10. Rujukan

Azizi Yahya dan Norlaily binti Ahmad (2006). "Teori – Teori Konsep Kendiri". Fakulti Pendidikan, Universiti Teknologi Malaysia

Farahiyah Murni Binti Harun (2013). " Keyakinan Diri Pelajar Diploma Pendidikan Lepasan Ijazah Di Universiti Tun Hussein Onn Malaysia Terhadap ICT."

Jafar Afshinfar, Hassan Aminpoor dan Ali Mostafaei (2012). " Standardization of Fear of Success Scale in College and High school students". ISSN 0976-1233. Payame Noor University, Tehran, (4287-429)

Mazlina Jamaludin dan Nek Kamal Yeop Yunus (2008). " Kesan Lawatan Lapangan Industri Terhadap Program Pemasaran di Politeknik : Satu tinjauan".

Rohany Nasir & Fatimah Omar (2006). "Kesejahteraan Manusia".Universiti Kebangsaan Malaysia

StageofLife.com Trend Report: Teens Overcoming Fear (2013). Retrieved from http://www.stageoflife.com/StageHighSchool/WhatDoTeensFear.aspx

Vansteenkiste, Mouratidis dan Lens (2014), " Fear of Failure Can Affect Student's Motivation and Attitude to Learn, Study. British Journal of Educational Psychology.

Wan Zarina Wan Zakaria, Azizah Rajab, Norfazila Sayuti dan Nur Al-Huda Hashim (2007). "Kajian Tahap Keresahan Di Kalangan Pelajar Bahasa Asing Terhadap Mata Pelajaran Bahasa Asing". Universiti Teknologi Malaysia

Yohan Kurniawan, Thana Abdullah, Khuzaiton, Zakaria dan Wan Muhamad Wan Abdullah (2012). " Mengapa Pelajar Takut Bertanya Dalam Kuliah?: Sebuah Kajian Di Universiti Malaysia Kelantan. Jurnal Personalia Pelajar 15 (2012), (13 – 22)

TAMIL PROVERB TRANSLATION TO ENGLISH USING GOOGLE TRANSLATOR

Nithiyaa Muniandy[1], Sathya Manoharan[2]

[1]General Studies Department,
Politeknik Metro Kuala Lumpur

[2] General Studies Department,
Politeknik Metro Kuala Lumpur

nithiyaa414@gmail.com, sathi_mano@yahoo.com

Abstract

Machine translation is a system that is vastly used by many people all over the world. Google Translator and Bing translator are the most viral translators found for free in the internet. Ancient rule based method is not effective in the process of finding the real meaning of words. The subject verb agreement in English and Tamil do not tally all the time. Thus, this paper focuses on hybrid method which combines Statistical Machine Translation (SMT) and Hybrid Machine Translation (HMT) to produce efficient Machine translation of Tamil to English. This paper will be the asset for future generation as they are the fundamental for human living. The action of storing all the information in digital enables these documents from destruction and also improves Google Translator's quality and efficiency to next level or next generation.

Keywords: Machine translation, Tamil Proverb, English, Corpora, Subject-Verb-Agreement

1.0 Introduction

Machine translation is a process of converting sentences or phrases from its source language to a preferred language for better understanding without any ambiguities. The most famous machine translation providers are Google Translator and Bing translator which work based on Statistical Machine Translation (SMT) method. This paper gives a deep focus on Google Translator with a stress on Tamil Proverb conversion to English. At most of the time Tamil to English translation is not accurate and the machine simply Romanize or rewrite the Tamil words in English if the machine is unable to find the equivalent words in dictionary. This spoils the translation process and makes the meaning unclear or meaningless.

Some of the Tamil words have some uniqueness where two distinct words can be combined as a word. This uniqueness is not found in English language and hence, it is difficult to match corpora for those Tamil words in English. Existing Google Translator is encountering this problem of splitting two combined Tamil words as two distinct words. Therefore, some of the new rules have been developed based on the literature components of Tamil. Two words in English may appear one word in Tamil. Likewise two different Tamil words in Tamil proverbs mostly combined as one word.

