Translating Tamil Adjective Words to Sign Gestures Using Heuristic Approach

D.Narashiman*, A. Shanmugapriya** and Dr. T. Mala[§]
* Teaching Fellow (dnarashiman@gmail.com)
** Student, Master of Computer Applications
§Assistant Professor (Sr Gr) (malanehru@annauniv.edu)
Department of Information Science and Technology, CEG, Anna University, Chennai

ABSTRACT

Sign Language is a mode of communication among hearing impaired people. Sign language is not common throughout the world and there are variations among them. These variations are due to the influence of the local linguistics and community in that particular region. This leads to the development of various sign languages like British sign Language (BSL), America sign Language (ASL), Australian Sign Language (AUSLAN), Indian Sign Language (ISL). Sign Languages are translated to Natural speaking languages and vice verse by translation system to enhance the communication between hearing impaired community and speaking community. It also enriches the basic knowledge of the hearing impaired community by interacting in their own mother tongue. A little amount of work has been exhibited in automatic translation of natural text to sign gestures. Most of the systems are designed based on the direct mapping. The draw backs of that system are limited usage and domain specific. Rule based approach enumerates certain rules and depending upon the rules the sign gestures are rendered

In the proposed approach Tamil text is parsed and tagged using the morphological analyser. Using heuristic approach each and every adjective word is analysed and classified into types like shapes, colors, quality. The various sign Language gestures are mapped to an equation based on the Hand shape, location, orientation and movement. These equations in turn mapped on to the respective classified adjective words and used to render sign gestures on to screen. This system can also be extended to render sign gestures of other parts-of-speech words and it also helps Tamil native hearing impaired community people to acquire their mother tongue easily.

1. INTRODUCTION

A study conducted by a medical team of the Madras ENT Research Foundation (MERF) found that six out of every 1000 kids were victims of severe to profound deafness. This is three times the national average and six times the international average. The study conducted over a period of 10 years (2003 - 2013) had a medical team screening more than 50,000 children below 12 years of age across the state. The number indicates that Tamilnadu would soon become the deafness capital of the country and also highlights the need for automatic communication system that translates Tamil text to Tamil sign gestures. Sign Language use Manual (hand shapes) features, Non-manual (face and body parts) features, Orientation, Location and Movements for communication between Hearing Impaired people. One of the alarming characteristics of Sign Language is that they won't follow any specific grammar order as English or any other Natural language does. The Sign alignment or order varies depending on the Signer and context. Secondly the signs are less compared to the words in the speaking languages. Third issue is the representation of part of speech like Nouns which is done mostly through Finger Spelling approach. Verbs are gestured depending on their tenses. The representation of adjective words depends on space and co-articulation.

The Translation of Tamil text to Tamil Sign Language gestures depends on the characteristics of Sign Language and the animation process. The final rendering should be more effective in such a way that each every minute detail of hand gestures are shown properly along with the Non-manual features. This paper is organized in such a way first it elaborates on various systems and then describes the architecture of the proposed system followed by the outcomes and finally ends in conclusion and enhancements of the system.

2. VARIOUS TEXTS TO SIGN LANGUAGE SYSTEM

Lots of system has been developed internationally for translating text to sign language but most of them are domain specific and of restricted usage. All the systems either uses pre-recorded videos or limited number of words.

S.No	SYSTEM NAME	DOMAIN	LIMITATION
1	TESSA	Post office	Pre-Defined Phrases[1]
2	SignSynth	Weather Forecasting	Non-Manual gestures are not properly rendered
3	VisiCAST	Television Broadcasting	Human Interpreter is needed

The following table list the work that is carried out in Indian sub continent.

S.No	SYSTEM NAME	DOMAIN	LIMITATAION
1	INGIT	RAILWAY RESERVATION SYSTEM	Classifier and animation are handled properly
2	Bangla text to BaSL	General	Limited to words and video clipping[4]
3	Tamil noun to TSL	Noun words	Only noun words[2]
4	Tamil Pronoun t TSL	Pronoun words	Only pronoun words[3]

Only two works has been carried out first of its kind in Tamil Sign Language a variant of Indian Sign Language. One of the main differences between Tamil and Tamil sign language is that Tamil is morphemically rich language. There are lots of lexical gap between Tamil and Tamil Sign language. This paper proposes a method that translates Tamil adjective words to sign gestures in Tamil Sign Language. The next section discusses about the translation process of Tamil adjective words to Tamil sign language variants.

