Sheffield Hallam University

A Sheffield Hallam University thesis

Predicting user acceptance of Tamil speech to text by native Tamil Brahmans

Raj Ramachandran

A thesis submitted in partial fulfilment of the requirement of Sheffield Hallam University for the degree of Doctor of Philosophy

In the Cultural Communication and Computing Research Institute (C3RI)

June 2018

ACKNOWLEDGEMENT

I extend my sincere and profound thanks to my Director of Studies Dr. Babak Khazaei and my supervisor Dr. Frances Slack. Without their guidance, motivation and constructive feedback, I would not have completed this thesis.

I am grateful to all the participants. Their responses and feedback were the guiding force of this research.

I thank my friends and colleagues at Unit 12, Science Park, Sheffield Hallam University for all their support.

I thank my *Amma* and *Appa* for their patience, faith, reassurance and support throughout my life, my wife for her unflinching faith in me and Ashik *Anna* for his advice and support throughout, especially in difficult times.

ABSTRACT

This thesis investigates and predicts the user acceptance of a speech to text application in Tamil and takes the view that user acceptance model would need to take into, the cultural constraints that apply in the context and underlines the need for a more explicit recognition. The user acceptance models such as Technology Acceptance Model (TAM) predominantly focus on the technological aspects to determine the acceptance. The cultural variables are considered as external but at the same time they acknowledge the influence of user acceptance due to external variables. The contribution to knowledge is, an empirical link between Tamil usage at a social level that indicates the ability to use and accept Tamil speech to text application. The economic value of Tamil, does not seem to warrant technology use and therefore, speech to text in Tamil was found to be less acceptable in the study samples.

In order to achieve the objective of predicting the user acceptance of speech to text in Tamil by the native Tamil speaking Brahmans, the researcher designed and evaluated a paper prototype of an iPhone iOS mobile representation of the paper prototype on the idea of 'what you speak is what you get'. As a result of the researcher's insider position, the idea was to convert the speech as spoken by the person into Tamil orthography without any technological interference such as auto correct, word prediction and spell check. Due to the syllabic nature of the language and the cultural tendency to code mix and code-switch, the investigation focused on three key areas- code mixing, pronunciation and choice of script . This thesis looks at the complexities involved in accommodating these areas. The user's choice of script was increasingly important as it cannot be assumed that all native Tamil speakers are able to read and write Tamil.

In order to bring in rich data, the researcher used the insider and outsider positionality alongside phenomenology. This was also to overcome any potential bias in analysis and interpretation. The multidisciplinary approach to answer the research question was inevitable owing to cultural variables like value and usage of language, social perception of language and its usage specifically code-switching, pronunciation and orthography in the native space.

Data gathering was done using quantitative study of transliteration and qualitative interviews of Tamil speaking Brahmans. The findings point to the Vedic philosophical texts and practices that influenced the attitude of the respondents on how words must be pronounced and how they ought to appear in text. The development of the speech to text application could be enriched by using a native approach that embeds cultural and philosophical values.

Based on the findings, this thesis has identified areas for further research which is to widely test the user acceptance model proposed in this thesis to aid development of speech to text and to further investigate on native perspective in the wider diaspora and also to investigate cultural and philosophical relevance in speech to text in other languages where technology is in developing stage.

TABLE OF CONTENT

TABLE OF FIGURES

Figure 2.1	The map of Tamil Nadu	18
Figure 2.2	Represents some of the Tamil syllables along with its transliteration	21
Figure 3.1	Technology Acceptance Model (Szajna, 1996)	38
Figure 3.2	Straud et.al(1997) cultural dimensions	38
Figure 3.3	Adapted TAM Paper& Simon (2007)	39
Figure 5.1(a-e)	Prototype representation	67
Figure 6.1	Proposed user acceptance model	75
Figure 7.1	Alai when written by the same set of participants in Roman orthography	101
Figure 8.1	A transliterated word showing a variety of conceptualised spellings.	125
Figure 8.2	A Mandarin – Taiwanese code-switching speech recognition	132
Figure 8.3	Example of Manipravalam (Sanskritised Tamil) in Granths script	134
Figure 8.4	List of words as dictated on google speech to text	135
Figure 8.5	Sentences dictated on google speech to text that includes code-swicthing	136
Figure 8.6	Some comments from the participants that relates to the user acceptance model.	137

LIST OF TABLES

Dedicated to Amma and Appa

CHAPTER 1 - INTRODUCTION

1.1 MOTIVATION AND BACKGROUND

It was in 2011, I switched to a smartphone in the United Kingdom. iPhone 4s was the first ever smartphone I bought. Upgrading to a smartphone was to some extent influenced by the social circle. It was at this point I noticed a little microphone symbol in the English keyboard. I slowly began to 'talk' to my phone in what I could call a reasonable British English. Tamil keyboard on iOS was unavailable in 2011. It was at this point, I started reflecting on Tamil's position in technology. At a time when I genuinely wanted to use the language in technology, the option of using the language was not available. Therefore, my messages in Tamil were in Roman orthography. When I upgraded to iPhone 5s, a couple of years later in 2013, Tamil keyboard, both in Tamil and Roman orthography (transliteration) were available. I then started to use Tamil in Tamil orthography. Nevertheless, my experience of using Siri in English motivated me to think of a similar application in Tamil. Siri was quite useful especially when I had to type e-mails, or make notes. When I reflected on the language ability, code switching, pronunciation and from own experience the dilemma in orthography were issues not just relevant but also inevitable. This eventually led me to investigate the user acceptance of speech to text application for native Tamil speakers.

1.2 INTRODUCTION

Speech recognition continues to be a fascinating area of research with unique and intriguing challenges. Speech recognition and speech to text in English made its humble beginning in the field of medicine about forty years ago. Shulman (2016) began the biography on 'Tamil' with focus on the syllbale '*Zha'*- a syllable that he referred to as a *proud characteristic of the language* as seen by its native speakers, which Srinivasan (2013) recognised as a challen. In the context of speech to text in Tamil, this research focussed on three phenomena - the accuracy of pronunciation of Tamil syllables especially *zha,* code switching and code mixing and choice of script or script complexity. The research took a multi-disciplinary approach to answer the overall research question on predicting the user acceptance of speech to text application in *Manipravalam Tamil* or simply *Manipravalam* - a hybrid language spoken

predominantly by the Tamil speaking Brahmans, also as referred by Ciotti (2017) and Shulman(2016)

Speech is the fundamental mode of communication for humans to interact amongst themselves. Speech to text involves a speech recognition component. Dalmiya et.al (2013) in their literature survey have discussed about the speech recognition system exclusively designed to recognise isolated numerals in the Assamese language. There are various techniques that can be used to implement a speech to text technology. Some of the techniques are:

- Neural network based approach (Pornpanomchai, 2012)
- Hidden Markov Model (HMM) approach (Radha, 2012)
- Microsoft's speech application program interface (SAPI) (Sultana et.al
- 2012)
- Dictionary based approach (Pornpanomchai, 2012)
- Deep learning based speech to text in Bangla (Tausif et.al, 2018)

Sankar & Nagarajan (2012) have done a comparative study on data compression on 'Tanglish' natural language text. The authors referred 'Tanglish' as a 'dialect' in which a sentence formed in Tamil liberally borrowed words form English. They further described it as a hybrid language which was a result of the code mixing Tamil and English. Auer (1999) has differentiated between codeswtiching and language mixing of some languages and indicated that the occurence of codeswitching was a result of limited knowledge of the 'other' language (**See appendix A.2**). Interestingly the code switching pattern was also observed in the Brahman Tamil which eventually became the focus of this research. Ciotti (2017) referred to Brahman Tamil as '*Manipravalam'*. However, the insider view suggested that Manipravalam is regarded closer to Malayalam language both in language style and in orthography. This research, for the sake of distinguishing Brahman Tamil from Standard Tamil and other Tamil dialects, has referred to Brahman Tamil as *'Manipravalam'* but has employed Tamil script instead of Grantha for the purpose of testing the output text. Goyal & Koolagudi (2013) have used the Mel-frequency cepstral coefficient (MFCC) technique in digit recognition for the Hindi language. They concluded that the performance of the system was dependent on a number of factors some of them being the effect of the speaker, age and gender. Bapat & Nagalkar (2008) indicated that the development of a speech to text application was much dependent on the target language, and that the number of phonemes the language contained.

For instance, English could be represented by a set of around 42 phonemes and Hindi with a set of about 46 phonemes.

Since this research focussed on the user acceptance of speech to text in Tamil in mobile devices, it was useful to consider See et.al (2010) finding on the user acceptance towards a personalised hands- free messaging application (iSay- SMS) by Malaysians in Malaysia. The application allowed the users to use their 'voice' to type without using a keypad. The authors considered the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and use of Technology. The authors suggested that most of the information system (IS) theories do not specifically look into the difference in technology acceptance between stationary and mobile environments. However, it was found that the determining factors identified by the Mobile Phone Technology Acceptance Model (MOPTAM) were similar to the one proposed in the Unified Theory of Acceptance and Use of Technology (UTAUT). In the end, study predicted that the user acceptance of the application would be fairly high amongst the consumers.

1.3 RESEARCH QUESTION

To what extent would the speech to text technology be accepted by native Tamil speaking Brahmans?

Objective

1. To investigate into speech to text technology.

2. To study the usage of Tamil in Roman script.

3. To design and evaluate a paper prototype for speech to text mobile application.

4. To propose an indigenous user acceptance model for predicting the user acceptance of Tamil speech to text.

1.4 ORGANISATION OF THE THESIS

The literature review is divided into three main sections. The first section discusses the literature on speech to text technology. The second section of the literature review deals with the models that have been used in the past to determine the user acceptance of technology and the third section deals with language and society.

The research, in order to answer the research question, has used an indigenous approach and has proposed a user acceptance model to predict the user acceptance of speech to text application which is based on the UTAUT model. The user acceptance model proposed in this thesis is broadly divided into language and technology for the purpose of predicting the user acceptance of speech to text. The language section, focuses on factors like the perceived usefulness of the language, actual use of language in the social sphere, government policy towards the language, opportunity to use the language and so on. Whilst in the technology section, it focuses on previous knowledge or experience with the technology, perceived usefulness of the technology, perceived ease of use of technology to name a few.

The user acceptance model was proposed specifically for the prediction of the user acceptance of language based technology with Tamil as a case study. It is foreseen that this model could be applied to any linguistic community for predicting the user acceptance or defining the requirements of a language based technology. Most of the user acceptance models were quite strongly focussed on the technology and surrounded around the technical ability of the user.

Lu et.al (2009) have used the theory of planned behaviour (TPB) and the Technology Acceptance Model (TAM) to examine the Chinese users' acceptance of instant messaging products (IM). Mathieson (1991) presented a comparison between the technology acceptance model and the theory of planned behaviour in predicting user intentions and concluded that there was very little on empirical grounds to suggest that one was better than the other. Whilst both took into account to certain extent the social influences and control issues, this research took the view that that it forms very little basis to evaluate technology acceptance for certain social groups. (Brown et.al 2002) in their work on 'Do I really have to? User acceptance of mandated technology' laid emphasis on the theory of planned behaviour but indicated that attitude, subjective norm and perceived behaviour control capture an individual's behaviour intention. (Sun & Zhang, 2006) cited Venkatesh's integrated model which was considered slightly more relevant in determining the user acceptance.

Extension of user acceptance model such as TAM can be seen in Gong et.al (2004). Although the integrated model has taken into account the individual factors and has indicated that it could potentially influence the user technology acceptance, it has given very little emphasis on factors like perceived usefulness of a language, behaviour intention to use the

language in technology. This research perceived that the technology acceptance models has always largely focussed on the technology itself. It was perceived that certain factors like subjective norms, perceived usefulness, behaviour intention, be applied to the target language as well in which acceptance of technology was being investigated.

Esenou & Egbue (2014) suggested that the rate of technology adoption varied and was dependent on various parameters such as government policy, societal receptiveness and a number of other social factors. These reasons have also been referred to by Iqbal et.al(2011) as reasons for failure of projects especially from the perspective of getting the correct user requirements. Esenou & Egbue (2014) further claimed that these social factors were quite often ignored in order for the product to win on the merits of technical superiority. The proposed model in this thesis takes into account the cultural influences and attitude towards the target language which could potentially influence the intention towards using speech to text in Tamil. In the case of Tamil society, it was important to consider the above factors in using the language in their daily lives as well as the intention to use the language in technology. The model proposed in this thesis to evaluate the user acceptance of speech to text technology was developed on the logic that language was the foundation on which a speech to technology could be used. UTAUT formed the basis of proposing the user acceptance model that was used to evaluate the user acceptance of speech to text technology in Tamil which could be extended to other languages as well.

This research has used phenomenology and brought in rich data which otherwise could have resulted in bias which is dealt in section 4.2 and chapter 8 respectively. And has further explained the methodological decision on abandoning quantitative approach. This research is unique in the sense that the researcher has used the native knowledge and Vedic philosophy (which will be discussed in section 2.3) which was also indigenous to the target audience and has presented an indigenous perspective of how technology was encountered and how it might influence the thought process on the application and user acceptance. It has further contributed to the diversity of the discipline through the model that has integrated various other disciplines such as philosophy, linguistics and sociology.

1.5 OUTLINE OF CHAPTERS.

Chapter 2 : The Tamil context: Provides an introduction to the Tamil context. It briefly introduces the language, the indigenous Vedic philosophy and the Brahmans.

Chapter 3: Literature review: This chapter presents a review of the literature on speech to text technology, various acceptance models, code-switching, language and politics, and on orthography.

Chapter 4: Methods: This chapter deals with the research methods. It introduces the research philosophy that this research has adopted. The chapter discussess employment of software engineering as a design method and explains how the prototype was conceptualised, developed and evaluated. It also briefly deals with the sampling and pilot which is crucial to provide a useful insight on decisions taken on methods.

Chapter 5: Evaluation of prototype: This chapter deals with the evaluation of the prototype and the rationale for some of the decisions in relation to prototype evaluation.

Chapter 6: Proposed Mode: The proposed model to predict the user acceptance of speech to text in Tamil is introduced in this chapter. The proposed user acceptance model is one the main contributions of this research.

Chapter 7: The studies: This chapter discusses the various studies and findings that were done in order to answer the main research question of predicting the user acceptance of speech to text in Tamil. It starts from the study of transliteration followed by a comparative study between two similar social groups which led to the observation of 'zha' pronunciation which in turn led to the target audience of this research- the Tamil Brahmans.

Chapter 8: Discussion**:** The chapter discusses on the findings and studies along with relevant literature. The chapter also provides a reflection on the user acceptance model proposed in this thesis. The section on researcher's own experience is in the first person.

Chapter 9: Personal reflection: This chapter provides a brief insight on researcher's own experiences, practices and observations as an insider. The chapter enables the reader to see through the researcher's lens. This chapter has used first person.

Chapter 10: Conclusion: In addition to predicting the user acceptance, it provides an insight on the feasibility of such applications along with some suggestions that could be incorporated when designing and developing such applications. Finally, it enumerates the contribution to the knowledge and further work.

1.6 PHRASES AND CONCEPTS USED IN THIS THESIS

The following section introduces key phrases used in this thesis along with its interpretation in the context of this research.

Indigenous knowledge

According to Khupe et.al (2017), there could be two views on describing an indigenous knowledge. The first is a universalist and the second one is a pluralist. According to a universalist, all knowledge could be indigenous but a pluralist would view knowledge as a sociocultural and historical construction and is dependent on the ways in which society process, code and assign meaning to their experience. This research has taken a pluralist view since sociocultural and historical experiences contributed to the prediction of the user acceptance. This thesis aligns with Parsons et.al (2017) proposal on accepting indigenous knowledge base especially oral tradition, history and culture.

Understanding and contextualising literacy

According to the Education for all global monitoring report (UNESCO, 2006) the national census of India defines a 'literate' person as one having the ability to read and write in any language. The Department for International Development (UK) defines literacy as a basic set of skills (reading, writing and counting) or competencies. Sri Lanka's definition of literacy is the ability to read and write simple sentence in a specified language (Tamil, English and Sinhalese). UNICEF's definition of literacy is the ability to use reading, writing and numeracy skills for effective functioning and development of individual and the community (p.158). Nevertheless, the common understanding of literacy points to the skill of reading, writing and speaking.

This research has adopted the UNICEF's definition of literacy and took into account the ability to read, write and speak Tamil.   According to the 2011 Government census, Tamil Nadu has a literacy rate of 80.33% and it's capital Chennai has a literacy rate of 90.33%. With reference to the definition of literacy in the broader Indian context, it is assumed that these literacy figures may not be reflective of the literacy in Tamil.

Manipravalam

Ciotti (2017) and Shulman (2016) have referred to Sanskritised Tamil as *Manipravalam*. However, the degree of Sanskrit vocabulary was highly variable and was dependent on a

number of individual factors. This research has used Manipravalam synonymous to Sanskritised Tamil.

Linguistic view

Dictionary defines linguistics as the study of the way in which language works. This research has adopted linguistic view of the Tamil language.

Orthography

The Oxford dictionary has defined orthography as "*the conventional spelling system of the language.*" This research sets its focus on Tamil orthography and predicted the user acceptance of speech to text on the basis of Tamil orthography.

Philosophy

The Oxford dictionary has defined philosophy as "*A theory or attitude that acts as a guiding principle for behaviour*". This research has perceived Vedic philosophy as a motivating factor for Brahman's ability of accurate pronunciation.

Native

The research has adopted the Oxford dictionary definition *2Originating or occurring naturally in a particular place; native.*". The user acceptance model proposed and used in this research and thesis was from the perspective of native Tamil speaking Brahmans in their native space.

Note on transliteration

All Tamil words have been transliterated using appropriate and equivalent Roman letters, but from the perspective of a user with no knowledge of formal transliteration.

Standard Tamil

The term standard Tamil is used to denote a spoken Tamil that is used in formal settings that is accepted by all native Tamil speakers regardless of dialect variations and philosophical affiliations.

Pure Tamil

The term pure Tamil is used to denote Tamil without the effect of code mixing and code-switching. Example: *Sangam Tamil*.

Proper Tamil

The term proper Tamil is used to indicate spoken Tamil without the effect of code-switching and with proper pronunciation of Tamil syllables.

Code mixing

Code mixing is phenomena where two languages are used in the same word. Examples of code mixing could be observed in Appendix.

Code switching

According to Oxford English Grammar dictionary (2014), code switching occurs when speaker changes from one language to another (Tamil to English or Tamil to Sanskrit or Tamil to Sanskrit and English from the perspective of this research) according to where they are, who they are talking to which speech community they identify with and so on.

Tamil Brahmans & Non Brahmans

Is used in the context of potential end users of the speech to text technology in Tamil. Section 2.3 extensively deals with Brahmans. The term 'non-Brahman' in the thesis is used to refer users who are not Brahmans. These terms must be viewed strictly from the perspective of end users of technology and acceptance of a language based technology with regards to the proposed model.

CHAPTER 2 - THE TAMIL CONTEXT

2.1 TAMIL –GEOGRAPHY, IDENTITY AND POLITICS
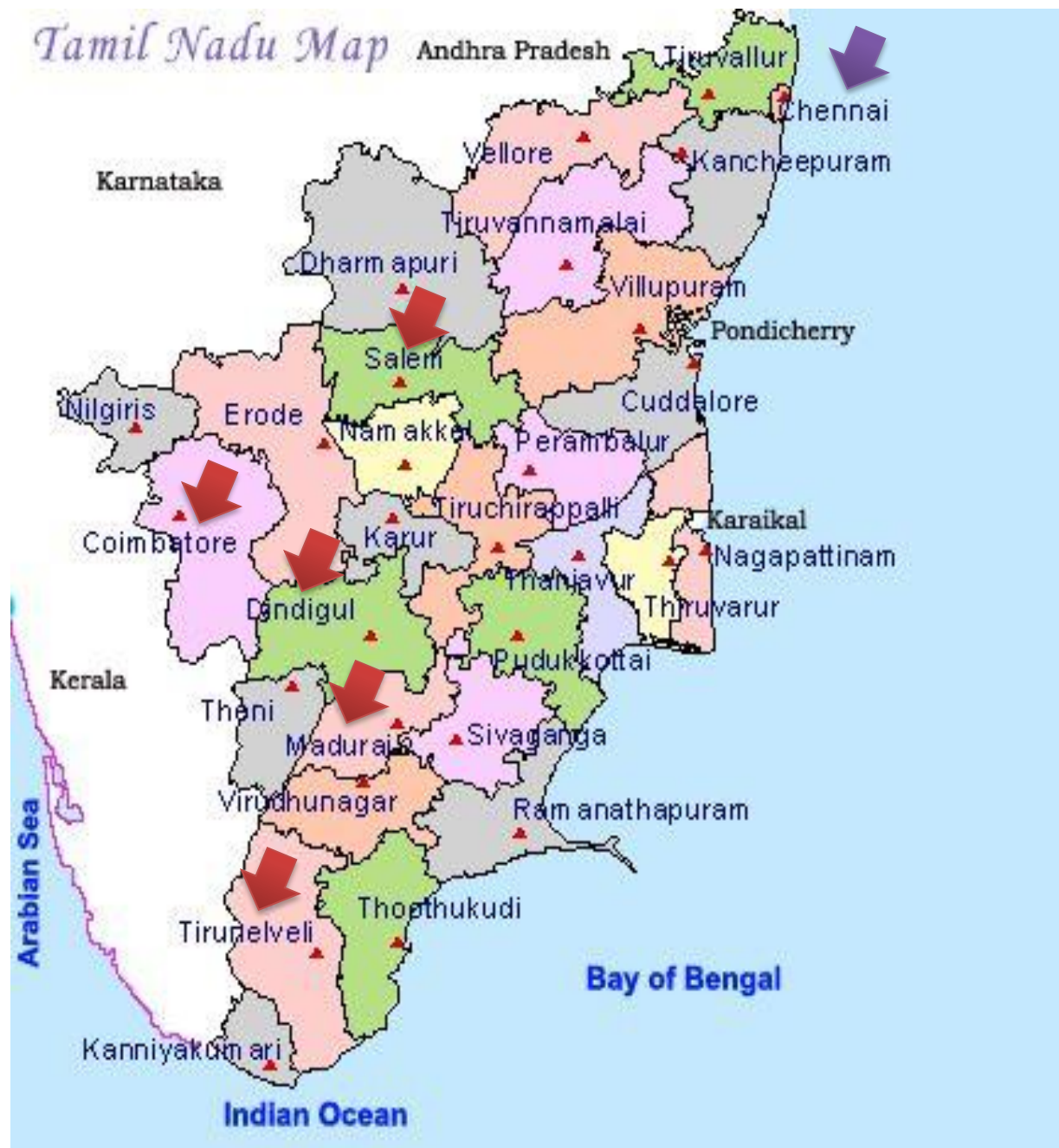
Figure 2.1: *The map of Tamil Nadu state (Kanniyakumari district website) within the Indian Union. The state of Tamil Nadu share borders with Kerala, Karnataka and Andhra Pradesh where Malayalam, Kannada and Telugu are the respective official languages.*

-    Quantitative   - Qualitative interviews

Tamil, is a language that is official in the Indian state of Tamil Nadu (Figure 2.1), Union Territory of Puducherry(Figure 2.1), Sri Lanka and Singapore .It is one of the recognised minority language in Malaysia, Mauritius and Reunion .However, the term could also refer to ethnicity. That is, a person born and raised in Tamil Nadu but is not a native Tamil speaker. The notions and identity of being Tamil could vary amongst the Tamil diaspora .For example, a native Tamil speaker who is a Canadian or British could probably identify himself as a Tamil Canadian or British Tamil as opposed to British Indian or a Canadian of Indian origin . At a diasporic level, the Tamil ethnicity cannot be restricted or assumed to be Indian, although the majority of the Tamils live in the Indian state of Tamil Nadu .Therefore, from the perspective of diasporic identity, Tamil is commonly used to represent one's linguistic identity, heritage and ethnicity with which people following various faiths from countries such as -India, Singapore, Malaysia, Sri Lanka could relate to. This is more prevalent amongst the Sri Lankan Tamils)Das, 2008 .(The findings and discussion chapter deals with how identity could also be used as a motivation to use the speech to text application .It also deals with the complex relationship with the Tamil identity that technology has to deal with, in order to determine the 'precise target audience 'who may be more receptive to use speech to text on the basis of identity .The findings and the discussion chapters bring to surface how identification with a language plays a crucial role in defining identity which in some cases, seem to be the only motivating reason to use technology .The section on Brahman would deal with identity of Tamil Brahmans and associated politics around Brahmans that would enable to critically appreciate the subtle differences within the Tamil society and how these subtle differences form the basis of identity which was essential to be identified at the requirement stage of the software engineering lifecycle .

2.2 TAMIL -LANGUAGE

Tamil, along with Latin, Greek and Sanskrit is one of the classical languages of the world that has a history of over 2,500 years .Contrary to certain claims, Tamil is perceived to have its origins independent of Sanskrit and other Sanskrit based Indian languages .This is apparent by the absence of a large set of syllables that can be found in Sanskrit and other Indian languages .*Tholkaappiyam* is the earliest and the most authoritative grammar .From the perspective of this research, it was important to distinguish between Tamil as a language and Tamil as an ethnicity, which diasporic Tamils identify themselves with, and may or may not speak the language .This research has considered Tamil as a language from the perspective of native Tamil speakers .The importance of differentiating language from ethnicity was also

partly to do away with attributing mispronunciation with dialects and region .The rationale was on the basis of *Tholkaapiyam* which defines the rules of the language and is common to everyone who speaks Tamil .However, the understanding of the complex question 'What is to be Tamil 'from an identity perspective was important to relate to the findings, that to an extent, answered the question of acceptance or rejection of a technology by a section of this social group and will be seen in chapter 5 Further, the argument that pronunciation of the syllables must be consistent across all Tamil speaking regions could be attributed to the feature of the language more than a product of a dialect and is dealt in chapter 8 It is in this regard, the speech to text application under study merits a cultural and indigenous philosophical approach . The discussion chapter deals with why consistency of pronunciation was important from a requirement perspective for this particular application. In this context, it is to be noted that the Tamil orthography mirrors the 'native 'phonemes.

BASICS OF TAMIL LANGUAGE

There are 18 consonants and 12 vowels in Tamil .The vowels are further classified into *Kuril* and *Netil* depending on whether they are short or long in pronunciation .The consonants are further classified into soft, medium and hard .The combination of the 18 consonants and 12 vowels form 216 compound characters .In total the Tamil language has 247)18+12+216+1 (letters .In addition to these, there are six characters borrowed from the Grantha script to represent sounds not native to Tamil and borrowed from Sanskrit .However, the usage of the additional characters that facilitate the representation of borrowed sounds from Sanskrit is not a socially acceptable practice in the context of Tamil .

	அ a	ஆ ā	இ i	ஈ ī	உ u	ஊ ū	எ e	ஏ ē	ஐ ai	ஒ o	ஓ ō	ஔ au
க் k	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ
ங் ṅ	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ
ச் c	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ
ஞ் ñ	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞௌ
ட் ṭ	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ
ண் ṇ	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணௌ
த் t	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தௌ
ந் n	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நௌ

Figure 2.2 : Represents some of the Tamil syllables along with its transliteration.

Figure 2.2 represents some of the 247 Tamil syllables. From the above figure, the combination of *k +a = ka; k + ai = kai (Hand)*. *k* is a syllable and *a* is a syllable with some unit of time measurement. Every unit of syllable needs to be pronounced to get a complete word. For example: *Give* in colloquial Tamil would be *Kudu or Kodu.* Therefore for someone to get this word right, the person should be able to pronounce *k+u = ku or k+o and t.+u = tu (du)* to collectively form *kudu*. The following is another example of Devanagri script (Hindi and Marathi), the approach of which, is similar to Tamil:

/र्+अ+ज्+अ+त्+अ (r+a+j+a+t+a)/

Dhore et.al (2013)

All syllables are equally stressed in Tamil and there are no accents or lexical stress . *Sandhi* may be defined as a set of rules for modification and fusion of sounds at or across the boundaries of grammatical units .*Sandhi* is quite common in Indic languages and in Tamil, one could observe a remarkable difference in pronunciation when the rules of *Sandhi* is observed .The boundaries of *Sandhi* merge in the spoken variety though .

An example of *Sandhi* :அரசு)Government + (பணம்) money =
அரசுப்பணம் (Thangarajan et.al, 2009)

Although it must be noted that in the above example the meaning of the word does not change before or after *Sandhi* the difference in pronunciation would result in a different spelling . Murthy & Kumar (2006) point to one to one correspondence between the sound .and orthographyThe above reason is yet another compelling reason for pronunciation to be studied in the context of Indic speech to text technology more specifically in the case of Tamil .

2.3 INTRODUCTION TO THE VEDAS AND THE VEDIC '*DHARMA*'

The word *Veda* is a Sanskrit word which means 'to know 'or knowledge .The closest possible translation of the word '*dharma* 'could be - eternal law of the cosmos that is inherent and the very nature of the things and creation to which Verma)2017 (agrees .This thesis, has made a conscious attempt to use the Sanskrit word '*dharma* 'instead of the more commonly used term 'religion .'Verma (2017) uses the word '*Bharata* ' to refer India and explained that amalgamation of spirituality and science as an inseparable aspect of the Vedic *Bharata.* The Vedas are *apaurusheya (*not invented by human). It is a vast body of knowledge that encompasses everything required for the human and includes but is not restricted to Science and Technology, Mathematics, Physics, Astronomy, Fine Arts in addition to discussing the means leading to the realization of the One and the absolute Supreme' Brahman) 'Not to be confused with the *Brahman varna* as explained in the following section .(!Therefore, under the context of investigation, this system merits to be reclassified into its original form as '*dharma',* which is distinctly different from that of a religion .The understanding of the difference between *dharma* and religion is important to understand the position of the Brahmans in the contemporary world and in their *dharmic* context   .The basis of this attempt was to provide an indigenous insight to the prevailing system which could provide a totally different, and in some cases, contrasting view of the same subject under investigation .The research, by focusing on indigenous knowledge, experience, philosophy and lifestyle in the process of predicting the user acceptance has also identified and elicited requirements for Tamil speech to text application and has proposed a user acceptance model relevant to the geography and target audience .It has further brought to surface, the prospects of indigenous knowledge in the context of software engineering and explored its employment as a means to more accurately predict the user acceptance of the Tamil speech to text technology   .

The *Vedas* are also referred to as *Sruti* – that which is transmitted through oral tradition and are used interchangeably .The *Vedangas* are classified as *Siksha)*Phonetics(, *Chandas)*Metrics(, *Vyakarana)*Grammar(, *Nirukta)*Etymology(, *Jyotisha)*Astronomy(, *Kalpa)*Ritual .(In order to understand the *Vedas* , it is important to have a profound understanding of the *Vedangas* .

2.3.1 THE VARNASHRAMA DHARMA

Contrary to the existence of the usage of the term 'caste 'F,uller & Narasimhan (2014) and Kroch (1986), this thesis, with a view to provide the intended meaning of the Scripture shall use the term '*Varnashrama dharma* 'for this Sanskrit phrase is best understood when used in its original form in the cultural context .It is also, a non translatable phrase which relates to cultural understanding and experience. The verses from the *Purusha Sukta* of *Rig Veda* forms the basis of the *Varnashrama dharma*

Brahmana, Kshyatria, Vaishya and the Sudra are the four *Varnas* .Every *Varna* is required to abide by their *dharma*, the rules of which are laid by the *Sruti* and is reiterated in the Manusmriti but their discussion is beyond the scope of this thesis .However, contrary to the popular perception and notion of hierarchy, class and superiority, philosophically, there is none .The *Sruti* orders these *varnashrama* to perform their assigned *dharma* in a collective sense for the wellbeing of the entire world .Study of the Vedas, austerity, self-restraint, teaching is the *dharma* prescribed for the Brahmans .Protection of the Brahmans, Vaishya and Sudra is the *dharma* of the Kshatriyas .Trade and business is the *dharma* of a Vaishya while service to , the Brahmans, Kshatriyas and Vaishyas is the *dharma* of a Sudra .It is the inter-dependence of these *varnashramas* that lead to successful, prosperous and a happy human society .The concept of *varnashrama dharma,* undoubtedly is one of the most mis-interpreted Vedic concept. An insider's understanding of *Varnashrama dharma* could best be related to the Myers-Briggs type indicator that is useful to form a software engineering group (Salter & Evans, 1997) . Just as the Myers-Briggs Type Indicator helps in identifying the dominant function of an individual, the philosophy behind the *Varnashrama dharma* is to indicate the personality type of a social group at large where the Brahmans are identified with the 'intellect' and 'knowledge'. Of the four *Varnas*, this thesis sets focus on the Brahmans who speak Tamil as their mother tongue. It is also relevant to note that the *Vedas* are in Vedic Sanskrit which is in Devanagri script and, in some cases, the Grantha script, which is significantly different from modern Tamil script .

2.3.2 TAMIL BRAHMANS -MAKING UP OF THE MIDDLE -CLASS CASTE!

The ethnography by Fuller & Narasimhan (2014) sheds some light on the journey, transition and transformations of Tamil Brahmans from the British India into the modern Indian society .The study also dealt with the relationship with the Non Brahmans and their perception about Brahmans .It focused on the religion and the negotiation of the identity in the contemporary world .It identified that politics, opportunity, aspiration and economy are factors that motivate this community to make compelling choices .The Brahman identity is as important as the Tamil identity but the politics of non Brahmans contend that the Tamil and Brahman identity can co-exist as also seen in Shulman (2016) and Solomon (2012) .The social transformation of this group, between 1950s and 1970s amidst strong Anti Brahman sentiments in Tamil Nadu explained the compromise this community had to make in order to integrate with the mainstream society .Physical appearance apart, the Brahmans are known for their distinct speech, pronunciation of some unique Tamil syllables such as '*zha*' and liberal Sanskrit vocabulary all of which are perceived as an imposition of the Brahman tradition and culture on the 'others as also seen in Shulman'(2016)

Vaitheespara & Venkatasubramanian(2015) provided a brief insight on the politics of language and culture in Tamil Nadu and aspects of Tamil 'purist' movement along with the relationship of the Brahmans with the 'other.' Politics, opportunities and the reservation system, where in order to provide social upward mobility, some percentage were allocated for the non Brahmans. These were perhaps the main reasons for this group to emigrate from Tamil Nadu to rest of India and abroad particularly to the United States .Life of a Brahman is strictly governed by the *Vedas* .A typical Brahman's life centers around spirituality and in pursuit of knowledge .Of course, in modern times, the life of a Brahman as suggested in the philosophy too, has undergone some transformation .The relevance of this ethnography could be related to the participants comments in the discussion chapter and the researcher's own experience as an insider to overcome bias and will be discussed in **section 2.4.** Although, this ethnography was conducted from a sociology perspective, it provided a finer insight of how social transformation over a period of time could change priorities which in turn could have a bearing on requirement and user acceptance .Ethnographic research methods have been employed in software engineering particularly to study the influence of human factors. Karn & Cowling (2006) brought out the benefits of ethnographic study within the software engineering field. They used the results of their ethnography to explain the conceptual

orientation of software engineers towards working through the project. In the context of this thesis, the findings of the ethnography of this social group coupled with the researcher's own findings, in addition to answering the larger question on predicting the user acceptance has also informed the requirements of the speech to text application.

2.3.3 RELEVANCE OF VEDAS AND BRAHMANS

The previous sections introduced sequentially, the Tamil language, the Vedas, *Varnashrama dharma* , Brahmans and Tamil Brahmans in the contemporary world .This section aims to explain the relationship between the Vedas, Brahmans and pronunciation .This section to a limited extent makes use of the 'insider's 'perspective of the researcher as seen in section **2.4 and figure 4.1**

The occupation of the Brahmans in the ancient Vedic times was learning and teaching .But many Brahmans have taken up professions according to the contemporary demands and are in constant negotiation with their 'Brahmanness) 'Fuller & Narasimhan, 2014 .(The relevance of Brahmans in the context of this thesis needs to be interpreted on the basis of their *Vedic* association, rituals and engagement in learning to recite a portion of Vedas, contrary to the concept of class and caste .This thesis takes the view that indigenous knowledge needs to be viewed from a native perspective. Swami Gambhirananda's (1972 (view on reliance of translated versions of the original Sanskrit texts merits contextual consideration which is also the basis of this thesis using the original Sanskrit terms unlike Fuller & Narasimhan (2014) who have, in many instances resorted to the translated equivalents of the certain native terms)such as caste (which this thesis perceives to provide a different view of the same social group under study in the context of technology.
*"Readers who are either unfamiliar with Sanskrit or, though fairly acquainted with it, prefer the medium of another language for better understanding .It is very difficult to present such an abstruse philosophy through translation, when the language into which it is rendered differs so widely in form and spirit from the original ."*Swami Gambhirananda (1972)

The first step to learning Vedas are the Siksha –phonetics .The accuracy of pronunciation and the Brahman's ability to accurately pronounce could be to a great extent, an influence as a result of the study of the Vedas with the following Scriptural explanation from the *Thaithriya Upanishad*

"One should spare no pains in learning the text; and according to the Upanishad proceeds with the lesson on phonetics .Here one may ask, what if one be careless? .Carelessness will lead to evil .It has been said, "The Mantras when wanting in rhythm or sound, or when wrongly used, conveys not the intended idea. That thunderbolt of speech will ruin the worshipper as the word 'indra-satru 'did owing to a fault in rhythm"

A Brahman perceives the 'method of reciting 'Vedic hymn and the Vedas as a Divine order where one should not err and is fearful of the consequences that might arise owing to mispronunciation of the text .The Scripture denies expiatory act in this connection in the following words:

"*If the yogin should unguardedly commit a sin, he should resort to yoga alone, never to any other thing such as mantra*".

It reiterates the importance of attention, effort and care in the study of phonetics so there is no defect in the knowledge acquired and that the Scripture is understood right .Further, the modulation   consists in reciting the text in pronouncing every sound according to its proper time, neither too fast nor too slow and according to the directions given in several sciences .In the Upanishad, 'siksha 'and other words, the recitation   happens in neutral accentless tone . These are some of the Scriptural evidences in support of the inclusion of accurate pronunciation as a requirement in the speech to text technology and also a rationale for choosing Brahmans as the subject for investigation .

2.4 POSITIONALITY

Court & Abbas (2013) presented an interesting approach on methodological and ethical challenges of insider-outsider research, multiple languages and dual-researcher cooperation. All of them except the dual-researcher cooperation was particularly relevant to this research since the researcher has taken both insider and outsider positionality as seen in Court & Abbas (2013). Colleen et.al (2016) supported the view that a researcher who has experience with the culture under investigation was an 'insider'. Chinn (2006) discussed the relevance of indigenous knowledge and the conclusion provided a useful insight on globalization and marginalization of indigenous, traditional and local knowledge and practice. Parsons et.al (2017) presented a different dimension of indigenous knowledge. This research took a contextual based approach that took into account, the narratives of the participants, cultural practices, history and knowledge of the environment that was specific and relevant to the

subject group under investigation as also supported by Parsons et.al (2017) and was consistent with the four key principles laid by the authors.

"All researchers can benefit from exploring the ways in which they are connected to their research" Keane & Khupe (2017)

The above quote of *Keane & Khupe (2017),* aligned well with this research because of the relationship, the researcher shared with the culture and geography. Brian (2017) threw some light on indigenous approaches to research and knowledge and discussed how significantly different was the approach of indigenous knowledge and method. One could infer from Brian (2017), the insistence on the West to acknowledge and recognize the indigenous methodologies. Contrary to the ontological separability, the concept of object/subject, knower/ known is questioned from a native framework (Brian, 2017). This research takes the position that the researcher having been a part of the phenomena, the boundaries of object/ subject, knower/ known merges, also as indicated by Djuraskovic & Arthur (2010) on heiristic inquiry. These boundary of object/ subject, knower/ known are inseparable from the perspective of the natives. For example, Frances (2001) concept of observation as a method, in the native framework would be inherent rather than using them as a separate methodological tool for data collection. They however, do raise the issue of the role of the researcher. Their reference to contextual knowledge and situation in their work, was already 'natively' available to the researcher however, the role and positionality of the researcher was vital from the perspective of how the 'native' knowledge was viewed- insider or outsider or both. Karan & Cowling (2006), warn about the *Hawthorne effect* in the context of ethnography and the researcher's position as an insider by virtue of the phenomena could potentially be a solution to the *Hawthorne effect*, in the larger context. Court & Abbas (2013) pointed how the role of an insider could appreciate the nuances and that an insider role evolves out of passion, getting deeply involved and perhaps a bit more personal. The view of an outsider as a researcher in the same context, in their case would have provided a totally different approach and would have potentially missed the finer points that an insider was able to receive. Although, in their case, they were two researchers, but a sole researcher switching the role of an insider and outsider would be indeed a daunting task. Colleen et. al (2016) recommended the researcher's experience as a source of data for a holistic view of the phenomenon that was under investigation. From this perspective, this thesis therefore takes the view that in order to answer the research question by employing the method of 'phenomenology', the status of insider was inevitable. Van Manen (2017) suggested that

lived experience are data for phenomenological research and carefully distinguished between the two forms of data- the phenomenological data and the coded data obtained for analysis. But, it does take into cognisance, the potential bias it could bring along with it which could be overcome with reference to the findings and arguments made in ethnography study and is dealt in the following section **;** Fuller & Narasimhan, (2014), Shulman (2016) relevant to the subject group and context, in addition to the findings of this thesis.

2.4.1 INSIDER- OUTSIDER POSITION

This section briefly deals with the contextual relationship of the researcher. Negotiation of power, position and relationship and a critical awareness is an important aspect of positionality (Parsons et al., 2017), (Dwyer & Buckle, 2009). The researcher's upbringing in the context and subject group under study in its own merit classifies the researcher to be an 'insider'. But at the same time, the researcher being a part of the diasporic community, reclassifies to be an 'outsider'. Therefore, in the fieldwork, the researcher by virtue of an 'insider' was challenged by the participants who viewed an 'insider' to be an 'outsider' merely on the account of the absence in the native space for a specific length of time, the details and context of which are discussed in section 9.2. Nevertheless, this view was taken in the best spirit, consciously upholding the identity of an 'insider' but with a view of an 'outsider'. The understanding of Tamil culture, interpretation of Tamilness and positioning of an insider as a Tamil Brahman, in the native context was controversial (Fuller & Narasimhan, 2014). Yet, this was a conscious choice of being identified as a Tamil Brahman and an 'insider' from the perspective of knowledge, practice, rituals and way of life of the Tamil speaking Brahmans as also suggested by Dwyer & Buckle (2009). And, an 'outsider' from a diasporic perspective provided a whole new dimension to the same context as previously 'experienced' by the researcher as discussed in chapter 8. The insider position helped in not just evaluating the responses but also to revisit the 'lived' experience that contributed to the analysis to answer the larger research question. The familiarity with the cultural nuances helped in taking intuitive decisions in search of a more 'natural' and 'genuine' data rather than a data for the sake of being 'correct' which could have resulted in a biased finding which is discussed in **section 4.16**. The result of being in the phenomena was a motivation to adopt a native framework to predict the user acceptance of the speech to text application. In the process, it also helped in identifying the requirements and feasibility of the application using the indigenous experience and knowledge base.

Since the conceptualisation of the research question was motivated by personal experience, according Colleen et.al (2016), the researcher was already immersed in the theme. The choice of dual positioning of the researcher as an insider and an outsider qualified more than just being either an insider or an outsider.

2.4.2 BENEFITS AND CHALLENGES

The adopted positionality was beneficial, and at the same time, challenging. It could be argued that the choice of positioning was crucial to the quality and depth of data and knowledge that this thesis bears. The 'insider' status from the perspective native Tamil speaking Brahman contributed to the phenomenally rich data in the form of experience , observation and a native perspective to the subject group and application under study. The role as an outsider from the perspective of a diaspora, enabled the researcher to proactively, analyse and relate to the participant's comments which increased the credibility, trustworthiness of the research.

By assuming the role of an insider, preconceived notions, prejudices and biases need to be acknowledged. However, these were addressed by using relevant literature (eg. Fuller & Narasimhan (2014), Shulman (2016). It also allowed to take a more neutral and a 'conscious' approach towards the data as the 'outsider' prism enabled to view the other side of the 'lived' experience. The dual role, to a great extent, was beneficial to understand and appreciate the prevailing social environment within which the negotiation of language and technology took place. It enabled to identify and challenge certain beliefs and experience primarily related to the use of language in the native sphere. The most common being the notion of what constitutes a 'Pure Tamil' , the 'relevance of Tamil today' in the native sphere as inferred by participants comment in section 6.1 and the association of technology with the English language. Multidisciplinary approaches especially from technological to philosophical and historical perspectives further strengthened in identifying and reducing prejudices and biases. This thesis takes the view that through knowledge, the biases would constantly be reduced.

2.5 SUMMARY

This chapter has introduced the thesis, along with the core focus group of this thesis -the Brahmans who speak Tamil as their mother tongue .The chapter has given an outline of Tamil language from the perspective of speech to text technology .It has provided some introduction about the social group under investigation and has explained the rationale for

selecting this social group – accuracy of pronunciation and code switching. It has discussed the positionality adopted by the researcher along with rationale. The accuracy of pronunciation as a requirement, was taken from the philosophical point of view from the Vedic Scripture (not in its entirety). That is, it has taken into account, the discipline and rigour that the philosophy sets for learning. But it has taken   into account the attitude, discipline and tone that the philosophy sets for the social group that leads to the accuracy of pronunciation. This chapter takes a hypothetical view that if pronunciation is a social habit and philosophy is a motivation for accuracy of pronunciation, then why, cannot it be a requirement to use a speech to text application in the context of technology?

In conclusion, this chapter contributed towards the following:

- Accuracy of pronunciation, as a requirement in speech to text application which complements the concept employed in the research 'what you speak is what you get'!
- Using knowledge in native form to elicit requirements of speech to text application .and to predict the user acceptance of speech to text application

CHAPTER 3 - LITERATURE REVIEW

 The first section of this chapter deals with the review of technology , the second section deals with the acceptance models and the third section deals with how Tamil language is used in the society and focuses on the code switching aspect.

3.1 INTRODUCTION

Transliteration in Tamil was first introduced in early 1980s to facilitate typing Tamil using Roman script. In 1999, the Government of Tamil Nadu accepted the Tamil phonetic keyboard. Two of the most widely used keyboard styles for inputting Tamil text in the computer are the Tamil typewriter keyboard and the English transliteration keyboard (known as முரசு அஞ்சல் *Murasu Anjal).* Kumar Cheran, (2004) predicted that the phonetic keyboard might become the standard keyboard for future generation. Speech -to- text technology in Tamil could potentially eliminate the usage of keyboards to a great extent which could empower a not so tech savvy native Tamil speaker to input in Tamil.

Anuraj Sivaraja, (2013) discusses about existing software available in Tamil for people to use and their attitude towards these software.   Research in the field of voice recognition began as early as 1974 (Dreyfus-Graf, 1976). French, German and English were among the languages chosen for speech recognition research. Since 1974 there have been two subjective and objective questions within voice recognition one of which is "Which may be the best language" Dreyfus-Graf, (1976). Although, it was unclear on the criteria for the best language. The work of Malarvizhi (2013) suggests that speech recognition technology, was used in radiology as early as 1981. s It further suggested that speech recognition although successful need to address issues such as ease of use, user acceptance, cost and exact time saving (Herman, 1995).

There is ongoing research in speech –to- text technology in other Asian languages like Bangla (Sultana et.al, 2012), Thai (Pompanomchai et.al, 2012), Devnagari script, which is common for languages like Hindi, Sanskrit and Marathi (Bapat & Nagalkar 2008). On the basis of Microsoft's Speech Application Program Interface (SAPI) only eight languages including English has speech to text technology and the authors Sultana et.al (2012) in their work had predicted that speech to text conversion using SAPI was an area that was yet to be explored in other languages especially the Asian languages which are very different to the European languages. Nuance - a speech and imaging company have launched a voice based

search engines exclusively for songs in 11 Indian languages one of which is Tamil.

(Medianama, 2009)

Author	Language	Type of work
Sultana et.al (2012)	Bangla	Bangla orthography (speech to text)
Sridhar et.al (2013)	Tamil	Ambiguous (Tamil or Roman) speech to text
Pornpanomchai et.al (2012)	Thai	Thai orthography (speech to text)
Sandanalakshmi et.al (2013)	English	Roman orthography (speech to text)
Lamel et.al (2009)	Arabic	Transliterated (Roman) (speech to text)
Kumar & Wei (2003)	English and Tamil	Speech recognition
Liao et.al (2014)	Chinese	Pronunciation training
Rallabandi & Black (2017)	Hindi, English, Tamil, Telugu, Marathi	Mixed lingual speech synthesis
Raghavendra et.al (2008)	Tamil, Telugu, Hindi	Global syllable speech synthesis for Indian language
Srinivasamurthy &Narayanan (2003)	Persian (Farsi)	Speech recognition
Cieri et.al (2004)	English Chinese and Arabic	Non US English and English of foreign accents. (speech to text)
Walker et.al (2003)	English, Mandarin Chinese, Egyptian Arabic	Roman orthography for English. Others not specified. (speech to text)
Hanani et.al (2013)	Regional and ethnic accents of British English	Speech recognition
Lyu et.al (2006)	Chinese dialects	Speech recognition on code-switching
Chan et.al (2009)	Cantonese- English	Speech recognition

Vu et.al (2012)	Mandarin-Englsih code-switch	Speech recognition

Table 3.1: Speech recognition and speech to text in various languages

Table 3.1 illustrates the work of different authors in the area of speech recognition and speech to text in various languages. Dey et.al (2009) acknowledge the lack of a good system to input Indian language text. Although their research was focussed on a voice keyboard, it was largely targeted towards Devnagari script (Hindi). They also briefly discussed about the script complexities in languages like Tamil, Thai and Bangla. A like-to- like comparison in terms of script complexities even amongst Indian languages would be of little use. Tamil for instance has 12 vowels and 18 consonants. The combination of these vowels and consonants give a total of 247 characters whereas the Devnagari script (Hindi) has 11 vowels and 33 consonants. Ahmed et.al (2011) explored the challenges in designing input method for Indian languages especially via transliteration.

3.2 SPEECH TO TEXT

One of the aspects to consider while implementing speech to text technology in any language is the accent and emphasis laid on syllables. Tamil has no accent and every syllable is pronounced with the same emphasis (Keane, 2006). Absence of accents on one hand reduces the complexity but other factors such as pronunciation increases it. One of the challenge that has been identified was dealing with phonemes native to Tamil language such as the ழ, ள, ன, ற (Roman equivalent: 'zha', 'LLa', 'na' and 'RA').

There is ongoing research in dealing with recognition of the ழ 'zha' phoneme (A Phonetic term: Any of the perceptually distinct units of sound in a specified language that distinguish one word from another, for example *p*, *b*, *d*, and *t* in the English words *pad*, *pat*, *bad*, and *bat*.) in Tamil (Srinivasan , 2013). This is important especially in speech to text technology where the recognition of syllables needs to suit the native speaker and certain phonemes may not necessarily be pronounced the way it has to be pronounced for various reasons. The three main aspects that were considered in predicting the user acceptance of speech to text application in Tamil were:

a) Accuracy of pronunciation

b) Choice of script- Tamil or Roman

c) Dealing with code mixing Tamil- English (Krishnaswamy, 2015)

The above considerations were based on the positionality as discussed in **section 2.4**

Despite above challenges and complexities, Tamil could be the most suited language for a speech to text technology because of its large character set (247) and very little ambiguity. About 300 phonemes in Tamil form the basis of Tamil speech.

There are ongoing researches in the field of speech to text technology in Tamil. Sridhar et.al, (2013) designed the தமிழ் வாயாடி (Tamizh vaayaadi), software that focussed on a chat system. Sridhar et.al (2013) stated that their software was designed to accept voice input from the user and produced a text output in Tamil. Although they mentioned Tamil as output, it was suggestive that the text produced as output was a transliterated version and not in the form of an actual Tamil orthography. It was ambiguous in terms of the output. Although the sketch and the objective visualises the output to be in Tamil, the actual output appeared to be a transliterated version. It identified that the phoneme (A phonetic term: Any of the perceptually distinct units of sound in a specified language that distinguish one word from another, for example *p*, *b*, *d*, and *t* in the English words *pad*, *pat*, *bad*, and *bat*.) detection was perhaps the main reason for performance degradation. Raghavendra et.al. (2008) discussed the nature of Tamil scripts, letter to sound rules, syllabification of rules and development of unit selection voice for Tamil. But Thennarasu recognises the 'diglossia' (where two or more languages are used by a single community) of Tamil that the technology has to deal with. Raghavendra et.al (2008) proposed a global syllable set, combining multiple Indian languages with the idea that the speech synthesizer could be used to borrow units from a foreign language when the syllable is not found in the target native language. Renganathan(2014), explained the many-to- onSe relationship of Tamil linguistic sounds and orthography symbols. In the context of this thesis, the following observation was crucial:

"*Learning to read and pronounce a wide variety of sounds correctly using any such limited orthographic symbols commonly becomes part of native speakers linguistic intuition and language skill in the respective language*" Renganathan (2008)

The issues identified in Tamil by Renganathan (2008), Raghavendra et.al(2008) Rama et.al (2002) was comparable to that of the issues faced in text to speech of Mandarin. Meyyappan et.al (2013), in the context of automatic speech recognition proposed a cloud based large speech recognition system for Tamil. Automatic Speech Recognition (ASR) is challenging for Tamil because of complex *Sandhi* rules, free word ordering and not strict syntactical rule for word order. Together with the pronunciation model, acoustic model and language model make up the speech recognition system. Similar to Meyyappan et.al (2013) is Schultz (2002) , work on globalphone data bases that contains a speech and text database covering about 15 world languages, Tamil being one of them. The recommendation for multilingual speech recognition was on the basis of the following:

a) Size of the speaker population
b) Economic relevance and politics
c) Phonetic coverage
d) Orthography script variety
e) Geographic coverage and
f) Morphologic variety. Schultz (2002)

Chen et. al (2014) discussed the Chinese pronunciation training system. The prototype was modelled with primary focus on the Chinese pronunciation and tone. Waibel et.al (2002) recognised the increasing interest in multilingual applications and proposes a universal speech recognition that accepted code-switching utterance.