To refine the idea, this paper has come up with an idea of combining Statistical Machine Translation (SMT) and Hybrid Machine Translation (HMT) to produce efficient Machine translation of Tamil to English. In the first part, Rule based approach is used. Collection of rules called grammar rules, morphological analysis, part of speech (POS) tagging, bilingual dictionary, source to target transliteration, and reordering rules has been used to refine the translation. In the next part, Statistical Machine Translation (SMT) is used. SMT is used to find the most accurate translation for the input sentence which is also called as decoding process. There are three different Statistical MT approach

called word based, phrase based and hierarchical based. Since this research is about translation of Tamil proverbs to English, phrase-based translation is most preferred than word-base or hierarchical phrase based translation. Phrase-based translation enables well-formed sentences and correlation between both Tamil and English languages. (Anthony, 2013)

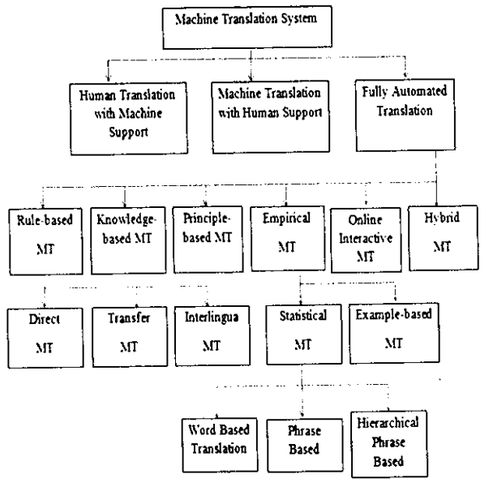
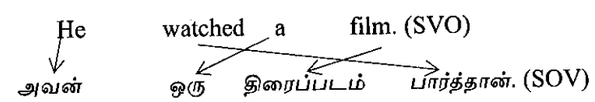


Figure 1 Machine Translation System Summary

2.0 Problem Statement

Early Tamil language translation was about using rule based approach to translate words from European Languages to Tamil Language. Unfortunately, most of the translation from Tamil to English does not give a desirable output since Tamil is morphologically rich in features. Moreover, rule based approach does translation only at the word level by performing word-by-word translation because European languages are lack in linguistics. Most researches follow statistical and hybrid method to develop machine translation system for Tamil or Indian language. Yet rule based morphological analysis is still crucial to form some of the rules that have not been added to the grammar rules. This paper utilizes rule based morphological analysis and phrase-by-phrase translation by combining both rule base and statistical base under hybrid.

Main verbs in Tamil can only found at the end of sentence. Object and verb are arranged in a reversible order in Indian languages. European Languages uses Subject Verb Object (SVO) order to construct a sentence structure, whereas, in Tamil verbs appear at the end of the sentence with an order of Subject Object Verb (SOV). In some cases, word orders are changed and this make Tamil as word order free language (Vinotha, Sriraman & Yugehwar, 2013).



Tamil affixes are mostly suffixes that comprise of inflection and derivation. Inflections in Tamil words modify the root word. Tamil has a morphological construct as Root + Sandhi + Inflections. Root words are the basic words that appear in the dictionary either verb or action. Whereas, Sandhi denotes tense and inflection denotes gender and number. Normally in Tamil we consider this as two different words. But most of the times they tend to appear together due to some inflection processes. These inflection words are rarely appear in the dictionary and thus,

morphological analysis is done by a lookup with partial stem after substring matching and extraction. This process requires reconstruction of the rules in order to make the clauses more meaningful. Tamil grammars are naturally agglutinative. Suffixes are often used with the verb. (Mallamma & Hanumanthappa, 2013)

3.0 Literature Review

3.1 Google Translator

Google translator is a free tool that can translate words, sentences, documents or even one whole website within blink of seconds. It is an automatic translate machine from one language to another. Google translate machine can translate nearly 51 over languages that most people speak around the world and offers translations for 2550 language pairs. (Och, 2009). All translation works come from the computer from time to time whenever a translation work is needed. Computer generates its own translation based on the pattern found from lager corpora or large amount of text. The computers were feed in with billions of vocabularies and a set of rules.

But most of the time we may have to give some exceptions for the language rules due to some complications. This is because language is not fixed and evolves from day by day, unlike Mathematics which has specific formula. When exceptions increase, then the quality of translation started to reduce. To overcome this problem, google translator takes a different approach. Instead of teaching computers, Google Inc makes the computers to discover rules for their own by analyzing millions of documents that already been translated by human translators where the sources are come from books, organizations and websites.