3. SYSTEM ARCHIECTURE

Some of the previous work has been carried out for translating Tamil noun and pronoun words to Tamil sign Language gestures. This architecture gives an overview of translating Tamil adjective words to Tamil sign Language gestures. Translation can be carried out in three different approach direct, transfer and Interlingua. These approaches vary depending upon the complexity of the source and target languages. If the lexical gap and grammar structure of the sentences in both the languages is more or less same direct approach is applied. In Transfer approach the source language sentences are analyzed first syntactically and then transfer from source to target language is performed finally using morphological analyzer of the target language the translation is done. Interlingua method generates an

intermediate form from the source language and translates that to the target language. In this approach translation of Tamil adjective words to Tamil sign language Variant is done using Interlingua techniques.
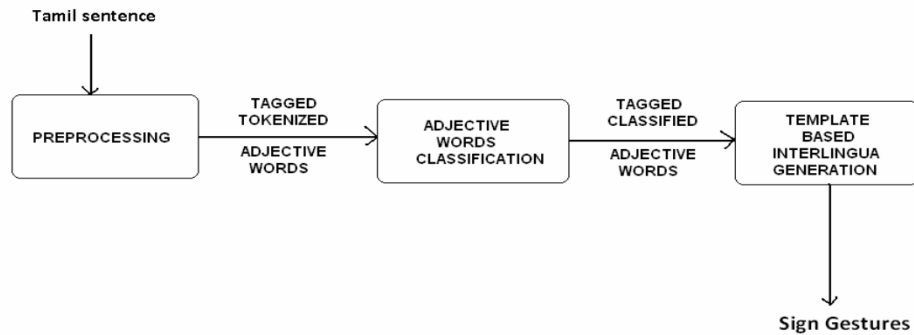

Figure 1: Architecture diagram of the proposed system

The Figure 1 depicts the overall design of the system. The system has three major modules. First module tokenizes and tag Tamil adjective word from the Tamil sentence. Second module classifies the list of Adjectives words into different classes using heuristic rules. Third module renders the Sign Gestures for the classified Adjectives.

PRE-PROCESSING: The input to this module is Tamil sentence. This module consists of two sub-modules namely Tagger and Tokenizer. Tagger is used to Tag each word of the given Tamil Sentence with its appropriate Parts of Speech such as Noun, Adjective, verb, Adverb and so on. Atcharam is a Morphological Analyzer tool for Tamil Language. Atcharam Tag each word of the Tamil sentence as Noun or Adjective or Adverb and so on and also splits them as paguthi, viguthi, idainilai, sandhi, saariyai and so on. Tokenizer module separate each tagged word of the Tamil sentence based on specific delimiters. Tokenizer parse each tagged word and separate all the words tagged as < Adjective > and < Noun>. In some cases certain Nouns are Adjectives so Nouns are also considered.

ADJECTIVE WORDS CLASSIFICATION: The tokenized adjective words are classified into different classes. Classification is done based on its common characteristics. Common characteristics are color, shape, quality and general. Heuristic Rule is used for classifying the adjective words. The tokenized Adjective words are taken as input and the corresponding class of the input adjective word is identified dynamically. The various classes of Adjectives are specified in Figure 2.
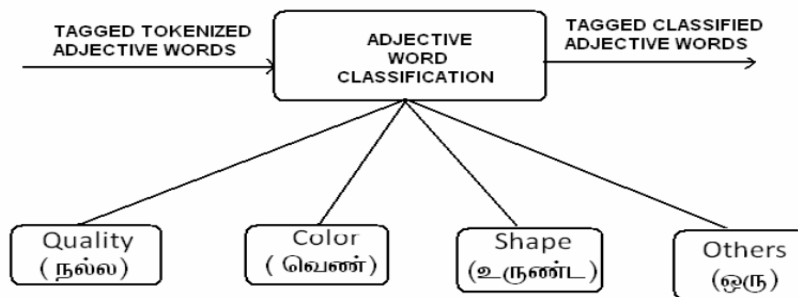