Author and year	Type of work	Corpus details
Rama et.al (2002)	Text to speech	Database from a native Tamil speaker over one month. Demographic details unknown
Yuvaraja et.al (2010)	Building a Tamil voice using HMM segmented labels	Database from a female native Tamil speaker. Demographic details unknown
Rallabandi et,al (2017)	Mixed lingual speech synthesis	Languages consisting of English, Hindi, Marathi, Tamil and Telugu from two males and females.

		Demographic details unknown
Raghavendra et.al (2008)	Global syllable set for speech synthesis	Female speakers in Telugu, Hindi and Tamil. consisting of 2393 sentences in Tamil. Demographic details unknown
Schultz (2002)	Globalphone	Thinaboomi, Tamil Daily. Information about the speaker whether native Tamil speaker is unknown. Demographic from South East India.
Radha et.al (2012)	Speaker independent speech recognition for Tamil using HMM	50 utterances of isolated speech by 10 females. Demographic details unknown
Kumar & Wei (2003)	Bilingual speech recognition in English and Tamil	American English and Tamil. 400 utterances in Tamil which is unclear. Demographic details unknown
Thangarajan et.al (2008)	Word and triphone based approaches in continuous speech recognition for Tamil	75 males and 75 females and training data of 25 speakers. all uttered the same sentences. But the type of corpus is unknown. Demographic details unknown
Schultz et.al. (2014)	Automatic speech recognition for under-resourced languages: A survey	100 native speakers per language. Demographic details unknown
Lamel et.al (2009)	Automatic speech to text transcription in Arabic	Unknown
Karthikadevi et.al (2014)	Syllable based text to speech for Tamil	A selection of 10 words. Demographic details unknown
Saraswathi & Geetha (2007)	Improvement in performance of Tamil phoneme recognition.	Tamil texts of about 200 sentences by 10 male and 10 female speakers.

		Demographic details unknown
Geetha et.al. (2015)	Automatic phoneme segmentation	100 Tamil speech utterances, 25 unique Tamil words consisting of 124 phonemes by 4 native speakers. Demographic details unknown
Vijayarajsolomon et.al (2016)	Acoustic similarities between Tamil and English	One hour speech data by a female from a Tamil novel *Ponniyan Selvan.* English from CMU Arctic database. Demographic details unknown
Chen et.al (2014)	Adaptive Chinese pronunciation training system	MAT 2000 corpus, native Chinese speakers. Demographic details unknown

Table 3.2: Speech recognition and speech to text work with focus on demography.

Table 3.2 provides an insight on sources of data for the corpus in used in the work and demography. Although, the work, in most cases specify that the corpus were collected from the 'native' speakers of the language, in the context of this research it is important to consider the accuracy of pronunciation even by the native speaker as brought out by Chen et.al (2014). Prabhakar & Sahu (2013) have dealt various speech recognition techniques and have summarised speech recognition techniques in other languages including English, Romanian, Urdu, Tamil, Malayalam to name a few. The work of Vijayarajsolomon et.al (2016) and Schultz (2002) suggested that the Standard Literary Tamil was used for the corpus. But the work of Sankar & Nagarajan (2012) in *Tanglish*, suggested the consideration of the aspect of code-switching and the issue of script complexity in Tamil. Since this research has adopted a linguistic view for the application and will be seen in section **4.8.1** and has emphasized on the accuracy of pronunciation as a requirement, it is imperative to consider that these are considered in the corpus. Solomi et.al (2013) have made references to 'appropriate pronunciation' however, from their work on mixed lingual synthesis on Tamil and English, it was ambigious whether the accuracy of pronunciation was considered in both languages. Stolcke et.al (2006) have emphasised on accuracy of speech to text transcription which this research perceives is reliant on the accuracy of pronunciation and speech corpus. This

research took the view that the language skills (especially on accuracy of pronunciation) of the speaker contributed to the corpus.

3.3 ACCEPTANCE MODELS

Over the years, there have been many acceptance models to evaluate the acceptance of technology. The Technology Acceptance Model (TAM) examined the role of perceived ease of use and perceived usefulness in their relation between systems characteristics (external variables) and the probability of system use.
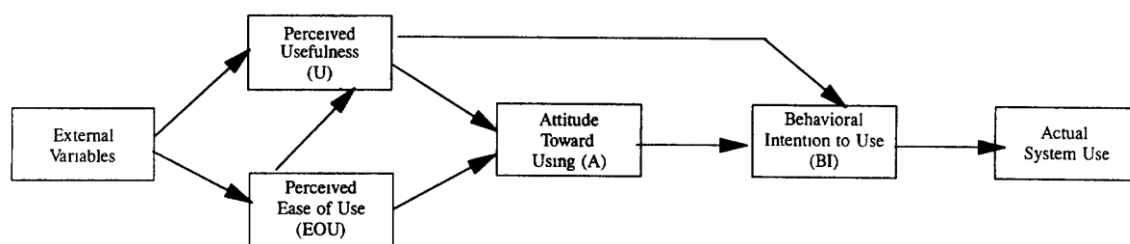
Figure 3.1 *Technology Acceptance Model (Szajna, 1996)*

The above represents the original Technology Acceptance Model (Szajna, 1996) as proposed by Davis in 1989. Hannay et.al (2007) have extensively dealt with usage of theory in software engineering experiments that included but not limited to TAM. However Straub et.al (1997) suggests that technology acceptance model was ineffective across cultures. The following table summarises Straub et.al. (1997), who argued that prediction of user acceptance on the basis of TAM would not necessarily hold true in different cultures as represented in figure 3.2.

Cultural dimensions for selected countries and corresponding TAM predictions

Country	UAI	PDI	MAS	IDV	CMSI	Supportive of TAM?
U.S.A.	46	40	62	91	157	Strongly supportive
Switzerland	58	34	70	68	204	Moderately supportive
Japan	92	54	95	46	287	Non supportive

Figure: 3.2 *Straub et.al (1997) cultural dimensions*

See et.al (2010)'s model of Mobile Phone Technology Acceptance Model in the Malaysian context was similar to Venkatesh et.al (2003)'s UTAUT model. Park et.al(2013) have tested TAM's suitability in mobile sector with specific focus on pyschology.

According to See et.al (2010) and Burton- Jones & Hubona(2006) and Venkatesh et.al (2003)'s UTAUT model was a consolidation and integration of eight models.

1. Theory of reasoned Action
2. Motivational Model
3. Technology Acceptance Model
4. Theory of Planned Behaviour
5. Model of PC Utilisation.
6. Innovation diffusion theory
7. Social cognitive theory
8. A combined theory of TAM and TRA.

Paper & Simon (2007) adapted the Technology Acceptance Model to predict the user acceptance of voice recognition technology where social and subjective norms contributed to the behaviour intention.
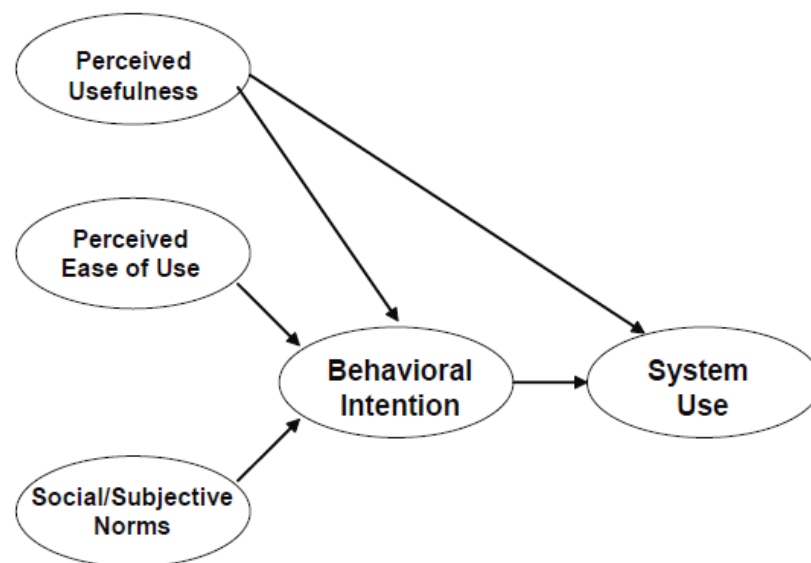
Figure 3.3 *Adapted TAM by Paper & Simon (2007)*

Figure 3.3 illustrates the adapted version of TAM by Paper & Simon (2007). It suggests that the social and subjective norms contribute to behaviour intention to use a system which in turn leads to the actual use of the system. Nistor et.al (2014) suggested that the UTAT model had been insufficiently validated in different cultural settings and has validated the UTAT model. Griffith (1998) model brought out the difference in technological

adoption owing to cultural difference in the case of Bulgaria and the United States. Rappa & Garud (1994) proposed a socio-cognitive model on technology evolution which was based on the basic definition of technology - technology as beliefs, artefacts and evaluation routine. The fundamental definition and understanding of technology became important as it was increasingly linked to individual's imagination of technological possibilities (Rappa & Garud, 1994).

Acceptance Model	Focus of the model
Theory of Reasoned Action	Behaviour behind adoption of technology
Theory of planned Behaviour	Behavioural intention taking into account attitude, subjective norm
Motivational Theories	Psychology. Recognises the difference between extrinsic and intrinsic motivation
Technology Acceptance Model	Perceived ease of use and excludes influence from social, personal and institutional factors.
Domestication approach	Sociology and integration of technology in everyday life
Innovation diffusion Theory	Perceived characteristics coupled with individual attitude, belief and society around.

Table 3.3: Acceptance model and their focus

The above table summarises various acceptance models and their focus. The table provides an insight on various factors that could be taken into account to predict the acceptance of a technology. Wixom & Todd (2005) proposed a model that theoretically integrates user satisfaction and technology acceptance. It has been identified that predicting behaviour has been a challenge and advocates that the attitude and beliefs must be contextual with the behaviour of interest, consistent in time and target. Wixom & Todd's (2005) model suggest that technology acceptance and user satisfaction are complementary rather than competing. TAM lacks contextual and temporal variations (Wixom & Todd, 2005) which is also to some extent consistent with the views of Straub et.al. (1997). The Unified Theory of Acceptance and Use of Technology (UTAUT) was formulated with four core determinants of intention and usage and up to four key moderators of relationship

(Venkatesh et.al, 2003). Venkatesh et.al (2003) has identified direct and indirect determinants of user acceptance and usage behaviour as listed in **Table 3.4**.

Direct determinant	Indirect determinant
Performance expectancy	Attitude towards using technology
Effort expectancy	Self-efficacy
Social influence	Anxiety
Facilitating conditions	

Table 3.4: Direct and indirect determinants as identified by Venkatesh et.al (2003)

Venkatesh et.al (2003) identified that little or no research has addressed the link between individual and or organisational usage outcomes and user acceptance. The role of social influence or construct was controversial and its inclusion in the user acceptance model remained inconsistent. Burton-Jones & Hubona (2006) observed that the UTAUT model integrate most of the acceptance model. Venkatesh et.al(2003) asserted the importance of social influence and that these social influences could change over time. Further, it suggested that the technology acceptance models could feed into the application designers. Faulds et.al (2013) explored the link between the user acceptance and requirement fulfilment. The work of Hwang et.al (2016) suggested that the Theory of Reasoned Action and Theory of Planned behaviour was within the domain of social psychology and that the ultimate goal for these theories is not just to predict but also to understand the human behaviour. Hwang et.al (2016)'s observation on social influence and its complication was consistent with Venkatesh et.al (2003)'s UTAUT model. Collerette et.al (2003)'s summary of different models and their findings provided an insight of various user acceptance models. They suggested that the very work included the social influence process such as Agarwal & Prasad (1997), Venkatesh & Davis (2000) and Spitler & Lucas (1999).

3.4 GAPS IN LITERATURE

The review of the literature around various user acceptance models and their underlying theory provides a generic framework to predict the user acceptance of a technology or system. However, these models were suggestive to have been developed from a perspective of organisational conditions as opposed to a more generic condition. This research aims to predict the user acceptance of speech to text by proposing a model for a specific application - speech to text and is built upon Venkatesh et. al (2003)'s UTAUT model of user acceptance.

The *Table 3.2* brings to the surface the ambiguity on the aspect of accuracy of pronunciation in the respective target language of the corpus which this research views extremely important from the perspective of the linguistic view.

3.5 CODE SWITCHING, LANGUAGE AND POLITICS

Code-switching and code- mixing between Tamil and English, among various other languages especially in countries like Singapore and Tamil Nadu (Krishnawsamy, 2008) in the Union of India (where Tamil is natively spoken), was an important aspect that was considered for speech to text in Tamil. This was also quite apparent from the Tamil medium educated students who have had a fairly high degree of exposure to the English language (Krishnawsamy, 2015). Schiffman (2009) brought out the variety of Tamil both Spoken and Literary Tamil, noting that the two are very different from each other. His work further suggested that Tamils in Singapore didn't even tend to speak in Tamil with their kids at home. His comparison of people's approach towards the Tamil language between Singapore and Tamil Nadu did not vary to a great extent. Tamil spoken in Tamil Nadu does make use of a good English vocabulary partly because of Anti Hindi, Sanskrit stance and English in Tamil Nadu is viewed more as a "buffer" to counter imposition of Hindi and Sanskrit and is considered as a language of the elite and educated. However, the loyalty of Tamils towards the Tamil language is hardly reflected in their speech (Schiffman, 2009). Even amongst the Tamils, English is quite a dominant language. Sivaraja(2013) contextualises the social attitude of native Tamil speakers in buying and using Tamil software.

The south east Asian countries were colonised by the British for over two hundred years. Therefore, the diaspora research suggested that the diaspora Tamils show a greater affinity and fluency in English than in their native language (Das, 2011). Manuel (2013) in her research on linguistic anxieties on the impact of English on politics, mother tongue and creative writing in India clearly expressed her position on the perception that the natives possess in the post-independence. Her research reported that the colonial language, English 'naturally' dominates the country which the natives on one hand view as a threat to their language but on the other hand, English has positioned itself in the economic front not just globally but also within the Indian Union. Manuel (2013) concluded that the natives, as a result of colonisation for many centuries have indigenized English. According to Fanon, post-colonial experiences could be quite traumatic and the transition may not be very easy (Sefa Dei &Asgharzadeh, 2003). In the

case of the Ghana, the Ghanian authorities argued that English was perhaps the only language through which they could unite the country which speaks over sixty different languages (Sefa Dei &Asgharzadeh, 2003). It echoed Fanon's argument on the dominance of colonial language. Colonial languages such as French, English were largely viewed as 'progressive', 'impartial' and 'civilised' languages compared to the native tongues. Fanon explained that it was achieved mainly through economic value of the language. For instance, the French was given more prominence and economic value in Algeria. With reference to Ghanian society, in the post independence context, English was viewed to be 'neutral' and a 'link' language, they have in the process has unconsciously undermined the potential of their own language and culture and without their knowledge, submitted to English which Fanon referred to as a 'psychological' trauma. During British rule, English was the language of education, commerce and the symbol of 'civilised' – a term that Fanon used to indicate that the coloniser attempted to subjugate the natives by referring to them as 'barbarians' and 'uncivilised' and by projecting the coloniser's language and culture to be more civilized. Post independence, the Tamils saw opportunity in English, like Ghanian society. The non Hindi speakers especially the Tamils viewed English not just a 'neutral' language, but also to use it against the Hindi hegemony. In the early sixties, Ezekiel and Parthasarathy have hailed Ramanujan- a Tamil Brahman poet who used English as the main language to write- as the best poet in Indian English (Ramazani & Ramanujam, 1998). It was interesting to note that English was a language that met with stiff resistance before 1947 was more or less accepted as an indigenous Indian language by the 60s. Lionnet (1993) presented an interesting study on Mauritius which has had a history of multiple colonisers and a predominant Asian population. The majority of the Mauritians not only spoke *Creole* but also to some extent their 'native' language such as Tamil, Bhojpuri etc. However, the first priority of official language and preferred use of language is English and French followed by their own native languages, which was reflective of the influence of the colonial powers not just in the language policy but also towards the attitude towards one's mother tongue and is extensively dealt in the discussion chapter.

3.5.1 THE NOTION OF STANDARD ENGLISH

Milroy & Milroy (2012) in their work on 'Authority in language- investigating standard English' provided a useful insight on the types of English, the idea of 'correctness' in language and pronunciation, changes in the language over centuries and some focus on linguistic prescription and the speech community. The tradition of 'pure English' first

emerged in the society around the 16th century. Sapir-Whorf hypothesis proposed that the thought process were conditioned by the structure of the language that was spoken by the society. The objection of foreign borrowings into English emerged in the nineteenth century. Milroy & Milroy (2012)'s work was particularly relevant as they have dealt with the prescriptive ideology of English and have presented a comparison of Singapore English and British English. The example of Singapore was useful and relevant as Tamil along with English, Malay and Mandarin is one of the official languages of the country. The authors observed that Standard British English was viewed superior by the Singaporeans as opposed to Singapore English when the latter was socially acceptable in Singapore. Milroy & Milroy (2012)'s work was similar to Schiffman (1998) on the standardisation of Tamil. According to Kroch (1986), the difference in pronunciation could also be observed in words of French origin in English which was referred to as prestige dialect. Nevertheless, in the context of Tamil, and according to Keane (2004), the dialectic variation should perhaps not be a reason for mispronouncing a syllable, as also supported by Murthy & Kumar (2006).

3.5.1 ORTHOGRAPHY CHOICES: தமிழ் OR TAMIL?

There are work on Romanisation of non -European languages such as the Uzbek language in the Uzbekistan (Uzman, 2010). The Uzbek community had perhaps undergone one of the most radical changes in the recent history. The possible reason behind this could be social, cultural, economics and perhaps even political. In East Asia, the reason for Romanisation of languages like Korean, was to take it to the rest of the world. Romanisation of Korean, like Tamil faced the same problem of inconsistencies because the native pronunciation was often ignored (**see appendix B.1 to B.30 and section 5.1**). Transliteration was first introduced in early 1980s to empower the Tamil speakers to write in Tamil using the Roman script. And there were several attempts to Romanisation of Indian languages in the colonial era (Kurzon, 2010). The Roman script was first introduced in India by the Portuguese and languages like Konkani- a regional language of Goa in the Indian Union continues to be represented in the Roman script. The phenomena of 'Romanisation' extended well beyond the Indian sub-continent. The Malay language which traditionally used the Jawi script uses the Roman script. Furthermore, the Roman script is perceived to be more international than other scripts (Sebba, 2010). The study of Romanisation of Indic Scripts in Ancient Indonesia (Acri & Griffiths, 2014) suggested that transliteration has little or no regard for pronunciation.

Aytürk (2007), suggested that conservatism that favours an established script and spelling was a widespread phenomena and was understandable from a speech community and explained the resistance of Romanising the Hebrew script. Romanisation of Hebrew script was attempted for the convenience of outsiders and those familiar with the Roman script- who could speak Hebrew fairly fluently but could not read it which is discussed in section **8.1 and 8.2**

Technology	Contextual relevance
Schultz (2014) et.al , Thangavelu et.al (2016), Wei et.al (2003), Schultz (2002), Sankar & Nagarajan et.al (2012), Thakkar et.al (2016), Black et.al (2010), Chandra et.al (2015), Adel et.al (2015), Natarajan et.al (2009)	Natarajan et.al (2008), Das(2011), Rudisill (2012), Das (2008), Schiffman(2002), Canagarajah (2008), Kailasapathy (1979), Schiffman (1998), Srinivasan (2008), McDonough et.al (1997), Fuller et,al (2014), Krishnaswamy (2015), Ridge (2012), Ciotti (2017)

Table 3.5 : Studies of speech recognition and speech to text technology and associated social constraints

Table 3.5 brings to surface, some of the work done in Tamil speech to text technology, speech recognition alongside, the social challenges within the context that needed to be considered. Although, mixed language synthesis and bilingual synthesis indicates possibility of embedding aspects of code-switching and code-mixing element, the work of Kailasapathy (1979) has compelled to investigate into the issues around language in particular that the technology is built upon. Srinivasan (2008), Krishnaswamy (2015) work were particularly important from the perspective of the attitude of native Tamil speakers in the native space and is explored in detail in Chapter 8 Das (2008), Canagarajah (2008), Schiffman (2002) Das (2011) and Ridge (2012) dealt with the issue of language maintenance and attitude towards Tamil at a diasporic level. The work of these authors point to a common issue of language maintenance, shift in language and attitude towards Tamil language. Rudisill (2012), Fuller & Narasimhan (2014), Ciotti (2017) McDonough et,al (1997) brought to surface, the Sanskritised Tamil of the Brahmans, and their ability to pronounce zha, la and La fairly distinctly and accurately. Their work also suggest a social difference towards the attitude of pronunciation and Tamil language and is discussed in the Chapter 8

3.6 SUMMARY

This chapter has provided an insight about speech to text technology, acceptance models, choice of orthography and has attempted to integrate key variables and focus of this research - technology and acceptance model proposed to predict the user acceptance of speech to text in Tamil.

The technology section has discussed the various methods by which a speech to text application could be developed including that of code switching and code mixing. The user acceptance model provided an overview of the different models along with their key focus. It identified that the integrated model consolidated all the eight models and more importantly recognised the social influence to a greater extent than the others.

The gaps in literature were identified in the context of user acceptance model since, a user acceptance model for speech to text proposed in thesis, was used to predict the level of acceptance of Tamil speech to text application. It further identified the importance of setting focus on the corpus demography and accuracy of pronunciation.

Code mixing and orthography choices provided some insight on the history of the target region. This was contextually relevant and important as the researcher has used insider and outsider position coupled with phenomenology and native knowledge in proposing the user acceptance model based on Venkatesh et.al (2003)'s UTAUT model. The choice of orthography was inevitable in speech to text and the choice between Roman and Tamil orthography could be attributed to the knowledge of reading and writing in Tamil. The insider positionality indicated the possibility of the receiver being unfamiliar with Tamil script in order to read Tamil in Tamil orthography cannot be underestimated.

CHAPTER 4 - METHODS

AIM

The aim of this chapter is to introduce the research philosophy and contextualise the philosophical considerations and approaches that this research has followed in order to answer the research question. This chapter deals with the research philosophy as opposed to indigenous philosophy (The Vedic philosophy as seen in section **2.3**).

4.1 RESEARCH PHILOSOPHY

Saunders et.al (2009) presented a comparison of four research philosophies, namely:

1. Positivism

2. Realism

3. Interpretivism and

4. Pragmatism

Both qualitative and quantitative or either one of them could be used to collect data. Saunders et.al (2009) referred to the process as 'research onion' where the data collection and data analysis forms the core of the 'onion'. The outer most layer is to decide the philosophies and to choose between inductive and deductive approach which in turn lead to choosing from one of the following strategies:

- Experiment
- Survey
- Case study
- Action research
- Grounded theory
- Ethnography
- Archival research

Once the above choice was made, the next layer of the research 'onion' was to choose between mixed methods, multi-method or mono method which led to the core of the onion-'data collection and analysis'.

Teddlie & Tashakkori (2010) indicated that mixed method approach puts more emphasis on humanistic conceptualisation of the research process. The authors insisted on *methodological eclecticism* - exploring a diverse methodological tool to answer the research question which can be related to Keane & Khupe (2017)'s view on individual connection to the research.

The **Ontology** represented the researcher's view of the nature of reality. But when this was applied into realism it implied that existence was independent of human thoughts and knowledge and could be interpreted through social conditioning which could also referred to as 'critical realist'.

The **Epistemology** qualified the nature of this thesis as the research question. That is, the idea of the research was based on the researcher's view regarding what constituted an acceptable knowledge which fitted in with the interpetivism philosophy that implied subjective meaning and focused upon details of the context which was again subjective. Arguably, it could also fit in well with Pragmatism philosophy but was beyond the scope of Positivism as it does not attempt to generalise. It could to some extent fit into Realism since it deals with explanation within a context but with small sample size, the results cannot be generalised and was open to misinterpretation.

The researcher in this research was value bound. Therefore the researcher was a part of what was being researched (see positionality) and cannot be separated. This position and view of the researcher was consistent with the axiology of interpretivism. The axiology of positivism could possibly not define the approach since the researcher was not independent of the data. On the other hand, the axiology of realism and pragmatism to some extent fitted into the context as the researcher's bias as a result of upbringing, cultural contexts and views did to some extent bore an influence on the findings of the research. Although, indigenous philosophical values were fundamental to this research, the researcher's subjectivity and bias in the interpretation of result could possibly rule out a pragmatist axiology.

This research has resorted to in-depth interviews, small samples and a qualitative method which was consistent with the interpretivism research philosophy.

4.2 PHENOMENOLOGY AS A METHOD

Aktay (1998) explored technology and phenomenology from the perspective of various philosophies including Zen Buddhism.   There have been criticisms in the choice of qualitative phenomenology as a method (Englander, 2016). Staehler & Lewis (2010), referred to Husserl's stress on intuition which, Husserl defined as a method of phenomenology. But, Husserl clarified that the word 'intuition' contextually referred to the direct contact of the object rather than mystical. Husserl's concept of *epoche* was central to his phenomenological method (Englander, 2016). Kinnunen & Simon (2012) provided a detailed account of using phenomenography and grounded theory as research methods in computing education research field. Some of the strengths, challenges, opportunities and threats of phenomenography and grounded theory. Some of them that are relevant to this research are listed in table 4.1:

Strength	Challenges	Opportunities
Models/ theories are likely to work since there is a close relationship between model, and actual data.	There is no "right way" to data analysis and there might be several ways and opportunities- all produce valid results and choosing the way that answers the research question might be challenging.	Both relate to phenomena that has not been investigated or studied before and these are not restricted by the already available models and therefore has the potential to discover something that is genuinely new.
Ground theory gives us the required tools to study and understand the context and consequences of something that happened or in other words to answer why something happened.	The researcher needs to be aware of presumptions throughout the research process and the data analysis takes time. Reporting the detailed qualitative analysis in limited pages is challenging.	There aren't strict rules on data analysis which allows the researcher to be creative.

Table 4.1 Strengths of phenomenography and grounded theory

The strength and challenge in this particular case complemented each other. The challenge was perceived to be an opportunity and lead to something genuinely new. Lukaitis (2013) provided a very similar impression to Kinnunen & Simon (2012) and suggested that using phenomenology along with the application of hermeneutic analysis of text could provide a research framework with robust philosophical and methodological underpinnings.

The following are the stages of phenomenological research process as provided by Lukaitis (2013)

- Epoche
- Textural representation
- Preliminary grouping
- Reduction and elimination
- Clustering into themes
- Invariant constituents
- Imaginative variation
- Synthesis and essences.

Other than the philosophical roots of grounded theory that was based on pragmatism, grounded theory as a method could also be held relevant to this research since the proposed user acceptance model was supported by the data and positionality of the researcher. It followed an inductive approach rather than a deductive approach and has used interview as the data collection method, which is strikingly similar to the stages of phenomenology. Kinnunen & Simon (2012) provided a brief account of the similarities between Grounded theory and Phenomenography:

- Both are non-positivist view of knowledge.
- Qualitative methods are used for data gathering although, both are flexible and open to other methods of data collection.
- Uses a purposive sample of interviews.
- Quality of the data has to be rich.
- Both use an inductive approach.
- Iterative process of data analysis is yet another common feature between the two.
- Boarder line cases help identify finer aspects of how one is different to the other.

4.3 SOFTWARE ENGINEERING AS A RESEARCH METHOD

Ghezzi et.al (2002) on the basis of IEEE defined software engineering as the application of a systematic, disciplined, quantifiable approach to the development, operation and maintenance of software. Vliet (2000) provided other definitions that include:

- Software Engineering concerns the construction of large programs.
- Software evolves
- The software has to support its users effectively
- Software engineering is a field in which members of one culture create artifacts on behalf of members of another culture.

The last two bullet points are of particular relevance to this research. Vliet (2000) indicated that the software engineers:

"*lack factual knowledge of the domain for which they develop the software, they lack the knowledge of its culture as well.*" . It further raised a very crucial question in relation to the system acceptance and the norms of the community for which the software was being designed and developed. The final point could be extended to the previous point that linked to supporting the user with a view to satisfy the users. Study of the target users would enable to elicit functional requirements. Software engineering lifecycle was used to develop the prototype and focussed on the requirement and feasibility of the application.

4.4 ABSTRACTION

Pfleeger & Atlee (2009) discussed Wasseeman's discipline of software engineering in which abstraction was a description of the problem at some level of generalisation. This research attempted to view the problem in its natural state (i.e. as it exists and also as informed by the study sample). For instance, the problem statement of speech to text application was based on the idea of *'what you see is what you get'* which to an extent was a generalised statement.

4.5 WHO IS THE CUSTOMER?

According to Stephen (2015), the customer was central to the requirement gathering process. Therefore, a customer's requirement forms the key document that directs a project. The essential question that need to be answered to identify the customer is:

"*For whom is this application being developed?"*

In some cases, the customers can be individuals developing the application. Although, initially, the prototype task developed for this research was aimed at native Tamil speakers, eventually it focussed on Tamil speaking Brahmans. The conceptualisation of the application was performed by the researcher. Since the researcher was 'clear' about how the application should function, contextually, the researcher was also a customer. The first prototype was almost entirely based on the researcher's conceptualisation. However, in the wider context and even within the study sample, the researcher along with other stakeholders (people speaking the same language within similar philosophical background) could be considered as customers for this application. Identifying the customer was important in order to determine the needs of the customers and to answer the broad research question on user acceptance.

4.6 SOFTWARE DEVELOPMENT LIFE CYCLE

The following are the stages involved in the development of software Pfleeger & Atlee (2009):
- Requirements analysis and definition
- System design
- Program design
- Program implementation
- Unit testing
- Integration testing
- System testing
- System delivery and
- Maintenance

The first and perhaps the most important stage in a software development life cycle is eliciting the requirements. Vliet (2000) suggested that it also includes a feasibility study which could again be sub-divided into the following:

- Economic feasibility
- Technical feasibility

The requirement phase recognised the diversity of the customer base, that is, not all the customers would come from the same background. It was particularly relevant to the study sample with whom the paper prototype was tested. Although, they could be grouped on the basis of language, geography and culture, they had diverse opinions, views and expectations on the functioning of the application (Geogy & Dharani, 2016). For instance:

VM: "*I want Tamil orthography to represent Tamil, Sanskrit to represent Sanskrit and Roman to represent English."* which was in sharp contrast to

VA: "*I want Tamil orthography to be represented in Tamil but then English word to be translated and represented. For example: normal as* பொது "

Although, the software development life cycle appears to be different stages, Vliet (2000) suggests that the requirement engineering also need to pay close attention to testing simultaneously and that it was not advisable to leave it till the end.

4.6.1 CATEGORISING REQUIREMENTS

Audience-oriented requirements: The primary focus of speech to text prototype were the native Tamil speaking Brahmans. The following requirements emerged in the process of interview and testing of prototype (will be seen in **chapter 8**):

1. The application should be able to cater to Sanskrit- Tamil code mixing.

2. The user must pronounce the syllable accurately and the corresponding syllable must appear in the Tamil or its equivalent Roman orthography.

 There were however, variations in the above requirement (will be seen in **chapter 8**)

Business requirement: There were no business requirements for this application from the perspective of commercial value. That is, the prototype in its entirety was developed from the point of view of the language and culture without a case for business. However Anuraja (2013) observed lack of market for a Tamil software application and attributed it to the perceived usefulness of the language and will be seen in **chapter 8**.

4.6.2 FORMULATING REQUIREMENT

This section formulates the requirement of a speech to text application based on the following as suggested by Stephens (2015)

- Clear
- Unambiguous
- Consistent
- Priotitised

The speech to text application was designed on the idea of '*what you speak is what you get*'. This meant that the application would merely translate the speech of the user in the orthography, that conformed with the scope of the target language - Tamil in this case. The paper prototype did not incorporate features such as auto correct and spell check.

MOSCOW Method:

M	Must	Users speech converted into Tamil orthography.
S	Should	An option for users speech converted into Tamil or Romanised Tamil orthography along with code switching and code mixing in two languages.
C	Could	Speech to text without the effect of code mixing or code switching.
W	Won't	Incorporation of spell check, auto correct features, word prediction etc.

Table 4.2 Illustrating the requirements of the speech to text prototype using the MOSCOW method.

4.6.3 REQUIREMENT DEFINITION

In this section, the requirement is expressed in the customer's vocabulary. It sets an expectation of what to expect from the system and makes the assumptions and environment clear.

General purpose and scope of the system	To convert the speech of the user to the target orthography- Tamil or Roman. As **seen in 3.2**
Rationale for the proposal of the system	The proposed system was based on the idea of '*what you see is what you get*'. In the proposed system, the system merely converted the user's speech into text and did not auto correct, predict the word for the user or check the spelling. The system used the language as a criteria for the orthography **(as seen in section 2.2)**.
Environment in which the system would operate:	The prototype was designed based on iOS. The prototype was designed for the native Tamil speaking Brahmans. For the system to be used, the prototype was designed with an expectation that the users were able to pronounce all Tamil syllables accurately. Brahman Tamil (Manipravalam) in itself is a mixture of Tamil and Sanskrit **(as dealt in section 2.2).** It was assumed that the Tamil Brahmans would have basic knowledge of Sanskrit and the Vedas. It was therefore expected that the Tamil Brahmans who use the speech to text application should be able to speak Tamil with an expectation to be able to read in Tamil orthography as seen in the literacy definition in the introduction chapter.
Proposed solutions by the users	Some of the participants suggested translation of English words into the equivalent Tamil orthography. And some

	suggested using the respective orthography for the words uttered by the user. For instance, Tamil words in Tamil orthography, English in Roman and Sanskrit in Devanagari **(see 6.1 & 6.2)**
Behaviour of environment	The native proficiency in language, pronunciation and choice of script by the users were perceived to play an important in the success and failure of this application **(4.12 and chapter 8)**. The application could almost be redundant if: • The users switched to other languages as their primary language for communication. (Linked to language maintenance and will be seen in section 8.1) • The application was designed in an orthography that the user is unable to read. • The user was unable to pronounce Tamil or Sanskrit syllables accurately.

4.6.4 REQUIREMENT CHARACTERISTICS

This sections aims to reflect on the characteristics of the requirement in conjunction with the previous section as recommended by Pfleeger & Atlee (2009)

Are the requirements correct?	The requirement elicited was reasonably correct from the point of view of the researcher and the stakeholder (study sample).
Are the requirements consistent?	No. There were discrepancies in the requirement which would be seen in section 8.1 & 8.2. The discrepancy was more prevalent in the choice of script.
Are the requirements unambiguous?	Yes. The requirements for the application was unambiguous and the requirements as suggested by the other stakeholders was clear. But there were differences in the expectation of the customers.
Are the requirements complete?	Yes. The requirement was that the user was able to speak in Sanskritised Tamil (Manipravalam) and the output in the Tamil orthography was produced. When a mispronunciation of a syllable occurred, the mispronounced syllable was produced as the output, instead of the actual syllable that was mispronounced.
Are the requirement feasible?	No. It was feasible to provide a compatible solution. It was feasible to cater to a code-mixed language (Sankritised Tamil) However, the incorporation of features primarily in the choice of script and translation of English words into Tamil, as suggested by the stakeholders was beyond the scope in the same application (which would be seen in **8.1 & 8.2**).

Is every requirement relevant?	No. The stakeholder's expectation of translation and displaying the output in a combination of scripts was not relevant to the specific type of prototype that was designed and investigated.
Are the requirements testable?	Yes. The requirements are testable under the defined conditions of the prototype as discussed in **section 4.6.3**.

4.6.5 CLARIFYING AS-IS SCENARIO

Google introduced speech recognition in Tamil in August 2017 (Sahoo, 2017). Sridhar et.al(2013) proposed Tamil vaayaadi, a chat system that attempted to implement speech to text technology. However, the output orthography in Sridhar et.al. (2013) was ambiguous. To date, there is little or no research or application addressing the speech to text, from the point of view of this research that caters to the native Tamil speaking Brahmans and that takes into account, the pronunciation and handles the Sanskrit Tamil code switching and code mixing speech.

4.6.6 CLARIFYING TO-BE SCENARIO

A speech to text application that is able to accept Sanskrit- Tamil code mixing (Brahman Tamil) and convert into its equivalent Tamil orthography. The application design would be consistent with the standard pronunciation of syllables in Tamil, Sanskrit and as spoken by the Tamil Brahmans. The application to be based on iOS iPhone and the technology should enable the user to use the application on any application within the iPhone such as notes, message, e-mail to name a few. Table 4.3 summarises the As-Is and To-Be scenario of the prototype scenario.

As- Is	To- be
Technology does not cater to the Brahman Tamil speech in iOS iPhone 4s/5s.	Brahman Tamil speech that accommodates Sanskrit Tamil code mixing was made available in the form of a paper prototype and its functionality was tested on messages.
Currently the two methods available for Tamil input are Standard Tamil keyboard (includes Grantha alphabets in Tamil orthography) and Roman Tamil keyboard that facilitates transliteration.	On the Standard Tamil keyboard, a speech input button was made available in the prototype.
The level of proficiency in Tamil orthography varies	The option of making available, the Tamil speech to text in Roman orthography was considered.
Limited or no use of Tamil language in technology for the purpose of communication (as seen in the study sample).	Some use of Brahman Tamil language in technology.
Code mixing and code switching in Brahman Tamil speech (Sanskrit- English- Tamil) is quite frequent. (As seen in the study sample and can be seen in the **appendix**)	Reduced effect of code switching that could contribute towards preserving the linguistic heritage.

Table 4.3 summarising As-is and To-be prototype scenario

4.7 SOFTWARE PROCESS MODEL

This section provides an overview of the various software process models and discusses the prototyping model employed in this research.

The various software process models are Waterfall model, V model, Prototyping model, Incremental and Rapid Application Development (RAD). The major drawback of the waterfall method was that it does not inform about the process that lead to the creation of the

final product. The V model focusses on the activity. It is derived from the waterfall model and demonstrates the relationship between analysis and design with testing.

4.7.1 PROTOTYPING MODEL

Pfleeger, S., & Atlee, J. (2009) indicated that the prototyping could potential be a process model of its own rather than following the waterfall model. Vliet (2000) discussed prototyping in the context of requirement engineering and the benefits it had to offer. Clarifying requirements at the initial stages of the software development lifecycle could be done more effectively through prototyping. Since prototyping was less time consuming, the process, itself could go through several iterations in order to clarify the requirements. Expression of precise requirement was quite often a difficult and prototype model was helpful to achieve the clarity of requirement before even proceeding to the development stage. Prototype offered the flexibility in design and development standards in the sense that it does not strictly enforce the design and development standards that was otherwise a part of the software development life cycle. It could also be used to test the user satisfaction at the initial stages of the software development. Prototyping model potentially involves the end users in the design and development process which in turn provides ample clarity on the requirements.

Prototyping is useful in the following scenarios :
  a) When the requirement is ambiguous and needs more clarity.
  b) When dealing with a system that lay emphasis on user interface.
  c) However, the potential pitfall of the prototypes could be:
     a. It is only a model or a representation of what the final product would look like.
     b. Although prototype offers flexibility and can be quickly updated based on the users feedback, it is essential to closely control and impose restrictions on the number of iterations in order to effectively manage the process.

4.7.2 TYPES OF PROTOTYPING

There are two main types of prototyping:
  • Throwaway prototyping
  • Evolutionary prototyping

In the throwaway prototyping, the prototype does not develop into an actual product. Rather it is explicitly thrown away after going through all the iterations.

When the prototyping leads to the evolution of a final product after a number of iterations, it is referred to as an evolutionary prototyping. Although both are in common usage, the latter is more commonly used in practice.

4.7.3 INCREMENTAL AND RAPID APPLICATION DEVELOPMENT

Ghezzi et.al (2002) and Vliet (2000) provide identical definition of incremental model. In the incremental model, as the name suggests, the end users are involved in planning the next steps, keeping in mind the desired outcome. Instead of doing nothing for a long period of time and then presenting with something all of a sudden, in the incremental model, the progression happens in small increments and it also involves the users in the process. The incremental model could help overcome the issue of 'over functionality'.

The Rapid Application Development (RAD) shares a lot in common with other development models. The fundamental difference is that the RAD lays emphasis on the time frame. The time limit set in RAD is non-negotiable. Therefore, the entire development process has to be completed by the fixed time. In certain cases, the functionality is sacrificed as opposed to extending the time. The objective of answering the main research question was achieved through throwaway prototyping.

4.8 PROTOTYPE DESIGN

The designed employed Carroll (2000)'s rationale on scenario-based design. Gould & Louis(1985) recommended three key principles of design one of which was to closely study the user group involved in using the application. A finer understanding was arrived by studying the user's behaviour and attitudinal characteristics during the evaluation of the prototype as recommended by Gould & Louis (1985).

Medhi & Toyama (2006)'s work on text-free user interfaces for illiterate and semi-literate users employ contextual and ethnographic design in the southern Indian city of Bengaluru. Their work suggest that the user's response was dependent on factors like cultural,

religious or psychological biases. This research has attempted to closely study the user group and taken into account the influences of culture and philosophy as seen in Gould & Louis (1985) and Medhi & Toyama (2006)'s work.

4.8.1 DESIGN CONSIDERATIONS FOR SPEECH TO TEXT APPLICATION IN BRAHMAN TAMIL

A PROTOTYPE THAT SUITS EVERYDAY LIFE

The possession of smart phones or iPhone in particular could be for various reasons. To a question on possession of an iPhone, VL commented:

"Social status. It is a social status. **நா மொதல்லையே வாங்கும் போது** , **நா மொதல் மொதல்ல** smart phone, iphone **வாங்கும் போது என்னத்துள்ள என்ன இருக்குனே தெரியாது**. **ஆனா வாங்கி குடுத்தா வாங்கிண்டாச்சு**. **அதுக்கப்பறம்** I just **நா வெறும் ஒரு தொலைபேசியா தான்** use **பண்ணேன்**. **ஆனா அதுல என்ன இருக்கு**? **இப்போ கூட என்னோட கைல இருக்கற** phone **ல என்னென்ன** provisions **இருக்குனு எனக்கு தெரியாது**. **நா எல்லான்லா** use **பண்ண மாட்டேன்**. **ரெண்டு மூணு இது தான்** use **பண்ணுவேன்**. **அது கூட இப்போ சுத்தமா விட்டாச்சு**."

*When I first got the iPhone, I didn't know what it had in it. It was bought and given, I had it. I just used it as a telephone. Even now, I don't know what it is in it, what provisions has it got. I don't use everything, just two or three and I have left them totally now - **approximate translation***

VM had very similar comments on transition from a basic phone to a smart phone.

"*Okay I will be honest. எனக்கு ஈர்ப்பெல்லாம் எதுவும் இல்ல. Okay there was a shift. There was a shift back when phones were introduced into the market, reliance, and then nokia started giving you the models at a cheaper price. The only thing that was hot were **sms text messages and calls**. Then came the android OS, and came the camera phones and one by one it started building up to a particular level. The reason, okay why this transition happened was, as I said, **it was partly due to the social pressure**. Nobody wanted to sit down and text*

anymore and spend fifty paise on a text message. They moved onto some instant messenger-whatsapp being the latest one. So, they moved onto an instant messenger, and they started sending- you know, let it be organisations, let it be schools and colleges, they started embracing this instant communication model."

These comments of the participants provided useful insight on how the users interacted with the technological device such as smart phones on a day to day basis to some extent informed the design . Beyer & Holtzblatt (1998) discussed the challenges of fitting into everyday life and suggested the system to support the way users want to work.

Vliet (2000) indicated that there are three different views from which one could look into the conceptual model. They are:

1. Psychological view
2. The linguistic view and
3. The design view

The conceptual model of the prototype took a linguistic view that was consistent with the researcher's idea of 'what you speak is what you get'. The prototype was conceptualised on the basis of linguistic view as discussed in section 2.2. By applying linguistic view to the design of the prototype, it meant that it imposed a way of working (Beyer & Holtzblatt , 1998). The design of the speech to text prototype was on the basis of language as a non negotiable component, with emphasis on accurate pronunciation of syllables, output in Tamil orthography, and some scope for negotiation for Roman orthography as part of refining the requirement. In the process, it however, did not in any way underestimate other Tamil dialects. Potentially, any native Tamil speaker with accurate pronunciation could use the speech to text application. But, the prototype was designed for native Tamil speaking Brahmans for reasons as discussed in section 4.6.3, with the assumption that it could be used by any native Tamil speaker with accurate pronunciation. Vilet (2000) views design as a multi-disciplinary collaboration where it has been argued that sociological, ethnographical psychological focus or any individual focus to the extreme could still provide unique contributions to the system under study.

The following were the considerations that informed the design of the prototype of speech to text application in Brahman Tamil dialect.

1. **Knowledge of language:** The knowledge of language refers to the ability of speaking Brahman Tamil dialect along with the ability to read Tamil in Tamil orthography. The design aspect included decisions on orthography for which the knowledge of language was essential. This also relates to the choice of script that was one of the three criteria for the prediction of user acceptance of speech to text application. VM and VL's ability to express in Brahman Tamil dialect and will be seen in section 7.1. The contrasting use of language could possibly question the point of developing a speech to text in Tamil. It is in this respect, the study of end users (also referred to as stakeholders in the previous sections) becomes not only relevant but also crucial as advocated by Medhi & Toyama et.al (2006). Gould & Louis (1985) emphasise the 'insider' view of user requirement for a superior user interface.

2. **Accuracy of pronunciation:** The Brahman Tamil is slightly different to the Standard Tamil. But the choice of Brahman Tamil dialect in speech to text application was on the basis of code switching (Sanskrit -Tamil) and the perceived indigenous philosophical influence towards accuracy of pronunciation. Studies such as McDonough & Johnson (1997)**,** Tola & Dragonetti, Fuller& Narasimhan (2014) have made references to accurate and clear pronunciation of the Brahmans. Fuller et.al. (2014) and McDonough & Johnson (1997) specifically refer to the Tamil Brahmans and Srinivasan (2013) emphasises on the importance of the recognition of Zha, la and La.

3. **Choice of script:** With specific reference to iOS keyboard, there are currently two keyboards available for input in Tamil orthography- Roman and Tamil. As a result of researcher's insider position and is dealt in , the government policy of Tamil Nadu state that follows a two language formula and would be seen in the chapter 8, it cannot be assumed that all native Tamil speaking Brahmans would be able to read and write Tamil in Tamil orthography. Further, with reference to VM and VL's comment, the presence of English words and to a sharp contrast, the presence of very few Tamil words as in the case of VM, were considerations that led to choice of script with higher priority accorded to Tamil orthography. The choice of according higher

priority to Tamil orthography was consistent with the linguistic view of the conceptual model. Sankar & Nagarajan (2012) have employed Roman script with specific reference to 'Tanglish' as a dialect, which according to the author was prevalent in Tamil Nadu. Ridge (2012) suggested that the Brahmans have largely shifted to English and retained Tamil only for cultural and religious reasons as dealt in Chapter 8.

CHAPTER 5 - EVALUATION

5.1 EVALUATION OF PROTOTYPE

Hendry et.al(2005) carried out a paper prototype evaluation by creating a paper model of the prospective system. The evaluation was done by using a scenario. A similar approach was followed by Rivero & Conte (2013). Vliet (2000) 'play a scenario' which was used to evaluate the paper prototype of speech to text application. The evaluation of paper prototype formed a part of the qualitative interview process. The following outlines the process followed in evaluation (**see appendix E4** for the words).

‹ Notes   **Done**

12 ஏப்ரல், 2016, 19:31

தொழிலாளி

ஆ ஈ ஊ ஏ ள ற ன ட ண ச ரு

அ இ உ ஐ எ க ப ம த ந ய

⇧ ஔ ஓ ஒ வ ங ல ர ழ ⌫

123 🌐 வெளி ◌் இடு

5.1 a : *ThozhilaLi* - shows an iOS Tamil keyboard and the word in Tamil orthography.

5.1 b: ThoLilaaLi - shows an iOS Tamil keyboard and the word in Tamil orthography. The difference between 5.1a and 5.1b is that this version shows the mispronounced version of Tamil syllable 'Zha'.

12 ஏப்ரல், 2016, 19:31

தொலிலாலி

5.1 c: Tholilaali - shows an iOS Tamil keyboard and the word in Tamil orthography. The difference between 5.1a and 5.1b is that this version shows the mispronounced version of Tamil syllable 'Zha'. The mispronounced syllable is different to that in 5.1a and 5.1b.

5.1 d : Tholiaali: Represented in Roman orthography.

5.1 e : Thozhilali[1]: Represented in Roman orthography. Mis spelling was intentional to get into finer details and the relationship between pronouncing syllable and its transliterated spelling.

---

[1] *Thozhilaali* was the transliteration spelling used in the dictionary. Mahadevan (2011).

The participants were presented words in Tamil orthography and in Roman orthography, in the interface similar to the above. The following table indicates how evaluation was carried out.

Criteria	What was observed?	Notes
Syllable pronunciation	Observed for accuracy of pronunciation.	Zha, la and La The usage of Zha la and La was not only observed when evaluating the prototype but was observed throughout the qualitative interview process.
Sandhi	Observed for representation in Roman orthography	Example: Thingakkazhamai / Thingakazhamai. (**See section 2.2**)
Syllable pronunciation	Observed to understand the expectation in orthography where mispronounced.	Example: Zha and la. Do the participants expect the right syllable to appear when mispronounced?
Code mixing and codeswitching	Observed to understand the expectation and choice of script when a different language is used.	**Refer discussion** What should be the ideal choice of orthography in speech to text in Tamil?

Table 5.1*: A table showing how evaluation was carried out*

The above figure summarises how the evaluation was carried out and how it informed in the prediction of the user acceptance of speech to text by the Tamil Brahman sample group. The user acceptance was predicted on the basis of accuracy of pronunciation, choice of script adopted by the participant, the spelling that appeared as the output text in conjunction with the pronunciation and code-switching. The aspect of pronunciation and code-switching and code mixing were observed throughout the interview.

5.2 REFLECTION: DESIGN AND EVALUATION

The design of the prototype took a linguistic view. Although, the evaluation of the paper prototype included the Roman orthography, in order to be coherent and consistent, Tamil orthography was accorded priority and responses of the study sample with regards to the prototype, was done in conjunction with the interview that was based on the user acceptance model proposed in this thesis in chapter 6. The design took into account, user -centred design (choice of script), consistency and simplicity of the application, feedback, individual user experiences and user experience levels (Hix & Hartson, 1993).

Since the design was based on the linguistic view, there was almost no scope for accommodating individual differences, especially when it came to pronunciation of words. The system to some extent has imposed ways of interaction with the application Holtzblatt & Beyer (1998). In doing so, it has classified the potential users on the basis of the ability to speak the language.

The ability to speak the language, coupled with the linguistic view of the design not only on one hand presumed the knowledge of Tamil orthography by the study sample, but on the other hand claimed legitimacy on the basis of the linguistic design of the application. The design employed was simple and consistent. The initial design of the application was done by the researcher as discussed in sections 4.5 & 2.4. While evaluating the user acceptance of the speech to text prototype, the study sample contributed to the enhancement. For example, translation of Non Tamil words and use of orthographies that represented the respective sounds. These represented the "mental models" of the users in the study sample. The possibility of embedding them into the application was in contradiction with the conceptual model of the researcher. In this specific case, the evaluation and design mutually complemented. The evaluation provided more insight to the user's "mental model" and opened avenues for exploring the possibilities as suggested by the users during the evaluation of the paper prototype. The design of the application was driven by the syntax and the semantics of the language. And as a result, the application was less user centric. This is indicative that the acceptance of the technology could also be influenced by the design view adopted by the designer.

5.3 ETHICS AND SOCIAL RESPONSIBILITY

Ghezzi et. al(2002) elicitate the ethics and social responsibility of software engineers. This research took the view that the design of the speech to text application in Tamil has to some extent fulfilled the social responsibility by taking a linguistic view. Though the design consideration impose the researcher's conceptual model, it was legitimate on the basis of 'quality and fairness from the perspective that any native Tamil speaker who accurately pronounced Tamil syllables and had the ability to read and write Tamil in Tamil orthography could use the system despite being specifically designed on the basis of Tamil speaking Brahmans for reasons discussed . A conscious decision of choosing language as a parameter for the application was made so all[2] users fluent in the target language are able to use the application as seen in section 4.6.3.

---

[2] All - Cannot be generalised. Study sample suggests that anyone with the knowledge of the language as per its syntax would be able to use it.