Computers scan these documents by looking for statistically significant patterns. For example patterns between the translation and the original text those are unlikely to occur by chance. When this translation happens billions of times these computers become smarter than human with instant translation. For some languages however, Google has fewer translated documents available and therefore fewer patterns that the google translator has detected. This is why sometime the translation quality vary by language and language pair. (Gwerin, 2010)

Google translate also provides some best features like detecting language, "romanization" and speech recognition. If a word has been translated to English then speech synthesis system which embedded in Google Translator helps us to pronounce or speak the word in English. Languages like Arab, Hindi and so on do not need any special keyboard to give input for translation. We can just simply type the word as how they sound in English and they will automatically transform into native script and translate as well. (Estelle, 2009)

Although Google Translator has been provided with so many features, yet the translation is still not considered as a complete successful process. Google translator has many weaknesses as well as the machine lacks of important documents of other languages in the corpus. Most of the Tamil rules are remained undiscovered in computer languages when translating Tamil words to English.

3.2 Tamil Proverbs

Tamil is an ancient language where the earliest inscriptions were discovered back at least 500BC. Tamil has variety of treasures or literary text called Thirukkural, Tolkappiyam, Aimperumkappiyam, Proverbs, idioms and so on. Tamil is best known as Dravidian language spoken around 52 million people in all over the world including India, Malaysia, Singapore, Sri Lanka, Vietnam, Myanmar, Canada, USA, UK and Australia. (Ager, 1998)

Tamil has a total of 247 characters, which includes 12 vowels, 18 consonants, 1 special character and vowels and consonant combination of 216 characters. For example த is ta with the inherent அ (aa) and த் (th) without a vowel (Tamil Language, 2014). Tamil is rich in morphological rules. Tamil nouns and also pronouns are classified into two super-classes rational (tinai) and irrational (akrinai) with five main classes called feminine (penbaal), masculine (aanbaal), rationalplural (palarpaal), irrational singular (otranpaal), irrational plural (palavinapaal).

Tamil has agglutinative grammar. Suffixes are often used to identify noun, class, number and case and other grammatical categories (Vinotha, 2013). Suffixes are used to perform the functions of cases and postpositions. Since Tamil uses SOV order, the SOV languages have a strong tendency to use postpositions rather than prepositions, to place auxiliary verbs after the action verb, to place genitive noun phrases before the possessed noun, to place a name before a title or honorific ("James Uncle" and "Johnson Doctor" rather than "Uncle James" and "Doctor Johnson"), and to have subordinators appear at the end of subordinate clauses. (Subject-Object-Verb, 2014)

Suffixes are further grouped into eight cases called nominative, accusative, dative, sociative, genitive, instrumental, locative and ablative. Adjective and adverbs are classified under as uriccol but, recently the grammar can be distinguished with the existence morphological and syntactical grounds. To cut short, this paper utilizes all the Tamil morphological-rules to improve machine translation from Tamil Proverb to English using Google Translator.

4.0 Methodology

This projects follow system development life cycle start from planning the arrangement of Tamil and English words, analysis of translation process, designing of new prototype and finally to implementation and maintenance for betterment of the system.

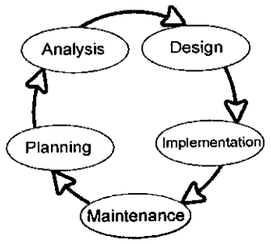


Figure 2 System development life cycle
(Retrieved from : https://en.wikipedia.org/wiki/Systems_development_life_cycle)

5.0 Proposed Solution

In this paper to resolve the problems stated, a proper module is suggested to improve the translation of Tamil to English. The framework for the translation is provided below. A total of seven modules are discussed in this paper including language identification, preprocessing, segmentation and tokenization, training and tuning, language model, grammar checker and statistical error correction. (Vinotha, Sriraman & Yugehwar, 2013) Each module is discussed as below but, the special focus is given to the segmentation and tokenization part.