Figure 2: Adjectives Classification

Adjective words are classified into Four classes. They are Shapes, Color, Quality and Others. Since the Shapes and Colors are few in number they are written in separate files and maintained as Dictionary. The tokenized Adjective words that ends with "ò", "ô", "ì" are classified into Quality and other words into Others class

INTERLINGUA APPROACH: In this approach, the source language is translated to an intermediate Interlingua form and target language is then generated from the interlingua. The advantages of this is that it requires fewer components in order to relate source language to target language, it takes fewer components to add a new language, it supports paraphrases of the input in the original language, it allows both the analysers and generators to be written by monolingual system developers, and it handles languages that are very different from each other. Based on the identified class the adjective words are rendered into sign gestures using Java. Each Adjective word is defined with basic shapes or pre-defined formula. The sign gestures for the adjective words are generated dynamically by combining the basic primitives, changing the orientation, location and movement of the primitives based on pre-defined formula. In addition to Hand Shapes, facial expressions are also included to render the sign gestures for certain words.

4. RESULT AND CONCLUSION

The Figure 3 is a Bar Chart which tells how much percentage the Tagged Tokenized Adjective words are correctly classified into their corresponding classes based on similar characteristics. Since Dictionary is used for classifying colors and shapes almost 90% colors and shapes get classified correctly into their classes respectively. In case of Adjectives of Quality 75% of the words get promptly classified. Among 64 Adjective of Quality words 49 words are correctly classified remaining words are classified into the Adjective of General class. For Adjectives of Quality classification it is difficult to form more generalized rule. The rule formulated for Adjectives of Quality is not able to classify all the Quality depicting words correctly. In case of general remaining words other than previous three classes get classified correctly. At present sign gestures are rendered only for Tamil Adjective words. In future it can be rendered for Adverbs, Verbs etc.
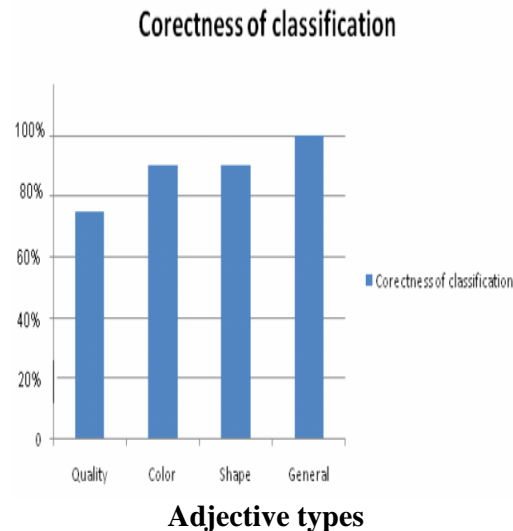
Corectness of classification

Adjective types
Figure 3 Correctness of Classification

REFERENCE:

[1]K.Datta, B.Sarkar, C.D.Datta, D.Sarkar, S.J.Dutta, I.Das Roy, A.Pual, J.U.Molla and A.Paul,"Interface between Bangla Text and Sign Language:,in the proceeding of Seminar on Applications of Computer & Embedded Technology,2009

[2]D.Narashiman, S.Vidhya and Dr.T.Mala "Tamil noun to sign language-A machine translation approach" in proceeding of 11th Tamil Internet Conference,2012,pp 175-179.

[3]D.Narashiman, Bavatharani Suriyan and Dr.T.Mala "An avatar rendering hand gesture for Tamil words" in proceeding of 12th Tamil Internet Conference,2013.pp 70-75.

[4]Tithankar dasgupta, Sandipan danpat and Anpam basu,"prototype machine translation system from text-to-Indian sign language", in proceedings of the International Joint Conference on NLP workshop for Less privileged Languages,2008,pp 19-26.