:**

*"I studied Tamil as a second language "*

The motivation to study in a school where Tamil was the medium of instruction lacked among both the Brahmins and the non-Brahmin participants. But it appeared as though the Brahmins had a choice between Tamil and Sanskrit when it came to learning a second language and the knowledge in the latter facilitates to understand the religious scripture. Whilst Tamil seemed to be a natural choice for a second language to a Non Brahmin participating attending an English medium school.

Brahmin Tamil dialect could serve as an excellent example when dealing with the aspect and issue of code mixing in Tamil language. The need and urge to speak Tamil without using 'foreign' words (more specifically Sanskrit vocabulary) was observed amongst the non-Brahmin participants while it made no significant difference to the Brahmin participants.

## **D. Discussion**

Initially it was decided to exclude religion, taking into account the social and cultural norms of the geographic area in which the experiment was being conducted. But the motivation to include religion as a parameter was based on the following comment obtained in various qualitative interviews at different stages:

*"Now days, its primarily the Brahmins who can accurately pronounce ழ, ஶ (zha, La)"*

It was observed that the accurate pronunciation of basic syllables in Tamil language which is expected out of native Tamil speakers is now being viewed as an ability of one particular community. The issue of language maintenance amongst the Brahmins seems to widen when they settle abroad (Ridge, 2014) but the lack of motivation to use Tamil language was quite visible even in the home region of Tamil Nadu. In general, some researches suggest that the Tamils abroad largely do not exhibit high functionality in the language and increasingly use English in domestic domain whilst the Tamil Brahmins abroad have largely shifted to English and retain a functional competence in Tamil for religious and cultural reasons (Ridge, 2014). We argue that in the case of Tamil Brahmins, the retention of Tamil cannot be for religious reasons because the religious scriptures are in Sanskrit but it is possible that it could be for cultural reasons. Further Tamil is perhaps quite distinct from other Indian languages. It has a much

smaller number of characters compared to any other Indian language (Murthy& Kumar, 2007). (Coelho, 1997) suggest that in popular Indian usage the term '*mother tongue*' does not have a linguistic meaning. It is primarily used to mean a "community language", the language a person claims as a marker of membership in a cultural community. This was visible amongst the non-Brahmin participants. A participant who migrated to Tamil Nadu from a different region felt he was Tamil because he was born and raised in Tamil Nadu. Whilst for the Tamil Brahmins, "mother tongue" referred to a language that traced their lineage regardless of the migrations.

The Tamil Brahmin participants exhibited a certain degree of clarity in pronunciation when compared to the non-Brahmin participants. Therefore, the findings of Joyce McDonough and Keith Johnson hold true for the pronunciation of ழ, ள (zha, La). Traditionally the user acceptance of a technology or a software have been quite focused on the design and about the actual technology itself rather than focusing on the user and the ability of the users to use the system in a prevailing social condition. For language-related technologies such as the speech-to-text, and more specifically for syllabic languages like Tamil, it is opined that the ability to speak the language through correct pronunciation could potentially be a moderating variable in using and accepting such systems. It is in this context we propose that language maintenance by the linguistic community regardless of the religious affiliation is important in order to retain functional competency in the language so as to take advantage of the technology. Therefore, language maintenance of Tamil language and the ability to pronounce every syllable properly is complementary.

### Conclusion

The findings from the experiment strongly suggest there exists a relationship between religious affiliation and pronunciation more specifically with the Brahmins. In the *Cullavagga*, the Buddha clearly expresses His position in relation to the language:

*"In that same occasion there were two monks, brothers, Yamelu and Tekula by name, brahmans by birth, of pleasant voice, of clear pronunciation."* (Tola&Dragonetti, n.d.)

The clarity of pronunciation by the Brahmins has also been referred to in the Buddhism scriptures. It further suggests that religion could perhaps plan an important role. Our findings through this experiment is consistent with the previous experiments and other literature (McDonough &Johnson, 1997).