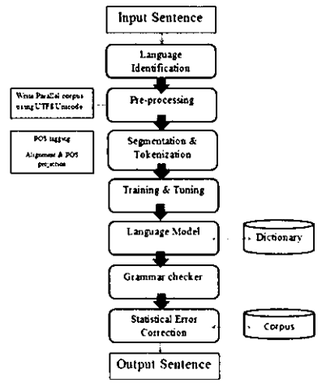


Figure 3 Machine Translation Modules

5,1 Language Identification

Some suggestions are provided in this module to improve the Tamil to English translations. As the initial step, once the input sentence was inserted the system should be able to identify the input language. This is knows as language identification module.

5.1.1 Preprocessing

After the identification, pre-processing of data has to be done. In this preprocessing step, we have to develop parallel corpus using UTF-8 Unicode character. In this Unicode character the number of consonants and vowels in Tamil use the "UTF-8" western windows encode and form vocabulary mainly from Sanskrit and English transliterations. (Mallamma & Hanumantappa 2013)

5.1.2 Segmentation & Tokenization

The next step is segmentation and tokenization. Segmentation is the process of splitting a sentence structure into distinct words. Tokenization is the process of dividing the input text into smaller components called tokens or chunks. These tokens can either be word, punctuation or a number. The machine translator sees these tokens as symbols separated by blank spaces. Segmentation and tokenization is available for all the European languages as this step separates a word into root and inflection or derivation form and simplifies the dictionary look-up. But, non-existence of tokenization script for Tamil language makes it difficult to perform translation. Tokenization, segmentation and tagging process make the input text ready for training.

5.1.3 Part of Speech Tagger (POS)

Once tokenization has been done, POS tagging is applied to word lists. POS tagging is the process of identifying the part of speech and classify whether the identified word falls under verb, noun or preposition and so on and assign an appropriate tag for the part of speech like English, Tamil also has eight regular parts of speech. This paper highlights some proposed POS rules for efficient Tamil to English translation. (Selvam & Natarajan, 2009)

Tag	Description	Examples and meaning
ADJ	Adjective	ஆனந்த (happy)
ADJAP	Adjective Past Participle	முடித்த (done)
ADV	Adverb	விரைவாக (as fast)

CON	Conjunction	அல்லது (or)
CVCN	Verbal Conditional Negative	முடிக்காவிட்டால் (if not done)
DET	Determiner	அந்த (that)
INT	Interjection	ஐயோ (alas)
NNSN	Noun singular Neutral	தேர்வு (exam)
NOSM	Pronoun Singular Feminine	அவள் (she)
NAPC	Adjective Noun Singular Common	ஆத்திரக்காரன் (angry person)
ORD	Ordinal	முதலாவது (third person)
PRP	Preposition	வெளியே (out)
QNT	Quantifier	எல்லாம் (all)
V	Verb	வாசி (read)
VC	Verb Causative	கவனி (concentrate)
VFPA	First Person Phrase Past Tense Verb	பார்த்தோம் (we saw)
VI	Intransitive Verb	திரும்பு (turn)
VIF	Infinitive Verb	வர (to come)
VSPAN	Second Person Plural Past Tense Negative Verb	வரவில்லை (did not come)
VT	Transitive Verb	திருப்பு (will not bite)
VTSN	Third Person Singular Neutral Future Tense Negative Verb	கடிக்காது (will not bite)

Table 1 : Proposed POS Tag for Tamil Language

5.1.4 Morphological Analysis and Generation

Once done with POS tagging now it's time for morphological analysis. Morphology includes recognition, analysis and generation of words. Commonly used morphological processes in Indian languages are inflection, derivation, affixes and combining forms. However, this morphology is not often used in all the words in Tamil. Some inflection processes are explained in details in the following examples.

After tokenization process a word will be checked with the direct word list to detect the existence. If the word exists in the word list, then a tag will be assigned. For example adverb (விரைவாக - fast) and adjectives (அன்பான - lovely) have a closed set of suffix (ஆக, ஆன). Direct prepositions (மேலே- on), interjection (ஐயோ - alas), conjunction (உடன்– with, அல்லது-or) and pronoun (அவன்-he, அவள்-she, அது-it) also have a closed data. Morphological analysis is not necessary for the closed set of words and the POS tags are applied directly with the look-up table. Normally for the adverbs and adjective, morphological analysis it is done with the separation and suffixes.