CHAPTER 6 - PROPOSED APPROACH

6.1 USER ACCEPTANCE MODEL

The following user acceptance model is proposed in this thesis in order to answer the research question.
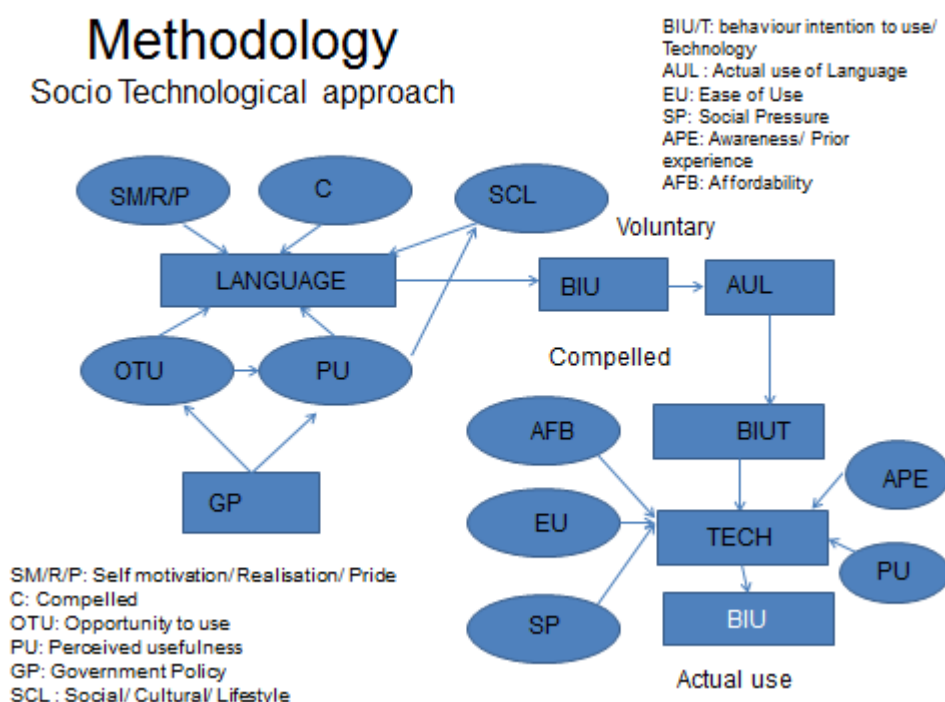
Speech to text user acceptance model

Figure 6.1: *Proposed user acceptance model to predict the user acceptance of speech to text application in Tamil*

In order to answer the research question, the user acceptance model as in figure 6.1 was proposed, which is one of the main contributions of this research and was based on the The Unified Theory of Acceptance and Use of Technology (UTAUT) model. The model consisted of two parts namely language and technology. In the language part of the model, it focussed on factors like the perceived usefulness of the language, actual use of language in the social sphere, government policy towards the language, opportunity to use the language to name a few. Whilst in the technology part, it focused on previous knowledge or experience with the technology, perceived usefulness of the technology, perceived ease of use of

technology. This acceptance model was proposed specifically for predicting the user acceptance of language based technology, with Tamil as the target language. It was assumed that this model could be applied to any linguistic community for predicting the user acceptance or defining the requirements of a language based technology such as speech to text. Most of the user acceptance models were quite strongly focussed on the technology and surrounded around the technical ability of the user.

Lu et.al (2009) have used the theory of planned behaviour (TPB) and the Technology Acceptance Model (TAM) to examine the Chinese users' acceptance of instant messaging products (IM). Pancer et.al (1992) indicated that examination of belief and underlying attitudes towards those beliefs were important to predict the attitude towards using a system. Mathieson, (1991) presented a comparison between the technology acceptance model and the theory of planned behaviour in predicting user intentions and concluded that there were very little on empirical grounds to suggest that one was better than the other. Whilst both took into account to certain extent the social influences and control issues, it was suggestive that it did not form the basis to evaluate technology acceptance for certain social groups. Brown et.al, (2002) in their work on 'Do I really have to? user acceptance of mandated technology' laid emphasis on the theory of planned behaviour but have argued that attitude, subjective norm and perceived behaviour control capture an individual's behaviour intention. (Sun & Zhang, 2006) cited Venkatesh's integrated model which was considered slightly more relevant in determining the user acceptance.

It was initially assumed that the Technology Acceptance Model (TAM) would be sufficient to predict the user acceptance of speech to text technology in Tamil. However, further review of literature introduced various other theories such as:

1. Theory of planned behaviour (Mathieson, 1991)
2. Theory of reasoned action (Madden et.al 1992)
3. Integrated model (Sun & Zhang, 2006)

The integrated model proposed in 2006 was put forward after a critical evaluation and comparison of TAM with TAM2 (Venkatesh and Davis 2002). The model further introduced the moderating factor that could potentially influence the user acceptance. The moderators considered in their model were:

a. Organisation factors,

b. Technology factors and

c. Individual factors (cultural background, experience, gender)

The user acceptance model proposed to predict the user acceptance of speech to text technology in Tamil was based on the integrated model. Although the integrated model has taken into account the individual factors and has indicated that it could potentially influence the user technology acceptance, it has given very little emphasis on factors like perceived usefulness of a language and behaviour intention to use the language in technology. It was suggestive that the technology acceptance model has always largely been focussed on technology itself. It was therefore important that certain factors like subjective norms, perceived usefulness, behaviour intention, be applied to the target language as well in which acceptance of technology was being considered. In doing so with the user acceptance model (see figure 4.1), this research has contributed towards further refinement of the user acceptance model for a more accurate prediction of user acceptance primarily on the basis of language ability of the users and secondarily on the basis of the user's experience with the technology as indicated by the study samples.

Esenou & Egbue (2014) suggested that the rate of technology adoption varied and was dependent on various parameters such as government policy, societal receptiveness and a number of other social factors. The authors further claimed that these social factors were quite often ignored in order for the product to win on the merits of technical superiority. The proposed model takes into account the cultural influences and attitude which could potentially influence the intention towards using a technology. In the case of Tamil society and with specific reference to the Brahmans, the work of Fuller et.al (2014), Kailsaspathy (1979), McDonough (1997), Ghafournia (2015),Schiffman (1998), Ridge (2012), Fuller & Narasimhan(2010), Lopez (2014), Asgharzadeh (2003), Martin (2017), Krishnaswamy (2015), Nagarajan (2012), Srinivasan (2013) compelled to consider societal factors that were either helpful or detrimental in using the language in daily lives. Milroy (2012) and Schiffman's (1998) work on the notion of standardisation of languages provided similar views in different social contexts. The model by considering language aspect has embedded Milroy (2012) and Schiffman (2012)'s work in the context of user acceptance of speech to text application. The model proposed to predict the user acceptance of speech to text technology was on the foundation of the language in which a speech to technology

application could be built. UTAUT forms the basis of developing a model that was used to predict the user acceptance of speech to text technology in Tamil which could be extended to other languages as well.

6.2 SAMPLING

There are various sampling techniques such as random sampling, stratified sampling Convenience and snowballing as a sampling technique was used to select potential participants for the qualitative interview and prototype testing. Convenience and purposive sampling technique was most suitable as the participants needed to be a Brahman and speak Tamil as their mother tongue in order to participate. Snowballing technique was quite common in qualitative studies. The snowballing technique together with convenience sampling did not threaten the privacy of individual. The snowballing technique also managed the bias in the recruitment of the participants. It was advantageous especially when the objective was to study and analyse rather than testing a hypothesis.

6.3 ETHICS

Ethical considerations were made throughout the research process. The studies and the interview complied with the University's ethics policy (SHU Ethics, 2018[3]) **.** All the participants in the study group were provided with an information sheet and the consent was obtained prior to the experiment and interview. It was ensured that they had enough time and opportunity to read and ask any questions they may have with regards to the interview or prototype testing**.** The informants were made aware that participating in the research was optional and voluntary and they had the option to withdraw from the study within the timescale specified in the information sheet**.** Reasonable care and attention were given to the requirements and adjustments of the participants**.**

6.4 PROTOTYPING

Towards the end of the interview, the paper prototype was tested with each participant. The participants were presented with the paper prototype and were asked to read aloud in Tamil,

---

[3] https://www.shu.ac.uk/research/ethics-integrity-and-practice

the name of the day from Sunday to Saturday. The researcher then presented the most appropriate spelling based on the pronunciation and allowed the participant to agree or disagree with the output text based on their pronunciation. The prototype was tested for the script (orthography), pronunciation and code switching.

6.4.1 PROTOTYPE

A paper prototype was developed that resembled an iPhone 4s/5s. The choice of iPhone 4s/5s was based on the researcher's experience and familiarity with the device. The evaluation was done with a paper prototype as it offered a lot of advantages such as:

- Less time consuming.
- Could be changed/ updated fairly quickly
- Inexpensive

The following were some of the limitations:

- The participants got a limited feel of what the 'actual' system would look like.
- It was manual and therefore there were scope for ambiguity especially when dealing with technologies like speech to text or speech recognition. But this was overcome by using a voice recorder, insider positionality in the context and language, which helped the researcher to revisit for more clarity and to 'play back' the recorded clip to the participant in case of any disagreement.

Screen shots of the researcher's phone was transformed into the paper prototype. Since the focus was on basic words such as the days of the week to look into issues like mispronunciation and choice of orthography, the 'message application' within the phone was chosen for the prototype.

6.5 QUANTITATIVE PRE-TEST AND PILOT

Initially, mixed method approach seemed to be the most appropriate method to answer the research question. But after going through the process of the quantitative survey, it was identified that since the core focus was on issues like pronunciation and code mixing, quantitative methodology did not enitrely serve the purpose. Factors such as 'treating everyone equally and fairly', ' Is a question likely to cause offense to certain participants?', 'Is the questionnaire inclusive?', ' Is it reasonable to expect the participant to complete the

questionnaire?' were considered while designing the questionnaire. The first version of the questionnaire had twenty five questions based on Cresswell (2003). For example, the following questions were taken into account:

1. Does the survey question measure some aspect of atleast one of the research questions?
2. Does the question provide information needed in conjunction with some other variable?
3. Will most respondents understand the questions and in the same way manner?
4. Will most respondents have the information to answer it?
5. Will most respondents have the information to answer it?
6. Is any other information needed to analyse the question?
7. Should this question be asked of all respondents or of a subset?

6.6 PRE-PILOT AND PILOT

The questionnaire was divided into two main parts. The division was however not made apparent to the participants but was done from a research and analysis perspective. The first part of the questionnaire captured data on the participant's language skills, education, occupation and age group. The latter part of the questionnaire captured data relating to the speech to text technology experience (if any), the awareness about the speech to text technology, the likeliness of the participant using voice to text in technology, choice of orthography- Tamil or Roman. The questionnaire had a mix of open ended , close ended and descriptive questions. The questionnaire was prepared in English and Tamil. Both the questionnaire had the same questions and in the same order to ensure that the participants were treated equally and fairly.

Evaluation and limitations of version 1

The first version of the pilot study questionnaire was not very successful in categorising the age group. For instance, the following is the question from version 1 of the pilot study questionnaire:

1.   Please circle the appropriate age group:
18-30   31-40   41-50   51-60   61-70   71-80   81-90   91-100

The question from the supervision team, *'Do you consider an 18 year old has a lifestyle and attitudes in common with a 30 year old?'* prompted to reorganise the age group in the subsequent version. It was realised that reorganising age group was more meaningful as some of the analysis in later stages especially in qualitative interview involved age group and lifestyle. It was then re-grouped as 18-25; 26-40 and 71-100 in version 2.

Question 6 of version 1 asked: "Which of the following language would you normally prefer to use in a social context. For example, whilst talking to a waiter?"

The question did not perhaps consider that while it may be within a social context for the participant answering the question, it was a business context for the waiter who served. Some of the suggestions from the supervision team were: *talking to a friend, talking with the members of the family and friends, talking to friends at a family event*.

Question 14: Is an iPhone your dream phone? Yes/ No. In either case, please give your reason.

It was felt that this question was intruding the participant's privacy although the intention to ask this question was to understand the affordability of the participant. One of the aspect that needed to be considered was the economic inequality which to a certain extent influenced the affordability of some of the gadgets with advanced technology such as smartphones. The other reason to ask this question was that the prototype was based on the iPhone and iOS because of the researcher's familiarity and experience with the smart phone. This question was deleted in version 2.

Question 22: generally speaking, if you want to use speech to text, how would you like the output text to appear?
The use of abbreviation confused the participant. In the second version, it was changed to speech to text technology.
Other observations

The first version was inconsistent in terms of numbering, and layout. Some of the other improvements that were made in version2 were regrouping certain questions, combining certain questions in a way that it reduced the effort of the participant and, at the same time, captured the information required to analyse.

Version2 of the pilot study questionnaire

   The English version of the second questionnaire were presented to the colleagues for their feedback. It was found that some of the questions were misinterpreted or had more than one answer. For example:

Which of the following is your mother tongue?

English   French Malay   Tamil

The suggestion received was to replace the question with '*Is Tamil your mother tongue?'* as the data was collected only from the native Tamil speakers. Based on anecdotal observations, some of the native Tamil speakers when migrated to a non Tamil speaking region, tend to adopt the local language to such an extent that it was sometimes spoken at home as the primary language. Some of the anedotal example regions include the Northern states of the Indian Union where Hindi is spoken, English speaking countries like the UK, Canada, Australia and the US, French speaking countries like France, Reunion. Over a period of time, 'Tamil' was retained only for the purpose of identity and origin but the language is not used in the capacity of mother tongue and this supported the views of Ridge (2012), Krishnaswamy (2015), Fuller & Narasimhan (2014).

What was the medium of instruction at school?

Tamil   Tamil mixed into English

English   Tamil mixed into French   Other (please specify)

It was felt that although one could choose 'English' as the medium of instruction in Tamil Nadu, some of the tutors depending upon their proficiency in the language, used 'Tamil' to explain a particular module. Therefore the participants were of the opinion that both Tamil and English 'qualified' as medium of instruction when there could only be one.

What is your most preferred choice of input?
Tamil script   English   Tamil in Roman script

In the next version, this question was modified to *'What is your most preferred choice of input on your mobile phone or smart phone?'*. The participants felt that the latter was very specific and related to the research question.

The second version did not include open ended questions although the first version at the end did provide an opportunity to the participants to share their experience, views and ideas for further improvements. The participants felt that it was important to ask open ended questions quite specifically that related to their experience with speech to text on a smart phone. Therefore the following question was included in the subsequent version:

'If you would like to make a change in the speech to text technology that you have already used. What would it be?'

The second version had a total of twenty two questions against twenty five question that the first version had. The participants suggested to capture their interest on whether they were interested to participate in qualitative research. The question

'If I were to invite you for a focus group session, would you be interested?' was added to the subsequent version.

Version 3

The third version was presented to two Sheffield based native Tamil speakers on a trial basis. A meeting room at the university's Science Park was used for this purpose. They were given a choice between Tamil and English and both chose Tamil as their 'preferred' language. They were given the information sheet in Tamil that explained the research background, the research question, who should participate, what should the participant do should they decide to participate, benefits of participating along with the researcher's contact details for any questions or clarifications. The fourth and the final version also included a question on gender to ensure there was no gender bias.

Participant 1

It was found that the participant was reasonably confused with questions relating to the usage of language.

For instance, to a question that asked:

'Please choose the language that you speak at home'	
Tamil	Tamil mixed into English
English	Tamil mixed into French

The participant initially chose 'Tamil'. The idea behind this question was to understand the effect of 'code mixing' when speaking the language. Since this was a pilot study, the researcher chose to intervene whilst progressing with the questionnaire so the participant could give the most suitable response and at the same time, it provided an opportunity to understand the participant's perception more accurately. Although the participant admitted that they spoke Tamil mixed into English, the participant's choice of 'Tamil' in the first instance despite providing with the most suitable option suggested that Tamil with the effect of 'code -switching' was perceived to be a part of the language itself for the individual answering the questionnaire or it suggested that the participants has perhaps not read the other options! Further probing showed that the participant 'assumed' certain answers from the nature of the questions and the nature of the research study.

"*I speak Tamil with my friends so I chose Tamil. But it is understood that Tamil means mixing English words as nobody speaks 'pure' Tamil these days"* (**Translated from Tamil)** suggested that the participant's perception towards Tamil as a language implicitly includes English words which further indicated that code mixing and code switching were important aspects that needed to be considered whilst developing a speech to text application for Tamil language. It was found that the participant who had been living in the UK considered 'Tamil' as his first language.

It was found that the participant has had English medium education throughout. The participant preferred 'English' as the choice of input but had not used speech to text technology as the participant didn't feel the need to use it. But it was interesting to observe that questions relating to Tamil and its usage, the participant had chosen the Tamil option. However, the comments at the end brought to surface, areas of focus with regards to its acceptance.

"The questionnaire was really good. After a lot of years I have spoken and written in Tamil and that too in 'proper Tamil' I wish such kind of questionnaire gain wide acceptance in future!"

The pilot of the quantitative survey was conducted at the Singapore conference (Ramachandran et. al 2015). Most of the participants were delegates at the International Tamil Internet Conference and therefore, the participants had a natural inclination towards the language and were found to respond in a more favourable manner to the questionnaire when the objective was to capture natural responses.

For the purpose of data collection and analysis, the geographical locations was initially divided into three based on 'nativity', 'official status' of Tamil language and the presence of Tamil diaspora in regions where the research was being carried out:

Group A: Singapore and Malaysia Group B: Tamil Nadu and Sri Lanka Group C: United Kingdom (Diaspora)

Majority of the Tamil population reside in Tamil Nadu followed by Sri Lanka, Malaysia and Singapore. Although Tamil has no official status in the region mentioned in Group C, a significant Tamil population have made this region their 'home' over a period of time, most of whom are believed to be of Lankan origin.

In the initial stages of the research,   the plan for data collection included 'voice' samples of native Tamil speakers from these regions- that could capture aspects of pronunciation under investigation, in addition to online questionnaire and qualitative interview.   It was then decided to follow the mixed method approach. The quantitative data from Group A, B and C was to be collected through an online questionnaire. Predicting the user acceptance required to understand the behaviour intention of using the language in the technology and therefore it was important to capture the evidence of 'behaviour intention' of using the language on technological devices which the paper questionnaire did not. This could be observed from the choice of language by the participants in certain questions in the questionnaire. Therefore, it was felt that an online questionnaire would perhaps be the most suited method to understand and interpret the behaviour intention of using the language in the technology as also reccomended by Turner et.al (2010). Vallejo et.al (2007) observed that the questionnaire , whether online or paper based provided similar results.

The questionnaire was divided into two main parts. The division was however not made apparent to the participants but was done from a research and analysis perspective. The first part of the questionnaire captured data on the participant's language skills, education, occupation and age group. The latter part of the questionnaire captured data relating to the speech to text technology experience (if any), the awareness about the speech to text technology, the likeliness of the participant using speech to text in technology, choice of script- Tamil script or Roman script in Tamil. The questionnaire had a mix of open ended, close ended and descriptive questions. The questionnaire was prepared in English and Tamil. Both the questionnaire had the same questions and in the same order to ensure that the participants were treated equally and fairly.

Before the pilot study was conducted in Singapore, it was decided to conduct a pre pilot study at Sheffield with a minimum number of people for maximum feedback. The pre pilot study was conducted in Sheffield with seven people. Of which six people were native Tamil speakers. The questionnaire was updated depending upon the participant's response, feedback and comments and went through four iterations before conducting the pilot study in Singapore.   The pilot study was conducted at the venue that hosted the 14th International Tamil Internet Conference. The pilot study questionnaire was paper based. At one point it was noticed that the participants who decided to voluntarily take part in the pilot study shared similar views, perhaps influenced by the environment. Almost all participants who took part were very excited and readily accepted to take part in quantitative and qualitative research. Taking into consideration the nature of the event, it was decided not to proceed with the quantitative survey as the analysis of responses could result in a biased finding and anecdotal observations confirmed the same. There was little co-relation between the responses on survey and the acutal behaviour (such as in using the language) and was consistent with Ajzen's (2011) observation on theory of planned behaviour.

6.7 OBSERVATIONS

It was observed that participants were conscious about the nature of the research and therefore in many places, the participants were found responding to a certain question contrary to their practise which according to the researcher's perception could influence the findings and result. Therefore, whilst analysis, it was felt that the literature could serve as a guide to determine if the participant natural responses to the questions based on a number of

observation such as their social background, environment, aspirations, anecdotal evidence of speaking the language as responded in questionnaire and so on.

6.8 PRE-PILOT AND PILOT FINDINGS

Most of the participants were reasonably confused with the word 'Tamil'. For some Tamil was merely an identity and did not serve as a 'fully functional' language which supported Brian (2012) view.

To most of them in the study sample, the script really did not matter – Tamil in Roman script worked perfectly fine. Participants especially amongst the diaspora Tamil in the UK considered Tamil in Roman script as Tamil. The ability to read and write in Tamil amongst the participants varied to a great extent from as little as zero percent to as high as one hundred percent. The pre pilot qualitative study suggested that the diaspora Tamils in the UK as seen in the sample tend to learn Tamil as 'language' to preserve their identity and heritage but that did not in most cases translate to the use of language in the technological devices. It was found that there was a very strong social element and an economic element that motivated a native Tamil speaker to switch to other languages primarily to English. The participants were unable to distinguish between 'code mixed Tamil' and Tamil with no effect of code mixing. Most participant felt that they were quite capable of speaking the language. They responded that 'code- switching and code -mixing' was natural and therefore it suggested that the code mixed Tamil was considered as 'proper Tamil' although anecdotal observations suggested that there were groups that strongly oppose code -switching and code-mixing of not just English but also Sanskrit in order to maintain language purity. The work of Kailasapathy (1979), was consistent with the contemporary Tamil society in this specific case because the opposition to code mixing was more often for political reasons than a genuine attempt to approach the language. The script complexity was another issue that needed to be carefully dealt with. Participants were comfortable in Tamil but preferred English or Romanised Tamil to communicate in hand held devices. Some admitted that the reason for choosing English or Romanised Tamil was their lack of proficiency in speaking proper Tamil- without the effect of code -switching and code-mixing. It was also found that proficiency in language does not necessarily lead to acceptance of the language in technology. For instance, many participants responded that Tamil was their most preferred

language but in practice, they used English despite the availability of resources such as Tamil keyboard . A few commented on affordability of gadgets that facilitate Tamil typing and suggested that lack of will and self- motivation was perhaps the prime factors in using the language in technology. The link between society, education, economic value of the language, opportunity to use the language in a social sphere and technology emerged as themes for focus in the context of prediction of the user acceptance of speech to text . It was interesting to find that in the quantitative analysis, an overwhelming majority of the respondents felt that both pronunciation and the ability to speak Tamil without the effect of code mixing was important to use such applications. The qualitative interview confirmed the same and the issue of mispronunciation and code- switching and code-mixing was seemingly apparent. The disconnect between Tamil language and technology was visible to some extent and the reasons identified were: economic value and scope of the language, society's attitude, educational system and politics.

6.9 DATA COLLECTION

The main data collection was done in Chennai between July 2016 and September 2016. The data collection was qualitative interview and adhered to the Sheffield Hallam University ethics policy. The quantitative questionnaire was similar to the one in the pilot study **(see appendix E.1**) but Lonsdale, Hodge (2006) suggested that there was not a significant difference between an online and paper based. The online questionnaire could actually show the evidence of 'behaviour intention' to use the language if the participants responded their preferred language was Tamil. Some of the key issues such as behaviour intention to use the language, mispronunciation, code mixing and the script were dealt with at the qualitative interview.   The findings of the qualitative analysis during the pilot were consistent with the previous findings which confirmed that the issues identified in the beginning were genuine and could potentially influence the user acceptance of speech to text application in Tamil. It was at this stage of the research, more specifically while dealing with the issue of pronunciation, the interaction in various districts responded fairly similarly on the Brahman Tamil. Similar comments to the following were received:

"Nowadays its only the Brahmins who are able to pronounce the syllables ழ,ல, ள properly."

The above comment was undisputed and led into the Vedic philosophy and Brahman Tamil as previously discussed. The choice of prototype design to suit the Tamil Brahmans also arose from the findings and observation at the qualitative interview. It indicated a relationship between social, cultural, lifestyle (SCL) and the language.

6.9.1 QUALITATIVE INTERVIEWS : TAMIL BRAHMANS

The research eventually focused on *Manipravalam* (Sanskritised Tamil) – a Tamil community that speaks a 'code switched' language and researches suggested that they have a fairly accurate pronunciation especially the syllables unique to the Tamil language which was also central to this research. Recruiting participants required the disclosure of their identity and their explicit identification as 'Brahmans' which was politically and socially a sensitive subject in Tamil Nadu and therefore, a convenient, purposive and snowballing technique was followed for selecting the participants for the qualitative interviews. Sampling technique such as the random sampling as suggested by Munn & Drever (1990) was not feasible in this particular case because of geographic, social and political constraints. This implied that the findings and analysis cannot be generalised and extended to the whole Brahman community who speak Tamil as their mother tongue. It is indicative and would aid in pursuit of accurate prediction of the acceptance of this application within this community.

In order to answer the research question, the interview consisted of two parts:

1. Interviewing the participant around language, lifestyle and experience with technology based on the above methodology.
2. Testing the paper prototype.

6.9.2 ICEBREAKER QUESTIONS INCLUDED THE FOLLOWING

1. Why do people use a phone/ smart phone?
2. What is your opinion on studying Tamil as a compulsory language that was recently introduced by the Government of Tamil Nadu?
3. If you were a software developer owning a software company, in what language would you develop an application and why?
4. When I utter the word 'technology', to which language do you immediately associate it with?

5. How would you sell Microsoft a Tamil speech to text application?

6. How would you motivate/ persuade someone to use technology in Tamil?

7. Should technology cater to the ways in which people speak a language or the way a language is spoken?

8. However not all questions were asked to all the participants. But all the questions were asked at the pilot to identify any potential pitfalls.

Not all ice breaker questions were asked to all the participants. The ice breaker questions were asked based on participants' interest and knowledge and the researcher has used insider positionality to make a decision on what icebreaker questions would suit a participant, which could also contribute towards the larger research question.

6.9.3 MAPPING QUESTIONS TO USER ACCEPTANCE MODEL

The following questions were asked to the participants. The rationale behind asking the question is shown the bracket and is mapped to the method adopted (**Refer: figure 4.1**) in answering the research question.

1. Can you take me through an experience where you've used Tamil in technology *(OUT, GP, SM, R, P)*

2. Being a Tamil and a Brahmin, which of the two languages have you traditionally viewed important and why- Tamil and Sanskrit *(SCL, P)*

3. Tell me a time when you were compelled to use a technology in a particular language and how did you feel about it? *(C, AUL, APE)*

4. If the participant doesn't feel English as a 'compelled' language in a 'native' space then - *Between Tamil and English, in which language are you confident of expressing yourself as well as carrying out daily activities? (OTU, PU, SCL)*

5. Proficiency in a language is an indication that the technology in that language was more likely to be accepted by its native speakers. What is your view on this?

6. Do you think you have embraced Tamil as a language and a natural choice for using technological devices like smart phones? (If no)

    Take me through your experiences and why do you think so.

7. Please tell me about your journey or transition from using a basic phone to a smart phone. *(APE, AFB)*

8. Have you heard about speech to text technology? *(APE)*

9. Have you ever used a speech to text application in any language? *(BIT, PU)*

10. Could you think of a time when you used a technology because you were influenced by someone using that particular technology? (Any example) (*SP, C*)

11. What does owing an iPhone or an expensive smartphone mean to you? Passion for technology/ social status/ other reasons *(BIT/AFB/ AUL/ Actual use of technology)*

12. Have you ever used Siri? (looked for consistency, repetition of question 6 and 7)

13. Have you ever used any application in Tamil? (looked for consistency. Repetition of question 1)

14. Between Tamil and English, which language would you accord more priority over other in Tamil Nadu and why? (*R/P/C/PU/ SCL/AUL/ SP*)

6.9.4 PROTOTYPE TESTING

As a part of the interview and data collection, the participants tested a paper prototype of the speech to text application in Tamil. Mispronunciation, code switching and the choice of script were the areas that were identified as critical factors that could influence the user acceptance of this technology. In order to closely study its effect, the prototype was tested on the days of the week which consisted of the frequently mispronounced Tamil syllable 'zha' and 'La'.

Method:

The participants were asked to read the day aloud and the researcher presented the most appropriate spelling on the paper prototype screen. The participants were then asked questions around their experience. Some of the questions asked are listed below:

What do you think about speech to text in Tamil? Would you use it?

Would you prefer the output in Tamil script or Roman script and why?

How could this application be further improved?

What did you like the most about speech to text in Tamil and what did you dislike the most?

Role of the researcher: Interviewer and moderator

Paper prototype was based on iOS and Apple iPhone 4s.

Plan

Description of activity	Time allotted	Notes
Revisiting information sheet/ ethics/ consent	15 minutes	Pre session
Ice breaker	05 minutes	Data
Interview	40 minutes	Data
Prototype testing	20 minutes	Data
Feedback and wrap up	10 minutes	For further improvements Audio/ video recording was stopped after this.

Table 6.1: Table that shows the plan of the interview.

The above table summarises the plan of the interview. The participants were made aware of the tentative schedule and the researcher ensured a timely finish, to the best possible and practicable extent.

Setting:

Choosing the right setting for interview was important for the interviewer as well as for the interviewee. The main purpose of choosing a setting was to make sure that the participant was comfortable and relaxed and was able to provided the best possible response to the questions asked.

6.10 SUMMARY

This chapter has briefly discussed the software engineering models, and has clarified the technicality of the paper prototype that was designed, developed and used to answer the research question. This chapter has also explicitly identified the customer and stakeholder from the software engineering point of view. It has provided a rationale for the design and reflected on the design process and evaluation of the prototype. It further clarified the conceptual view adopted in the design and how this may influence the outcome of the finding.

It has introduced and discussed various research philosophies and identified the one that suits the most to this research. It has also discussed phenomenology as a research method and compared and contrasted with the grounded theory. Based on this chapter's discussion,

this research took an inductive approach underpinned with the interpretivism research philosophy and phenomenology as a method for data collection. Although pragmatism and realism could have been the philosophical basis, the extremely small data sample and in-depth investigation makes interpretivism a more convincing choice of research philosophy.

CHAPTER 7- STUDIES AND FINDINGS

INTRODUCTION

This chapter sets focus on the studies and some its key findings. The findings of transliteration exercise, focus group and 'zha' pronunciation led to the qualitative interviews. One of the key issue and focus that was under investigation in this research was 'script complexity'. The script complexity in this case, referred to the choice of script by the user. The two main scripts that were in popular usage amongst the native Tamil speakers were the Tamil orthography and the Roman orthography.

The focus group that involved native Tamil speakers and Malayalam speakers led to the In- depth qualitative interviews of the Tamil Brahmans who, as a result of the observations, research study and anecdotal evidence indicated to satisfy two of the main areas of investigation – accuracy of pronunciation and code-switching and code- switching and code-mixing[4] of language.

7.1 STUDY 1

THE TRANSLITERATION EXPERIMENT

Objective: To study the consistency of using the Roman orthography in Tamil and to evaluate the feasibility of using Roman orthography in Tamil speech to text application.

Subjects: It involved a total of 177 samples from the age group 18-25.

Location: One location each in Northern, Western, Central and Southern Tamil Nadu**.**

Age : 18- 25 years

Material: Pen and paper.

Sampling: Convenient and purposive

Design: The researcher dictated thirty words and the participants were expected to '*write as they heard*' in both Tamil and Roman orthography. This was slightly different to the method followed by Ahmed et.al (2011) however, this research shares similarity in the objective of

---

[4] Linguistic study on Manipralavam by K Retinamma as indicated by Bright(1978) suggested that code mixing has been in existence and usage in the modern Southern India at least for the last 1000 years. The Tamil speaking Brahmans speak Tamil with heavy borrowings from Sanskrit.

the study in looking into the orthography. They were not made aware of the mispronunciation of the word. They were also not told of the possibilities of mispronouncing a word. This was because, the researcher, as a result of insider position wanted to avoid possible manipulation of the word. As a result of observations made in the pilot, the design adopted convincingly yielded a more naturalistic response which could be constituted as a genuine and trustworthy data for the purpose of this research.

Limitation: The participants were from a mixed socio-religious background. Although, standard design process was followed in all the regions, the pace of dictation varied owing to local factors such as allocation of time and individual commitments.

Role of the researcher: Dictating the words.

Positionality: Insider as a native Tamil speaker.

Ethics: Participants signed an informed consent and were advised that participating was voluntary and optional. The information sheet was provided in Tamil and English and can be seen in the appendix. Local ethical and cultural conduct was observed. For instance, gender segregation, planning seating arrangements that accommodated cultural considerations such as allowing women to leave early, enabling them to reach home before sunset.

Rationale: The experiment was designed by the researcher in order to observe the relationship between pronunciation, and accuracy of its appearance in Tamil script and Roman script and the complexities involved with the latter. Since the researcher's conceptualisation of a Tamil speech to text centered around Tamil orthography, this experiment provided some insight towards researcher's own perception towards using Roman orthography for Tamil in a speech to text application. Quite a few Asian languages like Uzbek, Malay, Bhasa Indonesia use Roman script and transliteration, was not a new phenomena in the Indian sub-continent either. Kurzon (2010) explained various attempts and reasons for Romanisation of Bengali and other Indian scripts before the Indian Independence. The author explained how culture and scripts that have been in existence and use for thousands of years would counter attempts of changing the scripts. But interestingly, the language based computing research in Tamil have used Roman script to test their application (Sankar & Nagarajan, 2012), Geetha et.al (2003), Sridhar et.al (2013).

Anecdotal evidences further suggest that, code- switching and code mixing of English into Tamil (**appendix A.2**) makes Roman script a popular choice amongst the native Tamil speakers- Brahmans[5] and Non Brahmans[6] alike. Further, an assumption was made that the code-switching element could possibly favour the Roman script as it is convenient to switch between the two languages – English and Tamil. This assumption was made in the absence of a speech to text application in Tamil and in a scenario where the user used the keyboard as the medium of input.

Method : The researcher adopted the 'dictation method' and chose thirty Tamil words (**see appendix**) for the purpose of the study. The words were chosen with focus on Tamil syllables under study in the context of speech to text . A few words were deliberately mispronounced and the participants were not made aware of the possible mispronunciation prior to the start of the exercise. However, the researcher clearly asked the participants to "Write as they heard!" This was done in order to capture the accuracy of the words in the written form and also to observe the usage of Tamil syllables (la, La and zha) in the Roman script, the importance of which was indicated by Srinivasan (2013) and Geetha et.al (2003). The words in Roman orthography was manually entered onto SPSS for statistics. Since it was not possible to enter Tamil script into SPSS for statistics, it was done manually on excel as seen in table 7.2

Tamil orthography	Roman orthography
அலை	Alai
முகநூல்	Muganool
அலை (அழை)	Alai (Azhai)
அளைப்பிதல் (அழைப்பிதழ்)	ALaippithal (azhaippithazh)
விளுப்புரம் (விழுப்புரம்)	ViLuppuram (Vizhuppuram)
கல்லு	Kallu
கள்ளு	KaLLu
பிளை (பிழை)	PiLai (pizhai)

---

[5] Kumar (2016) in a book review titled From Hagiography to biography: Ramanuja in Tradition and History points out that Ramanuja's spiritual work was key in the emergence of *Manipravalam* – in which Ramanuja intersperse Tamil words with Sanskrit, which the author claimed to be one of the first. The Brahmans till date largely retain the *Manipravalam* genre in the day to day communication which is significantly different to Standard Tamil and the form of Tamil dialects spoken by Non Brahmans.

[6] The term Non Brahman is used to denote native Tamil speakers who aren't Brahmans which was essential from a finding perspective as the research, as a result of pilot, primarily focussed on the Tamil speaking Brahmans.

வாலு	Valu
வால்க்கை (வாழ்க்கை)	vaalkkai (vaazhkkai)
வன்னம் (வண்ணம்)	vannam (vaNNam)
விலக்கு (விளக்கு)	vilakku (viLakku)
பலனி (பழனி)	Palani (Pazhani)
மாட்டு	Maattu
மீட்டு	Meettu
காலை	Kaalai
தொழிலாளி	ThozhilaaLi
கலை (களை)	Kalai (kaLai)
முட்டு	Muttu
கிளை	KiLai
குட்டு	Kuttu
கிளி (கிழி)	KiLi (Kizhi)
மேட்டு	Maettu
தொகுப்பலார் (தொகுப்பாளர்)	Thoguppaalar (ThoguppaaLar)
எண்ணை	ENNai
ஏற்றுக்கொள்	EttrukkoL
எளுதுகோள் (எழுதுகோல்)	ELuthugoL (Ezhuthugol)
மனப்பான்மை	Manappanmai

Table 7.1 *: List of words used in dictation.*

The above table is a list of words used by the researcher in the dictation experiment. The words in the bracket indicates the correct pronunciation and spelling. And those words were mispronounced by the researcher.

Findings

All the participants were familiar with writing Tamil in Roman script which they referred to as '*Tanglish*' – a colloquial name for Tamil mixed into English or vice versa and written using the Roman script. Almost all the participants, during the course of interaction revealed that they used Roman script to represent Tamil in the context of technology for example while sending a text message. The Tamil script was almost one hundred percent accurate (excluding the ones that were misheard, not legible and not attempted) and consistent as seen in figure 7.3.   The confusion on mispronounced words was apparent when

the participants requested the researcher to repeat the words (that were mispronounced) and looked perplexed as they attempted to write in Roman script! It was also observed that the mispronounced words were written with the right spellings in Tamil despite asking them to *'write as they heard'*. This was indicative of their expectation of appearing a word right in spite mispronunciation[7]. This was contrary to the fundamental design of the application as envisaged by the researcher. Whilst this suggested incorporating an 'auto correct' feature to do away with spelling errors but it potentially undermines the role of pronunciation[8] which is given paramount importance in most of the Indic languages, Tamil in particular. Sone words for example *kallu, kaLLu* were written in colloquial form in Tamil. Since the objective was to look into the consistency of spelling in orthography, they were also taken into account. In having done so, it brought to the surface, that colloquial speech did change the spelling and orthography and that it was not a rationale for accepting mispronunciation on the grounds of colloquial speech. **Figure 7.1 and table 7.2** show that there is a significant difference in the consistency of spelling in Roman orthography for the same word. It suggested a relationship between pronunciation, accuracy of pronunciation and Tamil orthography. The word Alai (Azhai) in **table 7.2** in the context of relationship of pronunciation and orthography could be regarded as an exemption. It suggested incorporation of auto correct feature in speech to text application and that the participants have used their discretion in writing the word. Despite mispronunciation, the word was written correctly in Tamil. However, the interaction with the participants showed that they exhibited the pronunciation of the transliterated version in Roman orthography and not what appeared in Tamil orthography.

---

[7] The mispronunciation refers to inaccuracies of pronouncing a syllable. In the wider context of this research, it needs to be interpreted in the social context that pertains to attitude of a community that is either reluctant to accurately pronounce or justifies mispronunciation on the basis of dialect- both of which cannot be accommodated within the language.

[8] Kurzon (2010) recognized the importance of accuracy of pronunciation of the Vedic texts which are mastered and chanted only by the Brahmans. Although, it falls within the purview of socio-religious studies and is beyond the scope of this research, it brings to light, the ability to accurately pronounce syllables at a socio-religious level. And facilitates a logical argument as to why pronunciation should not be a requirement to use a speech to text application.

1	அலை	(முகநூல்	அலை	(அழை)
2	அலை	(முகநூல்	அழை	
3	அலை	முகநூல்	அழை	
4	அலை	முகநூல்	அழை	
5	அலை	முகநூல்	அழை	
6	அலை	முகநூல்	அழை	
7	அலை	முகநூல்	அழை	
8	அலை	முகநூல்	அழை	
9	அலை	முகநூல்	அழை	
10	அலை	முகநூல்	அழை	
11	அலை	முகநூல்	அழை	
12	அலை	முகநூல்	அழை	
13	அலை	முகநூல்	அழை	
14	அலை	முகநூல்	அழை	
15	அலை	முகநூல்	அழை	
16	அலை	முகநூல்	அழை	
17	அலை	முகநூல்	அழை	
18	அலை	முகநூல்	அழை	
19	அலை	முகநூல்	அழை	
20	அலை	முகநூல்	அழை	
21	அலை	முகநூல்	அழை	
22	அலை	முகநூல்	அழை	
23	அலை	முகநூல்	அழை	
24	அலை	முகநூல்	அழை	
25	அலை	முகநூல்	அழை	

Table 7.2: *The words Alai, Muganool and Azhai hundred percent consistent in Tamil orthography.*

A	B	C		D		E		F	G
அலை	முகநூல்	அலை	(அழை)	அளைப்பிதல் (அழைப்பிதழ்)		விளுப்புரம் (விழுப்புரம்)		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு
அலை	முகநூல்	அழை		அழைப்பிதழ்		விழுப்புரம்		கல்லு	கள்ளு

Table 7.3 : *A few other words that shows consistency in Tamil orthography and mirrors the pronunciation.*

Statistics

Alai

N	Valid	177
	Missing	0

Alai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aalai	2	1.1	1.1	1.1
	aali	2	1.1	1.1	2.3
	alai	125	70.6	70.6	72.9
	alaiy	1	.6	.6	73.4
	alay	1	.6	.6	74.0
	ali	2	1.1	1.1	75.1
	alie	2	1.1	1.1	76.3
	allai	22	12.4	12.4	88.7
	alli	2	1.1	1.1	89.8
	azalai	1	.6	.6	90.4
	azhai	5	2.8	2.8	93.2
	azhlai	1	.6	.6	93.8
	MH	2	1.1	1.1	94.9
	NA	9	5.1	5.1	100.0
	Total	177	100.0	100.0	

Figure 7.1 : *Alai when written by the same set of participants in Roman orthograph*y

However, when it came to sentences, significant differences were observed. The participants were given two sentences- one had English words in it and the other did not. The rationale was to observe how the users translate the sentences in words. Do they use Roman script or Tamil script or a mix of both? The participants were advised to use their discretion and the choice of script to represent either the whole sentence, or a part of the sentence in Tamil or Roman were left to the participants.

The first sentence was tricky and the results were quite mixed. Most of them resorted to Roman script, a few used Tamil script (even for the English words) while a few used Roman script for English words and Tamil script for Tamil words. From a wider perspective, it was suggestive of the complexities involved in the choice of script as the negotiation of the script was apparent. The second sentence that had only Tamil words were written using Tamil

script and the choice was consistent among all the participants. A minority of the participants had written in both Roman and Tamil scripts. It further pointed that the choice of script to some extent was dependent on what was being spoken and how was it being spoken. The transliteration indicated an impact on the pronunciation which was seemingly apparent during the interaction with the participants. The syllables ழ(Zha) ல (la) ள (La), were the ones that reflected the transliterated pronunciation. These syllables were transliterated as *Pilai (*Pizhai), *Alai* (Azhai) to cite a few examples (**see appendix E.4 for all words**). Of the two words *Alai (*Waves) and *Azhai (*Invite or call), the latter was mispronounced by the researcher. The participants were perplexed because according to them, the word was already dictated. A few wanted to know the context of the word so they could arrive at the right spelling to which, the researcher declined to oblige since the objective was to look into the impact of mispronunciation on the output text and the user's perception and expectation. The aim was not to get the spelling right or to incorporate the 'spell check' or 'auto correct' feature which seemed to be the expectation of the participants. From the **figure 7.1,** if one were to look into the spelling, both are identical when mispronounced. But the spelling in the bracket indicates the correct spelling. Therefore, it was suggestive that a person intending to pronounce a syllable must get it right for it to appear in the text. The participants themselves were found to be slightly lost when words containing similar syllables were mispronounced[9]. When a participant, at random (not in a particular order), was asked to read the transliterated version in the order in which they were dictated, they were confused and had to resort to the Tamil script for the pronunciation of the respective syllable. This provided some light, on the importance of pronunciation and a reason why the choice of orthography needed to be considered in the context of speech to text. It also indicated that the pronunciation and orthography were mutually related.

The focus of this research was on commonly mispronounced syllables like ழ(Zha) ல (la) ள (La). However, this study brought to surface, combination of certain other syllables which needs to be investigated in the context of speech to text and was beyond the scope of this

---

[9] அழை - Azhai – Alai; அலை -Alai- Alai, காலை - kaLai- KaLai; களை -kalai- kalai, கல்லு - Kallu- Kallu; கள்ளு- kaLLu- Kallu.

research. Some of the syllables identified for further study are: ர (ra) ற[10] (Ra), ந(na), ன(na) ண[11] (Na) as also identified by Geetha et.al (2003)

7.2 STUDY 2

TAMIL SPEAKERS VS MALAYALAM SPEAKERS – A COMPARATIVE STUDY

Rationale: To observe the general usage of language, aspects of code-switching and attitude towards pronunciation in respective languages.

Location: One city in southern Tamil Nadu

Age : 18- 25 years

Material: voice recorder.

Design: The researcher conducted a semi-structured focus group on the basis of Marrelli (2008) focussing entirely on language, usage of language in social sphere, pronunciation and attitude towards pronunciation in the respective language.

Limitation: The participants were from a mixed socio-religious background. All participants in the focus group were male which contributed to gender bias.

Role of the researcher: Moderator

Positionality: Insider as a native Tamil speaker, Outsider in the context of Malayalam and Insider from the perspective of a former resident of Thiruvananthapuram- the capital of the south western state of Kerala.

Sample size: 20

Sampling: Convenient and purposive

Ethics: Participants signed an informed consent and were advised that participating was voluntary and optional. The information sheet was provided in Tamil and English (**see appendix**). Local ethical and cultural conduct was observed. For instance, due the allocation

---

[10] சொற்கள் –SoRkaL- Words;
[11] பண்ணு – paNNu- Do

of resources, it was practicable only to involve male participants. The researcher arranged light refreshment as a courtesy.

Rationale: The findings of the previous study coupled with researcher's insider positionality and work of Schiffman (2002), Schiffman (1998), Krishnaswamy (2015), McDonough et.al (1997), Kurian (2016) was the basis of a comparative study on pronunciation and mispronunciation of two social groups – the native Tamil speakers and the native speakers of the Malayalam language. Malayalam is a Dravidian language derived from Old Tamil and Sanskrit. One of the main reasons to choose native Malayalam speakers is the presence of the 'Zha' 'la' and 'La' syllables in the language which was under study. Interestingly, the literature on speech recognition of Malayalam does not raise the issue of pronunciation (Kurian, 2016). The researcher's primary observation as a result of the positionality, discussed in 2.4, in the south western Indian state of Kerala was used as an anecdotal evidence in conjunction with the focus group to suggest that the issue of pronunciation by a section of the native Tamil speakers (as also seen in the focus group sample). However, it is in this regard, the speech to text technology could be used as a tool to improve one's pronunciation in order to use the application.

Findings: The participants were native Tamil and Malayalam speakers and all of them knew to read write and speak Tamil and Malayalam respectively. The Tamil participants defended the mispronunciation on the basis on 'dialect'. For instance, it was 'okay' for them say *Pilai (*Pizhai) mispronouncing the 'zha' syllable (also as observed in transliteration exercise). It was suggestive that mispronunciation do not account for inaccuracies in contextual understanding and that, the syllable was 'fixed and understood' within the context. This was in sharp contrast with the native Malayalam speakers who accorded utmost importance to pronunciation. The native Malayalam speakers were not only able to correctly and accurately pronounce the syllables ழ(Zha) ல (la) ள (La) but also were able to speak Malayalam without the effect of code switching to a larger extent than their Tamil counterparts in the focus group. Overall, the code switching between English and respective mother tongue remained a challenge but, the frequency of code switching was observed to be less amongst the native Malayalam speakers in the focus group. The participants in the focus group whose first language was Malayalam did lay emphasis on pronunciation at a social level:

"Every time I used to mispronounce, someone used to correct me."

The position on pronunciation of the native Malayalam speakers was comparable to the Tamil speaking Brahmans who emphasized on correct and accurate pronunciation and is dealt in **sections 2.4 and 9.4**. Jayan et.al (2011) indicated that the Malayalam language was brought to Kerala by the Brahmans. The Malayalam speakers in the focus group to a large extent were found to be less tolerant on mispronunciations of Malayalam syllables by native Malayalam speakers. On the contrary, the native Tamil speakers in the focus group were okay with syllables being mispronounced despite the awareness of the correct pronunciation.   It was suggestive that the social attitude towards the language formed unwritten convention of how the language ought to be spoken. The interpretation of deliberate mispronunciation of a syllable on the basis of dialect could possibly be dismissed in favour of preserving and promoting the language in its original form in technology.

7.3 OBSERVING THE PRONUNCIATION OF 'ZHA'

Rationale: To observe the accuracy of 'Zha' pronunciation

Sample size: 30

Sampling: Convenient and purposive

Location: One city in western Tamil Nadu

Age : 18- 25 years

Material: Voice recorder

Design: The researcher asked the participants to repeat

"Vyaazhakkizhamai, yezhaikkizhavan vaazhappazhath thol vazhukki, keezhae vizhundhaan"

Limitation: The participants were from a mixed socio-religious background with one Tamil Brahman in the group. Since the recording was anonymised, it was not possible to identify the accuracy of pronunciation of the Tamil Brahman participant.

Role of the researcher: Moderator

Positionality: Insider as a native Tamil speaker.

Ethics: Participants signed an informed consent and were advised that participating was voluntary and optional. The information sheet was provided in Tamil and English. Local ethical and cultural conduct was observed. For instance, gender segregated seating arrangements that accommodated cultural considerations were followed. It was also not possible to explicitly request self identification of Tamil Brahmans owing to political and cultural constraints.

The scope of this study was primarily limited to observe the pronunciation 'zha', 'la' and 'La'. The findings from the above led to the evaluation of the comments which was referred to as the 'Zha' test! The focus group- all native Tamil speakers were asked to slowly and clearly record the phrase '*Vazhaippazham ezhaikkizhavan vaazhaippazha thol vazhukki keezhae vizhudaan*'. The phrase was written in Tamil orthography to avoid any discrepancy and the participants were able to see the sentence while they said aloud. The participants were given some time to get the phrase right. Only 20% of the 30 participants got the 'Zha' syllables accurate. This observation was consistent with the observations in the previous studies and anecdotal evidences obtained during the field work.

7.4 STUDY 3

QUALITATIVE INTERVIEWS WITH TAMIL BRAHMANS

Rationale: The target group satisfy code-mixing, code-switching and accuracy of pronunciation. The claim of attributing Brahmans to accuracy of pronunciation emerged in study 2 and also supported by Fuller & Narasimhan (2014), Fernando & Dragottei, Kroch (1986). Runeson & Host (2009) suggested that the multi disciplinary nature of software engineering allowed degree of flexibity to explore the research question in a manner that was appropriate to answer the question and provided guidelines on case study research in software engineering. The choice of Tamil Brahmans as the ultimate target audience for speech to text evolved as a result of the findings described in the previous sections. There was a general awareness   of the mis pronunciation of the 'Zha' syllable amongst the native Tamil speaking non Brahmans.

"People these days don't pronounce 'zha' properly. It is mostly the Brahmins who pronounce it right."

The above comment in the focus group was undisputed and supported by the work of Kroch (1986), McDonough(1997) and Fuller & Narasimhan (2014), led to further

investigations on Tamil speaking Brahmans. Kroch (1986) work further suggested that the presevation of orignial sounds especially Sanskrit, was not just unique to Tamil Brahmans, but also Brahmans of other Indian languages such as Kannada. It was with the Tamil Brahmans, the paper prototype was tested, alongside interview. The Tamil speaking Brahmans were found to satisfy two of the three criteria that was being investigated in the context of user acceptance of speech to text which also fed into the requirement and feasibility analysis of such applications in Tamil and they are :

a) Accuracy of pronunciation.
b) Code switching.

These indicated the suitability of Tamil Brahmans as the potential target audience who could be more receptive to the speech to text application. And that this application could find more takers within this community based on the indication that this user group could pronounce Tamil syllables with reasonable accuracy and that the language spoken by the Tamil Brahman had a natural element of code-switching as seen in Ridge (2012), Kailasapathy (1979), Ciotti (2017), Fowler(1954), Rudisill (2012) which made them more relevant to this research than the other social groups. Based on the methodology, the findings in this section could be further classified into:

a) Findings related to behaviour intention of using the language in technology (and)
b) Findings related to the feasibility of this application.

Sample size: 8 participants.

Sampling: Purposive, convenient and snowballing

Location: Chennai, Tamil Nadu

Age : 25- 55 years

Material: Voice recorder and the paper prototype

Design: Refer design in methods chapter for the design of the prototype. The first part was in-depth interview that focussed on language, culture and questions around technology and speech to text. The second part involved the testing of paper prototype.

Limitation: All participants were native Tamil Brahmans belonging to the *'Vadama'* sect. Therefore the findings cannot be generalised even within the broader Tamil speaking Brahmans.

Role of the researcher: Interviewer.

Positionality: Insider from the perspective of knowledge and practice as a Tamil and Brahman. Outsider from the perspective of diaspora.

Ethics: Participants signed an informed consent and were advised that participating was voluntary and optional. The information sheet was provided in Tamil (**see appendix D.1**). Local ethical and cultural conduct was observed. For instance, the interview and recording was paused when an elderly person arrived or when a guest arrived.

Findings:

Behaviour intention to use the language

- Behaviour intention to use the language was linked to the usefulness of the language in terms of opportunity and more specifically, the economic value the language brings with it.
- Proficiency in language seemed to have a relationship with the user acceptance of speech to text application.
- Possession of smart phone in the geography of investigation and within the cultural context was found to be a symbol of 'social status' more than 'utility'.
- Tamil and Sanskrit were given equal importance. However, the researcher interpreted that Tamil was sandwiched between Sanskrit and English! Sanskrit was given more importance in comparison with Tamil and English for the purpose of surviving.
- Constructive government policies on language could shape the attitude of people towards using the language at a social level which could potentially improve language proficiency. Improvement in language proficiency could aid in using the speech to text application with ease.
- Usage of English or use of Sanskrit words along with English was observed to be the trend amongst the native Tamil speaking Brahmans.

- There was some awareness about speech to text technology among all age groups of the study group however none has had the experience of using it in any language. Two participants from the study sample have tried using speech to text in English.

- Expressing in a language was dependent primarily on factors such as the level of education and occupation. Consequently, the language of expression varied between English and Tamil even in day to day conversations.

- There was no apparent display of behaviour intention to use Tamil language in technology.

- None felt the imposition or compulsion of English in any sphere and more specifically within the technology. This was more apparent when they responded in negative to a specific question on compulsion to use technology in a language. This was an interesting observation in a state that vehemently oppose Hindi [12] on the grounds of 'imposition'.

- Government bureaucracy, policies and politics indicated an influence on the attitude towards behaviour intention to use a language at a social level which then translated into using in technology (as it is currently with English). Examples include Government Order (GO (Ms) 145 dated 18.09.2014 of the Tamil Nadu Tamil Learning Act 2006) were yet to be fully implemented. The issue of language policy in the Indian Union is complex and the federal structure of the Union makes it more challenging.