In the research on morphological analyzer and morphological generator for Malayalam- Tamil machine translation, the authors have suggested that Malayalam has achieved the dynamic synthesis which no Indian language has achieved. In fact, the Malayalam language, though evolved from Tamil, uses a good amount of Sanskrit vocabulary and so is the language's phonology. They further highlight that the language itself was brought to Kerala by the Brahmins. (Jayan, RR &Rajendran, 2011). It is opined that the authors have perhaps overlooked the fact that the Brahmin Tamil dialect has also achieved the dynamic synthesis that Malayalam as a language has achieved. The argument is that because of little importance given to Brahmanism as a religion and Brahmin Tamil dialect, the dialect doesn't seem to attract enough credit and its potential is yet to be tapped and explored.

The choice of considering religion as a moderating variable is perhaps justified in this case. There have been other linguistic researches, which have taken into account the religion factor. (Coelho, 1997)

The Brahmin Tamil dialect could be an excellent starting point to investigate and deal with the code-mixing effect as the dialect deals with two different languages: Tamil and Sanskrit. The results of the experiment further suggest that although the Tamil Brahmins pronounce Tamil and Sanskrit syllables with greater ease than the non-Brahmins, the question of whether a speech-to-text technology should be used by the community needs further research. At this stage, the empirical investigation along with other literature suggests that if pronunciation be accorded a higher priority in using a voice-to-text technology, it is predicted that it might be best suited for Tamil Brahmins.

### **Further scope**

This experiment has brought out features that are unique to the Brahmin Tamil dialect- which may be considered as a code-mixed version of Tamil. However, this could be considered as an empirical investigation with very limited subjects. Yet we consider it to be one of the fundamental and important observations that could help shape the requirements of language based software and technologies. But it could be extended to other areas to make a more general conclusion. For instance, carrying out similar experiments in different Tamil speaking regions could provide a better understanding.



## References

- i. Bronkhorst, J., 2015. *Review Essay: Greater Magdha and the New Brahmanism. Religious Studies Review*, 41(3).
- ii. Coelho, G. M., 1997. Anglo- Indian English: A Nativized Variety of Indian English. *Language in Society*, 26(4), pp. 561-589.
- iii. Droixhe, D., 1978. *La Lingistique et l'appel de l'histoire (1600-1800)*. Librairie Droz, p. 445.
- iv. Dutt, N., 1970. Brahmanism and Buddhism. *Bulletin of Tibetology*, 6(3), pp. 1-11.
- v. Jayan, J. P., RR, R. & Rajendran, S., 2011. Morphological Analyser, and Morphological Generator for Malayalam Tamil Machine Translation. *International Journal of Computer Application*, 13(8), pp. 1-18.
- vi. McDonough, J. & Johnson, K., 1997. Tamil Liquids: An Investigation Into The Basis of The Contrast Among Five Liquids in A Dialect of Tamil. *Journal of the International Phonetic Association*, 27(1-2), pp. 1-26.
- vii. McGovern, N., 2012. Bramha: An Early and Ultimately Doomed Attempt at a Brahmanical Synthesis. *Journal of Indian Philosophy*, 40(1), pp. 1-23.
- viii. Murthy, K. N. & Kumar, G. B., 2007. Language Identification from Small Text Samples. *Journal of Quantitative Linguistics*, 13(1), pp. 58-80.
- ix. Raj, R. S. V., Ali, A. & Khazaei, B., 2015. *Attitude of Tamil Nadu Tamils Towards Tamil Computing*. 14<sup>th</sup> International Tamil Internet Conference proceeding, Singapore.
- x. Raj, R. S. V., Ali, A. & Khazaei, B., 2015. *Tamil Transliteration and Suitability of Roman Script*. 14<sup>th</sup> International Tamil Internet Conference proceeding, Singapore.
- xi. Rama, G. L. J., Ramakrishnan, A. G., Muralishankar, R. & Prathibha, R., 2002. A Complete Text To Speech Synthesis System in Tamil. *IEEE*, pp. 191-194.
- xii. Ridge, B., 2014. National Language Planning and Language Shifts in Malaysian Minority Communities: Speaking in Many Tongues. *Journal of Multilingual and Multicultural Development*, 35(5), pp. 522-523.
- xiii. Tola, F. & Dragonetti, C., (n.d.). *Brahmanism And Buddhism: Two Antithetic Conception of Society in Ancient India*. Institute of Buddhist Studies Foundation, Argentina.