Inflection is the most productive morphological process in Tamil language. Inflection in Tamil includes tense, gender, number, mood, person, case and aspect (Parakh & Rajesha, 2011). The pronoun (அவன்-he, அது-it) in Tamil language is embedded with some of the nouns called adjective noun (ஆத்திரக்காரனுக்கு – ஆத்திர + காரன் + க்கு), verbal noun (எய்தவன் – எய்த+அவன்), and participle noun (வளையாதது – வளை + ஆ + அது). This string is identified and separated by substring match. POS tagging is done with look-up table after separating pronoun, adjective, verbs, or participle and case ending.

From the phrase: அகத்தின் அழகு முகத்தில் தெரியும்

Case suffix (Genetive) with noun: அகத்தின் (inner) – அகம் + இன்

Another type of POS tagging is by doing substring match for the tense verbal affixes. Present tense and future tense verbal affixes (இருக்க- to have, ஆது-not) must be separated from root words (சென்றிருக்க – சென்று + இருக்க).

From the phrase: குரைக்கும் நாய் கடிக்காது

Verbal suffix (Imperative): கடிக்காது (will not bite) - கடிக்க + ஆது

Another inflection word like pronoun with preposition can be checked by substring match with case endings. This is the result of pronoun (ஆத்திரக்காரன், கடவுள்) and the ending is called inflectional prepositions (ஐ, க்கு) form.

From the phrase: ஆத்திரக்காரனுக்குப் புத்தி மட்டு

Noun with dative: ஆத்திரக்காரனுக்கு (to angry person) – ஆத்திரக்காரன் + க்கு

From the phrase: கடவுளை நம்பினோர் கைவிடப்படார்

Noun with accusative: கடவுளை (God-indicate to Hisself) - கடவுள் + ஐ

In Tamil, adjacent words are joined and pronounced as one word. This happens for all the Indian languages. Two different morphemes combine to make as a long word. Verb with conjunction is mostly found to have this kind of joined morphemes in Tamil phrases. Verb (மின்னு, சிறுத்தல்) can be separated from the conjunction (எல்லாம், ஆனாலும்) by doing substring match with case ending like preposition pronoun.

From phrase: மின்னுவதெல்லாம் பொன்னல்ல

Conjunction with verb: மின்னுவதெல்லாம் (all glitters) – மின்னுவது + எல்லாம்

From phrase: கடுகு சிறுத்தாலும் காரம் போகுமா?

Conjunction with verb: சிறுத்தாலும் (even small) –சிறுத்தல் + ஆனாலும்

Verbal inflection is another type of POS tagging that should be implemented for efficient machine translation. POS tagging is done with suffix pattern which includes numbers, person, gender, tense, voice, negative and question. But this verbal inflection may be ambiguous in nouns and verbs. Numbers are checked with look-up and partial pattern matching.

From the phrase: ஒரு காசு பேணின் இரு காசு தேறும்.

Verbal suffix (Past tense): பேணின் (if preserved) –பேண் + இன்

Verbal suffix (Infinitive): தேறும் (will accumulate) -தேறு + உம்

After all the morphological process, new nouns, verbs, adjectives, adverbs and prepositions that left untagged will be added to dynamic memory learning technique in various look-up tables. Words that are not matched with the lookup table will be added to dynamic memory learning technique. Finally disambiguation module is needed for identifying noun and verb disambiguation. Disambiguation occurs when multiple tags are assigned to a word. Numbers are checked in look-up or with partial matching.

Phrases	**Action**
Adverb & Adjective	Separation > Suffixes
Direct Preposition, Interjection, Conjunction & Pronoun	Lookup table
Pronoun with adjective noun, verbal noun and participle noun	Separated by substring > Case ending (Inflectional prepositions)
Verbal Inflection	Suffix pattern matching (number, gender, tense)
Dynamic Memory Learning (Noun, Verb, Preposition, Adjective, Adverb)	Add new words to various lookup table
Nouns & Verbs (Numbers)	Lookup & Partial Pattern Matching

Table 2: Summary of morphological processes

Some of the proposed tags are given in the table below.