- The Vedic philosophy, and Sanskrit seemed to have some relationship with the attitude towards accuracy of pronunciation of the Brahmans. On the basis of accurate pronunciation , it was predicted that the Tamil speaking Brahmans shall be able to use the speech to text application. It further indicated that perception towards pronunciation was a social attitude and the ability to accurately pronounce was a possibility.

- The proficiency in Tamil varied to a large extent and was dependent on age group, level of education and occupation. There might be some acceptance by the older

---

[12] Tamil Nadu in 60s vehemently opposed the idea of making Hindi the National language of India. As a consequence, India does not have a National language. The state from its previous name of Madras Presidency was renamed to 'Tamil Nadu' to assert the Tamil identity. The state government since 1969 has been following the two-language formula (Tamil and English) as opposed to three language formula (Regional language, Hindi and English) in the rest of the Union. The Anti- Hindi agitation was in favour of preserving the Tamil language and culture. However, most Tamil Nadu government website continued to be almost only in English. Therefore, the participant's argument on the utility factor of Tamil language in Tamil Nadu seemed logical.

generation on the basis of language proficiency but there might be little or no acceptance by the younger generation on the basis of the following excerpt from the interview:

"*Ok இதுக்கு நா எப்படி பதில் சொல்லனும்ன்னா அடையாளம் அப்படினு தேடறது தான் எனக்கு தெரிஞ்சு actually தமிழை forward ஆ எடுத்துண்டு போறவாளோட disadvantage னு சொல்லுவேன் . நீங்க ID தனியா வேணும்னு சொல்லும் போது நீங்க என்ன பன்றேள்னா you are starting to alienate your own language. You are starting to draw borders. there has to be integration, integration has to be seamless but, who is integrating whom is the question. You are joining two pieces together, one is actually going to encompass the other. So, if the thing that encompasses the other language is Tamizh, well and good but that is not the case. English and Tamil were merged together but English took the upper hand. அதாவது English இருக்கு English ஒரு language ஆ இருக்கு, English உள்ள வரத்த, it took pieces from Tamizh to make English itself seamless.*"

- The disproportionally large amount of English vocabulary (in Tamil grammatical syntax) was foreseen to be a major challenge in the development of this application. Again on the basis of complexity in the choice of orthography, it was not feasible from the readability point to view to resort to the Tamil script. However, the older participants preferred Tamil orthography while the younger participants were okay with the Roman orthography. Therefore, this research predicts that a speech to text application in Tamil in either script would not be universally accepted by the Tamil Brahmans in the study sample.

- Accurate pronunciation of the syllables could be a requirement to use the speech to text application

- Mispronunciation was a social phenomena and a social attitude towards the language. Speech to text application could be used to educate interested native Tamil speakers on their pronunciation.

- Tamil as a language was largely restricted to literature, entertainment and identity. The language that the study sample related with technology was English.

- Participants have had no experience in using any application in Tamil.

- English mixed into Tamil popularly referred to as *Tanglish* in Roman orthography seemed to be in popular usage in informal settings. However, this research as a result of adopting linguistic view, suggests that any research on *Tanglish* could not be carried out with reference to Tamil unless *Tanglish* was recognized as a separate language.

- The conversation in *Tanglish* or English in Tamil grammatical syntax was perceived to be Tamil by the native speakers which leads to the complex process of understanding the Tamil language and identity especially when the study sample indicated the use of technology to preserve the Tamil Brahman identity.

- The Tamil Brahmans appreciation of speech to text was from the perspective of the linguistic identity as it is distinctly different from Standard Tamil.

Prototype findings

- Participants usage and familiarity of *Tanglish* was apparent in their speech and in their experience of using Tamil in Roman orthography in handheld devices.

- Pronunciation was important to use speech to text application. The pronunciation also extended to Sanskrit words and not the Tamilised versions of Sanskrit words.

- Sanskrit syllables cannot be represented in the Standard Tamil script as there is no equivalent representation. Therefore from the output perspective, the closest possible representation could be the Tamil Grantha script as seen in 1.3.

- Neither the Tamil script nor the Roman script were universally accepted by the study sample.

- The spoken variety of the language must be preserved as it appears in the text.

- Participants in the study sample exhibited greater knowledge on Tamil transliteration i.e. writing Tamil in Roman script. This was in sharp contrast with the findings on the transliteration exercise. However, this need to be studied from a quantitative perspective to arrive at a more precise observation.

7.5 SUMMARY

Based on the model and three key parameters for the prediction of user acceptance, namely: accuracy of pronunciation, code mixing and choice of script, the overall concept of speech to text did appeal to the native Tamil speaking Brahmans. However, the functional value and usage of the language remained a key issue and an impeding factor to use the

application. The technical feasibility in speech to text also needs to be explored in a context where multiple languages are interspersed. If language is the fundamental unit on which such a technology is built (which in this case it is, at a conceptual level), then the aspects of language such as pronunciation, code- switching and code mixing and script needs to be a requirement. If on the other hand, a technology such as the speech to text is built based on the phenomena of a society disregarding the aspects of language, then it is suggestive of distorting a language, culture and the identity with which people relate and associate. Therefore, the findings could possibly lead to the argument of the implications on developing an application to satisfy a section of society for commercial reasons. The findings further indicate a strong sociology connection deeply embedded into the requirement and feasibility stages of the software development lifecycle and how a deep understanding or knowledge of the language, culture and lifestyle helps in arriving at an informed consensus on why certain parameters should form key requirements. It is argued that the feasibility of the application and acceptance of the application, based on the native requirements may well be incompatible with technology but could provide an interesting insight and an adventurous path to further research interests. Finally, it highlights the importance of adopting an indigenous   methodology to solve a technological problem.

CHAPTER 8- DISCUSSION

AIM

The aim of this chapter is to present with an interpretation and discussion of the findings. The section titled 'Researcher's own experience' has used first person.

8.1 THE LANGUAGE DILEMMA

Schiffman (1998) discussed on standardisation or restandardisation: the case for "Standard" Spoken Tamil. The work of Ridge (2012), Kailasapathy (1979), Saravanan et.al (2009) provided a useful insight and a compelling reason to look into the perspective and notion of 'Pure or proper' Tamil. The introduction chapter clarified that this research maintained a careful distinction between the user acceptance aspect of the social group involved, the application under study and the multi disciplinary approach that this research took in order to answer the broader research question on the prediction of user acceptance. The question on 'Standard' Tamil or 'Pure' Tamil still remains to be addressed, even after approximately two decades from Schiffman's (1998) study. The notion of 'proper Tamil' or 'Pure Tamil' varied substantially from the researcher's framework to the Brahmans and the non Brahman Tamil speakers (as observed in the study samples). The understanding of the perception of 'pure' or 'proper' Tamil was explored in the context of application. This research did not attempt to analyse its cause but merely reported the observation in order to appreciate the code-switching and code mixing perspective in the application. The participant's view on how Tamil ought to be spoken could advise the language model that could be adopted for such application and could aid in critical review its technical feasibility. Although the work of Schultz (2002), Kumar & Wei (2003), Bali et.al (2009), Prahallad et.al (2008), Rallabandi & Black (2017), Adel et.al (2015) indicate the possibility of accommodating the aspect of code-switching and code mixing, in the context of speech to text, the orthography also needed to be taken into account, in the event of occurence of code-switching and code-mixing. Whilst the literature in the context of speech recognition and speech to text was only from the point of view of language, the larger question on the choice of orthography, still needs to be addressed. In some cases, this was overcome using the Roman script as in Sankar & Nagarajan (2012) but the same was not possible to apply in the context of this research owing to the linguistic view.

VL noted: "*What we speak is not at all Tamil. We take some words from English, some from Sanskrit. If you talk about proper Tamil, then we need to go to the Sangam Tamil"* (Most approximate English translation)

The above remark from the interview displayed inherent code mixing and the notion of what constitutes a 'proper or pure' Tamil. Although the discussion of the above comment in the context of technology may be termed irrelevant, it cannot be, in its entirety be dismissed as seen in **section 4.6.3**. The emergence of the Brahman Tamil (Manipravalam) could be attributed to the philosophical text (Kailasapathy, 1979). The mixture of Sanskrit and Tamil was called '*Manipravalam*'. Historically, the code switching of Sanskrit and Tamil was not a Brahman privilege, although the work of Fulleret.et.al (2014) suggested the presence of disproportionately large Sanskrit vocabulary was unique to Brahmans. Fowler (1954) explained how Sanskritised Tamil was a result of one's heritage and that it was a living language subject to strict grammar rules. And in the 19th century, the non Brahmans had also adopted the '*Manipravala*' style until the quest for reviving the 'Pure Tamil' began (Kailasapathy, 1979). Nevertheless, *dharma* plays a key role in a Brahman's life and often forms the basis of relationship with the non Brahmans (Fuller & Narasimhan, 2014). The Vedic association of Brahmans meant, that Sanskrit could be safely assumed as a non negotiable entity of a Brahman. The following comments from the qualitative study sample exemplify the claim. The inference of the notion of a 'pure Tamil' in the participant's comment was indicative of a Tamil without the usage of foreign words or sounds as seen in Kailasapathy (1979), Fowler (1954), Schiffman (1998).

VL:

"*Sanskrit and Tamil are both equally important. Both are like two eyes. All our mantras are in Sanskrit."*

VM:

"*I would give more importance to Sanskrit because it is Divine and all the mantras are in Sanskrit. Tamil is my mother tongue and is required to interact with people in the place where I live."*

This comment of VL and VS especially, was strikingly similar to Fuller & Narasimhan (2014) interview of Tamil Brahmans in Chennai. The emphasis on accurate pronunciation, reference to the absence of Sanskrit phonemes in Tamil and the oral teaching method and will

be seen in section 9.4, was an indication of a connection to the Vedic philosophy and attitude towards accuracy of pronunciation. This has also been identified from the perspective of pronunciation by Ghafournia (2014), McGonough et.al (1997) and Keane(2004) observation of "Zha' being restricted to formal settings. From a native view, the presence of Sanskrit vocabulary in Tamil was 'natural' for a Brahman and was not viewed as code switching as opposed to the 'others' and will be explored in section 8.5. But, deciding on language model was crucial when it came to speech recognition and conversion of the speech to the target orthography. The issue was not occurence of code switching or code mixing in two languages but rather determing its limit. The work of Kumar & Wei (2003), Prahllad et.al (2008), Schultz(2002), Rallabandi & Black(2017) indicated the effort towards a bilingual and mixed lingual speech processing. Thangavelu et.al (2016) recognised the absence of equivalent phonemes for mapping between Tamil and English. On the basis Besacier et.al (2014) definition of under-resoursed languages, it was suggestive that Tamil could perhaps be classified as an under-resourced language from the perspective of human language technology. Although, this classification could potentially be contradicting to Besacier et.al (2014) definition of well-resourced language that took into account, the languages in core technologies such as Siri ASR, Google voice search etc.

The investigation of spoken language, in this context was incredibly important as the application of technology involved a speech recognition component and an output (text) component. By investigation, this research meant observing the usage of language in the social sphere (particularly in the native region) and the aspect of accurate pronunciation. Both of these were language dependent which further relied on the social framework. A more generic comparison of the attitude towards Tamil language in Tamil Nadu, Northern Sri Lanka and Singapore could be relevant in order to develop a finer understanding of the role of politics in the context of language and identity and how this may influence the choice of language in using the application as seen in **figure 4.1**.

Tamil Nadu, in its early sixties saw a fierce Anti Hindi agitation, in support of Tamil language, culture and identity. English replaced Hindi in Tamil Nadu and it continues to be in the forefront of administration, education and courts (Kailasapathy, 1979). Schiffman (2002) and Muniandy et.al (2010) provided similar account in other Tamil regions- namely Sri Lanka and Malaysia. The insider's view of this account was still largely prevalent in the Indian state of Tamil Nadu. Historical events indicated a fierce competition between English, Sanskrit and Hindi (Ramaswamy, 1998). The Brahmans have traditionally been the

scapegoats of the fiercely proud Tamils, who questioned the loyalty of Brahmans towards the language (Ramaswamy, 1998). Ridge's(2012) observation indicated an increasing language shift of diasporic Tamil Brahmans. The idea of a Pure Tamil movement, in the historical context arouse from Robert Cadwell's observation on Tamil and its relationship with Sanskrit. His observation was on the Brahmans and usage of Sanskrit words, as expressive religious words in the Tamil context (Kailasapathy, 1979). Despite Tamil being an official language of Tamil Nadu, and with all the pro- Tamil sentiments that prevailed, the economic value of the language in the native space, was still largely questioned:

VM:

"*What do I gain by learning Tamil? Okay even if I study completely in Tamil till 12th standard, I will still need to switch to English for my higher studies.*"

In some cases, the parents decided the language their children should be learning (Srinivasan, 2008). In schools, where the medium of instruction was English, Tamil was taught as a 'second' language even if it was the student's mother tongue. The students could choose between Tamil, Sanskrit, French among few others as their 'second language'. In schools where Tamil was the medium of instruction, the students learnt English as a compulsory second language. Tamil Nadu, as a state follows a two-language formula as opposed to the recommended three language formula. This meant that a student born and raised in Tamil Nadu could formally complete his education with almost little or no ability to read and write Tamil. The choice of language by parents were driven by the perceived notion of getting good grades (Srinivasan, 2008). And the Tamil Nadu Tamil Learning Act (2006) was implemented only from the academic year 2015-16, almost a decade after it was introduced. The comment of VM points to the perception of economic value of the language in the native sphere. This was significant in providing some insight on factors such as perception towards Tamil language, possible language shift , both of which is perceived to be language factors influence the acceptance of the application.

"And indeed Tamils everywhere have not treated the status of English as problematic. They have embraced English and continue to embrace it, as a barrier or buffer against Hindi, Sinhala and Malay. The problem now is that this group has relaxed its guard against English, and too much knowledge of English now means that this group now knows too little Tamil, and in fact not committed enough to Tamil. In fact, many of my informants though

committed to Tamil declared that they would not put their children in Tamil schools because Tamil schools are a dead-end, both professionally and socially."

Schiffman (2002)

Similar were the responses of BS, VS, VM while being in the native space. The variation in the usage of language, extent of code switching could be seen in the interview transcripts (**Appendix A.1 & A.2**)

Through the work of Canagarajah (1995), one could infer the effectiveness of politics, especially in 'imposing' or enforcing the rule with which the value of a language could be realised. Although, his work relates to the period when the country was in turmoil, the transition of a community from bilingualism to monolingualism is not just noteworthy but worthy to be underscored in the context of appreciation for language, its usefulness in the native sphere. This research argues that the usage of language at a social level could possibly lead to behaviour intention of using the language in the technology. However, insider's experience suggests that, in the absence of a political compulsion, it was still possible to achieve that through a self -sustained effort as discussed in chapter 8 and in the personal experience section of this chapter. And that, the shift from English to Tamil or retaining bilingualism in technology is a possibility. In which case, it would be the case of an individual versus the larger society. The then leadership insisted on 'Tamil only' which included, a form of a pure Tamil sans English and Sanskrit. The social pressure, could be seen as a compulsive motivation of using the language in the defined geography as seen in **figure 4.1**. Canagarajah (1995) further observed that interviews for jobs and selection tests held in Tamil. Under these circumstances, VM's perception on the usefulness of the language might have been different. Rao & Troshani (2007) from the perspective of perceived usefulness in the context of technology acceptance model defines 'usefulness' as a

"degree to which a person believes that using a particular system would enhance his or her job performance"

This research has acknowledged the perceived usefulness aspect within the user acceptance framework but was compelled to further narrow it down to the target language with a view that 'perceived usefulness' of a language led to its increased use which then had the potential to translate into behaviour intention to use the language in the technology. Eseonu (2014) contend that the technology increasingly intersect with the social systems. And it is in this

context, this research lays emphasis on the prevailing social system from the perspective of an insider in the context of technology. Stephan et.al(2013) explored the link between fulfilment of the requirement and user acceptance. But, they discussed about the system meeting the requirements of the users whilst this research partly dealt with the idea that the user also, to some extent need to meet some requirements (such as pronunciation and language skills) to use the application. Nevertheless, Stephan et. Al(2013) terminology of 'true requirement' and the user's mental acceptance of the system could be contextually applied to speech to text application. The true requirement in the context of this research was constructed as a requirement that the user fulfilled in order to use the system. And therefore, in the context of speech to text, the researcher's conceptualisation of the system and the analysis was composed of non negotiable component on pronunciation, but was open to negotiate on the adoption of the script. This was a conscious decision taken by the researcher taking into account, the varying level of proficiency in Tamil (to read and write) by the native Tamil speaking Brahmans. However, the insider view was that the code switching and code mixing element, needed to be understood from a social perspective to explore technological feasibility and to predict the user acceptance of speech to text in Tamil. Therefore it was in this context and from the point of view of technology, the dilemma of the language needed to be addressed. This reserach observed that the official definition of literacy in the Indian context meant that a person who resided in Tamil Nadu (where the official language is Tamil), and possessed no skills of reading and writing in Tamil as defined by UNICEF could still be offically a 'literate' Tamil citizen despite being illtierate in the state's official language. Hence, this research takes the view that literacy figures of Tamil Nadu may not reflective of literacy in Tamil. It further brought to surface that the issue of script complexity, that was important in the speech to text, could be attributed to the literacy in Tamil as defined by the UNICEF. Closely related to the language was identity. Schultz(2014) suggested that language was not just used for communication but also for other reasons such as preservation of culture and empowerment. One cannot inherently assume the functional knowledge of Tamil (to read and write). Familiarity of written Tamil in Tamil script was safely assumed as one of the pre requisites to use the application, the other being the ability to accurately pronounce the syllables. The Brahman's disdain for the Tamil language and a preference for English and Sanskrit could possibly be causes for the lack of quest for using technology in Tamil (Fuller & Narasimhan, 2014).

But VMM's reason for a need of a speech to text application seem to differ from Fuller & Narasimhan (2014) point on the Brahman's appreciation of Tamil but related to the reason provided by Schultz (2014) . But an insider would agree on Fuller & Narasimhan's observation to a greater extent at a macro level.

"Brahman Tamil needs to be preserved. It is unique and it has its own charm. And from that perspective, this technology will be useful" VMM

The shift in the use of language as seen in Ridge(2012), Canagarajah(2008), Das(2008), Schiffman (2002) and language preference possibly challenges the existence or development of such applications.

VMI

"If it is older people in 40s and 50s, Namaskaaram as a salutation is automatic, but if it is youngsters, then hello has to be said!"

On one hand, excerpt suggested that there was a desire to uphold and preserve the linguistic identity but the same time, the empirical findings suggested an increasing shift towards using English in domains where Tamil could have been used including salutations. VMM's comment were similar to the observations made by Keane (2017).

VMs comment on the presence of disproportionately large English vocabulary in conversational Tamil.

"You get used to a language especially when you have to draft an agreement or convey something to different kind of people. Okay, in inner mind you may think in Tamil but when expressing it, it goes in the form of English and therefore you have to kind of have a , you are forced to have a proficiency to excel in what you do"

challenges the need for a speech to text application in Tamil. Hinglish (Hindi English code switching) on Apple's iOS is yet another example of technology meeting the language requirements of the users. Eseonu (2014) in his work on socio-cultural influences on technology adoption and its sustainability took a view that a number of social factors were overlooked to develop a technology on the merits of technical superiority. This research has explored the socio-cultural aspect from the perspective of technological sustainability that genuinely commits to the preservation of language in its original form and identity. And further argues that these form a core part of the feasibility study of the software engineering

process with specific focus on Tamil. Wee (2008) explained how English, slowly emerged as the *lingua franca* replacing Malay in Malaysia despite the latter being an official language. The user acceptance model's prime focus was on language and the ways in which languages were used socially. Wee (2008) contributed to social, cultural and lifestyle of a community. Ramanujan, a Tamil Brahman wrote not in Tamil but in English. He was an English language poet in the United States who was devoted to South Asian studies (Ramazani & Ramanujan, 1998). His positionality of being an insider and outsider at the same time mirrored the approach of this research. Sefa Dei & Asgharzadeh (2003) asserted that language was a crucial instrument to preserve the indigenous living and a development of individual and collective identity. His work further explained how the English language was viewed as a social and economic currency to further one's aspiration. This was applicable in the Tamil context as well. According to VM's response, one could be punished for speaking in Tamil in schools. BS also shared a similar response. English as in the case of Ghana , was viewed as a 'neutral' language as it was not an indigenous language while in Tamil Nadu, it was viewed as a 'link' language. India's federal structure and it's article 351 and 345 in the context of language needed a focus:

*"351. **Directive for development of the Hindi language.**—It shall be the duty of the Union to promote the spread of the Hindi language, to develop it so that it may serve as a medium of expression for all the elements of the composite culture of India and to secure its enrichment by assimilating without interfering with its genius, the forms, style and expressions used in Hindustani and in the other languages of India specified in the Eighth Schedule, and by drawing, wherever necessary or desirable, for its vocabulary, primarily on Sanskrit and secondarily on other languages. "*

*"345. **Official language or languages of a State.**—Subject to the provisions of articles 346 and 347, the Legislature of a State may by law adopt any one or more of the languages in use in the State or Hindi as the language or languages to be used for all or any of the official purposes of that State:*

Provided that, until the Legislature of the State otherwise provides by law, the English language shall continue to be used for those official purposes within the State for which it was being used immediately before the commencement of this Constitution."

The article, whilst guaranteeing linguistic rights and equality to all the citizens of the country, encourages extensive use of one language, which the non Hindi speakers perceive it as an

attempt to homogenise the Indian identity. This was strikingly similar to Iran's concept of 'one country-one nation- one language' (Sefa Dei & Asgharzadeh, 2003). India's Article 345 and 351 was comparable with Iran's Article 19 of the constitution that guaranteed all citizens equal rights but Article 15 singled out on Persian as the legitimate script and Farsi as the 'only official language' that benefited perhaps the majority citizens of the country (Sefa Dei & Asgharzadeh, 2003). Through the model adopted to evaluate the user acceptance of the speech to text, the findings indicated a link between government policies and perceived usefulness of the language and opportunity to use that language socially which then translated to behaviour intention to use the language in technology (see figure 4.1). However, participants from the same linguistic background with similar exposure had contrasting views on forcing a language at a social level.

VL

"*There is nothing wrong in compulsorily learning Tamil in Tamil Nadu. We learn other languages when we go out. Therefore making Tamil compulsory in Tamil Nadu is not wrong*"

VH contradicts:

"*Language should not be forced. They must be learnt voluntarily.*"

VL and VH have both lived in Non Tamil environment for a considerable length of time, and over the years have developed the ability to speak a third language. The 'perceived usefulness of language' , in their case varied from region to region where the usage of Tamil was severely restricted. VL's comment by Non Tamils may come across as particularly 'chauvinistic'. Fuller & Narasimhan (2014) supported the view that the multi-lingual nature of Tamil Brahmans, make them more objective and were less chauvinistic compared to their Non Brahman counterparts. Nevertheless, both VL and VH indicated English as their choice of language when it came to engaging with the technology.

This research took the view that speech to text application was more likely to be accepted and better received within a homogeneous linguistic environment to which government initiatives such as the Tamil Nadu Tamil Learning Act (2006) among others could positively contribute. This could be inferred from VM's comment on the Chinese who according to the participant, were relatively more successful in using their mother tongue in technology. VS linked it to the perceived homogeneous Chinese society.

The research argues that when a language has a limited economic and social currency in a region, the behaviour intention to use that language in technology shall also be extremely limited . The views of Anita (2013) on indigenisation of English in the Indian context was reasonably justified on the grounds where words of English origin were perceived to be 'native' and 'normal' to be used. This seemed to erase the notion of code switching from a native perspective but happens without the conscious knowledge of the speaker but was identifiable by an outsider. Requirement elicitation was the most important activity in requirements engineering. And it brought back to Eseonu's (2014) view on developing applications to win on the merits of technical superiority. Just as elicitation of requirements was critical to develop a software, on the converse, this research argues that elicitation of requirements to use the application was also equally important. The rationale for this argument was from the perspective of the language shift as indicated by Fuller & Narasimhan (2014), Krishnaswamy (2015) and Ridge (2012).

The following observations by Caplan (1995) on Anglo Indians of Madras (Chennai), provided some insight on contrasting attitude towards one's mother tongue in the target region of this research:

" There was an Anglo-Indian English-medium railway school and a Tamil-medium school. We all went to the English-medium school. You knew all the teachers; in school they were 'miss' and after school they were 'auntie'. Everyone was Anglo-Indian. We were totally enclosed in an Anglo-Indian atmosphere. We spoke our own language and ate our own food"

Caplan (1995)

The code-switching language shift, and a contrasting response to the above can observed in the study sample in the following excerpt :

" *In a way, Tamizh has become foreign. Ask anybody, it is like asking a person on stage, do you support women empowerment? The obvious answer is going to be yes. In the same way, ask do you support Tamizh and love Tamizh, any Tamizh Nadu guy would say yes. But does he really, is the question? It has become foreign in several ways rather than one- simple reason is that we have adopted something and we have started moving in a path, and right now, we are reluctant to move, not reluctant, we have boxed ourselves into corner. if we* **choose** *to make Tamizh as the primary language, you know in all the industry, the transition is going to be costly at this day. Assume that everybody has this group - I want to make Tamizh as the 'happening' right now. If everybody wants to change, it is going to be costly.*

*We have embraced something, and **two to three generations** have been cultured in the same manner. So in a way, we have alienated our own language into something exotic, antique piece of history rather than something that is **used for day to day life**." - VM*

Shulman's (2016) recognition of the penetration of English words not just as borrowing but also lexically and in syntax, and recognised the conceptual consequences that Tamil has to deal with. VM's response supports Schiffman (2002) observation on the position of English amongst the native Tamil speakers.

Complexities relating to language such as code switching and code mixing, pronunciation were additional challenges and influences the language model used in the technology (Besacier et.al, 2014) as in the case of automatic speech recognition of under-resourced languages. Therefore, in the context of Tamil speech to text application and its acceptance, a detailed study and understanding of the language usage and attitude towards the language by a community could result in a more accurate prediction of user acceptance. Vazhenina et.al, (2012) brought out the complexities and difficulties in conversational Russian speech that involves varying pronunciation, speakers and dialects. Coelho (1997) brought out the language shift from Konkani- an Indian language, predominantly spoken in Mangalore (south western coast of India) and also indicated the choice of language was also influenced by philosophical affliation (as in the case of Brahmans in this research) and Christians in the case of Coelho's (1997) work. From the perspective of user requirements, the contextual relevance of technology is dependent on language and exhibition of characteristics that suggest a possible shift in the language could possibly question the need for the application. In this context, Coelho (1997) further indicated that despite English being a non native second language, it was possible that the speakers exhibited greater functionality in English and equally, identification of 'mother tongue' did not necessarily meant fluency in that language as seen in the study sample.

The user acceptance model of this research was largely influenced by the Unified Theory of Acceptance and Use of Technology (UTAT) (Venkatesh & Davis, 2000). But it has been adapted to suit the native framework as discussed in **2.4.1**. *Flexibility* as defined by Wixom & Todd, (2005) in the context of theoretical integration of user satisfaction and technology acceptance refers to the adaptation of the system to that of the dynamic demands of the user. This research has regarded language, as a non- negotiable component on which the paper prototype was built and evaluated. It was done so,   to support the aspiration of the users in

preserving their linguistic identity in its native form as supported by VM. Therefore from this perspective, it was reasonable to treat language as a fixed entity and predict the acceptance of speech to text technology that was under investigation in relationship with the language and prevailing social conditions. The above discussion relating to language, government policies, social interaction, perceived usefulness of a language in a defined geography was unarguably in the context of technology that could help in predicting the 'behaviour intention' that formed the core of user acceptance framework. Re reading the above discussion from the perspective of technology acceptance model with focus on social conditions as key factors to behaviour intention would provide a whole new dimension, in order for its relevance to be appreciated in the context of predicting the user acceptance.

8.2 TRANSLITERATION AND SCRIPT

The quantitative transliteration exercise findings suggested that there was a difference in the ability of transliteration between Brahmans and Non Brahmans. However, more samples are needed from the Brahmans to be able to comment on the comparative ability. This was inferred by the effective usage of appropriate letters by the Brahmans. For example: zh for its Tamil equivalent of ழ. Although, the findings suggest a possible link between transliteration, pronunciation and choice of script, this needs to be further investigated. The Brahman's accuracy of pronunciation was reflected in their speech and choice of transliterated spellings at the time of testing the paper prototype. Interestingly enough, they were quite accurate to spell the transliterated version that involved *sandhi* rules as **seen in 2.2**. This was evident when VL paused for a moment to reflect on the Tamil spelling before confirming:

*Thinga**kk**azhamai (Monday)*

And affirmed on the basis of the *Sandhi* rules. The presence of double 'kk' in the transliterated version was not a matter of choice but a necessity from the perspective of *Sandhi* and pronunciation. This was an empirical indication of a possible relationship between pronunciation and its reflection in the transliterated version. However, the inability to accurately transliterate by the non Brahmans was noteworthy in the context of pronunciation. In some cases, they were unable to attempt transliteration as in **appendix B.1 and B.4**. Further, multple spellings can be seen for the same word and can be seen in appendix B.1-B.30 in Roman orthography when transliterated as also identified by Dhore et al. (2013)

Statistics

Kalai2

N	Valid	177
	Missing	0

Kalai2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	galai	2	1.1	1.1	1.1
	kaalai	1	.6	.6	1.7
	kalaai	2	1.1	1.1	2.8
	kalai	116	65.5	65.5	68.4
	kalaie	1	.6	.6	68.9
	kalaiey	1	.6	.6	69.5
	kalaii	1	.6	.6	70.1
	kalay	3	1.7	1.7	71.8
	kalei	1	.6	.6	72.3
	kalhai	1	.6	.6	72.9
	kali	2	1.1	1.1	74.0
	kallai	13	7.3	7.3	81.4
	kalli	2	1.1	1.1	82.5
	kalzhali	1	.6	.6	83.1
	kazai	1	.6	.6	83.6
	kazhai	7	4.0	4.0	87.6
	kazhi	1	.6	.6	88.1
	kazi	1	.6	.6	88.7
	kazlai	1	.6	.6	89.3
	kelai	2	1.1	1.1	90.4
	kilai	2	1.1	1.1	91.5
	klai	1	.6	.6	92.1
	MH	6	3.4	3.4	95.5
	NA	8	4.5	4.5	100.0
	Total	177	100.0	100.0	

Figure 8.1 : A transliterated word showing a variety of conceptualised spellings. It is also noteworthy that the Tamil syllable ள was also represented by 'l' and hence had to be called 'kalai2' MH: Misheard; NA: Not attempted

The issue of mispronunciation of this syllable and the inclusion of accurate pronunciation as a requirement for speech to text was consistent with Shanmugalingam

(2002), Das (2008), Srinivasan (2013). This was crucial from the perspective of the design of the application as **seen in 4.6.3**. As the paper prototype of the application was designed from the view point of '*what you speak is what you get*'. On that basis, the prototype design purposefully did not incorporate features such as auto correct for correcting the misspelt words. The design of the prototype was more user dependent, in which the user output was dependent on the user's input without any technological interferences such as spell check and auto correct.

Despite the observation of accuracy of transliteration by the Brahmans, the confusion over the Tamil syllables ல, ள prevailed in the transliterated versions, and as pointed by Srinivasan (2013)- both represented by the same letter 'l' . But its accuracy and distinctness were noticeable in their pronunciation (**qualitative interview and prototype testing**). This seemed to contradict earlier observations on the generic relationship between transliteration and pronunciation. The Technology Acceptance Model (TAM) was originally formulated to understand the phenomena of why people choose to accept or reject technology. The model was later used to predict the user acceptance after the subjects briefly interacted with the system. TAM acknowledged the influences of external variables and its power to influence the acceptance of the system. Szajna, (1996) suggested that attitudes are formed from beliefs a person holds about the system. In this context, the usefulness of the application was not questioned. The perceived usefulness of the application was questioned only in the context of language (Tamil) and its contextual usage and contextual relevance within which the application was designed.

"*ஒரு எடத்துக்கு போ**யிண்டி**ருக்கேன் நா வந்து ஒரு ticket purchase பண்ணறேனோ இப்போ ஒரு automated service இருந்தா கூட the interface is in English. நீங்க use பண்ணறது and I work basically with **computers** which again orient towards **English.** இது தான் culture, heritage everything is associated with Tamizh Nadu and its culture but the fact is learning Tamizh alone, is it going to help me, me as a person in the long run? I can, at the end of the day, chestpounding and jargon only takes you so far. Practical viability is something."- **VM***

Therefore, the adapted version of the user acceptance model, laid greater emphasis on the external variables of TAM and was based on conditions such as social interaction, history, politics, government policies, opportunities and an individual's perception on the basis of

these. More specifically, the belief a person holds about a language was assumed to play a significant passive role in the acceptance of technology in Tamil.

There are two types of keyboards available in iOS and other applications including Gmail. The first keyboard is transliteration keyboard and the second one is a keyboard that has Tamil scirpts on it. One of the main purpose of the quantitative transliteration exercise was to determine the usage of Tamil script in technological devices despite the availability of transliteration keyboards and familiarity with the script. Therefore, in this case, availability of technology in the language was not an issue but using it for the purpose of communication was, as indicated by the study sample. The Brahman participants used Roman script despite their ability to accurately transliterate and can be seen in the following response:

VMM

"*I use Tanglish to communicate with my son who is in the US.*"

There were various reasons for using Roman script and the most common point put forward by the study group were the non familiarity of Tamil script to the other Tamil person with whom the communication took place. However, if the output was made available in Tamil script, then there was a potential of excluding those who might be interested in using the application but may not be able to read. In which case, the application has little or no value (**refer** literacy in introduction). On the other hand, it could be argued that speech to text application could be used to educate those native Tamil speakers who are not familiar with the Tamil script. The latter has educational value but demands self motivation or motivation by compulsion as **shown in figure 4.1**. Ihde (1995) discussed technology as a cultural instrument and argued that almost all technologies are non-neutral since they transform the human experience. He further reiterated the understanding of technology from a phenomenology perspective rather than merely viewing them as objects.

"Any technology may be used in a multiplicity of ways, limited only by individual and cultural imaginations of the people or user" (Ihde, 1995)

were relevant in the context of applications that were available in Tamil to use. VM's comment on using Tanglish to communicate could be interpreted as an individual limitation influenced by the cultural imagination of the user sending and receiving it. And this was, despite the availability of technological instrument to convey a message in Tamil language and script. The research shares similar views as that of Lhde (1995), that technologies must

be understood as human-technology pairing at a micro level and socio-cultural-technology pairing at a macro level and that they are culturally embedded. The user acceptance model adapted to answer the broader research question, from a phenomenological perspective, proposed six components that was perceived to contribute towards the usage of language at a social level which in turns, this research considered a factor that leads to behaviour intention to use the language in the technology as well. They are :

1. Self motivation/ pride
2. Compelled circumstances
3. Government policies (that dictates)
4. Opportunity to use which feeds into
5. Perceived usefulness that again contributes towards
6. Social and cultural lifestyle

However, further research is needed in this area to confirm the perception. The overall model could then be used to predict the attitude towards a language which this thesis argues to be a non negotiable component when predicting the user acceptance of a language dependent application. VM's comment on Brahman Tamil could be related to self motivation or pride on the account of preserving and promoting the dialect. But that alone may not be sufficient to safely assume that the application would be accepted. Rebman et.al (2003) work suggested that the acceptance of a technology could also be reliant on the age demography. But, the effect of age demography of the study sample did not appear to be an issue in acceptance of a technology. The self motivation could be an enormous force to use the language overriding other factors. But, insider's perspective suggests that self motivation could be a starting point to use the language but factors such as socio-cultural lifestyle and government policies emerge as a more stronger force that could be to the detriment of interest. Despite the availability of Tamil keyboard, familiarity and ability to reasonably accurately transliterate, the choice of Tanglish suggests lack of self motivation to use the language, influence of social and cultural lifestyle. In some cases, lack of awareness of existing technologies and prior experience of using them were also some reasons. Arguably though, the inference was that the quest for technology in mother tongue could possibly arise as a result of self motivation.

8.3 Rᴇsᴇᴀʀᴄʜᴇʀ'ꜱ ᴏᴡɴ Exᴘᴇʀɪᴇɴᴄᴇ

My use of computers and technological devices started in English. Like the participants, the communication with others what involved technology was predominantly in English and sometimes in Romanised Tamil with its equivalent transliterations. However, the urge to use Tamil script in technology and the motivation to communicate with others in Tamil script arouse as a result of my conscious Tamil identity. The shift from English to Tamil in technology was not easy either. It took me atleast four years to establish the fluency in technology and accept Tamil as a part of technology. In the process, language and code switching and code mixing indeed were factors that slowed me down considerably when I wanted to communicate with others. I noticed how I was 'differently' treated when I spoke in Tamil instead of English in Chennai or perhaps the socially well accepted '*Tanglish*'. In the process, I uncovered that language and identity played a vital role in associating technology with a particular language. In my case, it was the self - motivation with an attitude that 'it's okay to use Tamil in technology' helped me to rediscover myself in technology. I have, with great amount of effort been able to achieve a bilingual (English and Tamil) competency in technology. Equally interesting was the argument of language of opportunity and economy as a motivation to speak, write and promote one language. To me, use of mother tongue in technology should not have a monetary value attached to it. But I did realise that language economy has a consequence and it was making a conscious decision of prioritising one over the other in the native space.

In the user acceptance model, the behaviour intention consist of:

a. Behaviour intention to use the language at a social level leading to actual use of language and

b. Behaviour intention to use the language in technology.

The link between social and technology seemed unavoidable because, throughout my experience, English and *Tanglish* dominated my daily activities even as I remained in the native space. This could also be related to VL's response:

> "*I have never sent any message in Tamil to anyone*"

BS's view on Tamil particularly emphasises the importance of English:

"English is a globally recognised language. You live in the UK and Tamil is of little relevance to you. Therefore, a base and proficiency in English is important to survive and excel. " This was quickly linked to the career and relevance of language in it.

" I am not discouraging the language but it must support you and in your career"

This was quite similar to VS comment:

"What do I gain by learning Tamil?" and questions its contextual usage and relevance. BS has had no prior experience of using Tamil in technology. To a question- Have you ever used Tamil in technology? Or think of a scenario, where you have used Tamil in technology or may be perhaps forced to use Tamil in technology was responded in negative.

"I have not used Tamil in technology. I use only English even in computer"

With regards to Sanskrit, BS had the following comment:

"Sanskrit was my second language. And for Brahmins, it is in their blood. It is important to recite slokas which is a part of the culture and they are trained from the childhood. The base for slokas is Sanskrit. I was a Sanskrit student and it has helped me in many ways. But between Tamil and Sanskrit, both are equally important to me. They are like two eyes and English for survival."

These responses support Schiffman (2002) and Ciotti (2017). To a question on general script preference BS preferred Sanskrit script over Tamil. Even in the native space, BS had specific view on usage of English and Tamil at a social level as seen in **figure 4.1**.

"Depending upon the place, for example if I go to a hotel or a vegetable store then it has to be in Tamizh but if I go to bank or office or when I speak to someone senior, it has to be in English"

But on probing further on the usage of English when either of them are Tamils, BS responded *"It has nothing to do with respect but convenience"*. However, BS experience seemed to contradict earlier observation on forcing a language. BS later in one of the questions responded negatively:

"We speak Tamil only at home. I studied throughout in convent. In Convent you are not allowed to speak in Tamil so we are used to it". This was again a very similar response to VS. The relationship between the use of language at social level and behaviour intention to use the language in technology was evident as BS, VS, VL, VM, VH all had used code

switched and code mixed Tamil and that their daily activities were quite influenced and to certain extent dominated by English. Therefore, this research views that design of the application merits an approach that was within the scope of the language and perception of the native speakers on the accuracy of pronunciation as opposed to adopting a design view that accommodates all variations in pronunciation, that was not compatible with the language.

What should you be doing in order to choose Tamil in technology (involving keyboard) BS's response links to proficiency of language:

"*I should be familiar with the language and its script"* Therefore this response suggest that script and language complements each other but contradicted VMM who was familiar with both Tamil and Tamil script used Roman script in communication. BS's 'zha' pronunciation was accurate and was consistently reflected throughout the interview. BS was quick enough to reject citing a spelling mistake where 'zha' was replaced with 'la'. Infact, all the partcipant's pronounced 'zha', la and 'La' quite accurately and was reflected in their speech throughout the interview.

Should technology correct mis pronounced words or should people practice enough to change their pronunciation?

"*People should practice and should be able to pronounce correctly because you cannot change the language and you cannot also change the spelling"* This supported the researcher's conceptual design of the application – *'what you see is what you get'* BS disagreed to use Roman script to represent Tamil which was preferred with other participants like VL, VMM. However, BS did opt for a Tamil keyboard in the prototype version before responding to the choice of script. It was observed that choosing a Tamil keyboard had very little to do with behaviour intention to use the language in technology. This was indicative from participants admission of not using any application in Tamil. Although, this suggested that there was little link between choice of language and behaviour intention to use the language, it did however indicate that prior experience of using the language in technology could possibly strongly contribute to behaviour intention of using the language in technology. This was an empirical indication of the possible relationship between language, script and its preference when it comes to the output of speech to text.

8.4 REFLECTION ON USER ACCEPTANCE MODEL

As a result of the qualitative process, the user acceptance model could be further simplified on the basis on UNICEF's definition of literacy which also complemented the definition of the Department for International Development (UK). Arguably, the proposed user acceptance model, used to predict the user acceptance of speech to text technology did aid in answering the research question. This research takes the view that for accurate prediction of user acceptance, a detailed understanding of the culture, lifestyle and government policies and engagement with the language at a social level on a day to day basis is essential. However, from the linguistic view of the application and from the perspective of the ability to use the speech to text application, literacy in the target language (as defined by the UNICEF) was key. The phrase "*effective functioning and development of individual and the community*" with emphasis on reading, writing and numeracy was the rationale behind choosing UNICEF's definition for literacy. It is also predicted, that by adopting UNICEF's literacy definition, the issue of script complexity in speech to text could be overcome in one aspect. Nevertheless, the proposed model contributed in identifying precise areas that contribute to either inability to use the speech to text and was beyond the scope of the application (example: language shift, individual factor such as attitude towards the language) or acceptance of technology on the basis of linguistic identity which again related to the language part of the acceptance model. And having identified this, the prediction of the user acceptance could then almost entirely be based on language literacy coupled with experience with the technology. But in order to arrive to the point of language literacy in the context of technology, this research perceives the proposed user acceptance model to be crucial.

8.5 SPEECH TO TEXT IN TAMIL : THE CASE OF GOOGLE
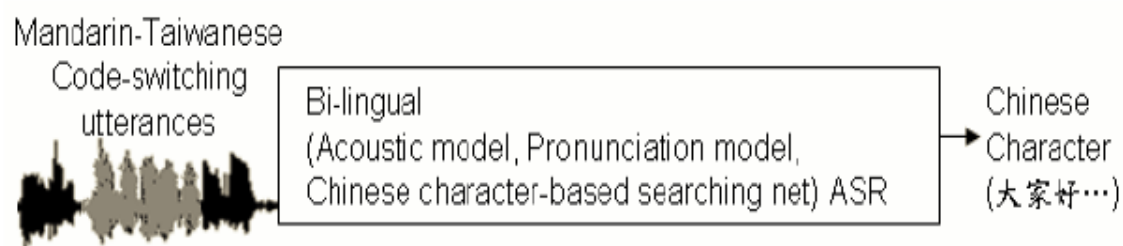
Figure 8.2 : A Mandarin- Taiwanese code-switching speech recognition Chan et.al (2006)

Chan et.al (2006) have attempted to overcome code-switching issue using automatic speech recognition. But, in their model, the output, regardless of the language was in Chinese character. Vu. et.al (2012) have proposed a bilingual acoustic model in Mandarin and English to overcome the issue of code-switching. This research perceives that the model adopted by Hsu et.al (2006) in the context of code-switching and Sridhar et.al (2013), may not be entirely acceptable on the basis of the responses provided by the study samples with specific reference to the preference of output orthography. Sultana et.al (2012), have resorted to transliteration to display the output in Bangla orthography. Although mixed language synthesis suggested the possibility of incorporating code-mixing and code-switching factors, this research perceives that it would only be useful from the perspective of speech recognition and may only partly solve the issue. But, the challenge of determining orthography still needs to be addressed which this research perceives could be done only by involving the relevant stakeholders such as the Tamil Brahmans, in this case. For example, if a mixed language synthesis is used as proposed by, Schultz et.al (2013), what script would be used to represent the syllable that may be found in multiple languages? For example, the word *Mazhai,* the syllables *ma* and *zhai* can be found in both Tamil and Malayalam. But when a mixed lingual model is used from the perspective of recognising code-switched or code-mixed utterances, in what script should that appear as the output? In cases where the participants like VMI have preferred respective script to represent the words in different languages, this needs to be further tested from a readability point of view. Further, adoption and usage of the Grantha script could assist in overcoming the script complexity associated with the Sanskrit Tamil code mixing and code-switching. But, the extent of familiarity of the Grantha script was not discussed with the participants. As an insider, it was assumed that the participants did not have adequate knowledge of the Grantha script for reasons described by (Sharma, L2/09-372 unicode.org). The participants in the study sample seemed to be more familiar with the standard Tamil script and in case of VM, Devanagari script. Nevertheless, Grantha and Tamil script share some similarities. Sharma (L2/09-372) suggested that this script was used by less than approximately 50,000 people primarily by Vedic scholars and students of Tamil Nadu and Sri Lanka.

"*Most people who can read Tamil cannot read Grantha and some people who can read Grantha cannot read Tamil properly*"

(Sharma, L2/09-372)
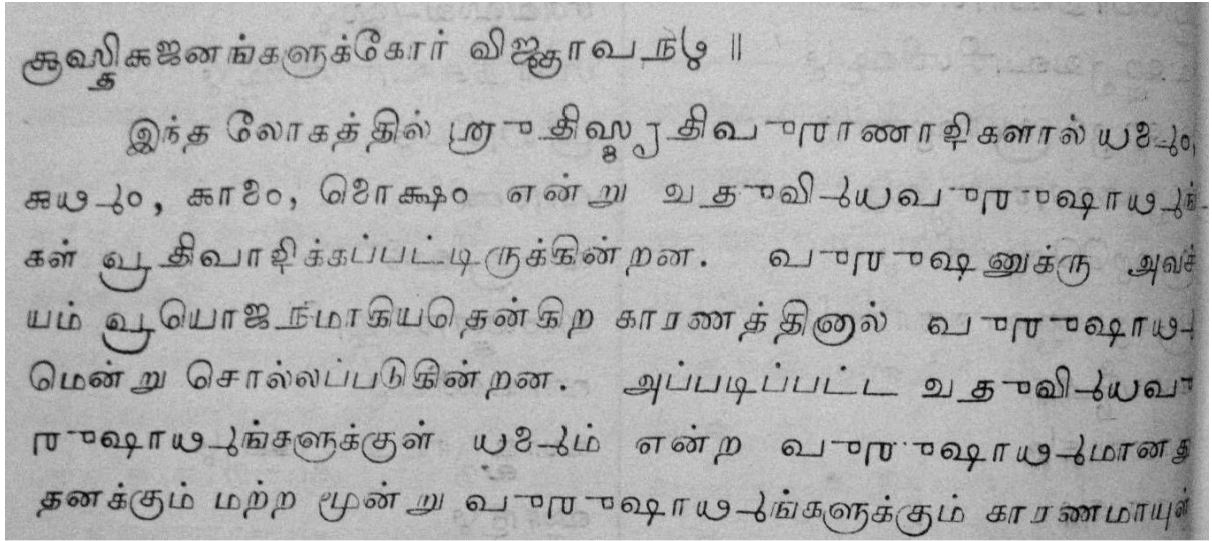
Figure 8.3 : Example of Manipravalam (Sanskritised Tamil) in Grantha script (Sharma, L2/09-372)

Figure 8.3 is an example of Manipravalam (Sanskritised Tamil) in the Grantha script. In order to read the text in the above figure, one has to be familiar with Tamil and Sanskrit. Although the linguistic view adopted in this research was from the perspective of Tamil, the Tamil Brahmans in the study sample acknowledged the Sanskrit influence in the day to day spoken Tamil. The code-mixing of Tamil Brahmans, and some expectations to accurately reflect the Sanskrit sounds in orthography compelled to consider the Grantha script which could benefit the study sample. In this context, Manivannan's (2013) view on orthography, its implication on the community and government's influence on orthography through language policies also needs to be considered. Although Tamil has borrowed six characters from the Grantha, it is insufficient to accommodate all the Sanskrit sounds. Further from the perspective of Standard Tamil, the borrowed Grantha characters are not required. It points to the work of Kailasapthy (1979), Schiffman (1998), Shulman (2016) on language purism and its influence on the orthography.

The words dictated to the participants were tested on Google's speech to text application. It was observed that mispronunciation of a syllable were not accounted for. Mispronounced words were still returned with the right spelling which was contrary to the conceptual model adopted in this research. Whilst the responses of study sample indicate that the study sample

would be able to use google-speech to text application, it was suggestive that the notion of accurate pronunciation as indicated by the study sample was not incorporated as a requirement to use the application. Correct spelling for inaccurate pronunciation in the case of google pointed to Eseonu's (2014) argument of technology overlooking social and language factors in favour of technology.

Figure 8.4: List of words as dictated on google speech to text

The words in figure 8.4 that are in bold indicate the mispronounced words. Some of them that were mispronounced that appeared were *Azhaippithazh, vizhuppuram, pizhai, vaazhkkai, vannam, vilakku, pazhani, kaLai, thoguppaaLar and yezhudhugol*. It was also interesting to observe that it came up with னinstead of ண, despite the correct pronunciation and as also identified in **section 4.1**. The word *Vaalu* despite the correct pronunciation , initially returned as '*paalu'* . This further indicated the nuances of accurate pronunciation and the possibilities of returning a different syllable. The proposed model lays emphasis on accurate pronunciation from the users as a requirement for the application according to the structure of the language which should resolve the output ambiguity.
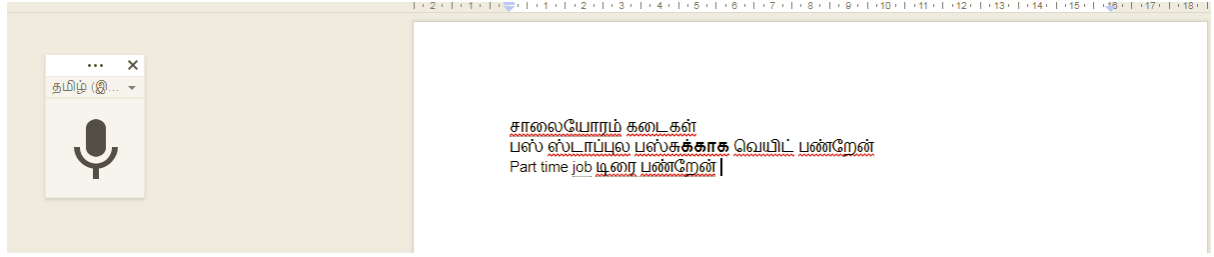
Figure 8.5: Sentences dictated on google speech to text that includes code-switching

Nevertheless, the acceptance of speech to text application as indicated by the study group was primarily dependent on accurate pronunciation, code-switching and code-mixing and its representation in orthography. This research takes the view, that the technology could to some extent accommodate the code-switching and code-mixing aspect, but views on pronunciation, orthographic representation in case of code-switching, language shift are factors that are largely dependent on the society and needs extensive research.

On orthography for Tamil speech to text

"Language and spelling cannot be changed."

"You have to pronounce the word correctly"

"People must change their pronunciation"

On pronunciation

"I am comfortable with Roman orthography"

"Tamil must be written in Tamil orthography. You either write in English or Tamil. Roman orthography is unacceptable"

"Tamil words in Tamil, English in Roman "

"Tamil, Sanskrit and English in respective orthographies"

Ability to use speech to text on the basis of pronunciation

"I have not used any application in Tamil"

"Tried using speech to text in English"

"Comfortable with English"

"Don't use any application on my phone"

"Will be useful to preserve Brahman Tamil"

Empirical link between society, language and use of technology

Influence of the society on language

"Society influences the usage of Tamil"

"What do I gain by learning Tamil?"

"You are not allowed to speak in Tamil in schools"

"You can be punished for speaking in Tamil in school"

"What we speak is not at all Tamil! Sangam Tamil is pure Tamil"

"All our mantras are in Sanskrit"

"Accurate pronunciation is important"

"Brahmans have accurate and clear pronunciation"

"Sanskrit is important to learn Vedas and Tamil is our mother tongue"

Figure 8.6: Some comments from the participants that relates to the proposed user acceptance model.

Figure 8.6 summarises the some of the responses of the participants and provides a holistic picture of the research findings.

8.6 SUMMARY

This chapter focussed on various aspects that this research perceived to be crucial in determining the overall user acceptability of a speech to text application in Tamil. It focussed on the responses provided by the participants and to an extent and has aligned it to the methodological framework that this thesis has adopted to answer the research question. The discussion suggested that, depending upon the user's experience with technology and language (both at social level and with technology), the user is more or less likely to receive a particular language in technology. The qualitative study sample unanimously agreed on accuracy of pronunciation which needs to be reflected in the Tamil script. The participants in the study sample were found to be split in relation to the choice of script therefore the unanimous acceptance, in a holistic sense of the application is questionable and needs further research. But, certain features of the application such as the input, accuracy of pronunciation were acceptable. The attitude towards accurate pronunciation stemmed out of Sanskritic culture and was crucial in studying the expectation of the output for this particular subject group. Interestingly enough, despite exhibiting a reasonably accurate pronunciation throughout the interview, ability to appropriately transliterate and prior experience with technology (English), the study sample did not exhibit behaviour intention to use Tamil language in technology. However, the link to usage of language to behaviour intention to use the language was inferred from the observation that almost no participant had ever regularly used application or technology in Tamil. Code mixing of Tamil and Sanskrit appeared to be integral to the study sample. Most participants however accorded equal priority to Tamil and Sanskrit alike but English vocabulary and in some cases most part of the sentence was almost entirely in English. Therefore it is predicted that determining a language model to accommodate the code-switching, code mixing and orthography would be quite challenging. Mixed lingual model could arguably be suitable but, it is perceived that it could only resolve the speech recognition component. Representation of orthography in case of code-switching and code mixing and readability are additional challenges that need extensive research. From the perspective of user acceptance of speech to text and as observed in the study sample, this research perceives that the prediction of acceptance of speech to text must also account for language maintenance within the cultural context.

Whilst the focus was on acceptance of speech to text, one of the indications as a result of the qualitative interview suggested that it was important to consider how the language was spoken and to what extent the user was able to exercise command over the language. This leads to the question:

Is there a need for a Tamil speech to text application when the users feel comfortable and convenient in English?

CHAPTER 9 - PERSONAL REFLECTION

AIM

The aim of this chapter is to provide an account of personal experience and reflection throughout the research process. This chapter is important since the researcher has also taken the positionality of an insider and phenomenology as a source of data and therefore it is important to gain an insight of the researcher's experience along with the journey throughout the PhD research process which would appreciate the persuasion and decisions taken at different stages. This chapter would use the first person. When describing observations and experiences, this chapter, in order to maintain anonymity has refrained to provide references to institutions, individuals and relationships. But points to the context within which the interaction occurred.

9.1 TRANSITION AND TAMIL NADU FROM 2014- 2018

In 2011, I started to occasionally use Tamil in Roman script in iPhone to communicate with my Tamil friends. English continued to be the language for communication via technology with a vast majority of my friends and relatives. In 2013, when iOS introduced Tamil keyboard, I gradually started to use Tamil orthography where possible. Usage of Tamil in Tamil orthography evoked mixed response. A few were very pleased, a few were unhappy, a few felt that I was not being 'professional' , a few pondered on how it benefited me, while some cautioned me against the usage of Tamil orthography in order to be perceived as 'educated' and less 'chauvinistic' . With these responses from well wishers, I chose to pause, reflect , observe and exercised a conscious choice of using the language at a social sphere. But this time, being an insider and in the native context, I did so as an 'outsider'. I observed and experienced that the usage of language was contextually very rigid. For instance, it was okay to speak in Tamil at home but it wasn't okay to speak Tamil at a posh restaurant. It was okay to speak in Tamil to a road side vendor but it wasn't okay to speak in Tamil to a bank manager or head of an institution as also seen in the discussion chapter. It was okay to codeswitch between Tamil and English but a Tamil only sentences with all technical words appropriately translated was perceived to be 'chauvinistic'. It was not socially acceptable to request a menu at a posh restaurant in Tamil because, of the underlying assumption that everyone visiting the restaurant should be knowing English. Most of the daily conversation

happened in Tamil-English code switching and this seemed to be socially acceptable and yet be a Tamil (referring to identity). Through the conferences, I had the opportunity to visit some universities and colleges to interact with the students. I was given the opportunity to deliver guest lectures on topics related to my research. The interactions in the field to a great extent helped to shape my research and also allowed me to challenge my own prejudices, notions and observations. It enabled me to develop the skills to articulate and further refine my research.

Some of the very interesting comments from the field that I did not expect were:

" Wow! You speak Tamil so fluently despite living in the UK"

"Why are you doing research in Tamil? What do you gain?"

"We want Tamil but no one speaks proper Tamil these days"

I attended and presented in four international conferences, three of which focussed on technology in Tamil. I made a conscious choice to present in a conference that focused on Tamil and technology to receive critical feedback on my work. It helped me not only in appraising my research from an audience familiar with the nuances of Tamil language but also to keep myself updated on the latest development in Tamil technology and the challenges the language faced in technology. The conference was particularly useful since it equally focussed on technology and language. These conferences enabled me to disseminate interim observations to the wider society for constructive and critical feedback.

9.2 CHALLENGE OF BEING AN INSIDER AND OUTSIDER

It was incredibly difficult to separate myself from the theme that I was already immersed in. As a result of my insider positionality, I decided to take a holistic approach with an aim to provide an overall view of the context within which the user acceptance of speech to text application was being studied. The insider status was very useful because, I was able to live through the transition and observe how 'others' viewed the change as an 'insider'. I was also able to observe how others responded to me as an 'outsider' when I displayed the traits of an 'insider'. The underlying assumption was that my insider positionality and the environment were inseparable. The 'outsider' positionality enabled me to question some of my own practices, beliefs, stereotypes and assumptions. My attitude of:

'Why should I use Tamil in technology' when I was an insider in the native context changed to

'*Why shouldn't I use Tamil in technology?'* when I approached the same context as an outsider.

The insider position helped me not just to reflect on my experiences but also helped me to observe if something had changed between then and now.

9.3 LANGUAGE, PRONUNCIATION AND MY EXPERIENCE

With Tamil as my mother tongue, I am able to read, write and speak in Tamil, English and Hindi. I possess some conversational skills and very basic reading skills in Malayalam, Kannada and Bengali. In addition to all these, I am able to read and write in Sanskrit. Indian languages like Sanskrit, Hindi, Nepali, Marathi and many other Indo- Aryan languages share the same script- Devanagari. And the Dravidian languages such Tamil, Telugu, Kannada and Malayalam have different scripts. Since Malayalam as language, is an offshoot of Old Tamil and Sanskrit, the script is strikingly similar to Tamil. I always had the urge to learn new languages which helped me not just to appreciate the cultural and linguistic diversity but also helped me to look from a native perspective when I visited these regions.

In the year 2000, I learnt to recite a portion of Vedas. It was at that time, I was first introduced to the oral tradition of transmitting knowledge. I did not have a book. The *Shastrigal* (The teacher . The same term could also be used for a priest.) taught me a portion of *Yajur Veda- Sri Rudram and Chamakam*. If there was a mispronunciation of a syllable in word, he would repeat it again, break the word which had to be repeated aloud for certain number of times before he moved to the next. In this process, it took me about 40 days to learn a tiny portion of the Vedas without the book, not compromising on the oral tradition and with great emphasis on pronunciation and intonation as seen in chapter 2.

I was also formally trained in Carnatic classical music. I had learnt classical music (songs composed in Tamil, Kannada, Telugu, Sanskrit and Hindi) for over a decade. The experience of learning Carnatic classical was an interesting one. Roman orthography was used to write the notation and lyrics. But the subtle differences in the pronunciation (depending upon the language) was corrected by the teacher. Therefore, although, the lyrics were in Roman orthography, depending upon the language, I knew how to pronounce the syllables accurately enough.

I had a natural flair for the English language since my childhood and just before my elocution competition, I mispronounced the word 'pronunciation'! At that point, I did not know – the right way of pronouncing the words in English and therefore, I followed my

English teacher's pronunciation at school. I did not know the difference between British and American English either for a very long time. For quite some time, I was involved in master of ceremonies, English dramas and speeches. English grammar and pronunciation were given a lot of emphasis. The focus on pronunciation was also observed in other languages such as Tamil, Sanskrit, Hindi, Kannada and Malayalam. The accuracy of pronunciation, in my experience had always been central in learning or practising a language. People (friends, teachers, native speakers of other languages) have always corrected when I mispronounced a syllable or word. The notion of accurate pronunciation, in my experience has always existed amongst speakers of various language including English teachers who spoke English as a second language. This is to suggest that accuracy of pronunciation and perception on accurate pronunciation is not only in Tamil but also in other languages and other fields such as music. But the idea, rigour and discipline of accurate pronunciation comes from the Vedic philosophy as discussed in **chapter 2.**

This experience of mine was consistent within the Brahman community (as seen in the sample) in Tamil Nadu, but my observation suggested a sharp contrast with the others. I observed that the accuracy of pronunciation was a Brahman attitude owing to the connection with the philosophy but that did not mean that they alone got to accurately pronounce the syllables in Tamil or Sanskrit or both. The native Tamil speakers of northern districts of Tamil Nadu, do accurately pronounce 'zha' and 'la'. In Chennai, sometimes, 'zha' takes the 'ya' form (Example: *Vaazhaippazham* becomes *vaayappayam*). The interesting observation was that the issue of pronunciation occured with the very native speakers of the language. A Brahman's view on their ability to accurately pronounce Tamil syllables was sometimes viewed as arrogance. But this view, in my experience was not usually conveyed in its entirety by citing the relevant philosophy that was adored and the rigour (as in my own case) that made it possible. Pronouncing the syllables, just the way they ought to be pronounced, especially by the native speaker, in my conviction was only 'natural'.

As a result of my experience and observations, applying a linguistic view in a software application like speech to text, with an intention to make it useful and available to all the native Tamil speakers regardless of their geography and philosophical affiliations, in my observation would serve little functional value owing to the different perceptions, notions and views on Tamil pronunciation. But at the same time, the aspect of pronunciation cannot be ignored. As a result of my experience with other Indian languages such as Malayalam, Kannada, Hindi, the aspect of pronunciation needs to be taken into account when dealing

with other Indian languages as well. My knowledge of Sanskrit- that is usually argued as the mother of all Indian languages- not only helped me to realise and use appropriate Grantha scripts in Tamil, but also subtly helped me to understand, how different is Tamil in its pronunciation and its representation in the orthography. In my view, I would argue that in order to take a holistic approach within the Indian context, atleast the basic knowledge of Sanskrit and Tamil are essential. My upbringing and practices as a Tamil Brahman contributed to the knowledge of Tamil and Sanskrit alike as seen in **chapter 2**.

9.4 CODE MIXING , CODE SWITCHING AND THE NOTION OF TAMIL

Ever since my childhood, I spoke a highly Sanskritised Tamil (Manipravalam). I was known for my incomprehensible Tamil by the non Brahmans! But that was Tamil to me. And that was because of my perception of what was Tamil. In my presumption, this could also hold true for others who code-switch between Tamil and other Indian languages not necessarily restricted to, but including English especially in border regions of Tamil Nadu. That was the kind of Tamil I grew up with. Of course, in the later years, I became familiar with the Non Sanskritised Tamil. But Sanskrit remained an integral part of my spoken and written Tamil. But for the other native Tamil speakers, I did not speak 'Tamil' owing to the usage of disproportionately large Sanskrit vocabulary.

Tamil-English code switching, in my observation was not only quite common but was also peculiar. A Tamil-English code switched sentence was 'acceptable' which was colloquially referred to as '*Tanglish*' but a Tamil-Sanskrit code switching and code mixing was unacceptable. This notion on what is Tamil or in other words, which forms of spoken Tamil were acceptable were mutually contradicting. Nevertheless, I observed that *Sangam Tamil* was constantly referred to as 'Proper and pure' Tamil both by the Brahmans and the Non Brahmans. But that is significantly different to the contemporary Tamil. However, in my experience, usage of *Sangam Tamil* or *Literary Tamil* significantly reduces the issue of code-switching and code mixing.

In a couple of Tamil seminars and conferences that focussed on technology, I observed that sessions that focused on the choice of script especially in case of Unicode were often divided on orthography with one group advocating Grantha (Tamil script that includes the borrowed Sanskrit syllables) while the other opposing the Grantha script. This is suggestive that an application in Tamil needs to carefully consider the language and orthography aspects.

I have also experienced and observed a definite perception or idea of who 'can' be a Tamil and who cannot. I have been in circumstances where I had to explain my Tamil roots and heritage only to be dismissed by the 'other' that a Brahman cannot be a Tamil! Nevertheless, in my experience, the Tamil Brahmans have taken great pride in both the Tamil and Brahman identity also as supported by Fuller & Narasimhan (2014). To me, both coexist and this identity merits a view without any prejudice. By explaining my Tamil heritage, I have acknowledged others perception on my identity.

9.5 TAMIL AND ROMAN ORTHOGRAPHY

I have used both Tamil and Roman orthography in various settings- for day to day communication and in the more formal fine arts. Because of the familiarity with the word and its pronunciation, the Roman orthography served more as a reminder. Nevertheless, the mispronunciation and confusion on pronunciation prevailed when I first encounter a word. For example: *kallu* and *kaLLu.* The choice of orthography was also dependent on many other factors in the native context such as the teacher being unfamilar with the orthography, or when presenting to a diverse audience who may not be familiar with the Tamil orthography. The transliteration has existed in Tamil Nadu for quite some time. For instance, the names of the street and roads in Chennai are transliterated. For example: *Anna Salai, Kamarajar Salai, Iyal Isai Nadaga Mandram,* and so on. I never felt compelled to use Tamil orthography in any setting while in the native sphere as an 'insider'. Nor as an 'outsider' did I felt the need to use Tamil orthography. Interestingly, in places where I did use Tamil orthography by choice especially in railway booking forms (in Chennai) and bank forms, I was requested to re-write in English. As an outsider I was able to observe the lack of need for Tamil even in native space. From my experience which may be subjective, one need not know to read, write or speak Tamil to survive in Chennai. In my case, I appreciate Tamil as my mother tongue and just as I use English in all the fields (technology, day to day communication, entertainment etc.), I tried extending the same to Tamil as well. I observed that Tamil was only well received in classical music, entertainment , literature and philosophy as opposed to a Tamil that I expected to pervade in all fields like education, day to day business transactions and more importantly in technology, in the native region. In seminars and conferences, the theme of 'challenges in selling Tamil software' emerged. It was interesting to observe that there were little takers for Tamil software despite pricing it less than English software.

9.6 SUMMARY

My responses would have been very similar to many of the interviewees. This indicates that I was not very different to my interviewees when I was an insider in the native space. However, I made a conscious effort to switch from using English to Tamil in Roman orthography to Tamil in Tamil orthography. It took me almost four years to get to a reasonably comfortable position of using Tamil in Tamil orthography with a positive attitude that '*Its okay to use my mother tongue in technology*'. I also noticed that it did bear subtle influences on the social circle. Friends and relatives who critiqued and questioned my position on using Tamil in Tamil orthography in technology, in four years time to my surprise, started to respond in Tamil in Tamil orthography. In the process, people interacted with me and brought to surface some of the very genuine issues in using Tamil in Tamil orthography which I initially assumed to be lack of interest. Some of the reasons for not using Tamil in Tamil orthography were:

1. Phone or system not supporting Tamil font (In some cases for economic reasons linked to affordability to buy a phone that does support Tamil).
2. Inability to transliterate or use a Tamil keyboard.
3. Readability issues.
4. Lack of awareness.

My perception and assumption on mis pronunciation was also corrected when I interacted with several people. I learnt that the issue of pronunciation was a social and educational one. In some places especially in Tirunelveli, I observed that the native speakers were unaware of the actual pronunciation of 'Zha'. They were neither taught in schools nor were socially corrected in course of their life and could be related to Eskenazi (1999). One of the Tamil teachers in Dindigul admitted the inability to pronounce 'Zha' but affirmed that it must be pronounced 'Zha' and explained how the teacher worked hard to get the right pronunciation. These experiences of mine as an insider has provided invaluable inputs that helped me constantly reflect and refine the research. To some extent, these have strengthened the argument of including pronunciation and adopting a linguistic view for the speech to text application. At the same time, it suggests that applications like speech to text, when designed and developed with a linguistic view could be beneficial in the realm of education. On the other hand, as native speaker, I am critical about technology recognising and accepting

mispronunciations which could perhaps justify the wrong usage of syllables that some people (with whom I have interacted) seem to slowly realise. Although, my strong views on pronunciation could perhaps be a bias, my view is that pronunciation ability at a social sphere contributes to the speech corpus for speech to text application. And when one takes a linguistic design of the application, I have realised that these become inevitably important. Nevertheless, my exposure to other languages and cultures has helped me to stay objective even within the subjective study and less chauvinistic. It has helped me to use my native knowledge and experience to propose a user acceptance model for a speech to text application in Tamil.

These reflections to some extent support the user acceptance model proposed by the researcher and further suggests that this research has dealt with a complex contemporary issue that needs extensive research in the field of Tamil speech to text user acceptance.

The PhD process has provided me with some invaluable learning experience. I have learnt some valuable lessons ranging from to avoid general assumptions, to avoid generalisation and most importantly regular back up of data! It has made a profound difference in my writing. The supervision team has provided me the confidence and support to take decisions pertaining to the research. I acknowledge that there are avenues for improvement in the work as indicated in the section that deals with further work and research. But the journey of PhD in itself has enriched my approach to the research process.

CHAPTER 10 - CONCLUSION

The initial aim of this research was to predict the user acceptance of speech to text application in Tamil. Initially, it was aimed at all native Tamil speakers. Eventually, owing to complexities , it was focused on the native Tamil speaking Brahmans.

10.1 CONCLUSION

The research provided an empirical finding of an inseparable link between language- its usage at social level, attitude of native speakers towards the language and the ability to use the speech to text application on this basis. The research took a multi-disciplinary approach in order to support the empirical findings. The findings of the research as seen in chapter 7 supports the idea of considering aspects of language such as code switching, code mixing and pronunciation as an intrinsic requirement of the speech to text application. Although language and culture could be considered as mutually related, from the perspective of development and user acceptance, these need to be treated as separate components. This means that the attributes of the language and its peculiarities would need to be considered at the requirement stage , while the cultural element would need to be considered to evaluate the user acceptance of the application.

The qualitative interviews and quantitative samples of transliteration (**appendix A and B**) provided a useful insight of how this technology could be used to the advantage of a linguistic group purely to preserve the unique hybrid language and heritage. And in this context, the elements of code-switching, code-mixing, pronunciation, which at one stage seemed detrimental, eventually became an indispensable part of the requirement. Equally contrasting was that despite the study sample possessing the attributes of requirement, the cultural lifestyle and social attitude towards using Tamil in technology and more specifically speech to text application seemed to dictate the ability to use Tamil in speech to text. The receptiveness to technology was more prominent in English than in a language which they claimed to be fluent and seemed to culturally associate with. Nevertheless, the idea of speech to text in Tamil was appreciated but its position was challenged in technology. Further, on the basis of the choice of script , code-switching, and associated complexities involved, this research concludes that a speech to text application in Tamil - either entirely in Tamil orthography or entirely in Roman orthography, was not universally appealing or acceptable to the study sample that consisted of native Tamil speaking Brahmans. While philosophical

considerations were highly advantageous to the study sample, varying priorities, experience, and more importantly attitude towards Tamil due to its limited economic value in the native sphere were found to be few causes for what this research perceives to impede the acceptance of the application. Furthermore, the use of phenomenology as a method and source of data contributed to the richness of the discussion and analysis. Phenomenology was key in leading to formulating the user acceptance model, which this research views to be culturally relevant. This work also brings to surface, the potential of native knowledge in technology and attempts to indicate why a native approach is essential to answer a technological question.

Of course, the conclusion cannot be generalised even within the Tamil Brahmans because of small sample size. But the findings provide an interesting perspective of how perceptions and experience in certain culture and language could intervene in acceptance of technology. Arguably, the nuances of language, and the cultural domain to an extent dictated the design requirements of application that was consistent with the syntax of the language. However, the research argues that the application in a particular language must be tested for its acceptance based on the syntax of the language as opposed to the user's usage of language, as adopted in this research. As a result of the findings, the gap between language and technology can be narrowed either by increasing the economic value of the target language through legislations and government policies or through sustained self-initiative with a genuine interest and passion to use technology in one's mother tongue. This research indicated an empirical link between usage of language at a social level and using the technology in that language. This work has suggested that the software feasibility study should include the study of the target users from a language and cultural perspective, with language as a fixed entity, in order to predict the acceptance of applications such as speech to text.

10.2 FURTHER WORK AND FURTHER RESEARCH

Although, this research has accomplished the overall objective, the following are some of the limitations to this work.

As a result of small study sample, the findings are rich, reflective, and contextual. However, the sampling technique adopted was purposive and convenient. Therefore, a different sampling technique for example a random sampling, where and if possible could be used to verify the findings of this research in order to be able to arrive at a more generic conclusion.

The target population chosen for this research were the Tamil speaking Brahmans. However, the usage of such application and the language could be further narrowed to Tamil speaking Brahmans who follow the Vedic occupation. This is from the perspective of knowledge of Sanskrit. The presumption is that the expectation could significantly vary between these two groups. The suggested user acceptance model in this thesis could be used with a different linguistic community to test the user acceptance of a speech to text application. The user acceptance model that is socio-technological in nature should arguably be relevant to any community in the context of user acceptance in language technology. Wider testing of the user acceptance model employed in this thesis and investigation on native perspective in wider diaspora. There also needs to be further work in understanding the cultural and philosophical relevance in technology especially in speech to text in other languages. This research could be further extended by using Hofstede's cultural dimension particularly the power distance, individualism vs collectivism and uncertainty avoidance dimensions within the cultural context. This could be done by considering participants who speak Manipravalam Tamil and are able to read the Grantha script and must be carefully done from an indigenous perspective as opposed to the perspective of a mainstream national culture and sub-culture. The findings of which could then be compared with the finding of this research. Another approach could be to adapt Checkland's Soft System Methodology into the cultural context. Checkland's Soft System Methodology could be relevant since it deals with complex real world issues and conceptions. It would also be useful to explore the extent to which both Hofstede's and Checkland's soft system methodology could be integrated in order to arrive at a more holistic finding.

10.3 SUMMARY

The broad aim of this research which was to predict the user acceptance of speech to text application by native Tamil speakers which was eventually narrowed to Tamil speaking Brahmans. This research has provided a subjective understanding on a holistic context and has answered the research question on the acceptance of speech to text application in Tamil by the Tamil Brahmans. Since the research was on the user acceptance, the need of a prototype was imperative to answer the main research question which is why the research chose a paper prototype. The research has used the software development lifecycle, as a method to develop the paper prototype that was relevant to the context. The evaluation was done through a paper prototype which was consistent with the conceptual design of researcher. Phenomenology and the insider status of the researcher were useful in the design

of the user acceptance model while the outsider status of the researcher was helpful in constantly reflecting, engaging and questioning the responses and experiences that has led to the identification of possible sources that contributed towards acceptance of speech to text application. The limitations of the research project has laid foundations for further work as outlined in the previous section and reiterates on the idea that in an application like speech to text, speech is an input for which the user exercises greater control than the technology that facilitates it.

The following chapter on personal reflection provides a consolidated account of the researcher's experience as an insider that led the researcher to take decisions during the course of the research.

REFERENCES

Aarts, B., Chalker, S. & Weiner, E.S.C. 2014, *The Oxford dictionary of English grammar,* Second edn, Oxford University Press, Oxford.

Acri, A., & Griffiths, A. (2014). The romanisation of Indic script used in ancient Indonesia. *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia*, *170*(2-3), 365-378.

Adel, H., Vu, N. T., Kirchhoff, K., Telaar, D., & Schultz, T. (2015). Syntactic and semantic features for code-switching factored language models. *IEEE Transactions on Audio, Speech, and Language Processing*, *23*(3), 431-440.

Agarwal, R., & Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision sciences*, *28*(3), 557-582.

Ahmed, U. Z., Bali, K., Choudhury, M., & Sowmya, V. B. (2011). Challenges in designing input method editors for indian lan-guages: The role of word-origin and context. In *Proceedings of the Workshop on Advances in Text Input Methods (WTIM 2011)* (pp. 1-9).

Ajzen, I. (2011). The theory of planned behaviour: reactions and reflections.

Aktay, y. (1998). Some remarks on martin heidegger's path of questioning technology and phenomenology. *selçuk üniversitesi edebiyat fakültesi dergisi (sefad)/selçuk university journal of faculty of letters*, (12), 167-184.

Auer, P. (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International journal of bilingualism*, *3*(4), 309-332.

Aytürk, İ. (2007). Attempts at romanizing the Hebrew Script and their failure: Nationalism, religion and alphabet reform in the Yishuv. *Middle Eastern Studies*, *43*(4), 625-645.

Bapat, A. V., & Nagalkar, L. K. (2008, December). Phonetic Speech Analysis for Speech to Text Conversion. In *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on* (pp. 1-4). IEEE.

Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, *56*, 85-100.

Beyer, H., & Holtzblatt, K. (1998). Contextual Design Defining Customer-Centered Systems Morgan KaufmanPublishers Inc.

Bright, W. (1978). A Linguistic Study of Early Manipravalam (Book Review). *Language, 54*(1), 248.

Brown, S. A., Massey, A. P., Montoya-Weiss, M. M., & Burkman, J. R. (2002). Do I really have to? User acceptance of mandated technology. *European journal of information systems*, *11*(4), 283-295.

Burton-Jones, A., & Hubona, G. S. (2006). The mediation of external variables in the technology acceptance model. *Information & Management*, *43*(6), 706-717.

Canagarajah, A. S. (1995). The political economy of code choice in a "revolutionary society": Tamil-English bilingualism in Jaffna, Sri Lanka. *Language in Society*, *24*(2), 187-212.

Canagarajah, A. S. (2008). Language shift and the family: Questions from the Sri Lankan Tamil diaspora. *Journal of Sociolinguistics*, *12*(2), 143-176.

Caplan, L. (1995). Creole world, purist rhetoric: Anglo-Indian cultural debates in colonial and contemporary Madras. *Journal of the Royal Anthropological Institute*, 743-762.

Carroll, J. M. (2000). Five reasons for scenario-based design. *Interacting with computers*, *13*(1), 43-60.

Chan, J. Y., Ching, P. C., Lee, T., & Cao, H. (2006). Automatic speech recognition of Cantonese-English code-mixing utterances. In *Ninth International Conference on Spoken Language Processing*.

Chan, W. C. (2003). Phenomenology of Technology: East and West 1. *Journal of Chinese Philosophy*, *30*(1), 1-18.

Chen, K. H. (1997). The decolonization effects. *Journal of Communication Inquiry*, *21*(2), 79-97.

Chinn, P. W. (2007). Decolonizing methodologies and indigenous knowledge: The role of culture, place and personal experience in professional development. *Journal of research in science teaching*, *44*(9), 1247-1268.

Cieri, C., Miller, D., & Walker, K. (2004, May). The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text. In *LREC* (Vol. 4, pp. 69-71).

Ciotti, G. (2017). Teaching and Learning Sanskrit through Tamil. *Indic Manuscript Cultures Through the Ages: Material, Textual and Historical Investigations*, 193.

Coelho, G. M. (1997). Anglo-Indian English: A nativized variety of Indian English. *Language in Society*, *26*(4), 561-589.

Court, D., & Abbas, R. (2013). Whose interview is it, anyway? Methodological and ethical challenges of insider–outsider research, multiple languages, and dual-researcher cooperation. *Qualitative Inquiry*, *19*(6), 480-488.

Cresswell, J. W. (2003). Research design. *Qualitative, quantitative, and mixed methods approaches*, *67*.

Dalmiya, C. P., Dharun, V. S., & Rajesh, K. P. (2013, April). An efficient method for Tamil speech recognition using MFCC and DTW for mobile applications. In *Information & Communication Technologies (ICT), 2013 IEEE Conference on* (pp. 1263-1268). IEEE.

Das, S. N. (2008). Between convergence and divergence: Reformatting language purism in the Montreal Tamil diasporas. *Journal of Linguistic Anthropology*, *18*(1), 1-23.

Dey, P., Sitaram, R., Ajmera, R., & Bali, K. (2009, November). Voice key board: multimodal indic text input. In *Proceedings of the 2009 international conference on Multimodal interfaces* (pp. 313-318). ACM.

Dhore, M. L., Dhore, R. M., & Rathod, P. H. (2013). Transliteration by orthography or phonology for hindi and marathi to english: case study. International Journal on Natural Language Computing (ijnlc) vol, 2.

Djuraskovic, I., & Arthur, N. (2010). Heuristic Inquiry: A Personal Journey of Acculturation and Identity Reconstruction. *Qualitative Report*, *15*(6), 1569-1593.

Dreyfus-Graf, J. (1976, April). Recognition of coded speech (phonocodes). In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'76.* (Vol. 1, pp. 198-201). IEEE.

Dwyer, S. C., & Buckle, J. L. (2009). The space between: On being an insider-outsider in qualitative research. *International journal of qualitative methods*, *8*(1), 54-63.

Englander, M. (2016). The phenomenological method in qualitative psychology and psychiatry. *International journal of qualitative studies on health and well-being*, *11*(1), 30682.

Eseonu, C. I., & Egbue, O. (2014). Socio-cultural influences on technology adoption and sustainable development. In *Proceedings of the 2014 Industrial and Systems Engineering Research Conference Y. Guan and H. Liao, eds*.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype.

Fernando Tola & Carmen Dragonetti Brahmanism and Buddhism: Two antithetic conceptions of society in ancient India (Institute of Buddhist Studies Foundation, Argentina)

Fowler, M. (1954). The segmental phonemes of Sanskritized Tamil. *Language*, *30*(3), 360-367.

Fuller, C. J., & Narasimhan, H. (2010). Traditional vocations and modern professions among Tamil Brahmans in colonial and post-colonial south India. *The Indian Economic & Social History Review*, *47*(4), 473-496.

Ganesan, N. Dravidian Letters in Tamil Grantha Script.

Geetha, K., & Chandra, E. (2015, January). Automatic phoneme segmentation of Tamil utterances. In *Advanced Computing and Communication Systems, 2015 International Conference on* (pp. 1-4). IEEE.

Geogy, M., & Dharani, A. (2016). A Scrutiny of the Software Requirement Engineering process. *Procedia Technology*, *25*, 405-410.

Ghafournia, N. (2014). Language as a Symbol of Group Membership. *Asian Social Science*, *11*(5), 19.

Ghezzi, C., Jazayeri, M., & Mandrioli, D. (2002). *Fundamentals of software engineering*. Prentice Hall PTR.

Gong, M., Xu, Y., & Yu, Y. (2004). An enhanced technology acceptance model for web-based learning. *Journal of Information Systems Education*, *15*(4), 365.

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, *28*(3), 300-311.

Goyal, H. R., & Koolagudi, S. (2013). Hindi Number Recognition using GMM. *International Journal of Computer Applications*, *63*(21).

Griffith, T. L. (1998). Cross-cultural and cognitive issues in the implementation of new technology: focus on group support systems and Bulgaria. *Interacting with Computers*, *9*(4), 431-447.

Hanani, A., Russell, M. J., & Carey, M. J. (2013). Human and computer recognition of regional accents and ethnic groups from British English speech. *Computer Speech & Language*, *27*(1), 59-74.

Hannay, J. E., Sjoberg, D. I., & Dyba, T. (2007). A systematic review of theory use in software engineering experiments. *IEEE transactions on Software Engineering*, *33*(2), 87-107.

Hendry, D. G., Mackenzie, S., Kurth, A., Spielberg, F., & Larkin, J. (2005, April). Evaluating paper prototypes on the street. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1447-1450). ACM.

Herman, S. J. (1995). Accuracy of a voice-to-text personal dictation system in the generation of radiology reports. *AJR. American journal of roentgenology*, *165*(1), 177-180.

Hix, D., & Hartson, H. R. (1993). *Developing user interfaces: ensuring usability through product & process*. John Wiley & Sons, Inc..

http://censusindia.gov.in/2011-prov-results/data_files/tamilnadu/3.Tamil%20Nadu_PPT_2011-BOOK%20FINAL.pdf   (last accessed May 2018)

http://www.unesco.org/education/GMR2006/full/chapt6_eng.pdf (last accessed May 2018)

https://www.unicode.org/L2/L2009/09372-grantha.pdf (last accessed April 2018)

http://oppidanlibrary.com/tamil-letters/tamil-letters-font/ (last accessed April 2018)

https://www.shu.ac.uk/research/ethics-integrity-and-practice (last accessed August 2018)

Hwang, Y., Al-Arabiat, M., & Shin, D. H. (2016). Understanding technology acceptance in a mandatory environment: A literature review. *Information Development*, *32*(4), 1266-1283.

Ihde, D. (1995). *Postphenomenology : Essays in the postmodern context* (Northwestern University studies in phenomenology and existential). Evanston, Ill.: Northwestern University Press.

Jayan, J. P., Rajeev, R. R., & Rajendran, S. (2011). Morphological analyser and morphological generator for malayalam-tamil machine translation. *International Journal of Computer Applications*, *13*(8), 0975-8887.

Johnson, J., & Henderson, A. (2002). Conceptual models: begin by designing what to design. *interactions*, *9*(1), 25-32.

Johnston, C. M., Wallis, M., Oprescu, F. I., & Gray, M. (2017). Methodological considerations related to nurse researchers using their own experience of a phenomenon within phenomenology. *Journal of advanced nursing*, *73*(3), 574-584.

Kailasapathy, K. (1979). The Tamil purist movement: a re-evaluation. *Social scientist*, 23-51.

Karn, J. S., & Cowling, A. J. (2006). Using ethnographic methods to carry out human factors research in software engineering. *Behavior research methods*, *38*(3), 495-503.

Karthikadevi, M., & Srinivasagan, K. G. (2014, April). The development of syllable based text to speech system for Tamil language. In *Recent Trends in Information Technology (ICRTIT), 2014 International Conference on* (pp. 1-6). IEEE.

Keane, E. (2004). Tamil. *Journal of the International Phonetic Association*, *34*(1), 111-116. Keane, E. (2006). Prominence in Tamil. *Journal of the International Phonetic Association*, *36*(1), 1-20.

Keane, M., Khupe, C., & Seehawer, M. (2017). Decolonising methodology: who benefits from indigenous knowledge research?. *Educational Research for Social Change*, *6*(1), 12-24.

Khupe, C., & Keane, M. (2017). Towards an African education research methodology: decolonising new knowledge. *Educational Research for Social Change*, *6*(1), 25-37.

Kinnunen, P., & Simon, B. (2012). Phenomenography and grounded theory as research methods in computing education research field. *Computer Science Education*, *22*(2), 199-218.

Kiruthiga, S., & Krishnamoorthy, K. (2012, January). Design issues in developing speech corpus for Indian languages—A survey. In *Computer Communication and Informatics (ICCCI), 2012 International Conference on* (pp. 1-4). IEEE.

Krishnasamy, K. (2015). Code mixing among Tamil-English bilingual children. *International Journal of Social Science and Humanity*, *5*(9), 788.

Kroch, A. S. (1986). Toward a theory of social dialect variation. In *Dialect and Language Variation* (pp. 344-366).

Kumar Cheran, S., Vaseeharan, T., & Cheran, E. Optimization of Thamil Phonetic Keyboard.

Kumar, C. S., & Wei, F. S. (2003, December). A bilingual speech recognition system for English and Tamil. In *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on* (Vol. 3, pp. 1641-1644). IEEE.

Kumar, M. (2016). From Hagiographies to Biographies: Ramanuja in Tradition and History.

Kurian, C. (2016). Speech database and text corpora for Malayalam language automatic speech recognition technology. *Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA), 2016 Conference of The Oriental Chapter of International Committee for,* 7-11.

Kurzon, D. (2010). Romanisation of Bengali and Other Indian Scripts. *Journal of the Royal Asiatic Society*, *20*(1), 61-74.

Lamel, L., Messaoudi, A., & Gauvain, J. L. (2009). Automatic speech-to-text transcription in Arabic. *ACM Transactions on Asian Language Information Processing (TALIP)*, *8*(4), 18.

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & management*, *40*(3), 191-204.

Li, Y., & Fung, P. (2012). Code-switch language model with inversion constraints for mixed language speech recognition. *Proceedings of COLING 2012*, 1671-1680.

Liao, H. C., Guan, Y. H., Tu, J. J., & Chen, J. C. (2014). A prototype of an adaptive Chinese pronunciation training system. *System*, *45*, 52-66.

Lionnet, F. (1993). Créolité in the Indian Ocean: Two models of cultural diversity. *Yale French Studies*, 101-112.

López C, C. (2014). Language is the Soul of the Nation: Language, Education, Identity, and National Unity in Malaysia. *Journal of Language, Identity & Education*, *13*(3), 217-223.

Lu, Y., Zhou, T., & Wang, B. (2009). Exploring Chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. *Computers in human behavior*, *25*(1), 29-39.

Lucas, H. C., & Spitler, V. K. (1999). Technology use and performance: A field study of broker workstations. *Decision sciences*, *30*(2), 291-311.

Lukaitis, S. (2013). Applying Hermeneutic Phenomenology to Understand Innovation Adoption. In *Social and Professional Applications of Actor-Network Theory for Technology Development* (pp. 103-116). IGI Global.

Lyu, D. C., Lyu, R. Y., Chiang, Y. C., & Hsu, C. N. (2006, May). Speech recognition on code-switching among the Chinese dialects. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* (Vol. 1, pp. I-I). IEEE.

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and social psychology Bulletin*, *18*(1), 3-9.

Manikumar, K. A. (2014). Impact of British Colonialism on Different Social Classes of Nineteenth-Century Madras Presidency. *Social Scientist*, *42*(5/6), 19-42.

Manivannan, M. M. (2013). Notes on Tamil Orthography–puLLi, kAl, ai, ja, etc. In *Tamil Internet 2013 Conference Papers*.

Manuel, A. 2013, "Linguistic Anxieties: Impact of English on Politics, Mother Tongue and Creative Writing in India", *Language in India,* vol. 13, no. 2, pp. 702-713.

Marrelli, A. F. (2008). Collecting data through focus groups. *Performance improvement*, *47*(4), 39-45.

Martin, B. (2017). Methodology is content: Indigenous approaches to research and knowledge. *Educational Philosophy and Theory*, *49*(14), 1392-1400.

Mathieson, K. (1991). Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. *Information systems research*, *2*(3), 173-191.

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), 1-26.

Medhi, I., Sagar, A., & Toyama, K. (2006, May). Text-free user interfaces for illiterate and semi-literate users. In *Information and Communication Technologies and Development, 2006. ICTD'06. International Conference on* (pp. 72-82). IEEE.

Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard English*. Routledge.

Muniandy, M. K., Nair, G. K. S., Shanmugam, S. K. K., Ahmad, I., & Noor, N. B. M. (2010). Sociolinguistic competence and Malaysian students' English language proficiency. *English Language Teaching*, *3*(3), 145.

Munn, P., & Drever, E. (1990). *Using Questionnaires in Small-Scale Research. A Teachers' Guide*. Scottish Council for Research in Education, 15 St. John Street, Edinburgh, EH8 8JR, Scotland, United Kingdom.

Murthy, K. N., & Kumar, G. B. (2006). Language identification from small text samples. *Journal of Quantitative Linguistics*, *13*(01), 57-80.

Nair, U. (2015). Soft systems methodology for personalized learning environment. *E-Learning and Digital Media*, *12*(1), 34-56.

Neela Das, S. (2011). Rewriting the past and reimagining the future: The social life of a Tamil heritage language industry. *American Ethnologist*, *38*(4), 774-789.

Nistor, N., Lerche, T., Weinberger, A., Ceobanu, C., & Heymann, O. (2014). Towards the integration of culture into the Unified Theory of Acceptance and Use of Technology. *British Journal of Educational Technology*, *45*(1), 36-55.

Pancer, S. M., George, M., & Gebotys, R. J. (1992). Understanding and predicting attitudes towards computers. *Computers in Human Behavior*, *8*(2-3), 211-222.

Park, E., & Joon Kim, K. (2013). User acceptance of long-term evolution (LTE) services: an application of extended technology acceptance model. *Program*, *47*(2), 188-205.

Parsons, M., Nalau, J., & Fisher, K. (2017). Alternative perspectives on sustainability: indigenous knowledge and methodologies. *Challenges in Sustainability*, *5*(1), 7.

Pfleeger, S. L., & Atlee, J. M. (2009). Software Engineering: Theory and Practice (4th.

Pornpanomchai, C., Ngamwongsakollert, P., Tangpitaksamer, P., & Wonvattanakij, C. (2012). Thai-Speech-to-Text Transformation Using Dictionary-Based Technique. *International Proceedings of Computer Science and Information Technology*, *57*, 65.

Prabhakar, O. P., & Sahu, N. K. (2013). A survey on: Voice command recognition technique. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(5).

Radha, V. (2012). Speaker independent isolated speech recognition system for Tamil language using HMM. *Procedia Engineering*, *30*, 1097-1102.

Raghavendra, E. V., Desai, S., Yegnanarayana, B., Black, A. W., & Prahallad, K. (2008, December). Global syllable set for building speech synthesis in Indian languages. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE* (pp. 49-52). IEEE.

Rallabandi, S., & Black, A. W. (2017). On building mixed lingual speech synthesis systems. *Proc. Interspeech 2017*, 52-56.

Rama, G. J., Ramakrishnan, A. G., Muralishankar, R., & Prathibha, R. (2002, September). A complete text-to-speech synthesis system in Tamil. In *Speech Synthesis, 2002. Proceedings of 2002 IEEE Workshop on* (pp. 191-194). IEEE.

Ramachandran R., Khazaei B, Ali A (2015, June). A study of Tamil transliteration and the choice of Roman script for Tamil input. Conference proceedings of 14th International Tamil Internet Conference, Singapore (pp 280-287).

Ramaswamy, S. (1998). Language of the People in the World of Gods: Ideologies of Tamil before the Nation. The Journal of Asian Studies, 57(1), 66-92.

Ramazani, J., & Ramanujan, A. K. (1998). Metaphor and postcoloniality: the poetry of AK Ramanujan. *Contemporary Literature*, *39*(1), 27-53.

Rao, S., & Troshani, I. (2007). A Conceptual Framework and Propositions for the Acceptance of Mobile Services. *Journal of Theoretical and Applied Electronic Commerce Research, 2*(2), 61-73.

Rebman Jr, C. M., Aiken, M. W., & Cegielski, C. G. (2003). Speech recognition in the human–computer interface. *Information & Management*, *40*(6), 509-519.

Renganathan, V. Computational Phonology and the Development of Text-to-Speech Application for Tamil. (INFITT conference proceeding 2008 & 2014)

Ridge, B. (2012). National language planning and language shifts in Malaysian minority communities: speaking in many tongues.

Rivero, L., & Conte, T. (2013). Using an empirical study to evaluate the feasibility of a new usability inspection technique for paper based prototypes of web applications. *Journal of Software Engineering Research and Development*, *1*(1), 2.

Rudisill, K. (2012). Everyday Flamboyancy in Chennai's Sabha Theatre. *Asian Theatre Journal*, *29*(1), 276-290.

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, *14*(2), 131.

Sahoo, S. (2017, August 14). Google launches voice search in eight more Indian languages. *Voice & Data*, p. Voice & Data, Aug 14, 2017.

Salter, D. W., Evans, N. J., & Forney, D. S. (1997). Test-retest of the Myers-Briggs Type Indicator: an examination of dominant functioning. *Educational and Psychological Measurement*, *57*(4), 590-597.

Sandanalakshmi, R., Monfort, V. M., & Nandhini, G. (2013). A novel speech to text converter system for mobile applications. *International Journal of Computer Applications*, *73*(19).

Sankar, S., & Nagarajan, D. S. (2012). A Comparative Study: Data Compression on TANGLISH Natural Language Text. *International Journal of Computer Applications (0975–8887)*, *38*(3), 33-37.

Sankaracarya, & Gambhirananda, S. T. (1972). *Brahma-sutra-bhasya of Sri Sankaracarya: Transl. by Swami Gambhirananda.[2d Ed.]*. Advaita Ashrama.

Saraswathi, S., & Geetha, T. V. (2007, February). Improvement in performance of Tamil Phoneme Recognition using Variable length and hybrid language models. In *Signal Processing, Communications and Networking, 2007. ICSCN'07. International Conference on* (pp. 11-15). IEEE.

Saravanan, V., Lakshmi, S., & Caleon, I. S. (2009). The debate over literary Tamil versus standard spoken Tamil: What do teachers say?. *Journal of Language, Identity, and Education*, *8*(4), 221-235.

Saunders, M., Lewis, P., Thornhill, A., & Wang, C. (2009). Analysing qualitative data. *Research methods for business students. 5th edn. Harlow, Essex, UK: Pearson Education Ltd*, 480-525.

Schiffman, H. F. (1998). Standardization or restandardization: The case for "standard" spoken Tamil. *Language in Society*, *27*(3), 359-385.

Schiffman, H. F. (2002). Malaysian Tamils and Tamil linguistic culture. *Language & Communication*, *22*(2), 159-169.

Schultz, T. (2002). GlobalPhone: a multilingual speech and text database developed at Karlsruhe University. In *Seventh International Conference on Spoken Language Processing*.

Sebba, M. (2010). Linguistic Landscapes: A Comparative Study of Urban Multilingualism in Tokyo Peter Backhaus. Writing Systems Research, 2(1), 73-76.

See, J., Yusof, U. K., & Kianpisheh, A. (2010, December). User acceptance towards a personalised hands-free messaging application (iSay-SMS). In *Science and Social Research (CSSR), 2010 International Conference on* (pp. 1165-1170). IEEE.

Sefa Dei, G. J., & Asgharzadeh, A. (2003). Language, education And development: Case studies from the Southern contexts. *Language and Education*, *17*(6), 421-449.

Shulman, D. (2016). *Tamil*. Harvard University Press.

Simon, S. J. (2007). User acceptance of voice recognition technology: an empirical extension of the technology acceptance model. *Journal of Organizational and End User Computing*, *19*(1), 24.

Slack, F., & Rowley, J. (2000). Observation: perspectives on research methodologies for leisure managers. *Management Research News*, *23*(12), 10-16.

Solomi, V. S., Christina, S. L., Rachel, G. A., Ramani, B., Vijayalakshmi, P., & Nagarajan, T. (2013, November). Analysis on acoustic similarities between Tamil and English phonemes using product of likelihood-Gaussians for an HMM-based mixed-language synthesizer. In *Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference* (pp. 1-5). IEEE.

Solomon, J. (2012). The decline of pan-Indian identity and the development of Tamil cultural separatism in Singapore, 1856–1965. *South Asia: Journal of South Asian Studies*, *35*(2), 257-281.

Sridhar, R., Madhavan, K. V., Nagarajan, S., & Nishanth, S. (2013, August). Incremental language model and dynamic decoder for Tamil chat system. In *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on* (pp. 2057-2062). IEEE.

Srinivasamurthy, N., & Narayanan, S. (2003). Language-adaptive Persian speech recognition. In *Eighth European Conference on Speech Communication and Technology*.

Srinivasan, A. (2013). Real time speaker recognition of letter 'zha' in Tamil language. *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on,* 1-5.

Stephens, R. (2015). *Beginning software engineering*. John Wiley & Sons.

Stolcke, A., Chen, B., Franco, H., Gadde, V. R. R., Graciarena, M., Hwang, M. Y., ... & Ng, T. (2006). Recent innovations in speech-to-text transcription at SRI-ICSI-UW. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(5), 1729-1744.

Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. *Information & management*, *33*(1), 1-11.

Sultana, S., Akhand, M. A. H., Das, P. K., & Rahman, M. H. (2012, July). Bangla Speech-to-Text conversion using SAPI. In *Computer and Communication Engineering (ICCCE), 2012 International Conference on* (pp. 385-390). IEEE.

Sun, H., & Zhang, P. (2006). The role of moderating factors in user technology acceptance. *International journal of human-computer studies*, *64*(2), 53-78.

Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model. *Management science*, *42*(1), 85-92.

Tashakkori, A., & Teddlie, C. (2010). Putting the human back in ''human research methodology'': The researcher in mixed methods research.

Tausif, M.T., Chowdhury, S., Hawlader, M.S., Hasanuzzaman, M. & Heickal, H. 2018, "Deep Learning Based Bangla Speech-to-Text Conversion", IEEE, , pp. 49.

Thakkar, J., Karthikeyan, G., Purohit, G., Thakkar, S., Sharma, J., Verma, S., ... & Singh, S. (2016). Development of macaronic Hindi-English 'Hinglish'text message content for a coronary heart disease secondary prevention programme. *Heart Asia*, *8*(2), 32-38.

Thangarajan, R., Natarajan, A. M., & Selvam, M. (2008). Word and triphone based approaches in continuous speech recognition for Tamil language. *WSEAS transactions on signal processing*, *4*(3), 76-86.

Thangarajan, R., Natarajan, A. M., & Selvam, M. (2009). Syllable modeling in continuous speech recognition for Tamil language. *International Journal of Speech Technology*, *12*(1), 47.

Thennarasu, S. Tamil Pronouncing Dictionary: A promising trend in Language Technology Development.

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology*, *52*(5), 463-479.

Uzman, M. (2010). Romanisation in Uzbekistan past and present. *Journal of the Royal Asiatic Society*, *20*(1), 49-60.

Vaitheespara, R., & Venkatasubramanian, R. (2015). Beyond the Politics of Identity: The Left and the Politics of Caste and Identity in Tamil Nadu, 1920–63. *South Asia: Journal of South Asian Studies*, *38*(4), 543-557.

Vallejo, M. A., Jordán, C. M., Díaz, M. I., Comeche, M. I., & Ortega, J. (2007). Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *Journal of Medical Internet Research*, *9*(1).

Van Manen, M. (2017). Phenomenology in its original sense. *Qualitative health research*, *27*(6), 810-825.

Vazhenina, D., Kipyatkova, I., Markov, K., & Karpov, A. (2012). State-of-the-art speech recognition technologies for Russian language. Proceedings of the 2012 Joint International Conference on Human-centered Computer Environments, 59-63.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Verma, S. (2017). Idea of Bhārata-The Amalgamation of Science and Spirituality.

VijayaRajSolomon, S. S., Parthasarathy, V., & Thangavelu, N. (2016). Exploiting acoustic similarities between Tamil and Indian English in the development of an HMM-based bilingual synthesiser. *IET Signal Processing*, *11*(3), 332-340.

Vliet, H. (2000). Software engineering : Principles and practice (2nd ed.). Chichester: Wiley.

Vu, N. T., Lyu, D. C., Weiner, J., Telaar, D., Schlippe, T., Blaicher, F., & Li, H. (2012, March). A first speech recognition system for Mandarin-English code-switch conversational speech. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 4889-4892). IEEE.

Wee, C. L. (2008). The Indigenized West in Asian multicultures: literary-cultural production in Malaysia and Singapore. *interventions*, *10*(2), 188-206.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information systems research*, *16*(1), 85-102.

Yuvaraja, S., Keri, V., Pammi, S. C., Prahallad, K., & Black, A. W. (2010). Building a Tamil Voice using HMM segmented labels. *International Institute of Information Technology, Hyderabad, India Language Technologies Institute, Carnegie Mellon University, USA, communication*.

அனுராஜ் சிவராஜா, தமிழ் மென்பொருள்களும் மக்கள் பாவனையும், 13 வது உலகத்த தமிழ் இணைய மாநாடு (2013) Anuraj Sivaraja, Tamil software and people's attitude, 13th International Tamil Internet Conference (2013)

APPENDIX

A.1 INTERVIEW TRANSCRIPT IN TAMIL AND ROMAN SCRIPT

An example of how output of speech to text in Tamil and Roman orthography would look like.

Transcript VL

நமஸ்காரம்னு சொல்லலாமா வணக்கம்னு சொல்லலாமா இல்ல hello னு சொல்லலாமா

*வணக்கம்

வணக்கம் சரி, okay இப்போ வந்து இந்த ஆராய்ச்சி வந்து ரெண்டு விதமா பிரிச்சிருக்கேன் ஒண்ணு வந்து அன்றாட வாழ்க்கையை சார்ந்த கேள்விகள் தேனோ நீங்க என்ன பன்றேள் அந்த மாதிரி இன்னொண்ணு வந்து மொழி சார்ந்த கேள்வி அதாவது இப்போ தமிழ் அப்படினா தமிழை வந்து எந்தந்த எடத்துல சின்ன வயசுலேர்ந்து எப்படி நீங்க உங்களோட அனுபத்தை எப்படி எப்படி பயன்படுத்தி வந்திருக்கேள் அப்படினு. மூன்றாவது வந்து தமிழும், தொழில்நுட்பத்துல தமிழோடிய பயன்பாடு அப்படினு மூணு விதமா பிரிச்சிருக்கேன் அதுக்கப்பறம், இது முடிஞ்சதுக்கப்பறம் ஒரு காகித முன்மாதிரி - paper prototype நீங்க evaluate பண்ணனும். என்னுடைய prototype எ நீங்க evaluate பண்ணி அதுல என்ன கருத்திருக்கோ அதை பதிவு செய்யறீங்க.

இப்போ வந்து முதல்ல வந்து உங்களபத்தி நீங்க சொல்லலாம் அதாவது நீங்க எங்க படிச்சேள், எந்த வழி கல்வில படிச்சேள் எப்படி இருந்தது அந்த மாதிரியான ஒரு just general introduction.

* நா வந்து +2 வரைக்கும் அரசு மேல்நிலையப்பள்ளி ல படிச்சேன் **ஆங்கில வழில தான் படிச்சேன் ஆனா தமிழ்ல தான் பேசுவோம்**. ஆங்கிலத்துல பேசத்தெரியாது அதே போல படிச்சது college பண்ணது வந்து BSc. Maths, MSc. Maths BEd பண்ணேன் எல்லாமே ஆங்கில வழி English medium தான் ஆனா தமிழ் அப்படிங்கறது வந்து தமிழ்ல தான் பேசுவோம். English தெரியாது வெளில போய் தான் நா English ஏ **கத்துண்டேன்**. ஆங்கிலம் கத்துண்டது வெளில போய் தான். அது வரைக்கும் தமிழ் தான். **தமிழ தவற வேற எந்த மொழியும் தெரியாது**.

செரி இப்போ வந்து நீங்க படிச்ச அந்த காலக்கட்டத்துக்கும் இப்போ படிக்கற பசங்களோட சூழலுக்கும்?
* நெறையா இருக்கு.