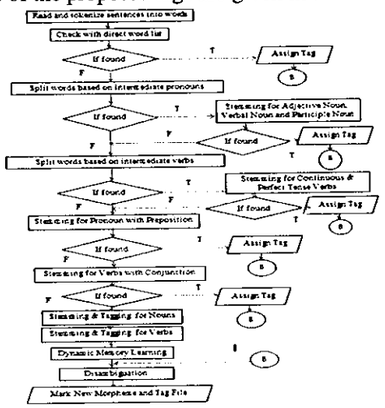


Figure 3 Flowchart for rule based morphological analysis and POS Tagging

5.2 Alignment and POS Projection

5.2.1 Alignment

After developing POS tagging, the next step is to do some alignment and POS projections. It is crucial to do alignment for Indian languages because Tamil language lacks in text corpora and tools. Hence, alignment and projection of parallel corpora with other resource rich language is very important. Projection and alignment with English language is highly recommended since English is rich in resources and is widely has been used. Bilingual parallel corpora that maps English to Tamil words are readily available in newspaper, magazines, television news and so. Projection techniques also need alignment with bilingual parallel corpora. To improve translation Tamil words are aligned to English words in a sentence.

Alignment tool and dictionary are used to align English and Tamil sentences. Alignment dictionary comprises of Tamil words and its equivalent word in English. The dictionary includes all the nouns, verbs, root words, adjective, adverb, conjunction, interjections prepositions and so on. Examples are given below.

In the example below "and" word is not exists in the sentence. Hence, it is not necessary to include the word there. Instead an expected outcome is provided to enhance the translation process.

Google translator results:

அகத்தின் அழகு முகத்தில் தெரியும் (SOV)

Know in the face and inner beauty (OSV)

Expected results:

அகத்தின் அழகு முகத்தில் தெரியும் (SOV)

Inner beauty knows in the face (SVO)

5.2.2 POS Projection

Now we have reached POS projection for a given sentence. Tamil phrases contain hidden meanings. Tamil Proverbs are like simplifying a long action within two or three word phrases to stress or indicate some kind of warnings. Within that two or three words, the phrase includes some realization concepts and these words trigger our thinking capabilities. Hence, it is very rare to find preposition, conjunction and determiner within the phrase. Instead they are more direct to the point and meaning. Hence, POS is needed to identify noun or verb, adjectives or adverbs and so on. Partial morphology is also needed for this phase.

Words in English:	Inner	beauty	knows	in the face
Aligned Tamil words:	அகத்தின்	அழகு	தெரியும்	முகத்தில்
POS projection:	ADJ	NN	VCF	NN
Categorical Information:	ADJECTIVE	NOUN	VERB	NOUN

Once done with the categorical information, further morphological processes can be done easily due to well defined suffix, stem, verb and noun. POS projection from aligned corpora is crucial. This increase the accuracy of rule based morphology and POS tagging.

5.3 Training and Tuning

Now we are moving to training and tuning module. Training and tuning is conducted based on phrase based morphological analyzer. In this phase corpus is cleaned by doing true-casing. Case sensitivity will be removed at this phase. Training is carried out using Hidden Markov Model (HMM) rule. Naive Baye's classifier can be used because the tokens are considered as symbols separated by blank spaces. (Vinotha, Sriraman & Yugehwar, 2013) Baye"s Theorem is represented as follows:

$$P(C|F1 \ldots . Fn) = \frac{p(C)p(F1 \ldots . Fn|C)}{p(\mathrm{F1} \ldots Fn)}$$

Once done with that, now we can focus on language model, grammar checking and statistical error correction. Language model contains the root words or best known as lemma which is defined using sentence based reordering. Language model has to incorporate with dictionary to get the right word and suggestion from dictionary for the input word. In other words it can be said that language model should give accurate output in both Tamil and English. Grammar and gender rules are used to check grammar in both Tamil and English. Spelling check and grammar check makes the translation more error free. Finally, statistical error correction is done based on the document collection found in the corpus to make the output sentence or phrase more accurate. (Vinotha, Sriraman & Yugehwar, 2013).