ஒரு எடுத்துக்காட்டு - example
* Example என்னனா நாங்க அங்க படிக்கறச்ச school ல என்ன படிக்கறோமோ அது தான். வீட்டுல எல்லாம் படிக்கறதெல்லாம் ரொம்ப கம்மி. இப்போ பசங்களுக்கு வந்து school**லையும்** படிக்கறாங்க **வீட்டுலையும்** படிக்கறாங்க பட்டிப்போட அந்த இது வந்து ஜாஸ்தியா போச்சு pressure வந்து ஜாஸ்தியா போச்சு. அதுனால வந்து நெறையா difference இருக்கு. பசங்களுக்கு இப்போ வளையாட time**ஏ** கடைக்க மாட்டேங்கர்து. எப்பப்பார் படிப்பு படிப்பு படிப்பு. படிப்ப தவற வேற எதுவுமே கடையாது. அந்த காலத்துல அந்த மாதிரி கிடையாது. நம்ப அந்த importance குடுக்கல படிப்புக்கு. படிப்பு வந்து just like that. அது ஒரு இதுவா தான் இருந்ததே வொழியே படிச்சா தான் முன்னுக்கு வரணும் அப்படினு கடையாது. இந்த காலத்துல படிப்பு படிப்பு படிப்புனு parents வந்து ரொம்போ pressure கொடுக்கறோம் பசங்களுக்கு. அப்படினு நா நனைக்கறேன்.

இல்ல அதையும் தாண்டி வந்து இப்போ வழிக்கல்வி பொழிப்பாடம். இப்போ நீங்க ஆங்கில வழில படிச்சேள்ணு சொல்றீங்க ஆனா அப்போ ரெண்டாம் மொழி தமிழா இருக்கும்ணு நா guess பண்றேன். Okay, இப்போ வந்து வழிக்கல்வி மொழிப்பாடம். வழிக்கல்வி ல படிக்கறதுனால அந்த மொழில புலமை இருக்கும்ணு நீங்க நனைக்கறேளா?
* இருக்கணும்ணு அவசியமும் இல்ல.

இப்போ for example நா English medium la படிக்கறான் so I must know English (participant fills in "know English")
* அது தான் அது தான் அப்படி கடையாது ஏன்னா நா படிச்சதெல்லாம் English medi**யத்தல** தான் படிச்சேன் ஆனா எனக்கு English ல ஒரு வார்த்த ஒரு sentence கூட பேசாத தெரியாது. அப்போ எனக்கு English medium ல படிச்சு என்ன ப்ரயோஜனம்? No use. ஏன்னா படிச்சது English medium ஆ இருந்தாலும் ஒரு sentence கூட என்னால English ல பேசமுடியாது ஏன்னா நம்ம எல்லாம் பேசறது தமிழ்ல தான் பேசிப்போம் teachers நடத்தும் போதும் கூட தமிழ்ல தான் நடத்துவாங்க. Maths ஆ இருக்கறதுனால எல்லாம் அப்படியே board ல எழுதி போட்டு போயிடுவாங்க நாங்க அப்படியே copy பண்ணிப்போம். So அதுனால வந்து ரொம்ப interaction ஏ கடையாது பேசரத்துக்கெல்லாம். நம்ப பேசறதுகூட எல்லாம் தமிழ்ல தான் பேசிப்போம். அது வந்து ரொம்போ என்ன சொல்லறது easy to mingle. தமிழ் நம்போட தாய் மொழிங்கற போது நம்ப இன்னும் freeயா இருக்கலாம். English ல பேசும் போது அப்படி தப்பா பேசிடுவோமோ grammar வந்து தப்பாயிடுமோ அப்படிலாம் கொஞ்சம் இதுவா இருக்கும் தமிழ்ல பேசும்போது நம்போட language அப்போ நம்ப language ல பேசும்போது இன்னும் கூட நம்ப நெறைய இன்னும் இதுவா இருக்கும் comfortable ஆ இருக்கும்.

இப்போ வந்து தமிழ், தமிழ் வகுப்புல நீங்க தமிழ் பேசர்துக்கோ அந்த வகுப்பிலிருந்து வெளில வந்து தமிழ் பேசறதுக்கு வித்யாசம் இருக்கா?
* நாங்க படிச்ச காலத்துல அப்படிலாம் ஒண்ணும் கடையாது. எல்லாம் ஒரே மாதிரி தான் இருக்கும்.

இந்த காலத்துல?
* இந்த காலத்துல கூட ரொம்ப ஒண்ணும் வித்யாசம் தெரியல. **இந்த காலத்துல தமிழை எப்படி பார்க்கறா நா ஒரு** subject **ஆ தான் பார்க்கறாங்க. அத வந்து அத பத்தி தெரிஞ்சுக்கணும் புரிஞ்சுக்கணும் புலமையா இருக்கணும் அது ஏன் அப்படி வந்தது எதனால வந்தது அந்த தேடலாம் வந்து ரொம்பக் கம்மி அது வந்து ஒரு** subject **படிக்கணும்** pass **பண்ணனும்** mark **வாங்கணும். வெளில போணும் அவ்ளோதான். அது வந்து** separate **ஆ எடுத்து படிக்கறவன் வந்து இங்கே ஒரு முட்டாள்அ. ப்படிங்கற ஒரு** brand **தமிழ் நாட்டிலியே இருக்கு.** அது ஏன் எடுக்கல அதாவது தமிழ் எப்போ எடுப்பானா எந்த இதுவும் கிடைக்கல எந்த college**லையும்** seat கடைக்கல அப்படினா அப்போ தான் தமிழ் எடுப்பாங்க. தமிழ் வந்து அந்த தமிழ் சொல்லிக்கொடுக்கற ஆசிரியருக்கே தமிழ் பாட நூல் கேட்டா பிடிக்க மாட்டேங்கர்து, இந்த காலத்ல. ஏன் mark கம்மி ஆயிடுத்து அதுனால தான் நீ இங்க வந்து சேந்தியா? அப்படிங்கிற இதுலா இப்போ தமிழ் இங்க இருக்கு தமிழ் நாட்டுல.

சரி இப்போ வந்து, தமிழை வந்து தமிழ் நாடு அரசு கட்டாயப்பாடமாக ஆக்கியிருக்கு அதுல வந்து நெறையா எதிர்ப்பும் இருக்கு வரவேற்பும் இருக்கு. ஒரு லறுத்து என்னன்னா வெளி மாநிலத்துலேர்ந்து வரவாள்க்கு இல்ல வெளி நாட்டிலேர்ந்து வரவாள்க்கு வந்து அது கஷ்டமா இருக்கும் இன்னொரு தரப்பு என்ன சொல்லறது அப்படினா தமிழ் நாட்டுல இருக்கோம் தமிழ் படிக்கர்துல தப்பில்லை, தமிழ் படிச்சாகணும் அப்படிணா. இது பற்றி உங்களோடிய கருத்து.
* Correct. தப்பே கடையாது. இப்போ நீ ஆந்திரா போனியேனா தெலுங்கு compulsoryயா படிக்கணும், கர்நாடகா போனியானா கன்னடா compulsoryயா படிக்கணும், Bombay போனியானா மராத்தி compulsoryயா படிக்கணும், ஏன் தமிழ் நாட்டுல மட்டும் தமிழ் படிக்கக்கூடாது?

வெளிலிருந்து வரவாளுக்கு கஷ்டமா இருக்கும்...

*ஏன் நம்ப அங்க போனா நம்ப பிடிக்கலையா? நமக்கு தனியா எதாவது குடுக்கறாங்களா? கஷ்டப்பட்டு நம்ப பிடிக்கலையா? ஏன் இங்க இருக்கும் போது படிக்கட்டுமே? அது தெரியாது வெளி மாநிலம்னு கடையாது **தமிழ் பசங்க தமிழ் கண்டிப்பா படிக்கணும் தமிழ் பசங்க, தமிழ் நாட்டு பசங்க எந்த ஊர்ல இருந்தாலும் அவங்களுக்கு தமிழ் எழுத படிக்க தெரியணும். கண்டிப்பா தெரியணும்.**

தொழில்நுட்பம் அல்லது technology அப்படினு சொல்லும் போது உங்களுக்கு முதல்ல நினைவுக்கு வரக்கூடிய மொழி எது?
*Obviously English.

Okay. Why is it?
ஏன்னா தொழில்நுட்பங்கறது it is a universal thing எல்லாரும் ..
தொழில்நுட்பம்ன்னு தமிழ்ல சொன்னாக்கூட English ஞாபகம் வருமா இல்ல technology னு சொன்ன English ஞாபகம் வரும் தொழில்நுட்பம்னா தமிழ்ல relate பண்ணுவேளா?
*தொழில்நுடம், technology எப்படி எதுல சொன்னாலும் அது English ல தான் comfortableஆ இருக்கும் ஏன்னா நமக்கு typing, main ஆ வந்து typing, type பண்ணறது வந்து English ல easy to type and easy to converse. ஆனா தமிழ்ல type பண்ணறது கொஞ்சம் கஷ்டம். என்னோட இதுல opinion ல.

Converse அப்படினு நீங்க சொல்லறது பேச்சு English ல சொல்றேளா இல்ல எழுத்து English ல சொல்றேளா?
* English ல நீ எப்படி எழுதினாலும் தப்பில்லை. இப்போ என்ன 'for' னு எழுதறோம், இப்போ என்ன 'for' என்ன 'f','o"r',ஆ, போடறோம்? 4து போட்டு போயிடறோம். So இப்போ எல்லாமே வந்து minimise ஆயிடுத்து. அதுனால வந்து நீ தப்பா அங்க வந்து type பண்ணும் போது grammarஎல்லாம் பாக்க போறதில்ல. சொல்ற content அங்க புரியர்தா அவ்ளோதான். **ஆனா தமிழ்ல அப்படி கடையாது correctஆ எழுதினா தான் நமக்கு புரியும். அது நம்ம correct ஆ எழுதர்துக்கு தேடணும்** typing **எப்படி இருக்கு** may be **நம்போ** speedஆ, type **பண்ணி** practice, practiceல வரலாம். ஆனா Englishல அப்படி கடையாது. நீ வந்து grammar தப்பா இருக்கலாம், either because za, use, என்னமோ ஏதோ பண்ணி எழுத்திற்றோம்.English நம்ப எப்படி டப்பா எழுதினாலும் நம்ப சொல்ல வேண்டிய விஷையத்த புரிஞ்சிக்கறாங்க, அதுனால communication நு பார்த்தா English better.

இப்போ வந்து தமிழ் நாட்டுல தமிழ் கட்டாயமா படிக்கணும் எல்லாருமே especially தமிழ்க்காரா தமிழ் பசங்க கண்டிப்பா படிக்கணும் அப்படினு சொல்றேள், ஆனா அப்போ அது வந்து ஒரு உரிமையா தமிழ்ல தப்பு செஞ்சாலும், தமிழ்ல எழுதி படிக்கலாமே அப்படிங்கற மனோநிலை ஏன் வரல? அது இருக்கா இல்ல அது வரணுமா இல்ல whats the status?
*தமிழ் படிச்சா அதாவது என்னனா தமிழ் எழுத படிக்க தெரிஞ்சிருக்கணும். அவ்ளோதானேவொழியே இப்போ இங்க இருக்கற எல்லா பசங்களுக்கும் தமிழ் படிக்க தெரியுமானா - தெரியாதுனு தான் சொல்லுவேன் நானு. **அது தமிழ் படிக்கணும்னு ஈடுபாடு வந்துட்டா அங்க பிழையே வாரத்துக்கு chance இல்ல. பிழையே வராது ஏனா அதுல ஒரு ஈடுபாடு ருக்கும் போது எதுலயும் ஒரு ஈடுபாடோட செய்யும் போது correctஆதான் இருக்கும் சரியாக தான் இருக்கும் , பிழையே இருக்காதுஅதுல. அப்படியே பிழை இருந்தாலுமே நம்ப சொல்லிக்கொடுக்கலாம்.**

தொழில்நுட்பத்துல தமிழை பயன்படுத்த நீங்க எப்படி ஊக்குவிக்க வெப்பீங்க? How will you motivate or persuade someone to use Tamizh in technology?
CLARIFYING: இப்போ ஒரு எடுத்துக்காட்டு என்னனா நானும் நீங்களும் தமிழ் இப்போ நீங்க ஒங்கக்கூட நா வந்து பேசறேன் அப்படினா பேசறச்ச தமிழ்ல பேசிடறோம் ஆனா எழுதும் போது

தொழில்நுட்பத்துல, அந்த மொழி இருக்கும் போது அந்த மொழியை தேர்வு செஞ்சு தமிழ்ல பேசுறோமே, அதே போல நம்ப தமிழிலே text அனுப்பலாம் தமிழ்ல message அனுப்பலாம் அப்படிங்கற அந்த ஒரு motivation நீங்க எப்படி how would you start?

*அது வந்து school லேர்ந்து பள்ளிலேர்ந்து தான் start ஆகணும். இப்போ schoolலலாம் என்ன ஆர்தனா every month, ஒரொரு மாசமும் தமிழ்ல வந்து project வெக்கறாங்க. அந்த project power point வந்து தமிழ் ல தான் இருக்கணும். So, அப்போ வந்து அது வந்த நம்ம வந்து தமிழ்ல சின்ன வயசுலியே நம்ம வந்து அது வந்து இப்போ வந்து ஒரு ரெண்டு project கஷ்ட படுவாங்க type பண்ணர்துக்கு அந்த இதெல்லாம் சேகரிக்கறதுக்கு எல்லாம் ஏன்னா இந்த fontஎல்லாம் வந்து மா**றிண்டே** இருக்கு அது தொழில்நுட்பம் என்ன தான் வளந்தாலும், இந்த fontஎல்லாம் correctஆ அமைய மாட்டேங்கர்து. அதுனால இந்த font correctஆ எங்க கடைக்கறது அதெல்லாம் ஒரு வாட்டி பசங்க setஆயிட்டாங்க நா, மூணாவது வாட்டிலேர்ந்து பசங்க தானா இப்போ வந்து பசங்க வந்து எல்லா பல்லிளியும் வந்து first ரெண்டு வாட்டி கொஞ்சம் தடுமாற்றமா இருக்கு thirdலேர்ந்து அவங்களே correctஆ பண்ண ஆரமிச்சுடுறாங்க. அவங்களே தமிழ்ல வந்து type பண்ணி அவங்களே வந்து presentation, power point presentation குடுக்கறாங்க. So இப்போ வந்து பள்ளிலேர்ந்து start பண்ணனும். school லேர்ந்து

சரி இப்போ தமிழ் அப்படி நா what is your definition of Tamizh? What do you consider as Tamizh? இப்போ அதுக்கொரு விளக்கம் சொல்லணும் அப்படினா, இப்போ நம்ம பேசற தமிழ் வந்து சமஸ்க்ருதம் கலந்த தமிழ். போர்த்துவாக வந்து ஆங்கிலம் கலந்த தமிழ் இப்போ நம்மளும் வந்து we use very liberal English words, பாதி sentence தமிழ்ல பேசுவோம் பாதி English ல பேசுவோம் அப்படியெல்லாம் இருக்கும். So இப்போ நீங்க தமிழ்ல பேசறேன் அப்படினா, what is Tamizh? How would you define the language?

* நம்ப பேசற தமிழ் எல்லாம் தமிழே இல்லனு நா சொல்லுவேன். நம்ப பேசற பேச்செல்லாம் எல்லாமே வந்து, நம்ப எல்லாம் நம்ப English லேர்ந்து கொஞ்சம் எடுத்திருக்கோம், Sanskrit லேர்ந்து கொஞ்சம் எடுத்திருக்கோம், Hindi லேர்ந்து கொஞ்சம் எடுத்திருக்கோம். நம்ப பேசற தமிழ் முழுமையான தமிழே கடையாது.

அப்போ எந்த மாதிரியான தமிழை, தமிழ்னு நீங்க சொல்றீங்க?
*சங்ககால தமிழுக்கு தான் போணோம் அப்படினா. சங்ககாலத்துல என்ன பேசினாங்களோ அது தான் தமிழ்.
Okay- Should technology cater to the ways in which people speak a language, or the way a language is spoken?
* எனக்கு புரில
CLARIFYING: அதாவது Should technology cater to the ways in which people speak or the way it is spoken? அதாவது
PARTICIPANT: Is it taking care of the languageனு கேக்கறியா ? இல்ல இல்ல என்ன கேக்க வறேன் அப்படினா, ஒரு மொழி வந்து சமூகம் வந்து ஒரு மொழியை மொழி நடை அமைச்சிருக்கும். அந்த நடைக்கு தொழில்நுட்பம் ஈடு குடுக்குமா? இல்ல அந்த மொழி எப்படி ஆதி காலத்திலிருந்து பேசிட்டு வரங்களோ, அதுக்கு ஈடு கொடுக்கணுமா?
* அதுக்கேத்தா மாதிரி தான் தொழில்நுட்பம் மாறணும். தொழில்நுட்பத்துகேதா மாதிரி மொழி நடையை மாத்த முடியாது. தமிழ் மொழிக்கேத்தா மாதிரி தான் அது மாறும்

Okay- Can you take me through an experience, where you have used Tamizh in technology?
* Tamizh used in technology CLARIFIES : Any experience, any point in life anywhere in the world. Or you might have seen somebody using it in any country.
PARTICIPANT CLARIFIES மொழி நால ? CLARIFICATION: இல்லல அதாவது smart phoneல இருக்கு இல, இப்போ நா ஓங்களுக்கு message பண்ணா தமிழ்ல பண்றேன் - I am using Tamizh in technological device. அந்த மாதிரி ஓங்களோடிய அனுபவம்?
*நா அந்த மாதிரியெல்லாம் யாருக்கும் அனுப்பினதே இல்ல.

Okay- Being a Tamil and a Brahmin, which of the two languages have you traditionally viewed important, and why- between Tamil and Sanskrit?

* Sanskrit வந்து நம்மளோட இந்த language. புராண காலத்திலிருந்து இருக்கற language. **அதுனால நம்பளோட அந்த இதெல்லாம் வந்து, மந்த்ரம் லாம் வந்து** Sanskrit **ல இருக்கு. அதுனால** mostly **எல்லாருக்குமே வந்து அந்த காலத்துல வந்து** they learn Vedas, **அப்போ** they should know Sanskrit (**கிரந்தம்**). ஆனா நம்ம இருக்கறது பழகறது எல்லாம் தமிழ் அதுனால we should learn Tamizh. அதுனால ரெண்டுமே வந்து important. Sanskritஉம் important Tamizhஉம் important. (15 **நிமிடங்கள்** 15 **நொடிகள்**)

Tell me a time when you were compelled to use a technology in a particular language. and how did you feel about it? இந்த தொழில்நுட்பம் இந்த மொழில தான் பயன்படுத்தணும் அப்படிங்கற சூழல் வரும் போது நீங்க அத எப்படி பார்த்தீங்க?
* நா ரொம்ப கஷ்டப் பட்டேன் ஏன்னா இங்க வந்த மொதல் power point project தமிழ்ல பண்ணும் போது அந்த font கடைக்காம, ரொம்ப கஷ்ட பட்டேன் ஆனா அது பண்ணி தான் ஆகணும், தமிழ்ல பண்ணி தான் ஆகணும் ஆனா அத cut, copy, paste பண்ணும் போது தான் அந்த script, java script, அது வந்து எடுத்துக்கவே இல்ல. அப்போ என்ன பண்ணறது னு தெரியாம என்னென்னமோ பண்ணோம் ஏதேதோ download பண்ணோம் அது ரொம்ப கஷ்ட பட்டோம் ஆனா அந்த script கடைக்கறதுக்கு. அதுக்கப்பறம் வந்து smart phone ல key pad ல வந்து தமிழ் select பண்ணி, key pad அ வெச்சு அதுல நம்ம power point பண்ணோம். இப்போ வந்து அதுலேர்ந்து தான் power point பண்ணிட்டிருக்கோம். ஆனா starting ல ரொம்ப கஷ்ட பட்டோம் அந்த font கடைக்காம.

But அந்த compulsion, கட்டாயமா செஞ்சே ஆகணும் வேற வழியே இல்லங்கற போது, do you see an acceptance of the language in technology?
*Technology ல இருந்தது இல ஆனா எங்கையோ samsung phone ல இருந்தது ஆனா கூகிள்ல்லாம் கிடைக்கல. Phoneல வந்து அந்த options இருக்கும் இல்லையா அந்த language options இருக்கும் இல்லையா? அந்த language options ல தமிழ் போட்டு அதுலேர்ந்து பண்ணேன் நா.
இப்போ தமிழ்ல கண்டிப்பா பண்ணியே ஆகணும்னுங்கறத நீங்க கட்டாய சூழல்னு சொல்றேளா அது இங்க ஆனா தமிழ் நாட்டுல தமிழ்ல பண்ணணும்ங்கறத ஏன் நீங்க கட்டாயம்னு பார்க்கறேள் அது இயல்பா இருக்கணும் இல? It should be natural.
* ஆமா . That is what அத தான் நானும் சொல்றேன். ஏன்னா இப்போ தானே தமிழ் இதெல்லாம் இங்க வரது. தமிழ் நாடுல இப்போ தான் power point பண்ணணும் இதெல்லாம் பண்ணணும், கணினி use பண்ணணும்ங்கறதெல்லாம் இப்போ கொஞ்ச காலமா தானே இருக்கு. இதுக்கு முன்னாடி இதெல்லாம் இல்லியே.

Between Tamil and English, in which language are you confident of expressing yourself as well as carrying out daily activities?
* தமிழ்.

Do you think you have embraced Tamizh as a language and a natural choice for using technological device like smart phone?
* No

மொதல்ல நம்ம ஒரு basic phone தன வெச்சிருப்போம் ஆரம்ப காலத்துல நீங்க பாத்தேள்னா, nokia, அந்த phone தான் வெச்சிருப்போம். இப்போ smart phone வந்திருக்கு. அந்த phone லேர்ந்து இந்த phoneக்குள்ளான பயணத்தை பற்றி உங்களோட அனுபவத்தை சொல்லுங்க. How was the transition? How did you feel about it? அதுக்குள்ள எப்படி நீங்க வளர்ந்து வந்தேள்? அந்த வளர்ச்சில அதுக்குள்ள மொழியை எப்படி பார்த்தேள்? How did you see the language?
* இப்போ வந்து வெறுமே first pager வந்துது இல்லையா? pagerல only நம்ப only message தான் send பண்ணலாம். குறுஞ்செய்தி தான் வரும். அதுக்கப்பறம் nokia வந்துது. Nokia ல வந்து - அப்போ - அப்போல்லாம், என்ன differenceனு கேட்டேயனா, அது phone வெறும் phone. (18 minutes 29 seconds) கூப்பிடலாம் கேக்கலாம் அவ்வோதான் pho**னுக்கு** உண்டான வேலையை மட்டும் அது செஞ்சுது ஆனா இப்போ உண்டானது pho**னுக்கு** உண்டான வேலையை மட்டும் அது செய்யல. நல்லது கேட்டது ரெண்டுமே செய்யுது. கைல இருக்கற phone எடுத்தா நம்ப ஒலகத்தயே பாக்கலாம். அது ஒலகத்த பாக்கற விதம் வந்து அது நல்லது தற்று கெட்டதும் தற்று. ரெண்டும். அது நம்ப என்ன பாக்கறோங்கறது பொறுத்து. **ஆனா என்ன கேட்டா ஆதி கால** phone **தான்**

நல்லது சொல்வேன். இந்த phone அ விட ஆதி காலத்து phone தான் phone phone இனத்துக்கு?- பேசறதுக்கு கேக்கற்துக்கு அவ்வோதான். இந்த phone எல்லாம் அவசியம் இல்ல. ஆனா தொழில்நுட்பம் வளந்ததுனால இப்போ நமக்கு இது தேவையா இருக்கு. (**தேவையாயிருக்கு**).

சரி. பேச்சுரையிலேர்ந்து எழுத்து வடிவத்துக்கு மாற்றக்கூடிய தொழில்நுட்பத்தை பற்றி நீங்க கீட்டிருக்கீங்களா?
*பேச்சுலேந்து எழுத்துக்கு? இல்ல நா கேள்விப்படல.

சரி அப்போ பேச்சுலேந்து எழுத்துக்கு மாற்றக்கூடிய தொழில்நுட்பம் நீங்க ஏதேனும் பயன்படுத்திருக்கேளா?
*இல்ல

Could you think of a time when you had to use a technology when you were influenced by someone to use that particular technology?
* No.

What is owing an iPhone or an expensive smart phone mean to you- Is it passion for technology, is it social status or have you got any other reasons?
* Social status. It is a social status.

Can you give me an example?
எல்லாரும் வெச்சிருக்கா நானும் வெச்சிருக்கேன். அவ்வோதான். If somebody is having that why cant I?

Okay. ஆனா அதுலியே வந்து தொழில்நுட்பமும் இருக்கு இல. So, there are two sides of the same coin. **நா மொதல்லையே வாங்கும் போது , நா மொதல் மொதல்ல** smart phone, iphone **வாங்கும் போது என்னத்துள்ள என்ன இருக்குனே தெரியாது. ஆனா வாங்கி குடுத்தா வாங்கிண்டாச்சு. அதுக்கப்பறம்** i just **நா வெறும் ஒரு தொலைபேசியா தான்** use **பண்ணேன். ஆனா அதுல என்ன இருக்கு?** இப்போ கூட என்னோட கைல இருக்கற phone ல என்னென்ன provisions **இருக்குனு எனக்கு தெரியாது. நா எல்லான்லா** use **பண்ண மாட்டேன். ரெண்டு மூணு இது தான்** use **பண்ணுவேன். அது கூட இப்போ சுத்தமா விட்டாச்சு**.

அப்போ அதுக்கு நீங்க basic phone**ஏ வெச்சிருக்கலாமே** why did you upgrade?
*That is what I am saying, it is a social status.

Okay. Have you ever used SIRI?
*Yes. Use **பண்ணிருக்கேன்** iPhone 5**ல**.

Okay. **அதனோட அனுபவம் பத்தி சொல்லுங்க**.
*அது நம்ப ஒன்னு சொன்னா (சிரிப்பு!) நம்ப ஒன்னு சொல்லுவோம் அது வேற ஒண்ணு கொண்டு வரும்.
அதுக்கப்பறம் அது சரியா வரல because **நம்போல்ட** slang **வந்து அதுக்கு** actual**ஆ** it is based on UK slang. **நம்ப பேசற** English **வந்து அதால** identify **பண்ண முடியல**. So it is **நம்ப எதாவது ஒண்ண சொன்னோம்னா அது வேற எதாவது ஒண்ணை கொண்டு வரும். அதுனால அது நமக்கு** set **ஆகல**.

சரி Have you ever used any application in Tamizh? எதான ஒரு application smart phone லயோ அல்லது computer ல whatever.
* பண்ணேன் ஆனா ஞாபகம் இல்ல இப்போ.

அது தொடர்ந்து செய்ய ஏதேனும் முயற்சி?
*இல்ல ஏன்னா அது தான் சொன்னேனிலையா அந்த font**எல்லாம்** correct**ஆ** கடைக்கல. ஆனா கொஞ்ச நாள் அது try பண்ணேன் ஆனா அது சரியா கடைக்கல.

Between Tamizh and English, which language would you accord more priority in Tamizh Nad and why?

* தமிழ் நாட்டுல தமிழ் தான். அது வந்து நமக்கு வந்து easyயா நமக்கு வந்து communication.Englishஓட நம்பளோட சொந்த நாட்ல தமிழ்ல பேசினா நமக்கின்னும் வந்து நம்பளோட ideas, நமக்கென்ன வேணுமோ அது நம்பளோட ideas வந்து freeயா சொல்லலாம்.

இப்போ what I will do is, I will give you scenarios, அதாவது we are going to debate or brainstorm a particular aspect, so அது பத்தி நீங்க என்ன நனைக்கிறேள். இப்போ வந்து, இப்போ everything is fine, whatever you have said so far. It was fantastic. இப்போ Dominos Pizza க்கு போறோம் அல்லது Forum Mallக்கு போறோம் Body shop, Nike, or Adidias போறோம் அப்படினா, நம்ம தமிழ்ல பேசுவோமா English ல பேசுவோமா?

* தமிழ்ல தான் பேசுவேன். நா போனது இல்ல. அவங்க English ல பேசினா கூட நா தமிழ் ல தான் பேசுவேன்.

Do you see an imposition of இப்போ தமிழ் நாட்டுல வந்து history எடுத்து பாத்தேள் நா we have Anti- Hindi agitation in 1969 in the pretext of நம்மளோட கலாச்சாரம் பாத்திக்கும், நம்மளோட மொழி பாதிக்கும் அப்படினு, which is fair enough according to me. ஆனா எதுக்கு அது செஞ்சாளோ அது English இன்னிக்கி செஞ்சுட்டிருக்கு. What is your take on that?

* இங்க வந்து அது பண்ணது .. எல்லாம் மொழியும் கத்துக்கணும். ஹிந்தியும் கத்துக்கணும் இங்கிலீஷூம் கத்துக்கணும் தமிழும் கத்துக்கணும். ஏன்னா தமிழ் தமிழ வந்து நா என்ன சொல்லுவேன் நா, தெரிஞ்சிக்கணும், தெரியாம இருக்கக்கூடாது. அதுக்காக மத்த language கத்துக்கக்கூடாதுங்கறது கடையாது. எல்லா languageஉம் கத்துக்கணும் ஏன்னா நம்ம தமிழ் நாட்டுலே இருக்க போறதில்லை. நம்ப வந்து we are going to travel all over the world. அப்போ எனக்கு அந்த ஊர்ல போயிட்டு எனக்கு தமிழ் ன்மட்டும் தான் தெரியும். நா Bombay ல போய் நா ரொம்ப கஷ்ட பட்டேன். ஏன்னா எனக்கு வேற எந்த languageஉம் தெரியாது. அதுனால நா அந்த ஹிந்தி கத்துக்கறதுக்கு நா ரொம்ப கஷ்ட பட்டேன். அப்போ அங்க போய் கொஞ்ச கொஞ்சமா ஹிந்தி கத்துண்டேன். அதுக்கப்பறம் வெளில போனோம். வெளில போனா ஹிந்தியும் தமிழும் போறாது. அப்பறம் கொஞ்ச கொஞ்சமா இப்போ கூட fullஆ, English பேசத்தெரியாது கொஞ்சம் பேசுவேன் அவ்ளோதான். English medium ல படிச்சாலும் English பேச தெரியாது. அதுனால எல்லா languageஉம் கத்துக்கணும் தமிழ் முக்கியமா கத்துக்கணும்.

So what would be your advice? உங்களோட advice என்னவா இருக்கும் இப்போ யாராவது so if you are motivated to use a smart phone நீங்க வந்து smart phone எந்த மொழில எதிர்பார்ப்பீங்க?

*English ல தான் எழுதவுங்க பசங்க.

இல்ல ஒங்களுக்குணு ஒரு எதிர்பார்ப்பு இருக்கும் இல, இப்போ வந்து it may be different (participant interrupts.. see..) இப்போ நா வந்து if I write something to a French native, அவன் வந்து French ல இருக்கணும்னு எதிர்பார்ப்பான், because that is his language, அவனோட native so அந்த மாதிரி நமக்குன்னு ஒரு சில எதிர்பாப்புகள் இருக்கும் இல.

* எனக்கு கண்டிப்பா இப்போ whatsapp ல நெறையா message வரது நா, English ல இருந்ததுன்னு நா just அப்படியே scroll down பண்ணிடுவேன். அதே தமிழல இருந்தா I will read. புரியர்தா. நா இப்போ தமிழ்ல இருக்கும் போது நா வந்து every sentence, I can read fast. That is the only reason. I can read fast. English நா I have to read first, then I have to understand, so I அது வந்து எனக்கு தேவையே இல்லை so I will just scroll. அதே தமிழ் ல இருந்தா அது நல்ல ரசிச்சு படிக்கறது, if it is a joke சிரிச்சிப்பேன், i Can share to someone. If it is in English இவ்ளோ பெரிய message ஆ இருந்தா I will just scroll and skip.

So why don't you insist on Tamizh? Compulsory situation ல நீங்களே இருந்திருக்கேள் So why don't you insist? that எனக்கு அனுப்பறதுna நா தமிழ்ல தான் படிப்பேன், தமிழ்ல இருந்தாகணும், why don't you insist? ஒங்களுக்கு எங்கெங்கெல்லாம் opportunity கடைக்கறதோ அங்கங்கேல்லாம் ஒரு insist

* எனக்கு அவ்வளுளாம் இல்ல, வந்தா படிப்பேன் வரலனா விட்டுருவேன். அவ்ளோதான். எனக்கு போய் யார் கிட்டையும் இதுல அனுப்பு அது அனுப்பு லான் இல்ல யார் அனுப்பறாங்களோ so

தமிழ்ல இருந்தா நா படிப்பேன், தமிழ்ல இல்ல நா யாருக்கும் அனுப்ப மாட்டேன். ந எந்த messageஉம் Englishளையும் சரி தமிழலியும் சரி, நா யாருக்கும் message இந்த மாறிலாம் அனுப்ப மாட்டேன். யாரான அனுப்பினா அத ஒக்காந்து படிப்பேன். If it is in Tamizh, enjoy பண்ணி படிப்பேன். English ல இருந்தா just I will scroll on.

PROTOTYPE

இப்போ நா வந்து I will show you the prototype, and ask a few questions and opinion around the prototype. நீங்க என்ன சொல்லணும் அப்படினா whats your opinion on that and feedback on so basically you are evaluating the prototype- paper prototype. இது வந்து I have based it on iPhone 5s ல base பண்ணிருக்கேன். So,

This is the starting that is the home screen. in iPhone 5s இப்படி தான் home screen இருக்கும். இப்போ வந்து, இது வந்து மொதல் திரை, மொதல் screen. இப்போ நம்ம வந்து message அதாவது குறுஞ்செய்தி வந்து let me select, it brings you to the next screen. so when you click on message, this one, ஒங்களுக்கு வந்து எந்த screen would you prefer, out of these two (One with Tamil keyboard and one with Roman?

* It depends on whom I am sending. If it is aTamizh person, I can send in this (Tamizh) ஆமா தமிழ்ல அனுப்புவேன் ஏனா this is more comfortable, I can நா வந்து எனக்கு என்ன வேணுமோ அது அப்படியே type பண்ணுவேன். ஆனா இப்போ தமிழ் தெரியாத ஒருத்தருக்கு தமிழ் ல அனுப்ப முடியாது ல So I will send அப்போ English use பண்ணுவேன்.
Ok

* ஆனா அந்த English ல தமிழ் type பண்ணறது I hate (transliteration)
English words ல தமிழ்ல type பண்ணுவாங்க இல்லையா that I don't like. ஒண்ணு இருந்தா English ல எழுது இல்லனா தமிழ்ல எழுது.

So நீங்க வந்து you are going to say, that you would prefer Tamizh (clarification)
*Yes.

Okay. சரி இப்போ வந்து, இந்த திரை வந்திருக்கு. இந்த திரை வந்த பிறகு, ஒரே நிமிஷம் நா இந்த இத எடுத்து வெச்சுக்கறேன். இப்போ, வியாழக்கிழமை:
இதை நீங்க விரும்புவீங்களா இல்ல இதை நீங்க விரும்புவீங்களா?
*This is correct ழ கிழமை.

This is correct? ஆனா இப்போ இந்த ழாவும் ளவும் உச்சரிப்பு. நெறையா பேருக்கு சரியா வராது அப்போ அவங்க viyaalakkilamai அப்படிம்பாங்க. அப்போ viyaalakkilamai னு சொல்லும் போது வியாழக்கிழமைனு வரணுமா? இல்ல அவா சொல்றா மாதிரி viyaalakkilamai னு தான் வரணுமா?
* வியாழக்கிழமை நா வியாழக்கிழமை தான் வரணும். (interrupts:இல்ல அவா சொல்றா மாதிரி தான் வரணுமா) ழ தான் வரணும், ள லாம் வரக்கூடாது. வியாழக்கிழமை.

தமிழுக்கு அழகு 'ழ' தான் அதுனால அந்தந்த wordக்கு என்னவோ அது தான் வரணும்.

சரி இப்போ வந்து நீங்க தமிழ் English ல எழுதநேதேயில்லை
* ம்ம்ஹ்ம் (No)

சரி நம்ம வந்து மொழி கலப்பு பத்தி பேசுவோம். அதாவது we mix two or three languages, and speak, So, அப்படி இருக்கும் போது எந்த எழுத்துருவத்துல எழுதினா இருக்கணும் இப்போ for example, "நா வந்து 10 o clock க்கு bus stand க்கு வறேன்" அந்த மாதிரி colloquial ஆ எழுதணும்னு வெச்சுக்கோங்களேன்
*colloquial ஆ எழுதும் போது தமிழுல எழுதினா தான் நமக்கு புரியும். நான் அப்படினா nan இல்ல naan போட்டு எழுத்துக்கூட்டி படிக்கர்துக்குள்ள அது English ல தமிழ English ல translate பண்ணி எழுதும் போது அது நமக்கு புரியர்த்துக்கு ரொம்ப நேரம் எடுக்கும். அதுக்கு இந்த

மாதிரி தமிழ்ல type **பண்ணிட்டா** easy**ஆ இருக்கும் படிக்கர்துக்கு. தமிழ தமிழா எழுதணும்** English **அ** English **ஆ எழுதணும்.** English **ல தமிழ் எழுதக்கூடாது.**

சரி அப்போ ஆங்கில வார்த்தை ஆங்கிலத்துல எழுதணும் சொல்றேளா?
*ஆமா

இப்போ for example, நான் 12 O Clock க்கு வரேன் அப்படினா அந்த 12 O Clock ஆங்கிலத்துல இருக்கணுமா?
*இல்ல அது ஏன் why நீ full**ஆ** English**ல** எழுதிட்டு போயேன்? I am coming by 12 O Clock**னு**?

பேசறது அப்படி வரது இல?
*எது?

நம்ம என்ன பேசறோமோ அது தானே அங்க வரணும்?
* அப்போ தமிழ்ல type பண்ணேன் எதுக்கு English**ல** type பண்ற?

நா என்ன சொல்றேன் நா, இது வந்து பேச்சுரை, நம்ப எப்படி பேசறோமோ, அது தானா வந்து தொழில்நுட்பம் மாத்தி குடுக்கும்.
*Ok

அப்போ நான் 12 O Clock க்கு வரேன் அப்படினு நம்ம தான் பேசறோம். அப்போ நம்ம பேசறா மாதிரி தான் வரும்
* (Overlaps) நான் 12 மணிக்கு வருகிறேன்னு சொன்னா போச்சு.
நா 12 மணிக்கு வரேன்னு சொல்லு.

அந்த முயற்சி நம்ம எடுக்கணுமா இல்ல தொழில்நுட்பம் எப்படி சொன்னாலும் மாத்தணுமா?
*நம்ம தான் எடுக்கணும். தொழில்நுட்பம் எப்படி மாத்தும்.
நம்ப பேசற்து அது தான் அந்த Siri க்கு தான் வரணும் அப்படினா back to siri. siri வந்து அவன் வந்து என்ன வெச்சிக்கிறானோ அது தான் வரும். அப்போ நம்ப என்ன பேசறோமோ அது தான் அது recognise பண்ணிக்கும். நீ தொழில்நுட்பம் என்ன தான் பண்ணாலும் நம்ப என்ன பேசறோமோ அது தான் அது recognise பண்ணிக்கும் அப்போ நம்ப correct**ஆ** பேசினா தான் அது correct**ஆ** குடுக்கும் அது.

சரி இப்போ வந்து so you,
*அதுக்கு தெரியாது இல நம்ம மன நிலை நம்ப என்ன பேச போறோம், என்ன topic ல பேசபோறோம் அது நமக்கு என்ன வேணுங்கறது தொழில்நுட்பத்துக்கு தெரியாது. நமக்கு தான் தெரியும். Si நம்ப குடுக்கும் போது நம்ப correct**ஆ** குடுக்கணும்.

Okay Okay so there are two aspects தமிழ்ல வந்து நெறையா இது இருக்கு slang நெறையா dialects இருக்கு பொது தமிழ்னு சொல்றோம், இலக்கிய தமிழ்னு சொல்றோம் அதாவது வட்டார வழக்கு மதுரை தமிழ் சென்னை தமிழ், இதுவும் தாண்டி ப்ராமின் தமிழ் அப்பறம் இஸ்லாமிய காரர் வந்து அரபிக் கலந்த தமிழ் பேசுவார் அந்த மாதிரி So, இப்படி இருக்கும் போது, நெறையா வந்து accuracy of pronunciation உங்களோடிய பார்வைல எப்படி இருக்கு how do you raise the accuracy of pronunciation amongst all of us?

* அதாவது ஒவ்வொத்தருக்கு ஒரொரு slang, இங்க வந்து நம்ப தண்ணின்னுவோம், வெளில Brahmin **வந்து ஜாலம் சொல்வோம் அந்த மாதிரி அந்த அவரு** particular **இதுக்கே வந்து** different **பேர் இருக்கும்**. **அதான் வட்டார வழக்கு தான் ஒரொரு இதுக்கு ஒரொரு பேர் இருக்கு**.

வட்டார வழக்குல பேசறேன் அதுனால its okay. For example, கீழ வா, கீள வா அந்த ள பயன்பாடு வந்து பெரும்பாலும் என்ன சொல்றா அப்படினா, கீள வாங்கறது its okay ஏன்னா நம்ம வட்டார வழக்குல பேசறோம். சொல்லறது புரிஞ்சா சரி அப்படிங்கற ஒரு போக்கு இருக்கு. இந்த போக்க நீங்க எப்படி பாக்கறேள்?
*அது அப்படியே பழகி போச்சு வேற ஒண்ணுமில்ல. இப்போ சென்னை தமிழ் எடுத்துண்டா கேக்கறதே இதுவா இருக்கும். ஆனா பேச பேச அவங்களோட சேர்ந்து பேச பேச நம்ப தமிழ் மறந்து போய்டும் அவங்க என்ன பேசறாங்களோ அது தான் நமக்கு வரும். அதுனால நம்ப எந்த எடத்துல எப்போ எப்படி பேசறோங்கறது நம்ம நம்மளோட நம்ம சுத்தி இருக்கறவங்க எப்படி பேசறாங்க அதுவும் ரொம்ப முக்கியம். இப்போ பல மொழிகள் தெலுங்கு பேசறவங்க தெலுங்கு கலந்த தமிழ் பேசுவாங்க. இருந்தாலும் அவங்க கூட கலந்து பேசினா அந்த தெலுங்கு கலந்த தமிழ் தான் நமக்கு வரும். நம்ப பாஷ நமக்கு மறந்து போயிடும். So நம்ப எந்த எடத்துல இருக்கோம், சுத்தி என்ன பேசறாங்க, அதெல்லாம் depend ஆயி தான் இருக்கு. அது தான் வட்டார வழக்கா மாறிருக்கு. அப்படினு நா நினைக்கறேன்.

Okay so how would you, what would be your recommendation to அதாவது நம்பளோடிய அந்த பாஷையை காப்பாத்தி பேசணும்ங்கறதுக்கு what's your take and recommendation?
* Schoolலேர்ந்து ஆரம்பிக்கணும்.

அப்போ do you recommend வீட்டுல ஆத்துல அந்த மாறி பாஷை பேசி இப்போ நம்ப Brahmin so brahmin school அந்த மாதிரி யா இல்ல..
* Not necessary,
எங்க இருந்தாலும் நம்ம நம்ம மொழியை விட்டுக்கொடுக்காம பேசினா பேசணுமா?
*அதாவது நம்ம வீட்டுல, நம்மளோட வழக்கு எப்படியோ பசங்க அந்த மாதிரி பேசணும். School க்கு போனா ஆசிரியர் என்ன சொல்லி குடுக்கறாரோ அத பண்ணணும். இது ரெண்டும் பண்ணாலே மொழி தன்னல வளரும்.

But what about the Brahmin Tamizh dialect? The emphasis on the Brahmin Tamizh
***நமக்கு அது வீல்ல வந்து** practice **பண்ணணும்**. **வீட்ல பேசறதே இல்ல**. **அப்போ அது எப்படி இதுவாகும் எல்லாமே வந்து**, **எல்லாம்** English **ல வந்தாச்சு** நம்ப இந்ததமிழ் தான். தமிழ் இப்போ வெளில போய் Brahmin பாஷ பேசினா நம்பள கிண்டல் பண்ணுவாங்க. So நம்ம வந்து அதான், நம்பள சுத்தி என்ன language இருக்கோ அது தான் நம்ம பேசறோம்.

So do you feel that your language changes and that you are influenced by the society?
*Yes.

Okay can you give me an example?
*இப்போ இப்ப நம்ம பசங்க கீழ போய் வளையாடறாங்க, கீழ போய் வளையாடும் போது சொல்றான் late ஆயிடுத்து நா ஆத்துக்கு போணும் so மத்த பசங்களெல்லாம் சிரிப்பாங்க So ஒரு நாள் அவன் சிரிச்சு ஆடி வாங்கிண்டு வந்தான் வையென், next day அப்படி சொல்ல மாட்டான். நா வீட்டுக்கு போறேன். எனக்கு வீட்டுக்கு போணும். தண்ணி. Starting ல ஜலம்னு வந்தது, இப்போ தண்ணி குடிக்க போறேன். அதுக்கப்பறம் வந்து அதான் சேர்ந்து வளையாடும் போது சேந்து பண்ணும் போது மத்த பசங்க என்ன பேசறாங்களோ அதே influence ஆறது அதே வீட்டுக்கும் வருது. இப்போ வீட்டுக்கு வந்தா அந்த மாறி பேசும் போது நரியா நரியா வார்த்தைகள் அந்த

மாதிரி. அப்பறம் நாலாம் சொல்லுவேன் அப்படிலாம் பேசாதே நல்லாவே இல்லகேக்கறதுக்கே நல்லாவே இல்ல. ஆனா அது அவனையும் அறியாமலே அந்த மாதிரி வரது. So அது வந்து கூட எப்படி இருக்கோ. முத்தியெல்லாம் அப்படி கடையாது ஒரு அக்ராஹாரம்னு ஒண்ணு இருந்தது so அந்த அக்ராஹார பசங்கள்லாம் எல்லாம் ஒரே மாதிரி பேச்சு இருந்ததுனால Brahmin தமிழ் வந்து அப்பிடியே வளைந்தது இப்போ அப்படி கடையாது இப்போ ஒரு flatல வந்து Muslimமும் இருக்கான் ஹிந்துவும் இருக்கான் so all religions இருக்கும் போது அந்த எடத்துல போய் நா வந்து நா ஆத்து போறேன் ஜாலம் /துத்தம் குடிச்சுட்டு வரேன்னு சொன்னாக்க நம்பள கேலியா பாப்பாங்க. அதுனால அவங்க நம்ம இதுக்கு வர முடியாது நம்ம அவங்க levelஊக்கு எறங்கிட்டோம் அவ்ளோதான்.

Okay how would you consider அதாவது உங்களோட religious identitiy னு பாக்கும் போது, do you consider yourself as a Hindu or do you consider yourself as a Brahmin? Is there a difference between the two?

A.2 INTERVIEW TRANSCRIPT IN TAMIL AND ROMAN SCRIPT

Transcript VM

நமஸ்காரம், வணக்கம் or Hello, what is your preference?
* Hello and நமஸ்காரம்

சரி. இப்போ உங்கள பற்றி கொஞ்சம் சொல்லுங்களேன்

My name is (Anonymised/ removed). I am a Software Professional. Born and brought up சென்னை ல தான் வளர்ந்திருக்கேன், படிப்பெல்லாம் சென்னைல தான் நடந்திருக்கு. and currently I work in the IT industry. That is all about me.

நீங்க வந்து பொறந்து வளர்ந்து, உங்களோட பல்கலைக் கல்வி அப்பறம் வந்து கல்லூரி படிப்பு, இதெல்லாம் எந்த வழி கல்விள இருந்தது? Talk me through your experience with the language.
* Ok language அப்பிடின்னு பார்த்தோம் நா basically, the moment,I was introduced to pre primary and kindergarten, the moment you actually start learning something at school எல்லாமே வந்து ஒரு English oriented ஆ தான் இருந்தது. So தமிழ் was a language that we use at home. It was something that we use to communicate with the society. இப்போ வெளிய போறோம் நாலு பேர பார்க்கறோம், நாலு பேர பார்த்து பேசறது தான் தமிழ் language. At the same time, when it comes to purely academical point of view இப்போ வந்து ஒரு school க்குள்ள போறோம் classroom ல ஒக்கார்ந்து நம்ம படிக்கர்துக்கு போறோம். At that time, **you are discouraged to talk in Tamil** because everything you look at even right from programmes in TV, I am talking about programmes that are meant for Kids, something that is considered good for kids. From that point till whatever you are taught at school, everything   at a subconscious level, it drives you to learn English. And even at school, there is a prejudice amongst teachers and students that if you communicate in the local language (Tamil), it sets you a bar lower than the actual so throughout the journey, I தமிழ் is a language that everybody is aware of இப்போ நா தமிழ் நாட்டுல தான் இருக்கேன், நா ஒரு தமிழ் பைய்யன் தான் but நா தமிழ் பேசுவேன் வெள்ள இருக்குறவனும் பேசுவான் அது casual say for example right now in offices you have a dress code- four days you wear formal clothes and one day you are free to wear casual clothes தமிழ் was treated like that. It is something that you do 'outside'. It something that you do in spare time. everything that is 'proper' everything that is considered formal from the academic point of view, it was all driven through English.

Okay. தமிழ் அப்பிடின்னா, How do you see Tamil? What is Tamil to you? அதாவது நரையா versions of தமிழ் இருக்கு இப்போ வந்து இப்போ நம்ம பேசற தமிழ்ல சமஸ்க்ருதம் கலந்திருக்கும் we mix English as well as speak அப்போ தமிழ்னு சொன்னா உங்கள பொறுத்த வரைக்கும் எது தமிழ்?

* Ok. Right off நா அதுல ஒரு expert இல்லை. At the same time, அதுல comment பண்ணறது நா பேசற தமிழ் வெச்சு தான் comment பண்ண முடியும். Yes English உள்ள வரது There are certain things that is integrated நம்ப

174

life ல வந்துடுத்து phonesலேர்ந்து இப்போ இங்கே use பண்ணற lighting, switch ஆ இருக்கட்டும் தமிழ்ல common conventions use பண்ணறது கிடையாது. When you go to ஒரு shop க்கு போனா கூட we use the nearest English term இப்போ ஒரு switch குடுங்க ஒரு light bulb குடுங்கன்னு தான் சொல்றோமே தவிர we don't use the actual Tamil word that is actually ascribed with the particular word or particular object that you are going to buy. So நம்ம வந்து தமிழ்னு பார்த்தோம்னா I was நா பேசற தமிழ் வந்து இந்த Sanskrit root இருக்கும் ஜ, **ஷ்ரா ஶ்ரீ** அந்த words எல்லாம் incorporate ஆயிருக்கு. So அந்த words உம் and for essential items and commodities whatever you are going to buy or something இதுல வந்து English include பண்ணற மாதிரி தான் இருக்கு. Even those words, there is this thing called *Tanglish* right? for example you go to a shop நீங்க எல்லாமே தமிழ்ல purchase பண்ணீங்கனா கூட at the end of the transaction you have got say எனக்கு change குடுங்க. எனக்கு bill குடுங்க னு தான் சொல்வோம். We wont say எனக்கு சில்லறையை கொடுக்கவும். **People start looking at you awkwardly.**

Okay so இப்போ வந்து சமீபத்துல வந்து தமிழ் நாடு அரசு ஒரு ஆணை கொண்டு வந்திருக்கு you have to study Tami**zh** (*reflects the native pronounciation*) as a language இது பற்றி உங்கள் கருத்தென்ன and how is it going to help? Is it going to help at all? What is your opinion?

* My opinion would be simple. I cant comment about why such a decision was taken இப்போ வந்து இப்படி நீங்க கண்டிப்பா தமிழ் கத்து**ண்டே** ஆகணும் There might be a political angle to it and I dont wish to comment on that. ஆனா fact அ பார்க்கணும் நம்ப. Okay - the government easily says- do this. You have to do this etc. அதோட practical viability யை பார்க்கணும். In ஏற்கனவே introduction ல சொன்னா மாதிரி I work in the IT industry. I am guy who works in IT. இப்போ நா வந்து என்னோட work place க்கு போரேன். நா ஒரு public transit எடுத்து என்னோட work place க்கு போரேன். அங்க 9/10 people வந்து English உம் தமிழும் கலந்து ஒரு hybrid பேசற்துல தான் பழக்கம் இருக்கு. நாங்க வெறுமனே இப்போ plain ஆ வந்து இப்போ நா ஒரு எடத்துக்கு போ**யிண்டி**ருக்கேன் நா வந்து ஒரு ticket purchase பண்ணறேனோ இப்போ ஒரு automates service இருந்தா கூட the interface is in English. நீங்க use பண்ணறது and I work basically with **computers** which again orient towards **English.** right from the beginning ஒரு inception point of view லேர்ந்து ஒரு industry ல if you'd have driven it along with the language parallel ஆ போச்சு நா problem இல்லை. இல்ல next generation க்கு வந்து நீங்க வந்து இப்போ வர பசங்களுக்கெல்லாம் நீங்க தமிழ் கண்டிப்பா கத்துக்கணும் சொல்லலாம். அந்த பசங்க ஒரே ஒரு கேள்வி இப்போ நாங்களே கேக்கக்கறோம் - நாங்க கத்துக்கறதுனால Do I have any long term value? I **respect** my language. இல்லனு சொல்லல. தமிழ் மொழி தான் எனக்கு நா தமிழ் நாட்டுக்காரன் தான் , பவீட்டுல தமிழ் தான் பேசறான். இது தான் என்னோட language, இது தான் culture, heritage everything is associated with Tamizh Nadu and its culture but the fact is learning Tamizh alone, is it going to help me, me as a person in the long run? i can, at the end of the day, chestpounding and jargon only takes you so far. Practical viability is something. yes you can mandatorily force the language into someone, into people but is the same as something. okay I will compare this with environmental sciences. So நா இது வந்து compare பண்ணறது வந்து environmental sciences, சுற்று சூழல் அது கூட EVS னு தான் school ல சொல்லுவாங்க யாரும் சுற்று சூழல் பாதுகாப்புன்னு சொல்ல **மாட்டா**. EVS. So, Environmental **scienceனு** பாத்தீங்கன்னா it says dont throw trash. வெளிய spit பண்ணாதீங்கன்னு சொல்றா so these kind of things are better realised than thought upon. Just by forcing people to learn a language, you are not going to make them love the language. Again சொல்றேன் the decision coming from the government may have many political angles to it. There might be stalwarts who say Tami**zh** is everything etc. Practical life, it doesnt, is it the way, the society views? Tami**zh,** நாட்டுக்குள்ளையே இப்போ city side ஆகா இருக்கட்டும். ஊர் பக்கமே இப்போ வந்து towards the country side move பண்ணறச்ச they may embrace but they embrace a different form of Tami**zh** and that is also there. When you take English, if you take English as a language, English is kind of homogeneous. You can phonetically make it different but at the same time, you dont have different words to represent- you have different words but all of these are documented. Its kind of closed. There is a way of pronouncing a word. அவ்வளவுதான். Phonetically it may be different and everything but here, the things are vastly different. If you go to the country side, the people will use for the same object for example bucket இங்க வந்து எல்லாம் bucket னு தான் சொல்றா. சில பேர் வந்து வாலிம்பா ஊர் பக்கம். அது நெறையா பேருக்கு **awareஆ** இருக்காது. So இந்த differences எல்லாம் arise ஆறது. So நா அது தான் சொல்றேன் practicality னு ஒன்னு இருக்கு. நீ வந்து கத்துக்கறது force பண்ணறது இருக்கற்தும் , population மொத்தமா realise பண்ணி , actually we are going to implement the change. Okay I want

to learn the language and நா பண்ற எல்லா விஷயத்திலும் அந்த language நமக்கு வரணும்னு இருந்தா அப்போ தான் அந்த government issue அதுல succeed ஆகும் or else it is nothing more than a plain order.

Okay- தமிழ் நாட்டுல பிறந்து தமிழ் நாட்டுல வளர்ந்து இதுவரைக்கும் நா ஏன் இங்கிலிஷ் ல படிச்சேன் அல்லது நா ஏன் தமிஸ் படிக்கல அப்பிடின்னு அந்த மாதிரியான இது வரைக்கும் வந்திருக்கா?
* அந்த மாதிரியான சிந்தனை எனக்கு வரத்துக்கான window எனக்கு வரல. ஏன் நா simple. நா வந்து ஒரு community apartment ல வளர்ந்தேன். என் கூடயே ஒரு பத்து பதினாலு பசங்க வளந்தாங்க almost என் age அல்லது என்னைவிட ஒரு ரெண்டு வயசு மூத்தவாளா இருந்தா. *எல்லாருமே everybody* நாங்க ஆரமிச்சதே இந்த மாதிரி English based இதுல தான். And இதுல ரெண்டு விஷய்யம் நா முக்கியமா கொண்டு வர விரும்பறேன். Within Tami**zh** Nadu there is a stigma. அது நம்ப ஒத்துண்டே ஆகணும். Any medium that instructs in Tamil is considered somewhat okay for lack of better word I will say low class. ஒரு corporation school ல தான் தமிழ். அது தமிழ் medium படிக்கற பசங்க அப்பிடிங்கறது ஒரு stigma. சின்ன வயசு லேர்ந்து பசங்கக்குள்ள integrate ஆறது. இது வந்து நெறைய people can choose to deny it. But nobody, negative ஆ, portray பண்ணலனாலும், nobody is going to say that **அவாளும்** அந்த மாதிரி படிக்கறதும் ஒரு viable course தான். நீங்க தமிழ் medium படிச்சாலும். அதுலயும் படிச்சு நீங்க வந்து graduate apart from pursuing it language based இல்லாம நீங்க வந்து மத்த field லையும் சாதனை புரியலாம் அப்பிடிங்கறதுக்கான ஒரு இது இல்ல. ஏன்னா for example நீங்க தமிழ் medium லியே படிச்சு ஒரு high school வரைக்கும் போய்ட்டு graduate பண்ணிட்டு, you cant expect to go to ஒரு பெரிய university க்கு போய்ட்டு ஒரு advance physics- நீங்க எந்த higher studies pursue பண்ணினாலும் again English comes into the picture. So ஒரு சின்ன வயசுல நா school போறது வந்து இல்ல இத்தனை வருஷம் நா படிச்சுட்டு ஏன் நா தமிழ் ல படிக்கல னு அந்த மாதிரி கேள்வி வந்ததா அப்படி அந்த மாதிரி கேள்வி கேக்கற்துக்கான எந்த மாதிரியான ஒரு scenario அமையல. To question what I currently have, எனக்கு வந்து opposite ல இது வரைக்கும் positive ஆனது இது வரைக்கும் நா பாக்கல. Because I have studied with people who come from a Tami**zh** medium. When I went to college என் கூட வந்து , in my classroom there were people from Tamizh medium, backgrounds. First two years அவா பட்ட struggle நா பார்த்தேன். See, they might have had a choice, to pursue in English medium, they might have chosen to pursue in Tamizh medium, or ஆரம்பத்திலையே அவா தமிழ் **mediயத்துல** இருந்திருக்கலாம் . So அவாளோட personal choice. Apart from that, the moment they transit into higher studies, ஒரு graduation levelக்கு அவா வர்ரச்ச, they face difficulty because, at the graduate level, we dont encourage this. நா வந்து engineering exam, இன்னி வரைக்கும் அண்ணா university papers எழுதியிருக்கேன். அங்க வந்து, everything is based on / published in English. There is no option giving you that you can also answer this in Tamizh, at least as far as I have seen. The only place where Tamizh is actually encouraged is in your government application form. You have these 20 pages in Tamizh and 20 pages in English and all. It is being encouraged over there but at the same time, ஒரு professional ஆ படிக்கணும், நீங்க ஒரு பெரிய, if you want to do R&D something, Tamizh க்கு ஒரு transition கிடையாது. நீங்க தமிழ் medium லியே படிச்சும் நா வரலாம், நா ஏன் கேள்வி கேக்கணும், நா ஏன் (என்) மொழியிலே படிச்சு வரலாங்கறது கேக்கற்துக்கான windowவே இங்க இல்ல. Because obviously we know that at some point, we have to embrace English as a language.

இப்போ வந்து "தொழில்நுட்பம்" அப்படினு சொன்ன உங்களுக்கு எந்த மொழி உடனே ஞாபகம் வரும்?
*"தொழில்நுட்பம்" அப்படினு சொன்னா தமிழ் தான் ஞாபகத்துக்கு வரும்.

Researcher clarifies: Let me say technology.

* Technology நா obvious ஆ English. It will either be English or from a pioneering point point of view.. Okay இப்போ language னு பார்த்தோம் நா it is automatically, by default going to be English.*அது தான் சொன்னேன்* like I said we are pre conditioned right from childhood to learn English, understand English and apply English in all aspects. இப்போ நா பேசறது கூட பாதி இங்கிலிஷ் ல தான் பேசிண்டிருக்கன். ஏனா இந்த மாதிரி தான் condition இங்க ஆயிருக்கு. ABCD தான் மொதல்ல இது பண்ணறோம் அ ஆ இ ஈ யாரும் சொல்லித் தரத்தில். இதுல ஒரு முக்கியமான இது சொல்லணும்னா ஒரு குழந்தையை primary school ல சேக்கறேள் அப்பிடினா okay அது எத்தனை schools ல அப்படி பண்ணறா னு தெரியாது may be

government schools ல இருக்கலாம் but normal cbse institutions இருக்கட்டும், இல்ல நீங்க எதான ஒரு நல்ல convent இதுல போய் சேக்கறேள் நா the first thing, the child is taught is to learn ABC and 123. Nobody will teach ஒண்ணு ரெண்டு மூணு and nobody says அ ஆ இ ஈ. அ ஆ இ ஈ comes into picture when they go into class 1 grade 1 or 2, where they take the language papers. When language as a separate subject comes in, they integrate the **local language** for heritage- where your heritage is from அந்த language அ integrate பண்ணறா but at the same time, your default language always points to English.

Is there a reason why we are being preconditioned? or why we are being preconditioned to study English? Do you see a reason why? Or what could be the reason?
* Yes there is a reason. We as Indians we have this habit, where in we say we are 'free', in 1947 we managed to obtain independence but the only thing we managed to do is form our own government. At the same time, **we have chosen to shun our culture**. Totally apart, the fact is in the name of , we have tied in culture and progress with the same thing. We have adopted a different principle. Yes. Globally, once upon a time, Britain has conquered more than half of the world. So naturally, their language started propagating everywhere else. அதே சமயத்துல பார்த்தோம்னா இப்போ china இருக்கு japan இருக்கு அங்கெல்லாம் வந்து they even for computer programming language, they have a separate UTF format for displaying their strings because they had ஏன் அவா succeed பண்ணானா அப்பிடின்னு பார்த்தோம்னு வெச்சுங்கோளேன் அங்க இருக்கற மனுஷா எல்லாம் ஒரு homogeneous type. இப்போ Chinese நா obviously everyone is Chinese. நம்ம multicultural ஆ இருக்கறது there is a plus point to it
at the same time, there is a negative to it. negative part as well. the negative is that we started questioning our own **heritage.** We started looking. Okay yes, நம்ப வந்து degrade ஆகல we have moved forward but at the same time, there is a cost. The cost is simple. We have chosen to embrace what is **current and happening** rather than innovate it by ourselves and grow it. The world was moving at a pace, we did not.. I would say we have just formed our government and we could have taken every region should follow their own regional language, the government could have given its decision right there. Embrace your own culture, grow on your own rather than following the norms set by someone else. No we didn't do that. We wanted to get into the market, we wanted to be the hero so what did we do? We changed the **identity.** Its the same kind. நீங்க ஒரு புதுசா office க்குள்ள போறீங்க. புதுசா ஒரு office குள்ள போயிட்டு 10 பேர், they are having their own work culture. you are going to go there and sit and say I am going to be like this its going to take time to get acclimated we didn't want to do that. அவள மாதிரியே we started **adapting.** அதுனால என்ன ஆயிடுத்து அதே syllabus அதே இது follow பண்ணிட்டோம். Right from இது இப்போ ஒண்ணுமில்ல law and order, we have Acts being framed and everything, அதுல இருக்கற content அ பாத்தீங்க நாலே it is so complicated. English லியே ரொம்ப complicated ஆ இருக்கும் ஏன்னா இது British standard ல தான் எழுந்திருக்கும். So, base லியே வந்து we started embracing the foreign culture னு சொல்லலாம். நா என்ன சொல்ல வரேன்னா <u>**we are not disrespecting Tamizh, but we are not regarding it in the way it is supposed to be regarded.**</u>

Okay, do you think, Tamizh has become 'foreign' in its own native land?
In a way, it has become foreign. Ask anybody, it is like asking a person on stage, do you support women empowerment? The obvious answer is going to be yes. In the same way, ask do you support Tamizh and love Tamizh, any Tamizh Nadu guy would say yes. But does he really, is the question? It has become foreign in several ways rather than one- simple reason is that we have adopted something and we have started moving in a path, and right now, we are reluctant to move, not reluctant, we have boxed ourselves into corner. if we **choose** to make Tamizh as the primary language, you know in all the industry, the transition is going to be costly at this day. Assume that everybody has this group - I want to make Tamizh as the 'happening' right now. If everybody wants to change, it is going to be costly. We have embraced something, and **two to three generations** have been cultured in the same manner. So in a way, we have alienated our own language into something exotic, antique piece of history rather than something that is **used for day to day life**.

இப்போ நா கேக்க போற கேள்வி வந்து தொழில்நுட்பம் சார்ந்தது அதாவது இப்போ நாமம் பேசிண்டிருக்கோம் இல, அதாவது ஆங்கிலம் கலந்த தமிழ் பேசறோம் எல்லாருமே for that matter, அதை வந்து தொழில்நுட்பம் support பண்ணனுமா அல்லது மொழிக்குனு ஒரு தனி அடையாளம் இருக்கு இல. இந்த மொழி இப்படி தான் பேசணும், பிற மொழி சொற்கள் கலக்காமல், அதை வந்து தொழில்நுட்பத்தில் கொண்டு வரணுமா?

* Ok இதுக்கு நா எப்படி பதில் சொல்லனும்னா அடையாளம் அப்படினு தேடறது தான் எனக்கு தெரிஞ்சு actually தமிழை forward ஆ எடுத்துண்டு போறவாளோட disadvantage னு சொல்லுவேன். நீங்க ID தனியா வேணும்னு சொல்லும் போது நீங்க என்ன பன்ரேள்னா you are starting to alienate your

177

own language. You are starting to draw borders. there has to be integration, integration has to be seamless but, who is integrating whom is the question. You are joining two pieces together, one is actually going to encompass the other. So, if the thing that encompasses the other language is Tamizh, well and good but that is not the case. English and Tamil were merged together but English took the upper hand. அதாவது English இருக்கு English ஒரு language ஆ இருக்கு, English உள்ள வரத்த, it took pieces from Tamizh to make English itself seamless. Okay இப்போ normal ஆ வந்து இப்போ வெளியே போர்த்த இப்போ நா பேசறது - normal ஆ வெளியே போர்த்தங்கறேன். **பொதுவா வெள்ள போர்த்தனு நா ஒரு வார்த்த சொல்லலாம்** but I am comfortable in saying normal ஆ வெளியே போர்த்த. இப்போ நா என்ன பணறேண்ணா normal எடுத்துண்டுட்டேன் 'Normally when heading out' ல normal எடுத்துண்டுட்டேன் வெள்ள போர்த்து தமிழ் லேர்ந்து எடுத்துண்டுட்டேன் . இது வந்து பாத்தேள்னா 80% English and 20% தமிழ் இருக்கும். இல்ல இன்னும் கொஞ்சம் forward ஆ போறோம். அதாவது இந்த 100% ல ratio பாத்தேள்னா more than 60% will always be in English ஒரு 40% தமிழ் இருக்கும் இல்ல இந்த ஒரு border ல தான் தாண்டிண்டிருக்கும். இப்போ நா local ஆ, என்னோட friends ஓட பேசறேன்னா நா தமிழ் தான் ஜாஸ்தி வரும் but the fact is saying my language should have its identity மத்த language உள்ள வரவே கூடாது அப்டிங்கறது இந்த கட்டத்துல possible ஆ அப்டிங்கறது அதோட feasibility யோசிக்கணும். அது தான் சொல்றேன் இது தான் first generation who is being exposed to the language நா ப்ரச்சனை கிடையாது. இது my grand father, great grand dad, my dad, அதுக்கப்பறம் நா எல்லாருமே இந்த இதுல தான் filter ஆய் இந்த strain தான் வந்திருக்கு. to change everything overnight ஏனா when my grand dad adopted that particular language, and he passed it on to his son, something might have been lost. The same, next level, something else might have been lost. So ஒரு cumulative ஆ, you would have lost something of the language. இப்போ இருக்கற காலக்கட்டத்துல நீங்க survive பண்ணனும் அப்டினா ஒரு particular language மட்டும் வெச்சுண்டு போர்த்துங்கறது feasible கிடையாது. We are moving to a state where borders are becoming blurred. So அதுனால integration அவசியம் but the fact is who is following who வந்து முக்கியம். **You have to retain your roots** அது தான் விஷயம் இங்க. உங்க roots இருக்கணும் but ஆனா என்னோட root மட்டும் தான் இருக்கணும்னு சொல்லறது அது தப்பு.

Okay சரி இப்போ, I am talking in Tamizh Nadu context ல you've said 'normal ஆ வெளிய போர்த்து' வந்து அந்த example லேர்ந்து, how do you see the word 'normal'? I know it is an English word ஆனா அதை நீங்க தமிழா பார்க்கறீங்களா இல்ல English ஆ பார்க்கறீங்களா?
* இது again it comes to the identity part which you said. நா என் மனசுக்குள்ள தமிழ்ல தான் பேசறேன் னு .. நா என் friend கூட பேசிண்டிருக்கறத்த, இங்க போலாம் அங்க போலாம் நம்ம என்ன பண்ணறது அப்படினு யோஷிச்சிண்டிருக்கோம் பேசிண்டிருக்கோம். எனக்கு **Whatsapp** ல நீ **ping** பண்ணனு தான் சொல்றோம். But at the same time, if anybody asks me, *நீ எந்த language ல பேசறேன்னு கேட்டா நா தமிழ்னு தான் சொல்வேன்*. இது automatic ஆ வந்துடுத்து. ஒரு example சொல்லணும்னா - normal ஆ இப்படி தான் டா பண்ணுவோம் அப்படினு ஒரு பையங்கிட்ட சொல்றேன் என் பக்கத்துல போய் பக்கத்தாத்து பையன் கிட்ட normal ஆ இப்படி தான் பண்ண போறோம்னு சொல்றேன். நீங்க என்ன language ல பேசினீங்க னு யாராவது என்ன வந்து கேட்டாங்க நா, நா அவாக் கிட்ட என்ன சொல்லர்துனா நா தமிழ்ல தான் பேசினேன்னு சொல்லுவேன்.

இது தொடர்பானக் கேள்வி அதை தான் நா முன்னாடி கேட்ட கேள்வி- தமிழ் நா என்ன?
* அதான் சொல்றேன், this again comes to the fact of **identity.** Right now, Tamizh is **not pure.** <u>You can</u>, if you want to purify it, you can make it as the language which is integrating other languages and slowly start culling out the other language part. இப்போ வந்து, இப்போ for example, you, இப்படி ஒரு இது இருக்கு, suppose, you want to provide pure Tamizh translation for anything, being said commonly right now, yes that is possible.

I'm just going to interject. நா இப்போ நெறையா pure னு நாமம் சொல்றோம். நெறையா பேர் பயன்படுத்தறா pure, pure அப்பிடின்டு. இப்போ வந்து, pure தமிழ் நா அதுல ஒரு perception இருக்கு சங்ககாலத்து தமிழ் நான் சற்று அமர்ந்து வருகிரேன் அது சங்ககாலத்து தமிழுக்கு போயிடறோம் ஆனா
* No what I meant by pure okay may be அந்த context தப்பா இருந்திருக்கலாம். Pure means something else actually. நா என்ன mean பண்ணேனா, ஒரு sentence when I frame it, in any object, subject, adjective, anything that is involved in it, அது வந்து தமிழிலே மாத்திரம் இருந்தால் i will call it as pure Tamizh. நான் மைதானத்தில் போய் விளையாடுகிறான் அப்பிடின்னு அவ்வளவு ஷூத்த தமிழ்ல பேசணும்னு அவசியம் இல்ல. நா

மைதானத்துல விளையாட போரேன். அப்டினு சாதாரணமா கூட பேசலாம் but the fact is எல்லா words, எல்லா syllables உமே தமிழ் ல தான் இருக்கு. இந்த foreign language interjection ஆறது உள்ள வந்து ஒரு ஒரு foreign language word இப்போ வந்து தமிழாக தான் identify ஆக பட்றது. So நம்ம வந்து பேசற விதம் வந்து அதான் சொல்ரேன்

phonetically or accent wise, difference இருக்கலாம், okay but at the same time, இங்க problem என்னனா, some other language is coming into the picture. அதான் இப்போ problem ஏ. ok when I am saying pure, I am saying only Tamizh words in a sentence. அது use பண்றோம் ஒரு 100% ratio ல 50% 45% போயிட்டே தான் இருக்கு its close. ஒரு பக்கம் advantage இருக்கர்துன்னு இல்லை at the same time, this language is being **decimated** slowly. ஏன் நா சொல்ரேன் நா previous questions அ recall பண்ணேன்னா long term viability, நா தமிழ் மட்டுமே practice பண்றவனா இருக்கேன் அப்படினா எனக்கு பின்னாடி, என்னால எந்தளவுக்கு heights என்னால reach பண்ண முடியும் apart from Tamizh Nadu. Tamizh Nadu ங்கறது ஒரு state. அதுல இத்தனை தான் இருக்க போறது. **இத தவிர்த்து ஒரு step வெள்ள எடுத்து வெச்சேன்னா** கூட, எனக்கென்ன கடைக்க போறது, என்னால **survive** பண்ண முடியுமா if I only embrace Tamizh? Would I be able to survive, the answer is obviously no. Yes I have to **embrace** a foreign language.

இப்போ இதே பழைய conversation, கருத்து நம்ம எடுத்தோம்னா, you mentioned about Japanese and Chinese, அந்த ஒரு கலாச்சாரமும் you talked about homogenity, அந்த ஒரு homogenity தமிழ் நாட்டுக்கு பொறுத்தினா, இதே பிரச்சனை தானே அவங்களுக்கும். அவங்க நாட்ட விட்டு வெள்ள வந்தாலும் they will face the same problem just as you mentioned to what extent can we get ஆனா அவங்க அடிப்படை எல்லாமே வந்து அவங்க தாய் மொழி ல இருக்கும். So how do you compare அவங்களோட இதுக்கும் நம்மளோட இதுக்கும்,?

* அது தான் it again goes back to what I say, we say that we are free but at the same time, we are not culturally free. We have gotten a taste of something which the foreigners have offered us and at the same time, we have chosen to embrace it. நம்மளுக்கு அது புடிச்சு போச்சு. புடிச்சுதுன்னா எனக்கு புடிச்சது இல்ல, உங்களுக்கு புடிச்சது இல்ல, முன்னாடி யாரெல்லாம் இது எடுத்தாலோ, சரி simple ஆ சொல்றேன் Freedom கிடைச்ச உடனே they started framing, education rules போட பார்த்தா எல்லா சட்டம் எல்லாம் போட்டா why was it in English? Why couldn't it have been in some local language. அந்த time ல வந்து நம்ம ஸ்வதேஷியா இருக்கணும், this is our country அவ்வளோ patriotism flow ஆர்த்த ஏன் you basic ஆ embraced a foreign concept and not your own அதான் first question.

ஏன்னா simple ஆ சொல்லனும்னா இப்போ china japan ல பண்றானா they have created their own identity. For example immigration எடுத்துக்கோங்கோ நம்மளுக்கு ஒரு இது இருக்கு நீங்க சீனக்கோ ஜப்பானுக்கோ போனீங்கன்னா ரொம்ப வருஷம் இருந்தா கூட to get the citizen of that country is very difficult. They will never accept you as one of them. You will always be a foreigner to a Japanese. They will always excuse you for any mistakes that you do regarding the cultural habits. so அந்த விதத்துல அவா, அவாளோட mindset ரொம்ப clear ஆ இருக்கு. This doesnt happen overnight. Its generations after generations of embracing their own identity. They have identified themselves as Japanese, they love their language and this is what we are. The person who comes in, அவன்கிட்ட ஒரு நல்ல விஷயம் இருந்ததுனா அது admire பண்ணலாம் but at the same time, we cannot be that. There are people, அவாளோதலையும் மாறாமல்லாம் இருக்க மாட்டா but the majority have chosen to embrace. நம்மளுக்கு எப்படின்னா இங்க வந்து இப்போ தமிழ் நாட்டே எடுத்தோண்டோம் நா நம்மளுக்கு ஒரு habit இருக்கு எப்போவுமே எதிர்கேள்வி கேக்கற பாசக்கம் இருக்கு. எதுவா இருந்தாலும் நம்ம ஒரு கேள்வி கேட்டு அதுக்கு answer வரணும் but the problem என்னனா we argue for the sake of argument rather than to get the solution. So, we have at some point, started to dilute our own identity. Homogenity நம்ம maintain பண்ணனும்னு விரும்பல. நா individual எ பார்த்தேன் but, நா என்னோட society அ பார்க்கல because, given the state of out country, survival was the key. Anything that I can use to survive at the end of the day was good. அது அப்படியே இப்போ filter ஆயி filter ஆயி வருது. நா வந்து இப்போ ஒரு language இதுக்காக பேசினேன்னு வெச்சுக்கோங்களேன் , I might take a stage. ஒரு few political stalwarts இருப்பா some people who want to embrace the language so badly அவாள்ளாம் தனக்கு எனக்கு இந்த மொழி தான் best சொல்லிட்டு to give their thoughts to the public. Anybody, ஒரு NGO வாக இருக்கட்டும் anyone who is into it எல்லாருக்கும் ஒரு individual gain தான் இருக்கு அதுல. யாருமே நம்ம ஒரு சமுதாயம் னு , ஒரு சமுதாயம்னு benefits க்கு சொல்ல நல்லா இருக்கும் but at the same time, nobody wants to see the ugly side of it. Homogenity க்கு ஒரு ugly side இருக்கு. You will be alienated in a very bad manner. We seamlessly integrate into any

environment we go to. it is the same reason why you are in the US / sorry UK . There is a difference between Indians coming to the UK and Asians coming to the UK. When they say Asians, it is not Indians. Indians, they claim Indians separately. Asians I believe are separate if I am not wrong. So, you are alienated in a different manner, if you embrace homogenity too badly. We have this habit of integrating - wherever we go and stand in, we integrate with that crowd and thats our habit.

Being a Tamizh and Brahmin, which of the two have you traditionally viewed important and why between Tamizh and Sanskrit?
* Well, primarily, I would say Sanskrit and Tamizh following that because, even though, நா இப்போ பூஜை room ல போய் pray பண்ண போறேன் நா my mantras and everything is based upon Sanskrit. So on a personal level, I believe, Sanskrit is something **divine.** Because everything is based on that. Tamizh, again, since I come from a Tamizh background, most of the text, that has been given to read or something has been in Tamizh. But you would see it as a rough translation from Sanskrit to Tamizh. There is no equivalent word in Tamizh. **It is just Sanskrit represented in Tamizh.**

Researcher: Example?
* Okay when you say श्री गुरुभ्यो नमः , the terminology ஶ்ரீ . If you go to mainstream Tamizh nadu, people will just ignore श (ஶ) क्ष त्र, ज्ञ because they say it is not Tamizh, it is Sanskrit. So, श्री गुरुभ्यो नमः, the pronunciation and all that, you start representing in Tamizh, and if you were to pronounce in Tamizh, you wouldn't say गुरुभ्यो, the भ wouldn't come in because that is Sanskrit type pronunciation. So when I give precedence, and what I have been taught, my spiritual conditioning and all, these things have been in Sanskrit. So I respect Sanskrit at a higher level. Tamizh, is the language which I was born and raised with so obviously that has a separate place. I am not drawing comparison with both of them.

Tell me a time when you were compelled to use a technology in a particular language and how did you feel about it?
* I will tell you a time when I was compelled to use technology in Tamizh- Never! By default it was English- there wasn't any compulsion. Its like this, you go to a school, and you graduate out, you be a chemist, you be a pharmacist, you be even a civil engineer, English is your bread and butter, you have to follow English. So இதுல compulsion வாரத்தில் this is how you are conditioned. There was this one time and this is something that I admire about the Chinese guys I was working in one particular application in the office. There was one particular application or software or report or something where our clients who are based somewhere in Hong Kong, they had wanted it specifically in Chinese. **(28 minutes 33 seconds)** And not in English. In that way, it was admirable. They know that the system has capabilities and we only follow the English set of characters and they were pretty adamant over there that was **one time I was forced to use, forced to translate into some other language.** Apart from what I do traditionally. **Such a kind of compulsion I have not faced locally.** Its the same thing- A man in Hong Kong would like to see his report on Chinese or Mandarin or Cantonese in whatever script that they want to see but you wont find people here saying - give me a Tamizh transcript I dont want to read it in English- people wouldn't say that, because they are comfortable.

You being in that scenario, did that motivate you at a personal level, when you get a smart phone, did you try and ask - எனக்கு இது தமிழில கொடுங்களேன், தமிழ் ல இருக்கா அல்லது எனக்கு தமிழ்ல கொண்டு வாங்க இல்ல நா தமிழ் ல தான் எடுப்பேன், I mean just drawing an inspiration kind of thing.
Ok at a personal level you are saying- yes I do have a Tamizh reader an all in *my smart phone at the same time,, did I change the interface into Tamizh- I wouldnt see the point in changing the interface into Tamizh. Yes, at a personal level, I might say that I am Tamizh and I am respecting it.But the moment I open a messenger, and start texting to somebody, I am going to do it in English. I am going to convert my Tamizh keyboard into English. This is the thing I was trying to- if I start embracing Tamizh by myself right, I am going to be something who is doing for public perception. its going to be like a trend rather than- it wont be default, it wont be taken normally.Its going to be like this guy has தமிழ் பற்றும்பா அது தான். இந்த ஊர்ல அது ஒரு common. எனக்கு தமிழ் பற்று ஜாஸ்தி அப்படினு வரும். தமிழ் தான் இங்க language but, at the same time, தமிழ் பற்று , you are loyal to your own language அப்படிங்கறது இந்த ஊர்க்காறாளேச் சொல்லுவா

Okay proficiency in a language is an indication that technology in that language is more like to be accepted by its native speakers, Whats your view on this?
*Well it depends on who these speakers are the targeted zone- when you say native there are different varities of native people. **You deploy something like this in the country side, it will be very useful.** Yes there are say for examples villages, they have these that 90% operates in Tamizh. When they do conduct operations in Tamizh, So giving a software for them that would capture what they say and reproduce it in Tamizh text itself. directly, that would be very useful. But, coming to mainstream, when you come to urban environment, would people actually embrace. Okay, they would embrace but would would they actually use it at day to day level? I believe that would be kind of minimal. they would use it for a translation

purpose they would, people here are more interested in translators, rather than reproduction of what they say into the Tamizh text.

* Do you think, if youve embraced Tamizh as a natural choice for using technological devices like smartphones?
Tamizh as a natural choice for using technological devices like smartphones. Yes I would embrace it, if , if, the interfaces provided by the systems are okay and seamless. I am using Tamizh as my basic smartphone, within my smartphone and local apps, I am using Tamizh- the moment I go to appstores, I am going to go back to English. If across the board I can use Tamizh, then no issues.

*Please tell me about the transition from using a very basic phone like nokia to smartphone.
Transition as in (clarification) Transition as in நீங்க ஒரு basic phone வாங்கிருப்பீங்க இல்லையா? அந்த phoneலேர்ந்து ஏன் நீங்க மாறிநேள் எதுக்காக? என்ன ஈர்ப்பிருந்தது towards smart phoen and how do you feel? அந்த மாதிரி. Okay I will be honest. எனக்கு ஈர்ப்பெல்லாம் எதுவும் இல்ல. Okay there was a shift. There was a shift back when phones were introduced into the market, reliance, and then nokia started giving you the models at a cheaper price. The only thing that was hot were sms text messages and calls. Then came the android OS, and came the camera phones and one by one it started building up to a particular level. The reason, okay why this transition happened was, as i said, **it was partly due to the social pressure**. Nobody wanted to sit down and text anymore and spend fifty paise on a text message. They moved onto some instant messenger- whatsapp being the latest one. So, they moved onto an instant messenger, and they started sending- you know, let it be organisations, let it be schools and colleges, they started embracing this instant communication model. Emails and everything, previously we used to say back into our 2004, 05, can you send me an e-mail I will send it to you today evening, நா வீட்டுக்கு போயிட்டு தான் system access பண்ண முடியும் இல்ல net centre போய் access பண்ண முடியியும். But now, இப்போ பார்த்தேள் நா e-mail அனுப்பணும்னா on the go i can access gmail app and send it. So, communication வந்து கொஞ்சம் speech up ஆச்சு. அதே சமயத்துல people also expected you to be scaling at that level. (**33 minutes 57 seconds**)

ஏன்னா இப்போ அனுப்பரேன்னா உடனே அனுப்ப முடியாதா? this is next question comes ஒரு புது phone வாங்கிக்கலாமே அப்படிங்கறது so ஒரு பக்கம் partly social pressure இருக்கு இன்னொரு பக்கம் technology is moving forward. Okay there are different உங்களுக்கு அதுக்கேத்தா மாதிரி இன்னும் scale பண்ணனும். Ultimately, yes இங்கே ஒரு factor என்ன involve ஆர்துனா ஒரு particular if society starts to accept something at a subconscious level, automatically things will move forward whether you like it or not.

* Have you ever heard about speech to text technology?
Yes   I   have   heard   about   speech   to   text   tech.

*Have you ever used speech to text technology in any language?
English- yes I have there was a (researcher in background: what was your experience like?) **I did spot something there okay - these speech to text translators were very very accent oriented**. I used to have fun with my friends you know- we used to say some complex sentences in English and these used to get translated into some xyz format- we used to have fun like who **used to pronounce in a valid way** and what sort of you know funny words it turn up what ever we were saying. So if I said 'watever' it would go wad something. So I have to say what ever (pronounced properly/ clearly) you know I have to stimulate or something I have to do to get a British accent something of that sort. So it was very accent oriented. **So speech to text translators were useful when they actually work.** But the fact is when you actually give the speech to text translator and you cant speak the same way all the time. **When people use a particular technology, they would want to use at any level and that's how the end end user wants it. He would use it in any way he wants to. There shouldn't be restrictions on the user saying that the user has to use it in this way and then the system is going to respond. I think that was primitive speech to text translators which I have seen out. Those things were represented like that- like you had to you know modify the way you say stuff for the system to recognise what you are saying properly- by the time I can actually type it! That was the problem.**

* So what is your opinion/ experience, very specifically on pronunciation of the Brahmins and the non Brahmins? (clarifies: தமிழ், தமிழை பொறுத்த வரைக்கும்)

Brahmins வந்து ஒரு விஷயம் ஒரு pronunciation ஒண்ணு பண்றோன்னா அதுல ஒரு ஸ்பஷ்டம்ணு ஒண்ணு சொல்வோம் - ஸ்பஷ்டமா சொல்லர்துனா அழுத்தி சில வார்த்தைல எங்க emphasis போடறது எங்கே emphasize **பண்ணக்கூடாது அது தான் திருப்பி குரூபியோ நமஹ அந்த ப (bha) ஒரு சின்ன gap இருக்கும் when you இது brahmin community க்குள்ள இந்த ஸ்பஷ்டமா pronounce பன்னரதுங்கறது வரது நெறைய பேருக்கு. வருது கொஞ்சம் அஸூத்தி pronounce பண்ணனும் ஒரொரு வார்த்தையை** ஒரு word இருக்குன்னா அது கொஞ்சம் phonetically structure ஆயி இருக்கும் because it again comes from the Sanskrit root. When you go to traditional rural Tamil Nadu அந்த பக்கம் போய் பார்த்தேள் நா they will speak very seamlessly ரொம்ப fast ஆ பேசுவா at the same time, they will make sense. **இப்போ ப்ராமணா கிட்ட ஒரு வார்த்தை பேசறோம்னா "அவா என்கிட்ட இது சொன்னா" இப்போ ஊர் ல மதுரை பக்கம்லாம் வந்து "என்கிட்ட சொன்னாப்பல" அப்படிஜு so அந்த வார்த்தையே மாறிடுத்து** it is not the way you pronounce it, the word itself changes. சொன்னாப்ல then அவா சொன்னா இவா சொன்னா we use அவா இவா the southern community அவிங்கம்பா இவிங்கெ பண்ணிவங்கே அப்பிடின்னு so வார்த்தையே மாறிடுத்து. So இதுக்கெல்லாம் ஒரு common base ஆ இருக்கணும் அதுலேயே ஒரொரு ஊர்க்காரா வேற வேற pronounce பண்ணுவா

* So இப்போ தமிழ் அப்படிஜு எடுத்துண்டோம் நா வட்டார வழக்கு இருக்கு and then we have the religious Tamizh அந்த வகைல வந்து Brahmin Tamizh would you want to see it as separate with a religious identity?

Okay as a personal preference, I have grown up with this Tamizh, I wouldn't see it in any other way. but the fact is we are, there is no question that we are not different, we are different but who is going to embrace whom is the question. **We have to find a middle ground, for it and then we would have to save obviously- I am used to Brahmin Tamizh, so I would prefer speaking in Brahmin Tamizh, I cannot say the same thing for another person. Some might even find it offensive, given the normal scenario here in Tamizh Nadu. Some might even find it offensive, the way you want to structure it has to be in Brahmin Tamizh. That itself would be a bit offensive here.** *The thing is, I grew up with Brahmin Tamizh and I am comfortable with it but at the same time, the level of pronunciation that Brahmin Tamizh would offer, I don't think any other form of Tamizh would offer - the eloquent part of it.* It is a bit difficult to come across.

Researcher: Example?

* I can't quote any, like I said people when they speak in the rural areas, they are a bit fast in their conversations, they use different words- At times we are complicating sentences and theirs are much more simpler as far as Ive heard.

Resarcher: Okay I am coming to the religious part of it. Would you see yourself as a Hindu or would you say you are a Brahmin? How would you introduce yourself to somebody?

I would say I am a Hindu- on a overall level. Simple. Okay. I work in a Software Company. What do you do in a software company, There are departments from HR to development. What do you do is a different thing. Like that, I am associating myself. I am Hindu basically. I follow the Hindu philosophy. Okay when you see the Indian religion as a whole, Hindu as a term is coined by a foreigner to identify the people of India who follow a particular standard of they follow the scripture and they follow their own gods and goddesses. That was the term associated. So for the rest of the world, I won't complicate it- I am a Hindu. Internally, here, I am a Brahmin. I was born a Brahmin, I was raised a Brahmin. I am not quoting this as a superior concept. This is what I am. I can't shun away from what I am. Hindutva or Hindu was first quoted by the British. I don't know it was quoted by the British or somebody else or the Mughal people quoted or something. This is the entire, it is the superset of all the Indian traditions and cultures. Our *Varnashrama* and everything is encompassed in this. And, Hinduism is this. Its like saying these guys are Muslims, they are Christians, There are protestants etc., And there are so many people out there. When they introduce themselves to someone, they might say okay I am a protestant or Catholic or something. Normally, when you go to foreign, they will just use the term Christian. Like that, from a global point of view, I am a Hindu but I am a Brahmin. Brahmin is actually what I am. Hindu is what you've chosen you are comfortable with its a word that the foreign people know that they - even us Indians right now in the government forms itself say - are you Hindu' or something else . I am Brahmin and internally I know that I am a Brahmin but for identifying purpose, to make it easy for others, I will call myself as Hindu.

How do you see the term Brahmanism int he context of technology and especially for Tamil Brahmins? (**41 minutes 41 seconds**)

From a language point of view and even from a spiritual point of view, and a s Brahmin, Brahminism and Hinduism, you cant actually separate them out. There are Vedas, we have our own set of bulky set of scriptures. Okay and Brahminism is more like a spiritual quotient - a level or something you can say. Okay, instead of associating it with groups and individuals, it is something that you associate with yourself. at a primary level. You can be spiritual enough, you can be brahmin at heart. **So Brahmanism it something that I would say that it is something sacred. At the same time, it has been subject to many different form of manipulations.** There are many different perceptions outside. Right now Brahmin is something akin to as I said, you are a protestant, you are a catholic, its is like a way, it is not a way that is the most primary part of it. Brahminism doesnt entail anything, it doesnt take a particular piece of a scripture and says 'this is what we have to follow' . No. Hinduism,

in general, the entire set of Vedas encompass it applies to all divisions of Hinduism. You cannot say, I am only going to take the Rig Veda. Okay, basically when you are doing any what do you say from a religious offering point of view, when you are doing yagna or something, you will take excerpt from the Vedas and you will start saying that but, பொதுவா இப்போ நா என்னைச் சொல்ல வறேன்னா வேத ஸாஸ்திரத்துல ஒரு பங்கு பிரிச்சு போட்டு, இது இவா தான் ப்ராமணா , இவா வந்து Brahmins utilise this part of Hinduism, the other set

Researcher intervenes: No no I am trying to - இப்போ நம்ம வந்து சமஸ்க்ருதம் கலந்தத் தமிழ் பேசறோம் இல, which is very different to the Non Brahmins,

Participant: அது தான் I am getting to that.

நா வந்து first நா இங்கேர்ந்து வரேன். So it is nothing that Brahmins had a separate way. அப்டிங்கறது கிடையாது. okay சமஸ்க்ருதம் ஏன் தமிழ்ல இப்போ நா என்ன சொல்ல வறேன்னா வேத ஸாஸ்திரத்தை entire ஆ, follow பண்ணறது தான் Hinduism and Brahmins are part of Hinduism. so and coming from that, சம்ஸ்க்ருதம் என் வரத்து அப்டினு பார்த்தோம் நா - You are a Brahmin and நீங்க ஒரு Brahmin ஆ, இருக்கறவன் வந்து ஒரு கோவிலுக்கு போயோ எதாவது ஒரு மந்திரமோ (*Mantram*) *ஷ்ஊலோகமோ* சொல்லணும்னா Sholka is written in Veda shastras, it is in Sanskrit. You can give the closest Tamizh translation to it. அதத் தவர, இப்போ நம்ம ஒரு ஷிவன் கோவிலுக்கு போறோம் there are ஷிவன் அடியார் இருக்கா எல்லாமே their language is purely ஷிவன் அடியார் சொல்ற எல்லாமே தமிழ்ல தான் இருக்கு. There is no Sanskrit and all in that. So அதெல்லாம் வந்து see அத தான் சொல்றேன், When you look Hinduism as a part of it, it embraces all languages especially in Tamizh. Sanskrit கலந்தத் தமிழ் அப்டினு பார்த்தோம்னா வேதத்திலிருந்து direct ஆ எடுத்து சொல்ற ஷ்லோகத்துக்கு வரும் and Brahmin community க்குள்ள communicate பண்றப்போ இந்த வார்த்தை ல சிலதெல்லாம் எடுத்து we use it for our daily terms. தமிழலியும் scriptural wise ஆ பார்த்தோம் நா தமிழிலும் pure text இருக்கு so Brahmin Tamizh அப்டிங்கறது இருக்கு **and at the same time, traditional Tamizh அதுக்கும் பார்த்தோம்னா differences வந்து you can easily filter out the difference. Not, எல்லா வார்த்தையும் Sanskritize ஆயிருக்காது ஆனா அதே சமையம் வந்து Normal Tamizh உம் இதுவும் ஒரே மாதிரி இருக்கும்னு சொல்லவும் முடியாது. There are ஒரு complication இருக்க தான் செய்யறது இங்க**

B.1 QUALITATIVE

WORDS TRANSLITERATED IN ROMAN SCRIPT

Statistics

Alai

N	Valid	177
	Missing	0

Alai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aalai	2	1.1	1.1	1.1
	aali	2	1.1	1.1	2.3
	alai	125	70.6	70.6	72.9
	alaiy	1	.6	.6	73.4
	alay	1	.6	.6	74.0
	ali	2	1.1	1.1	75.1
	alie	2	1.1	1.1	76.3
	allai	22	12.4	12.4	88.7
	alli	2	1.1	1.1	89.8
	azalai	1	.6	.6	90.4
	azhai	5	2.8	2.8	93.2
	azhlai	1	.6	.6	93.8
	MH	2	1.1	1.1	94.9
	NA	9	5.1	5.1	100.0
	Total	177	100.0	100.0	

B.2 - Muganool

Statistics

Muganool

N	Valid	177
	Missing	0

Muganool

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MH	14	7.9	7.9	7.9
	muganiik	1	.6	.6	8.5
	muganol	2	1.1	1.1	9.6
	muganool	68	38.4	38.4	48.0
	muganual	1	.6	.6	48.6
	muganul	41	23.2	23.2	71.8
	muganule	1	.6	.6	72.3
	muganull	4	2.3	2.3	74.6
	mugganul	1	.6	.6	75.1
	mughanool	2	1.1	1.1	76.3
	muhanuul	1	.6	.6	76.8
	mukanool	3	1.7	1.7	78.5
	mukanuil	1	.6	.6	79.1
	mukanul	17	9.6	9.6	88.7
	mukanule	3	1.7	1.7	90.4
	mukanull	1	.6	.6	91.0
	myganool	1	.6	.6	91.5
	NA	13	7.3	7.3	98.9
	NL	1	.6	.6	99.4
	nuganul	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.3 Azhai

		Frequency	Percent	Valid Percent	Cumulative Percent
			Azhai		
Valid	aalai	2	1.1	1.1	1.1
	aazhai	1	.6	.6	1.7
	alaai	3	1.7	1.7	3.4
	alaaii	1	.6	.6	4.0
	alai	62	35.0	35.0	39.0
	alaie	1	.6	.6	39.5
	alaiey	1	.6	.6	40.1
	alaii	2	1.1	1.1	41.2
	alaiy	1	.6	.6	41.8
	alay	1	.6	.6	42.4
	alaz	1	.6	.6	42.9
	alazi	1	.6	.6	43.5
	alhai	1	.6	.6	44.1
	ali	1	.6	.6	44.6
	alie	1	.6	.6	45.2
	allai	13	7.3	7.3	52.5
	allaii	1	.6	.6	53.1
	allei	2	1.1	1.1	54.2
	alli	3	1.7	1.7	55.9
	alzai	1	.6	.6	56.5
	alzhi	1	.6	.6	57.1
	alzie	1	.6	.6	57.6
	azai	8	4.5	4.5	62.1
	azali	1	.6	.6	62.7
	azallai	1	.6	.6	63.3
	azhai	38	21.5	21.5	84.7
	azhali	1	.6	.6	85.3
	azhi	1	.6	.6	85.9
	azhlai	3	1.7	1.7	87.6
	azi	1	.6	.6	88.1
	azlai	7	4.0	4.0	92.1
	MH	6	3.4	3.4	95.5
	NA	8	4.5	4.5	100.0
	Total	177	100.0	100.0	

B.4 Azhaippithazh

Azhaippithazh

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aazhaipethal	1	.6	.6	.6
	alaaipidhal	1	.6	.6	1.1
	alaibithal	5	2.8	2.8	4.0
	alaiipidhal	1	.6	.6	4.5
	alaipethal	2	1.1	1.1	5.6
	alaipethale	1	.6	.6	6.2
	alaipidhal	4	2.3	2.3	8.5
	alaipithal	42	23.7	23.7	32.2
	alaipithazh	1	.6	.6	32.8
	alaippathal	1	.6	.6	33.3
	alaippethal	1	.6	.6	33.9
	alaippidhal	2	1.1	1.1	35.0
	alaippithal	17	9.6	9.6	44.6
	alayppithal	1	.6	.6	45.2
	alazipithal	1	.6	.6	45.8
	aliepethal	2	1.1	1.1	46.9
	alipethal	1	.6	.6	47.5
	alipithal	3	1.7	1.7	49.2
	allaipedhal	1	.6	.6	49.7
	allaipethal	1	.6	.6	50.3
	allaiphithal	1	.6	.6	50.8
	allaipithal	13	7.3	7.3	58.2
	alleipithal	1	.6	.6	58.8
	allipithal	2	1.1	1.1	59.9
	alypithal	1	.6	.6	60.5
	azaipidhal	2	1.1	1.1	61.6
	azaipithal	4	2.3	2.3	63.8
	azaippithal	1	.6	.6	64.4
	azallaipidhal	1	.6	.6	65.0
	azhaipidhal	4	2.3	2.3	67.2
	azhaipithal	13	7.3	7.3	74.6
	azhaipithazh	2	1.1	1.1	75.7
	azhaipithzh	1	.6	.6	76.3
	azhaippathal	1	.6	.6	76.8
	azhaippidhal	1	.6	.6	77.4
	azhaippidhaz	1	.6	.6	78.0

azhaippidhal	1	.6	.6	
azhaippidhaz	1	.6	.6	78.0
azhaippithal	7	4.0	4.0	81.9
azhaippithazh	3	1.7	1.7	83.6
azhalipithal	1	.6	.6	84.2
azhapithal	1	.6	.6	84.7
azhipithal	1	.6	.6	85.3
azhlaipidhal	1	.6	.6	85.9
azhlaipithal	2	1.1	1.1	87.0
azibbethal	2	1.1	1.1	88.1
azlaipidhal	1	.6	.6	88.7
azlaipithal	3	1.7	1.7	90.4
azlaippithal	1	.6	.6	91.0
azphidhal	1	.6	.6	91.5
muganool	1	.6	.6	92.1
NA	14	7.9	7.9	100.0
Total	177	100.0	100.0	

B.5 Vizhuppuram

Statistics

Vizhuppuram

N	Valid	177
	Missing	0

Vizhuppuram

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NA	6	3.4	3.4	3.4
	valuppuram	1	.6	.6	4.0
	velluppuram	1	.6	.6	4.5
	vellupuram	2	1.1	1.1	5.6
	veluppuram	3	1.7	1.7	7.3
	velupuram	6	3.4	3.4	10.7
	velupurame	1	.6	.6	11.3
	villuburam	1	.6	.6	11.9
	villuppuram	3	1.7	1.7	13.6
	villupuram	37	20.9	20.9	34.5
	viluppuram	43	24.3	24.3	58.8
	vilupuram	45	25.4	25.4	84.2
	vizhlupuram	1	.6	.6	84.7
	vizhuppuram	12	6.8	6.8	91.5
	vizhupuram	13	7.3	7.3	98.9
	vizluppuram	1	.6	.6	99.4
	vizupuram	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.6 Ka<u>ll</u>u (லு) See B.7 for the difference

Statistics

Kallu

N	Valid	177
	Missing	0

Kallu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	kaalu	1	.6	.6	.6
	kaillu	1	.6	.6	1.1
	kailu	2	1.1	1.1	2.3
	kal	21	11.9	11.9	14.1
	kalilu	1	.6	.6	14.7
	kall	3	1.7	1.7	16.4
	kallu	131	74.0	74.0	90.4
	kalu	5	2.8	2.8	93.2
	kalzhu	1	.6	.6	93.8
	kellu	1	.6	.6	94.4
	MH	1	.6	.6	94.9
	NA	9	5.1	5.1	100.0
	Total	177	100.0	100.0	

B.7 Kaḻḻu (எ)

Kallu2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	gal	2	1.1	1.1	1.1
	gallu	3	1.7	1.7	2.8
	kaalu	1	.6	.6	3.4
	kaillu	1	.6	.6	4.0
	kailu	2	1.1	1.1	5.1
	kailyu	1	.6	.6	5.6
	kal	9	5.1	5.1	10.7
	kale	1	.6	.6	11.3
	kalelu	1	.6	.6	11.9
	kalhu	1	.6	.6	12.4
	kalilu	1	.6	.6	13.0
	kall	3	1.7	1.7	14.7
	kalloo	1	.6	.6	15.3
	kallu	105	59.3	59.3	74.6
	kallue	1	.6	.6	75.1
	kalluu	1	.6	.6	75.7
	kalu	11	6.2	6.2	81.9
	kaluzhu	1	.6	.6	82.5
	kalzhlu	1	.6	.6	83.1
	kalzhu	1	.6	.6	83.6
	kazazhu	1	.6	.6	84.2
	kazh	1	.6	.6	84.7
	kazhllu	1	.6	.6	85.3
	kazhlu	1	.6	.6	85.9
	kazhu	1	.6	.6	86.4
	kazhzhu	2	1.1	1.1	87.6
	kazllu	1	.6	.6	88.1
	kazlu	2	1.1	1.1	89.3
	kazu	1	.6	.6	89.8
	kazzhu	1	.6	.6	90.4
	kazzlu	1	.6	.6	91.0
	kazzu	1	.6	.6	91.5
	khallu	1	.6	.6	92.1
	kzhallu	1	.6	.6	92.7
	MH	2	1.1	1.1	93.8

khallu	1	.6	.6	92.1
kzhallu	1	.6	.6	92.7
MH	2	1.1	1.1	93.8
NA	11	6.2	6.2	100.0
Total	177	100.0	100.0	

B.8 Pi*zh*ai (மீ)

Pizhai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	kalzhu	1	.6	.6	.6
	MH	2	1.1	1.1	1.7
	NA	12	6.8	6.8	8.5
	NL	1	.6	.6	9.0
	palai	1	.6	.6	9.6
	pali	1	.6	.6	10.2
	pelai	10	5.6	5.6	15.8
	pelay	1	.6	.6	16.4
	pelei	1	.6	.6	16.9
	pellai	2	1.1	1.1	18.1
	pelli	1	.6	.6	18.6
	pezhai	1	.6	.6	19.2
	philai	2	1.1	1.1	20.3
	phizhai	1	.6	.6	20.9
	pilai	72	40.7	40.7	61.6
	pilaii	1	.6	.6	62.1
	pilay	2	1.1	1.1	63.3
	pilei	1	.6	.6	63.8
	pili	1	.6	.6	64.4
	pillai	13	7.3	7.3	71.8
	pillay	1	.6	.6	72.3
	pizai	8	4.5	4.5	76.8
	pizali	1	.6	.6	77.4
	pizhai	32	18.1	18.1	95.5
	pizhlai	2	1.1	1.1	96.6
	pizlai	1	.6	.6	97.2
	plai	1	.6	.6	97.7
	plazi	1	.6	.6	98.3
	pzhali	1	.6	.6	98.9
	pzlai	2	1.1	1.1	100.0
	Total	177	100.0	100.0	

B.9 Vaalu

Statistics

Vaalu

N	Valid	177
	Missing	0

Vaalu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MH	16	9.0	9.0	9.0
	NA	6	3.4	3.4	12.4
	NL	1	.6	.6	13.0
	vaal	1	.6	.6	13.6
	vaallu	1	.6	.6	14.1
	vaalu	50	28.2	28.2	42.4
	vaazhu	1	.6	.6	42.9
	val	1	.6	.6	43.5
	vallu	36	20.3	20.3	63.8
	valu	58	32.8	32.8	96.6
	vazhlu	2	1.1	1.1	97.7
	vazhu	2	1.1	1.1	98.9
	vazu	1	.6	.6	99.4
	vhazhu	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.10 Viḻakku (எ)

Statistics

Vilakku

N	Valid	177
	Missing	0

Vilakku

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MH	10	5.6	5.6	5.6
	NA	8	4.5	4.5	10.2
	NL	1	.6	.6	10.7
	velagu	1	.6	.6	11.3
	velakki	1	.6	.6	11.9
	velakku	12	6.8	6.8	18.6
	velaku	2	1.1	1.1	19.8
	vellakku	1	.6	.6	20.3
	vellaku	1	.6	.6	20.9
	vilakkam	1	.6	.6	21.5
	vilakku	77	43.5	43.5	65.0
	vilaku	28	15.8	15.8	80.8
	vilkku	1	.6	.6	81.4
	villakku	9	5.1	5.1	86.4
	villaku	7	4.0	4.0	90.4
	vilzakku	1	.6	.6	91.0
	vizaku	1	.6	.6	91.5
	vizhakku	10	5.6	5.6	97.2
	vizhaku	1	.6	.6	97.7
	vizhalaku	1	.6	.6	98.3
	vizhukku	1	.6	.6	98.9
	vlakku	1	.6	.6	99.4
	vlaku	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.11 Maattu