5.0 Conclusion

A total of seven modules have been discussed in this paper. Starting from language detection module, followed by preprocessing, segmentation and tokenization, to training and tuning module, language module, and grammar checker and finally to statistical error correction using some formulas. An insight focus was given to the segmentation and tokenization part which includes POS tagging and

rule based morphological analysis. To improve translation, hybrid machine translation is used by combining both rule base and statistical machine translation. The performance of these rules can be measured based on the alignment of words with the root word in dictionary. At the end of the process some disambiguation seems to exist in the case of identifying nouns and verbs in some context. This can be regard as future work.

Although many morphological rules have been researched and proposed yet not all the rules are applicable for all documents in the corpus. Some documents in the corpus may contain ancient characters and words which are hardly can found or use in latest era. It is difficult and impossible to translate them to English unless there are experts who can recognize old Tamil characters and words. It is still a challenging process to translate Thirukkural (Tamil Quran/Bible) and mantras because of hidden meaning and unusual words.

This module is only applicable for Tamil characters that are widely use nowadays all around the world. Also, latest poem, proverbs, phrases and clauses may follow these rules stated above. It is highly recommended to manually translate and store Tamil treasures like Thirukural and Tholkaapiyam in both English and other languages. These documents will be the assets for future generation as they are fundamental for human living. This action of storing all the information digitally enables these documents from destruction and also improves Google Translator's quality and efficiency to next level or next generation.

6.0 References

Ager, S. Tamil. Omniglot. (1998). Retrieved from http://www.omniglot.com/about.htm#what

Anthony, P.J., (2013 March) Machine Translation Approaches and Survey for Indian Languages. *The Association for Computational Lingiustic and Chinese Language Processing.* (Vol.18 pp. 47-78) Retrieve from http://www.aclclp.org.tw/clclp/v18n1/v18n1a3.pdf

Estelle, J. Text Translation on Google Translate, [Google]. (2009, Nov 16). Retrieved from https://www.youtube.com/watch?v=FijOWfO3Frk

Gwerin, Inside Google Translate, [Google Videos]. (2010, Aug 27). Retrieved from http://translate.google.com.my/about/intl/en_ALL/

Jurafsky, D., Martin, J. (2008) Words and Transducers: Speech and Language Processing, 2nd edition. Retrieved from http://www.cs.colorado.edu/~martin/slp.html

Jurafsky, D., Martin, J. (2008) Machine Translation: Speech and Language Processing, 2nd edition. Retrieved from http://www.cs.colorado.edu/~martin/slp.html

Mallamma, V.R., Hanumanthappa, M. (2013, Jan 3) Machine Translation Tools for NLP. *Indian Science Congress Association (ISCA).*

Och, F. (2009, Aug 31). 51 Languages in Google Translate. Retrieved from Google Research Blog website: http://googleresearch.blogspot.com/2009/08/51-languages-in-google-translate.html

Parakh, M., Rajesha, N. (5 May 11). Developing Morphological Analysers For Four Indian Languages Using A Rule Based Affix Stripping Approach. *Linguistic Data Consortium for Indian Languages, CIII, Mysore.* (pp. 13-16). Retrieved from https://www.academia.edu/909483/DEVELOPING_MORPHOLOGICAL_ANALYZERS_FOR_FOUR_INDIAN_LANGUAGES_USING_A_RULE_BASED_AFFIX_STRIPPING_APPROACH

Raman, S., Alwar, N. (1990 May). An AI-Based Approach to Machine Translation in India Languages. *Communication of the ACM.* (Vol.33, pp. 521-527). Retrieved from http://dl.acm.org/citation.cfm?id=78612

Selvam, M., Natarajan, A.M. (2009) Improvement of Rule Based Morphological Analysis and POS Tagging in Tamil Language via Projection and Induction Techniques. *International Journal of Computers.* (Vol.3 pp.357-367) Retrieved from http://www.naun.orgwww.naun.org/multimedia/NAUN/computers/19-095.pdf

Subject-Object-Verb. (2014, Oct 4). In Wikipedia, The Free Encyclopedia. Retrieved 08.00, November 22, 2014 from http://en.wikipedia.org/w/index.php?title=Subject%E2%80%93object%E2%80%93verb&action=history

Tamil Language. (2014, Nov 24) In Wikipedia, The Free Encyclopedia. Retrieved 19.50, November 26, 2014 from http://en.wikipedia.org/wiki/Tamil_language