Statistics

Mattu

N	Valid	177
	Missing	0

Mattu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	maaddu	1	.6	.6	.6
	maatdu	1	.6	.6	1.1
	maathu	4	2.3	2.3	3.4
	maattu	28	15.8	15.8	19.2
	maatu	13	7.3	7.3	26.6
	maddu	3	1.7	1.7	28.2
	maittu	1	.6	.6	28.8
	maitu	1	.6	.6	29.4
	matdu	2	1.1	1.1	30.5
	mathu	1	.6	.6	31.1
	mathyu	1	.6	.6	31.6
	mattu	108	61.0	61.0	92.7
	matu	3	1.7	1.7	94.4
	MH	2	1.1	1.1	95.5
	NA	8	4.5	4.5	100.0
	Total	177	100.0	100.0	

B.12 SorkaL **(எ)**

196

Statistics

Sorkal

N	Valid	177
	Missing	0

Sorkal

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	chorkhal	1	.6	.6	.6
	NA	10	5.6	5.6	6.2
	scorkal	1	.6	.6	6.8
	shorkkal	2	1.1	1.1	7.9
	soarkal	1	.6	.6	8.5
	soarkkal	1	.6	.6	9.0
	sorekall	1	.6	.6	9.6
	sorgal	8	4.5	4.5	14.1
	sorgkal	1	.6	.6	14.7
	sorkal	96	54.2	54.2	68.9
	sorkall	1	.6	.6	69.5
	sorkkal	46	26.0	26.0	95.5
	sorkkale	1	.6	.6	96.0
	sorrkal	1	.6	.6	96.6
	sorukal	1	.6	.6	97.2
	sourkkal	2	1.1	1.1	98.3
	soutkal	1	.6	.6	98.9
	sozhrkal	2	1.1	1.1	100.0
	Total	177	100.0	100.0	

B.13 Pazhani (பழனி)

Statistics

Pazhani

N	Valid	177
	Missing	0

Pazhani

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	balani	1	.6	.6	.6
	MH	12	6.8	6.8	7.3
	NA	5	2.8	2.8	10.2
	palane	1	.6	.6	10.7
	palani	107	60.5	60.5	71.2
	palanie	1	.6	.6	71.8
	pallani	5	2.8	2.8	74.6
	palni	1	.6	.6	75.1
	payani	1	.6	.6	75.7
	pazahni	1	.6	.6	76.3
	pazani	5	2.8	2.8	79.1
	pazhani	33	18.6	18.6	97.7
	pazhni	1	.6	.6	98.3
	pazlani	2	1.1	1.1	99.4
	phazni	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.14 Vaa<u>zh</u>kkai (ழ)

Vaazhkkai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MH	2	1.1	1.1	1.1
	NA	7	4.0	4.0	5.1
	NL	1	.6	.6	5.6
	vaalgai	1	.6	.6	6.2
	vaalkai	17	9.6	9.6	15.8
	vaalkhai	1	.6	.6	16.4
	vaalkkai	9	5.1	5.1	21.5
	vaazhgai	1	.6	.6	22.0
	vaazhkai	3	1.7	1.7	23.7
	vaazhkkai	3	1.7	1.7	25.4
	vahzkai	1	.6	.6	26.0
	val	1	.6	.6	26.6
	valgai	4	2.3	2.3	28.8
	valikai	1	.6	.6	29.4
	valiky	1	.6	.6	29.9
	valkai	44	24.9	24.9	54.8
	valkgai	1	.6	.6	55.4
	valkhai	1	.6	.6	55.9
	valki	1	.6	.6	56.5
	valkkai	27	15.3	15.3	71.8
	vallgai	1	.6	.6	72.3
	vallkaei	1	.6	.6	72.9
	vallkai	8	4.5	4.5	77.4
	valukki	1	.6	.6	78.0
	valzhkai	2	1.1	1.1	79.1
	valzhkkai	1	.6	.6	79.7
	valzkai	2	1.1	1.1	80.8
	vazhakai	1	.6	.6	81.4
	vazhkai	15	8.5	8.5	89.8
	vazhkkai	8	4.5	4.5	94.4
	vazhlkai	2	1.1	1.1	95.5
	vazikki	1	.6	.6	96.0
	vazkai	4	2.3	2.3	98.3
	vazlhkai	1	.6	.6	98.9

		Frequency	Percent	Valid Percent	Cumulative Percent
	vazhkkai	8	4.5	4.5	94.4
	vazhlkai	2	1.1	1.1	95.5
	vazikki	1	.6	.6	96.0
	vazkai	4	2.3	2.3	98.3
	vazlhkai	1	.6	.6	98.9
	vazlkai	1	.6	.6	99.4
	vhazhkkai	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.15 Vannam

Statistics

Vannam

N	Valid	177
	Missing	0

Vannam

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.6	.6	.6
	MH	64	36.2	36.2	36.7
	NA	7	4.0	4.0	40.7
	NL	1	.6	.6	41.2
	vaannam	1	.6	.6	41.8
	valnnam	1	.6	.6	42.4
	vanam	5	2.8	2.8	45.2
	vannam	96	54.2	54.2	99.4
	vanname	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.16 Meettu

Statistics

Meettu

N	Valid	177
	Missing	0

Meettu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	maatu	2	1.1	1.1	1.1
	maitu	1	.6	.6	1.7
	mattu	7	4.0	4.0	5.6
	meattu	1	.6	.6	6.2
	meatu	3	1.7	1.7	7.9
	meddu	1	.6	.6	8.5
	medu	1	.6	.6	9.0
	meedu	1	.6	.6	9.6
	meet	2	1.1	1.1	10.7
	meethu	1	.6	.6	11.3
	meettu	46	26.0	26.0	37.3
	meetu	47	26.6	26.6	63.8
	mett	2	1.1	1.1	65.0
	mettu	28	15.8	15.8	80.8
	metu	2	1.1	1.1	81.9
	MH	11	6.2	6.2	88.1
	mittu	12	6.8	6.8	94.9
	mittuu	1	.6	.6	95.5
	mitu	2	1.1	1.1	96.6
	NA	6	3.4	3.4	100.0
	Total	177	100.0	100.0	

B.17 Kalai

Statistics

Kalai

N	Valid	177
	Missing	0

Kalai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	galai	1	.6	.6	.6
	kalai	154	87.0	87.0	87.6
	kali	3	1.7	1.7	89.3
	kallai	8	4.5	4.5	93.8
	kalli	1	.6	.6	94.4
	kazhai	3	1.7	1.7	96.0
	MH	1	.6	.6	96.6
	NA	6	3.4	3.4	100.0
	Total	177	100.0	100.0	

B.18 Tho**zhila**a**Li**

This word contains all the three syllables (zha, la and La)

Thozhilaali

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NA	12	6.8	6.8	6.8
	tholali	1	.6	.6	7.3
	tholelale	2	1.1	1.1	8.5
	tholelali	6	3.4	3.4	11.9
	tholelalli	1	.6	.6	12.4
	tholiali	1	.6	.6	13.0
	tholilaali	4	2.3	2.3	15.3
	tholilai	3	1.7	1.7	16.9
	tholilale	4	2.3	2.3	19.2
	tholilalee	1	.6	.6	19.8
	tholilalei	1	.6	.6	20.3
	tholilali	67	37.9	37.9	58.2
	tholilalie	2	1.1	1.1	59.3
	tholilaly	2	1.1	1.1	60.5
	tholilazhi	4	2.3	2.3	62.7
	tholilazi	2	1.1	1.1	63.8
	tholizahi	1	.6	.6	64.4
	tholizalzi	1	.6	.6	65.0
	tholizhali	4	2.3	2.3	67.2
	thollalei	1	.6	.6	67.8
	thollilaai	1	.6	.6	68.4
	thollilali	1	.6	.6	68.9
	thollilaly	1	.6	.6	69.5
	tholzhizhazhi	1	.6	.6	70.1
	thoulelale	1	.6	.6	70.6
	thozhihali	1	.6	.6	71.2
	thozhilaali	4	2.3	2.3	73.4
	thozhilali	34	19.2	19.2	92.7
	thozhilaly	2	1.1	1.1	93.8
	thozhilalzhi	1	.6	.6	94.4
	thozhilazhi	1	.6	.6	94.9
	thozhiyali	1	.6	.6	95.5
	thozhlali	1	.6	.6	96.0
	thozilale	1	.6	.6	96.6
	thozilali	4	2.3	2.3	98.9
	thozlilali	2	1.1	1.1	100.0
	Total	177	100.0	100.0	

B.19 Mettu

Statistics

Mettu

N	Valid	177
	Missing	0

Mettu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	galai	1	.6	.6	.6
	maattu	1	.6	.6	1.1
	maettu	2	1.1	1.1	2.3
	maittu	2	1.1	1.1	3.4
	maitu	1	.6	.6	4.0
	mattu	12	6.8	6.8	10.7
	matue	1	.6	.6	11.3
	mayttu	1	.6	.6	11.9
	meaitta	1	.6	.6	12.4
	meattu	1	.6	.6	13.0
	meatu	2	1.1	1.1	14.1
	meddu	1	.6	.6	14.7
	meettu	1	.6	.6	15.3
	meetu	4	2.3	2.3	17.5
	metdu	1	.6	.6	18.1
	metidu	1	.6	.6	18.6
	mettu	115	65.0	65.0	83.6
	metu	6	3.4	3.4	87.0
	meyttu	1	.6	.6	87.6
	meytu	1	.6	.6	88.1
	MH	5	2.8	2.8	91.0
	mottu	5	2.8	2.8	93.8
	NA	10	5.6	5.6	99.4
	NL	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.20 Ka**L**ai (எ)

Statistics

Kalai2

N	Valid	177
	Missing	0

Kalai2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	galai	2	1.1	1.1	1.1
	kaalai	1	.6	.6	1.7
	kalaai	2	1.1	1.1	2.8
	kalai	116	65.5	65.5	68.4
	kalaie	1	.6	.6	68.9
	kalaiey	1	.6	.6	69.5
	kalaii	1	.6	.6	70.1
	kalay	3	1.7	1.7	71.8
	kalei	1	.6	.6	72.3
	kalhai	1	.6	.6	72.9
	kali	2	1.1	1.1	74.0
	kallai	13	7.3	7.3	81.4
	kalli	2	1.1	1.1	82.5
	kalzhali	1	.6	.6	83.1
	kazai	1	.6	.6	83.6
	kazhai	7	4.0	4.0	87.6
	kazhi	1	.6	.6	88.1
	kazi	1	.6	.6	88.7
	kazlai	1	.6	.6	89.3
	kelai	2	1.1	1.1	90.4
	kilai	2	1.1	1.1	91.5
	klai	1	.6	.6	92.1
	MH	6	3.4	3.4	95.5
	NA	8	4.5	4.5	100.0
	Total	177	100.0	100.0	

B.21 Muttu

Statistics

Muttu

N	Valid	177
	Missing	0

Muttu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	mattu	4	2.3	2.3	2.3
	meddu	1	.6	.6	2.8
	mettu	2	1.1	1.1	4.0
	Mh	1	.6	.6	4.5
	MH	10	5.6	5.6	10.2
	moottu	1	.6	.6	10.7
	mootu	1	.6	.6	11.3
	motidu	1	.6	.6	11.9
	mottu	30	16.9	16.9	28.8
	mouttu	1	.6	.6	29.4
	muittu	1	.6	.6	29.9
	muthu	1	.6	.6	30.5
	mutt	1	.6	.6	31.1
	muttu	104	58.8	58.8	89.8
	mutu	5	2.8	2.8	92.7
	NA	12	6.8	6.8	99.4
	NL	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.22 Kilai

Statistics

Kilai

N	Valid	177
	Missing	0

Kilai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	kalai	1	.6	.6	.6
	kalaie	1	.6	.6	1.1
	keelai	2	1.1	1.1	2.3
	kelai	25	14.1	14.1	16.4
	kelaie	1	.6	.6	16.9
	kelay	1	.6	.6	17.5
	keli	4	2.3	2.3	19.8
	keliy	1	.6	.6	20.3
	kellai	5	2.8	2.8	23.2
	kelli	1	.6	.6	23.7
	kezhai	1	.6	.6	24.3
	kilai	95	53.7	53.7	78.0
	kilay	1	.6	.6	78.5
	kile	1	.6	.6	79.1
	kili	3	1.7	1.7	80.8
	killai	14	7.9	7.9	88.7
	kizhai	5	2.8	2.8	91.5
	kizhlai	1	.6	.6	92.1
	MH	3	1.7	1.7	93.8
	NA	11	6.2	6.2	100.0
	Total	177	100.0	100.0	

B.23 Kuttu

Statistics

Kuttu

N	Valid	177
	Missing	0

Kuttu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	koottu	1	.6	.6	.6
	kottu	3	1.7	1.7	2.3
	kuddu	2	1.1	1.1	3.4
	kuittu	2	1.1	1.1	4.5
	kuthu	1	.6	.6	5.1
	kuttu	123	69.5	69.5	74.6
	kutu	7	4.0	4.0	78.5
	MH	26	14.7	14.7	93.2
	NA	12	6.8	6.8	100.0
	Total	177	100.0	100.0	

B.24 Kizhi

Statistics

Kizhi

N	Valid	177
	Missing	0

Kizhi

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	gili	1	.6	.6	.6
	keele	1	.6	.6	1.1
	keelhi	1	.6	.6	1.7
	keeli	2	1.1	1.1	2.8
	kelai	3	1.7	1.7	4.5
	kelei	1	.6	.6	5.1
	keli	17	9.6	9.6	14.7
	kelli	6	3.4	3.4	18.1
	kezhi	1	.6	.6	18.6
	kilai	1	.6	.6	19.2
	kile	5	2.8	2.8	22.0
	kilee	2	1.1	1.1	23.2
	kilhi	1	.6	.6	23.7
	kili	96	54.2	54.2	78.0
	kilie	3	1.7	1.7	79.7
	kilizi	1	.6	.6	80.2
	killi	13	7.3	7.3	87.6
	killy	3	1.7	1.7	89.3
	kily	1	.6	.6	89.8
	kizhi	6	3.4	3.4	93.2
	kizhli	2	1.1	1.1	94.4
	lkili	1	.6	.6	94.9
	NA	9	5.1	5.1	100.0
	Total	177	100.0	100.0	

B.25 Maettu

Statistics

Maettu

N	Valid	177
	Missing	0

Maettu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	maattu	2	1.1	1.1	1.1
	maatu	1	.6	.6	1.7
	madu	1	.6	.6	2.3
	maeitu	1	.6	.6	2.8
	maettu	10	5.6	5.6	8.5
	maetu	4	2.3	2.3	10.7
	maittu	1	.6	.6	11.3
	mateu	1	.6	.6	11.9
	mattu	18	10.2	10.2	22.0
	matu	1	.6	.6	22.6
	matue	1	.6	.6	23.2
	maydu	1	.6	.6	23.7
	mayttu	4	2.3	2.3	26.0
	maytu	1	.6	.6	26.6
	meattu	21	11.9	11.9	38.4
	meatu	5	2.8	2.8	41.2
	meddu	2	1.1	1.1	42.4
	meettu	3	1.7	1.7	44.1
	meetu	4	2.3	2.3	46.3
	meeyatu	1	.6	.6	46.9
	meto	1	.6	.6	47.5
	mettu	70	39.5	39.5	87.0
	meyttu	2	1.1	1.1	88.1
	MH	6	3.4	3.4	91.5
	mittu	2	1.1	1.1	92.7
	mottu	2	1.1	1.1	93.8
	NA	10	5.6	5.6	99.4
	NL	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.26 Thoguppaalar

Statistics

Thoguppaalar

N	Valid	177
	Missing	0

Thoguppaalar

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NA	10	5.6	5.6	5.6
	thappalar	1	.6	.6	6.2
	thogeppaalar	1	.6	.6	6.8
	thoghupalar	1	.6	.6	7.3
	thogubalar	1	.6	.6	7.9
	thogupaalar	5	2.8	2.8	10.7
	thogupahlar	1	.6	.6	11.3
	thogupalaar	1	.6	.6	11.9
	thogupalar	39	22.0	22.0	33.9
	thogupallar	5	2.8	2.8	36.7
	thogupazhalar	1	.6	.6	37.3
	thoguphalar	1	.6	.6	37.9
	thoguppalar	55	31.1	31.1	68.9
	thogupphalar	1	.6	.6	69.5
	thokubalar	1	.6	.6	70.1
	thokupaalar	1	.6	.6	70.6
	thokupalar	9	5.1	5.1	75.7
	thokupalare	1	.6	.6	76.3
	thokupallar	2	1.1	1.1	77.4
	thokupaller	1	.6	.6	78.0
	thokuppaalar	2	1.1	1.1	79.1
	thokuppalar	32	18.1	18.1	97.2
	thokuppalare	1	.6	.6	97.7
	thoppalar	1	.6	.6	98.3
	thoukupalar	2	1.1	1.1	99.4
	thukuppalar	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.27 Ennai

Statistics

Ennai

N	Valid	177
	Missing	0

Ennai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aaennai	1	.6	.6	.6
	ainnai	1	.6	.6	1.1
	anai	1	.6	.6	1.7
	anainai	1	.6	.6	2.3
	annai	9	5.1	5.1	7.3
	anney	1	.6	.6	7.9
	eannai	2	1.1	1.1	9.0
	enai	3	1.7	1.7	10.7
	ennai	89	50.3	50.3	61.0
	ennaie	1	.6	.6	61.6
	ennaiy	3	1.7	1.7	63.3
	ennanai	1	.6	.6	63.8
	ennay	1	.6	.6	64.4
	enneiy	2	1.1	1.1	65.5
	enney	2	1.1	1.1	66.7
	enneyi	1	.6	.6	67.2
	enni	1	.6	.6	67.8
	enniye	1	.6	.6	68.4
	iannai	1	.6	.6	68.9
	NA	15	8.5	8.5	77.4
	yainnai	1	.6	.6	78.0
	yannai	5	2.8	2.8	80.8
	yeannai	2	1.1	1.1	81.9
	yenai	3	1.7	1.7	83.6
	yennai	28	15.8	15.8	99.4
	yenni	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.28 Yaetrukkol

Statistics

Yaetrukkol

N	Valid	177
	Missing	0

Yaetrukkol

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aathrukkol	1	.6	.6	.6
	aatrukozhl	1	.6	.6	1.1
	aattrukol	1	.6	.6	1.7
	aatukol	1	.6	.6	2.3
	addrakol	1	.6	.6	2.8
	aerukol	1	.6	.6	3.4
	aetrukol	1	.6	.6	4.0
	aetrukul	1	.6	.6	4.5
	aitrukul	1	.6	.6	5.1
	arrugkol	1	.6	.6	5.6
	atrukol	1	.6	.6	6.2
	atrukul	1	.6	.6	6.8
	attrukkul	1	.6	.6	7.3
	attrukol	5	2.8	2.8	10.2
	attrukoll	1	.6	.6	10.7
	attukkol	3	1.7	1.7	12.4
	attukoll	1	.6	.6	13.0
	eadrukkol	1	.6	.6	13.6
	eaettrukol	1	.6	.6	14.1
	eatrukkol	6	3.4	3.4	17.5
	eatrukol	8	4.5	4.5	22.0
	eatrukoll	1	.6	.6	22.6
	eatrukul	1	.6	.6	23.2
	eattrukol	2	1.1	1.1	24.3
	eattukgol	1	.6	.6	24.9
	eattukol	2	1.1	1.1	26.0

eetrukol	1	.6	.6	26.6
errukkol	1	.6	.6	27.1
errukol	1	.6	.6	27.7
ethrukol	1	.6	.6	28.2
ethukkol	1	.6	.6	28.8
etrukkol	2	1.1	1.1	29.9
etrukol	10	5.6	5.6	35.6
etterukkol	1	.6	.6	36.2
ettrugol	1	.6	.6	36.7
ettrukkol	4	2.3	2.3	39.0
ettrukol	5	2.8	2.8	41.8
ettrukoll	1	.6	.6	42.4
ettukkol	1	.6	.6	42.9
ettukol	1	.6	.6	43.5
etturgol	1	.6	.6	44.1
etturukol	1	.6	.6	44.6
etturukul	1	.6	.6	45.2
NA	49	27.7	27.7	72.9
NL	3	1.7	1.7	74.6
yaetrukol	1	.6	.6	75.1
yaetrukole	1	.6	.6	75.7
yatrikol	1	.6	.6	76.3
yatrukkol	1	.6	.6	76.8
yatrukol	2	1.1	1.1	78.0
yattrugol	1	.6	.6	78.5
yattrukkol	1	.6	.6	79.1
yattrukol	2	1.1	1.1	80.2
yattukol	1	.6	.6	80.8
yatturgol	1	.6	.6	81.4
yeatrukol	1	.6	.6	81.9
yeatrukul	1	.6	.6	82.5
yeattrukol	3	1.7	1.7	84.2
yerrukol	1	.6	.6	84.7
yerukkol	1	.6	.6	85.3
yerukul	1	.6	.6	85.9
yetrukkol	5	2.8	2.8	88.7
yetrukol	12	6.8	6.8	95.5
yetrukole	1	.6	.6	96.0
yettrukkol	3	1.7	1.7	97.7

		Frequency	Percent	Valid Percent	Cumulative Percent
	yettrukkol	3	1.7	1.7	97.7
	yettrukol	3	1.7	1.7	99.4
	yettrukool	1	.6	.6	100.0
	Total	177	100.0	100.0	

B.29 Yezhudhugol

Statistics

Yezhudhugol

N	Valid	177
	Missing	0

Yezhudhugol

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	aaluthukol	1	.6	.6	.6
	alluthukol	1	.6	.6	1.1
	aluthugol	3	1.7	1.7	2.8
	aluthukkol	2	1.1	1.1	4.0
	aluthukoal	1	.6	.6	4.5
	aluthukol	10	5.6	5.6	10.2
	azhudhukol	1	.6	.6	10.7
	azhudukkol	1	.6	.6	11.3
	ealuthugol	1	.6	.6	11.9
	ealuthukol	3	1.7	1.7	13.6
	ealuthukoll	2	1.1	1.1	14.7
	eazluthugol	1	.6	.6	15.3
	elathucol	1	.6	.6	15.8
	elauthukol	1	.6	.6	16.4
	elluthugol	1	.6	.6	16.9
	elluthukoll	1	.6	.6	17.5
	elthugoal	1	.6	.6	18.1
	eludhugol	1	.6	.6	18.6
	eludhukoal	1	.6	.6	19.2
	eludhukol	3	1.7	1.7	20.9
	eludugol	2	1.1	1.1	22.0
	eluthogol	1	.6	.6	22.6
	eluthokol	1	.6	.6	23.2
	eluthugoal	1	.6	.6	23.7
	eluthugol	10	5.6	5.6	29.4
	eluthuhol	1	.6	.6	29.9

eluthukkol	2	1.1	1.1	31.1
eluthukoal	3	1.7	1.7	32.8
eluthukoel	1	.6	.6	33.3
eluthukol	36	20.3	20.3	53.7
eluthukole	1	.6	.6	54.2
eluthukoll	1	.6	.6	54.8
eluthukool	2	1.1	1.1	55.9
eluthukozhl	1	.6	.6	56.5
elzudhukol	1	.6	.6	57.1
elzuthukol	1	.6	.6	57.6
ezhluthukoal	1	.6	.6	58.2
ezhluthukol	1	.6	.6	58.8
ezhudhugol	3	1.7	1.7	60.5
ezhudhukol	3	1.7	1.7	62.1
ezhuthugol	1	.6	.6	62.7
ezhuthukel	1	.6	.6	63.3
ezhuthukoal	1	.6	.6	63.8
ezhuthukol	15	8.5	8.5	72.3
ezluthugoal	1	.6	.6	72.9
iluthukol	1	.6	.6	73.4
NA	10	5.6	5.6	79.1
NL	1	.6	.6	79.7
yaeludhugol	1	.6	.6	80.2
yaeluthukol	1	.6	.6	80.8
yaluthugol	1	.6	.6	81.4
yaluthukoal	1	.6	.6	81.9
yaluthukol	1	.6	.6	82.5
yeakuthugol	1	.6	.6	83.1
yealuthukoal	1	.6	.6	83.6
yealuthukol	3	1.7	1.7	85.3
yedhukol	1	.6	.6	85.9
yelludhugoal	1	.6	.6	86.4
yelludukkol	1	.6	.6	87.0
yeludhukol	3	1.7	1.7	88.7
yeludhukole	1	.6	.6	89.3
yeluthugoal	1	.6	.6	89.8
yeluthugol	1	.6	.6	90.4

216

yeluthugoll	1	.6	.6	91.0
yeluthukol	11	6.2	6.2	97.2
yezhudhukol	1	.6	.6	97.7
yezhuthukoal	1	.6	.6	98.3
yezhuthukol	1	.6	.6	98.9
yezhuthukool	1	.6	.6	99.4
yezuthukoal	1	.6	.6	100.0
Total	177	100.0	100.0	

B.30 Manappaanmai

Statistics

Manappaanmai

N	Valid	177
	Missing	0

Manappaanmai

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	manabanmai	1	.6	.6	.6
	manapaanmai	7	4.0	4.0	4.5
	manapaanmay	1	.6	.6	5.1
	manapanmai	52	29.4	29.4	34.5
	manapanmaiy	1	.6	.6	35.0
	manapanmi	3	1.7	1.7	36.7
	manapanmy	2	1.1	1.1	37.9
	manapanymai	1	.6	.6	38.4
	manappaanmai	9	5.1	5.1	43.5
	manappanimi	1	.6	.6	44.1
	manappanmai	68	38.4	38.4	82.5
	manappanmaiy	2	1.1	1.1	83.6
	manappanmay	1	.6	.6	84.2
	manappanme	1	.6	.6	84.7
	manappanmi	4	2.3	2.3	87.0
	mannapanmai	6	3.4	3.4	90.4
	mannappanmae	1	.6	.6	91.0
	mannappanmai	2	1.1	1.1	92.1
	manppanmai	2	1.1	1.1	93.2
	MH	1	.6	.6	93.8
	NA	10	5.6	5.6	99.4
	NL	1	.6	.6	100.0
	Total	177	100.0	100.0	

C.1 Words in Tamil (Alai, Muganool, azhaippithazh, vizhuppuram, kallu and KaLLu)

Consistency in spelling in Tamil orthography can be observed.

A	B	C	D	E	F	G
அலை	முகநூல்	அலை (அழை)	அளைப்பிதல் (அழைப்பிதழ்)	விஞுப்புரம் (விழுப்புரம்)	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு
அலை	முகநூல்	அழை	அழைப்பிதழ்	விழுப்புரம்	கல்லு	கள்ளு

D. 1 Information sheet in Tamil

தகவல் ஆவணம்

தமிழில் பேச்சுரையில் இருந்து எழுத்து வடிவத்திற்கு மாற்றக்கூடிய தொழில்நுட்பம் தமிழ் பேசும் பிராமணர்களால் எந்தளவிற்கு ஏற்றுக்கொள்ளப்பட்டு பயன்படுத்தப்படும்?

ஆராய்ச்சி நோக்கம்

*தமிழில் பேச்சுரையில் இருந்து எழுத்து வடிவத்திற்கு மாற்றக்கூடியத் தொழில்நுட்பம் தமிழ் பிராமணர்கள் மத்தியில் வரவேற்பை பெறுமா?

*மற்றும் சமூக அமைப்பு மற்றும் மொழியின் பயன்பாடு இத்தகையான தொழில்நுட்பங்களின் வரவேற்பை தீர்மானிக்கிறதா?

ஏன் பிராமணர்கள்?
இது வரையிலான தமிழ் மொழி சார்ந்த ஆராய்ச்சியின் அடிப்படையில் இத்தொழில்நுட்பத்திற்கு மொழியின் உச்சரிப்பு மற்றும் மொழியை பிற மொழி சொற்கள் கலப்பில்லாமல் பேசுவது முக்கியமானதாகும். நமது தமிழ்ச் சமூகத்தில் ஆங்கிலக் கலப்பின்றித் தமிழ் பேசுவது இக்காலக்கடத்தில் மிகக் கடினமாகிவிட்டது. தமிழில் தொடங்கினால் ஆங்கிலத்தில் முடிவடைகிறது அல்லது இடையில் ஆங்கிலச் சொற்கள் அறிந்தோ அறியாமலோ வருகின்றன. இது பிராமணச் சமுதாயத்திற்கும் பொருந்தும். மேலும் பிராமணர் அல்லாதவர்களுக்கு பரவலாக ழ, ள உச்சரிப்பில் சிக்கல்கள் இருப்பதாக சில ஆய்வுகள் மற்றும் ஆராய்ச்சியின் அடிப்படையில் கண்டறியப்பட்டுள்ளது.

இதன் காரணமாகவே இக்காலசூழ்நிலையில் இயற்கையாக மொழிக்கலப்புள்ள தமிழ் சமூகத்தில் இருந்து பிராமணர்கள் பயன்படுத்தும் பிராமணத் தமிழை (மணிப்பிரவளம்) ஆராய்ச்சிக்காக தேர்ந்தெடுக்கப்பட்டுள்ளது. மேலும், இச்சமூகத்தினர் ழ, ள, ல வை தெளிவாக உச்சரிப்பதாக பல ஆராய்ச்சிகளின் அடிப்படையில் கண்டறிகிறோம்.

மேலும் இச்சமூகத்தினரின் உச்சரிப்பு தெளிவிற்கு காரணமாக சமஸ்க்ருத மொழி மற்றும் வேத அத்யாயனம் செய்யும் போது உச்சரிப்பிறகு கொடுக்கப்படும் முக்கியத்துவம் என்று கருதப்படுகிறது.

பங்கேற்பாளரின் கவனத்திற்கு
18 வயதிற்கு மேல் 60 வயதிற்குள் உள்ள தமிழ் நாட்டை பூர்வீகமாகக் கொண்டத் தமிழ் பிராமணர்கள் பங்கேற்கலாம்.

பங்கேற்க என்ன செய்ய வேண்டும். என்ன நடைபெறும்?

நேர்முகப் பேட்டியில் பங்கேற்க 90 நிமிடங்கள் வரை ஒதுக்க வேண்டும்.

- உங்கள் அன்றாட வாழ்க்கையைச் சார்ந்த சில கேள்விகள் கேட்கப்படும். எப்போவும் வீட்டில் பேசுகின்ற இயல்பான மொழி நடையில் உங்கள் பேச்சு இருக்க வேண்டும். மொழி, கலாச்சாரம், நீங்கள் வாழும் சமூகச் சூழல் மற்றும் நீங்கள் பயன்படுத்தும் தொழில்நுட்பத்தில் தமிழின் பயன்பாடு குறித்து கேள்விகள் கேட்கப்படும்.

- அதன் அடிப்படையில் ஆராய்ச்சியின் ஒரு பகுதியாக உருவாக்கியிருக்கும் காகித முன்மாதிரியை சோதித்து பார்த்து சோதனைகளின் முடிவுகள் பதிவு செய்யப்படும்.
- உங்கள் அனுபவங்கள், கருத்துகள் மற்றும் ஆராய்ச்சியை மேம்படுத்த உதவும் கருத்துகள் பதிவு செய்ய வாய்ப்பளிக்கப்படும்.

உங்களுடன் நடைபெறும் பேட்டி **ஒலி/ ஒளிப்பதிவு** செய்யப்படும். இது பல்கலைக்கழகத்தின் விதிமுறைக்கு உட்பட்டதாகும். பேட்டி முடிந்த சில நாட்களில், நடைபெற்ற பேட்டி எழுத்து வடிவத்தில் உங்களுக்கு அனுப்பப்படும். பிழைகள் இருந்தால் உடனடியாக என்னுடைய கவனத்திற்கு கொண்டுவருவது அவசியம். உங்களுக்கு கிடைத்து ஏழு நாட்களுக்குள் தங்களிடமிருந்து பதில் அல்லது திருத்தங்கள் வரவில்லையென்றால் எல்லாம் சரியாக இருக்க என்று எடுத்துக்கொள்ளப்படும். அதன் பிறகு அதில் மாற்றங்கள் செய்ய இயலாது.

பதிவு செய்யப்பட்டுள்ள கருத்துகள் முனைவர் பட்டம் கிடைத்த ஆறு மாதத்தில் அழிக்கப்படும். உங்கள் கருத்துகள் கடுவுச்சொல்லால் பாதுகாக்கப்பட்ட என்னுடைய கணினியில் சேமிக்கப்படும். இந்த கருத்துகள் என்னுடைய மேற்பார்வையாளரை உள்ளடக்கிய ஆராய்ச்சிக்குழுவிடம் மட்டுமே பகிரப்படும்.

இந்த ஆராய்ச்சி எவரால் சீராய்வுசெய்யப்பட்டது?

பல்கலைக்கழகத்தில் உள்ள ஆராய்ச்சிக் குழுவின் சீராய்வுக்கு பிறகு அங்கீகரிக்கப்பட்ட ஆராய்ச்சியாகும்.

இந்த ஆராய்ச்சிக்கு யார் நிதியுதவி செய்கிறார்?

இந்த ஆராய்ச்சி ஷெபீல்ட் ஹாலம் பல்கலைக்கழகத்தில் அமைந்துள்ள கலாச்சாரம் மற்றும் கணினி ஆராய்ச்சிசாலையில் சுய நிதியை பயன்படுத்தி மேற்கொள்ளப்படுகிறது.

ஆராய்ச்சியின் முடிவுகள் என்ன செய்யப்படும்?

கொடுக்கப்பட்டுள்ள தகவல்களின் அடிப்படையில் ஆராய்ந்து அதற்கு பின்பு, பெயர் மற்றும் அடையாளத்தை அகற்றியப் பிறகு ஆய்விதழ்கள், மாநாடு படைப்புகள், கலந்துரையாடல்களில் பயன்படுத்தப்படும்.

பங்கேற்பத்தின் விளைவாக ஏதேனும் உள்ளனவா?

உங்கள் விருப்பத்தின் அடிப்படையில் பங்கேபதற்கு ஒப்புக்கொண்டப் பிறகு, ஒப்புதல் படிவத்தில் உங்கள் விவரங்களை பதிவு செய்ய வேண்டும். உங்கள் தனிப்பட்ட விவரங்கள் ஒப்புதல் ஆவணத்தில் மட்டுமே பதிவு செய்யப்படும். வேறு எங்கும் பதிவுசெய்யப்பட மாட்டாது.

ஏன் பங்கேற்க வேண்டும்?

உங்கள் தாய் மொழியில் உங்கள் நண்பர்கள் மற்றும் உறவினர் இடையே உங்கள் இயல்பான மொழி நடையை மாற்றாமல் 'ஆத்துல பேசும் பாஷையை' தொழில்நுட்பத்தில் பயன்படுத்தி பயனடையலாம். இந்த ஆராய்ச்சியில் பங்கேற்பதின் மூலமாக நீங்கள் மற்றும் தமிழ் பேசும் பிராமணர்கள் பயனடைவார்கள்.

D.2 Sample invitation letter

Invitation letter

4th September 20XX

Mr. ABC DEF
Unit 12 Science Park
Howard Street
Sheffield S1 1WB

Dear Mr. DEF ,

    Ref: Our recent telephone conversation and e-mail.

I am pleased to invite you to participate in my research. I am attaching the information sheet and the consent form. I request you to carefully read through the information sheet. Should have any questions or clarifications, please feel free to get in touch with me and I shall be happy to clarify.

I am keen to meet you on Saturday the **05th of August 2016** at North Usman Road, Thyagarayanagar, Chennai 600 017 at **10.00 AM**

Should you require any further information, please feel free to get in touch with me.

Sincerely,

Raj Ramachandran

E.1 Pilot questionnaire

Sheffield Hallam University

May 2015

Thank you for choosing to take part in my pilot study!

You were offered a choice between English and Tamil. Thank you for choosing English!

Please **answer** the following questions to the best of your ability. Remember that there is no **right or wrong** answer to any of the following. The questionnaire is anonymous and therefore your honest response is much appreciated!

1. Is Tamil your mother tongue?

Yes / No

2. Which of the following best describe your origin:

Malaysia

Singapore

Sri Lanka

Tamil Nadu

United Kingdom

Other please specify _____

3. Which of the following do you consider as your first language? (please choose only **one**)

English   French   Malay   Tamil

4. Please circle the appropriate age group:

18- 25   26-40   41-50   51-60   61-70   71-100

5. What language do you speak at home?

        Tamil   Tamil mixed into English   English

        Tamil mixed into French   French   Other (please specify)

6. Which of the following language would you normally prefer to use in a social context example at a family event.

        Tamil   Tamil mixed into English   English

        Tamil mixed into French   French   Other (please specify)

7. What was the medium of instruction at school?

        Tamil   Tamil mixed into English   English

        Tamil mixed into French   French   Other (please specify)

    Was any other language used to explain a concept other than the medium of instruction?

                Yes / No

    If yes, please specify : _____

8. What was the language of instruction at University?

        Tamil   Tamil mixed into English   English

        Tamil mixed into French   French   Other (please specify)

9. Please circle one of the following that best describes your job title?

Tamil Professor /teacher

Professor/ teacher (Tamil medium)

Professor/ teacher (English medium)

Customer service (call centre environment Non Tamil)

Customer service (other than call centre)

IT Professional

Tamil Nadu government employee

Self employed

Student

Other

10. Please choose the language that you think is most frequently used by you at work:

        Tamil   Tamil mixed into English   English

        Tamil mixed into French   French   Other (please specify)

11. Do you own a smart phone?

Yes / No

If no, then please go to question 15

12. Please circle the operating system and type of smart phone you own and use:

Apple (iOS)   Android Others (please specify)   Not applicable

13. What is the current language setting in your smart phone?

14. Which of the following application do you frequently use?

SMS (text)

Facebook

Twitter

All the above

None of the above

Other

15. What is your most preferred choice of input on your smart phone?

Tamil script   English   Tamil in Roman script

16. Have you ever used voice to text technology in any language on your smart phone ?

                Yes/ No

If no, then please go to question 18

17. If yes, how would you rate your experience (in terms of recognising your speech)?

                Excellent

                Very good

                Good

                 Fair

                 Poor

                Very poor

Comments:

If you would like to make a change in the voice to text technology that you have already used, what would it be?

18. If No, please choose from one of the following:

        There was no opportunity to use.

         I don't know what that is.

I didn't feel the need to use it.

I don't own a phone that supports voice to text in any language.

Other (please specify):

19. If you have used a voice to text technology in any language, did you find it useful?

Yes / No

20. Would you use speech to text technology in Tamil on a mobile?

Yes / No

21. Generally, which of the following would you consider as an important factor for someone to use speech to text: (please circle your response)

Pronunciation

Ability to properly speak in a language without code mixing

Both

Other (please specify)

22. Generally speaking, if you want to use voice to text, how would you like the output text to appear? Please circle your response.

வணக்கம் / *vanakkam*

23. If I were to invite you for a focus group, would you be interested?

Yes / No

Please share your experience in completing this questionnaire including any comments or suggestions that you would like to make for further improvement.

E.2 Pilot questionnaire in Tamil

Sheffield Hallam University

Tamil version
வினாவரிசை

என்னுடைய ஆராய்ச்சி படிப்பிற்கு பங்கேற்க ஒப்புக்கொண்டமைக்கு நன்றி!
ஆங்கிலம் அல்லது தமிழ் என்ற விருப்பதேர்வில் தமிழை தேர்வு செய்தமைக்கு நன்றி!

தயவு செய்து இந்த வினாவரிசையை நேர்மையாக பூர்த்தி செய்ய பணிவுடன் வேண்டுகிறேன். இதில் எந்த கேள்விக்கும் சரினா பதிலோ அல்லது தவறான பதிலோ கிடையாது.

க. உங்கள் தாய் மொழி தமிழா?

         ஆம் / இல்லை

உ. தங்களின் சொந்த இடத்தை வட்டமிட்டு குறிப்பிடுக

      மலேசியா   சிங்கப்பூர்

      இலங்கை   தமிழ் நாடு

ஐக்கிய பிரிட்டிஷ் பேரரசு மற்றவை. (தயவு செய்து குறிப்பிடுங்கள்)

ங. இவற்றில் தங்களின் முதல் மொழியாக கருதும் மொழியை வட்டமிடுக (ஏதேனும் ஒன்றை மட்டும் தேர்ந்தெடுங்கள்)

    ஆங்கிலம்   பிரெஞ்சு   மலாய்   தமிழ்

ச   பொருத்தமான அகவைப் பிரிவு வட்டமிடுக. (உதவி வேண்டுமா? ஆம் / இல்லை)

         ௧அ - உ௰௫   உ௰௫ -௩௰   ௩௰௫ - ௫௰   ௫௰௫   - ௬௰
    ௬௰௫ - எ௰   எ௰௫ -ா

ரு. தாங்கள் வீட்டில் பேசக் கூடிய மொழியின் மீது வட்டமிடுக

    தமிழ்   ஆங்கிலம் கலந்தத் தமிழ்   ஆங்கிலம்

    பிரெஞ்சு கலந்தத் தமிழ்   பிரெஞ்சு

சு. சமூகச் சூழலில் தாங்கள் விரும்பிப் பேசக்கூடிய மொழியின் மீது வட்டமிடுக (எடுத்துக்காட்டு : குடும்ப நிகழ்ச்சியின் போது)

    தமிழ்   ஆங்கிலம் கலந்தத் தமிழ்   ஆங்கிலம்

    பிரெஞ்சு கலந்தத் தமிழ்   பிரெஞ்சு   மற்றவை. தயவு செய்து குறிப்பிடுங்கள்

எ. இவற்றில் இருந்து தங்களின் பள்ளியின் 'வழிக் கல்வியை' தேர்ந்தெடுத்து வட்டமிடுக

    தமிழ்   ஆங்கிலம் கலந்தத் தமிழ்   ஆங்கிலம்

    பிரெஞ்சு கலந்தத் தமிழ்   பிரெஞ்சு   மற்றவை. தயவு செய்து குறிப்பிடுங்கள்

அ. இவற்றில் இருந்து தங்களின் பல்கலைக்கழகத்தின் 'வழிக் கல்வியை' தேர்ந்தெடுத்து வட்டமிடுக

    தமிழ்   ஆங்கிலம் கலந்தத் தமிழ்   ஆங்கிலம்

    பிரெஞ்சு கலந்தத் தமிழ்   பிரெஞ்சு   மற்றவை. தயவு செய்து குறிப்பிடுங்கள்

கூ. இவற்றில் இருந்து தங்களின் தொழிலை தேர்ந்தெடுத்து வட்டமிடுக:

    தமிழாசிரியர்
    ஆசிரியர் (தமிழ் வழி)
    ஆசிரியர் (ஆங்கில வழி)
    தமிழ் நாடு அரசு அலுவலர்
    வாடிக்கையாளர் சேவை (தமிழ் அல்லாத அழைப்பு மையச் சூழல்)
    வாடிக்கையாளர் சேவை (மற்றவை)
    தகவல் தொழில்நுட்ப வல்லுநர்
    சுய தொழில்
    மாணவர்
    மற்றவை குறிப்பிடுங்கள்

ஏ. நீங்கள் அலுவலகத்தில் அதிக பயன் படுத்தும் மொழியினை தேர்ந்தெடுத்து வட்டமிடுக

    தமிழ்   ஆங்கிலம் கலந்தத் தமிழ்   ஆங்கிலம்

    பிரெஞ்சு கலந்தத் தமிழ்   பிரெஞ்சு   மற்றவை. தயவு செய்து குறிப்பிடுங்கள்

௰க. தங்களிடம் நுண்ணறிபேசி இருக்கிறதா?

    ஆம் / இல்லை

இல்லையென்றால் பதினைந்தாம் கேள்விக்கு செல்லுங்கள்

௧௨. உங்கள் நுண்ணறிபேசியில் உள்ள தற்போதிய செயலியை வட்டமிடுக

     ஐ ஒ எஸ்   ஆண்டிராய்ட்   மற்றவை (குறிப்பிடுங்கள்)

௧௩. உங்கள் நுண்ணறிபேசியில் தற்போதுள்ள மொழி அமைப்பு எது?

௧௪. இவற்றில் தாங்கள் அடிக்கடி பயன்படுத்துவதை தேர்ந்தெடுத்து வட்டமிடுக (பொருத்தமானது அனைத்தையும் தேர்ந்தெடுக்கலாம்)

     குறுஞ்செய்தி
     முகநூல்
     ட்விட்டர்
     எல்லாமே
     ஒன்றும் இல்லை
     மற்றவை (குறிப்பிடுங்கள்)

௧௫. நுண்ணறிபேசியில் உள்ளீடுக்கான தங்களுக்கு விருப்பமுள்ள எழுத்துருவத்தின் மீது வட்டமிடுக

     தமிழ்   ஆங்கிலம்   ரோமன் எழுத்துருவத்தில் தமிழ்

௧௬. ஏதேனும் மொழியில் பேச்சில் இருந்து எழுத்துருவத்திற்கு மாற்றும் தொழில்நுட்பத்தை பயன்படுத்தி இருக்கீங்களா?

               ஆம் / இல்லை

இல்லையென்றால் பதினெட்டாம் கேள்விக்கு செல்லுங்கள்

     நீங்கள் பயன்படுத்தி இருந்த பேச்சில் இருந்து எழுத்துருவத்திற்கு' மாற்றும் தொழில்நுட்பத்தில் ஏதேனும் மாற்றங்கள் அறிமுகப் படுத்த வேண்டும் என்றால் யாவை?

௧௭. பேச்சில் இருந்து எழுத்துருவத்தை மாற்றக்கூடிய தொழில்நுட்பத்தை நுண்ணறிபேசியில் பயன்படுத்தி இருந்தால் தங்களுக்கு அது உபயோகமாக அமைந்துள்ளதா?

               ஆம்/ இல்லை

௨௦. பேச்சில் இருந்து எழுத்துருவத்திற்கு மாற்றக்கூடிய தொழில்நுட்பம் நுண்ணறிபேசியில் தமிழில் வந்தால் தாங்கள் பயன்படுத்துவீர்களா?

               ஆம்/ இல்லை

௨௧. பொதுவாக, பேச்சில் இருந்து எழுத்துருவத்திற்கு மாற்றக்கூடிய தொழில்நுட்பத்தில் இவற்றில் இருந்து முக்கியமானதை தேர்ந்தெடுத்து வட்டமிடுக:

     உச்சரிப்பு
     மொழியில் கலப்பில்லாமல் பேசுவது
     இரண்டும்
     மற்றவை (குறிப்பிடுங்கள்)

௨௨. பொதுவாக, பேச்சில் இருந்து எழுத்துருவத்திற்கு மாற்றக்கூடிய தொழில்நுட்பத்தில் எந்த எழுத்துருவத்தை காண விரும்புவீர்கள்? விருப்பத்தை வட்டமிடுக

               வணக்கம் / Vanakkam

தரம் சார்ந்த ஆராய்ச்சிக்கு தங்களை அழைப்பதாக இருந்தால் தங்களுக்கு அதில் பங்கேற்க விருப்பம் உள்ளதா?

               ஆம் / இல்லை

இந்த வினாவரிசையை தமிழில் பூர்த்தி செய்ய இயலவில்லை என்றால் தயவு செய்து காரணத்தை பதிவு செய்யுங்கள்:

இந்த வினாவரிசையை பூர்த்தி செய்யும் அனுபத்தை பகிருங்கள். ஏதேனும் கருத்துகள்/ ஆலோசனைகள் இருந்தால் மேம்படுத்த உதவும்! நன்றி

E.3 Participant consent form - bilingual

பங்கேற்பாளர் ஒப்புதல் படிவம்
PARTICIPANT CONSENT FORM

1. ஆராய்ச்சி குறித்த தகவல் என்னிடம் எழுத்துருவத்தில் வழங்கப் பட்டுள்ளது. அதை நான் படித்துவிட்டேன். ஆராய்ச்சியின் விவரங்கள் என்னிடம் விளக்கப்பட்டது .

I have read the Information Sheet for this study and have the details of the study explained to me

ஆம் Yes / இல்லை No

2. இந்த ஆராய்ச்சி குறித்த என்னுடைய கேள்விகளுக்கு திருப்திகரமான பதிலளிக்கப்பட்டது. மேலும் கேள்விகள் இருந்தால் நான் கேட்கலாம் என்பதை நான் புரிந்துக்கொண்டேன்.

My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point.

ஆம் Yes / இல்லை No

3. தகவல் ஆவணத்தில் குறிப்பிட்டுள்ள கால அவகாசத்திற்குள் சுய விருப்பத்தின் அடிப்படையில் இந்த ஆராய்ச்சியில் இருந்து காரணங்கள் விளக்காமலோ அல்லது ஆராய்ச்சியில் குறிப்பிட்ட கேள்விக்கு பதிலளிக்காமல் வெளியேறலாம். அப்படி செய்வதின் மூலம் ஆராய்ச்சியாளருக்கும் எனக்கிடைய உள்ள உறவு பாதிக்காது என்பதை நான் புரிந்துகொள்கிறேன்.

I understand that I am free to withdraw from the study within the time limits outlined in the Information Sheet, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher.

ஆம் Yes / இல்லை No

4. தகவல் ஆவணத்தில் குறிப்பிட்டுள்ள மறைவடக்கத்தின் அடிப்படையில் தகவல்களை கொடுக்க ஒப்புகொள்கிறேன்.

I agree to provide information to the researcher under the conditions of confidentiality set out in the Information Sheet.

ஆம் Yes / இல்லை No

5. தகவல் ஆவணத்தில் குறிப்பிட்டுள்ள நிபந்தனைகளுக்கு கட்டுப்பட்டு இந்த ஆராய்ச்சியில் பங்கேற்க விரும்புகின்றேன்.

I wish to participate in the study under the conditions set out in the Information sheet.

ஆம் Yes / இல்லை No

6. என்னை அடையாளம் அறியாத விதமாக மாற்றிய பிறகு, நான் கொடுத்திருக்கும் தகவல்களை இந்த ஆராய்ச்சிக்கு மற்றும் வேறு ஏதேனும் ஆராய்ச்சிகளுக்கு பயன் படுத்த நான் ஒப்புதல் அளிக்கின்றேன்.
I consent to the information collected for the purpose of this research study, once anonymised(so that I cannot be identified), to be used for any other research purpose.

ஆம் Yes / இல்லை No

உங்கள் **விருப்ப மொழியில்** (தமிழ் அல்லது ஆங்கிலம்) பூர்த்தி செய்யலாம்
You are encouraged to fill it in a language of **your choice** (Tamil or English)

பங்கேற்பாளரின் கையொப்பம்/ Participant's signature _____

தேதி/ Date: _____

பங்கேற்பாளரின் பெயர்/ Participant's name　　　: _____

தொடர்புக்கு/ Contact details:

_____

_____

_____

மின் அஞ்சல்/ E Mail:_____ **கைத்தொலைபேசி/ Mobile:** _____

ஸ்கைப் முகவரி/ Skype ID:_____

ஆராய்ச்சியாளரின் பெயர்: ராஜ் / **Researcher's name**: Raj Ramachandran

ஆராய்ச்சியாளரின் கையொப்பம்/ Researcher's signature:_____

E.4 List of words used in dictation (Word in the bracket indicate the correct spelling and pronunciation. Those were were mispronounced by the researcher)

Tamil orthography	Roman orthography
அலை	Alai
முகநூல்	Muganool
அழை	Azhai
அளைப்பிதல் (அழைப்பிதழ்)	ALaippithal (azhaippithazh)
விளுப்புரம் (விழுப்புரம்)	ViLuppuram (Vizhuppuram)

கல்லு	Kallu
கள்ளு	KaLLu
பிளை (பிழை)	PiLai (pizhai)
வாலு	Valu
வால்க்கை (வாழ்க்கை)	vaalkkai (vaazhkkai)
வன்னம் (வண்ணம்)	vannam (vaNNam)
விலக்கு (விளக்கு)	vilakku (viLakku)
பலனி (பழனி)	Palani (Pazhani)
மாட்டு	Maattu
மீட்டு	Meettu
காலை	Kaalai
தொழிலாளி	ThozhilaaLi
கலை (களை)	Kalai (kaLai)
முட்டு	Muttu
கிளை	KiLai
குட்டு	Kuttu
கிளி (கிழி)	KiLi (Kizhi)
மேட்டு	Maettu
தொகுப்பலார் (தொகுப்பாளர்)	Thoguppaalar (ThoguppaaLar)
எண்ணை	ENNai
ஏற்றுக்கொள்	EttrukkoL
எளுதுகோள் (எழுதுகோல்)	ELuthugoL (Ezhuthugol)
மனப்பான்மை	Manappanmai

F.1 Participant demography and observation

Participant	Age range	Experience of using speech to text	Accuracy of pronunciation observed	Code-switching and code mixing	Behaviour intention to use Tamil	Orthography preference	Accommodation of mispronunciation or pronunciation variants
BS	45-55	No	Yes	Yes	No	Tamil	No
VM	25-35	Yes (English)	Yes	Yes	No	Roman	No
VMM	45-55	No	Yes	Yes	No	Tamil	No
VMI	35-45	No	Yes	Yes	No	Tamil, Roman and Devnagari	No

VH	35-45	No	Yes	Yes	No	Tamil, Roman	No
VJ	35-45	Yes (English)	Yes	Yes	No	Tamil for Tamil Roman for English	No
SSV	35-45	No	Yes	Yes	No		No
ARM	25-35	No	Yes	Yes	No	Tamil	No

F.2 Participant interview setting

Participant	Interview setting	Interruption	Language Code switching Sanskritised
BS	Restaurant	No	Tamil- English
VM	Residence	Yes (minor)	Tamil- English
VMM	Residence	No	Tamil- English
VMI	Residence	No	Tamil- English
VH	Residence	Yes (minor)	Tamil- English
VJ	Residence	Yes (minor)	Tamil- English
SSV	Instituition	No	Tamil- English
ARM	Instituition	No	Tamil- English