

Problems related to Eng-Tam Translation

M.B.A.Salai Aaviyamma
Assistant Professor/ CSE
SRM university
Ramapuram
Chennai-600089
India

Dr.K.Kathiravan
Professor and VicePrincipal,
Easwari Engg.College
Ramapuram
Chennai-600089
India

Introduction:

Countries like India, where many official languages are used, and if there is a commonly used language, like English, then translating from that common language to regional language solves many purposes. Tamil, a South Indian language, not only used in Tamil nadu, a State of India, but is also used as one of the official languages in many countries like Singapore, Malaysia and Sri Lanka.

Stages:

There are 6 modules in this project like Tokenization, Parsing, Mapping, Word formation, Sentence structure changing, and Display. The English sentences(to be translated) are separated into words (Tokenization), each word is recognized as affiliations of root words (parsing) i.e. root words and morphemes are identified, root words are mapped into equivalent Tamil words using dictionary (Mapping,). Then according to the parser output, Thamizh words are formed (Word formation), and then structure of sentence is formed according to Thamizh sentence making rules (Sentence structure changing) and the output statement is displayed in Thamizh. Mainly concentrated module of the project is the word formation .

Problems found in the following areas:

1. when tense markers are added
2. when case markers are added
3. when proper nouns are translated

Problems in forming Verb Phrases:

Traditionally, a Tamil word is divided into a maximum of six parts, known as pakuthy (prime-stem), sandhi (junction), .viha:ram (variation), idainilai (middle part), sa:riyai (enunciater) and vikuthy (terminator) in that order.

(prime-stem)	(junction)	(Tense marker) middle part	(enunciater)	(terminator)
Pagudhi	Sandhi	Idainilai	Sariyai	Vigudhi

The sixth part is Vigaaram. This is the trouble making part of translation.

For example, if the root word is a finite verb, then changes of penultinating characters are of many type.

First case

Changes can be introduced, when root words are joined with tense markers or additional suffixes. For many root words, verb formed are different.

For example, “HE ACTED”,

“nadi” + th + th + aan --- nadiththaan
 + + + ———→

“ He walked”

“nada” + th(ndth) + th + aan --- nadandthaan
 + () + + ———→

To overcome this problem, Dr. Crowl² and M.Raagava iyangar³, in their books (Thamizhp peragaraathi and Vinaith thiribu Vilakkam), they divided the entire Thamizh verb family into 12 groups as

, , , ±, , , , , , , , , , ,

according to the last characters of root words, the tense markers they accept etc. Even though they are grouped, some root words, having same last character but grouped under different tables, makes the rule based translation, a problem.

For example,

“ ” - is grouped under “kol” table, group 3

“ ” - is grouped under “kal” table, group 10

Even though both are having the same last character “l” p, but when added with past tense markers, they are turned as

“cendraan” - and

“kattrraan” -

Second case

The same root word is kept under 2 or more table and it takes different tense markers under each circumstances.

For example,

“migu” -

under “ari” table, it is transformed as “ migundthaan”

under “nagu” table, it is transformed as “mikkaan”.

So, to inform the system, under which table, that root word is classified is so difficult.

Hence , it is a problem to form the verb phrase for these root words.

Third case

The same root word, even though kept under a same table, because of its different meaning, it is transformed differently, which creates problem to decide how to transform it. For example,

“madi” - under “ari” table,

when having meaning as ‘die’, transformed as “madindhaan”-

when having meaning as ‘fold’ , transformed as “madiththaan”-

For this problem, a solution is obtained as, for verbs denoting self deeds, it will be transformed with “thth”. And for denoting other’s deeds, it will be “ndhth”

But there is a problem on this solution, for some roots .

For example,

“vadi” -

(kanneer) vadiththaan -

(azudhu) vadindhään -

Both are denoting self deeds but are transformed in both the ways.

But, as a different case,

root word “pidi” – is transformed as

(“pidiththaan”) - , in both the cases.

Problems with case markers

Eight cases are there. They are Nominative, Accusative, Dative, Benefactive, Instrumental, Sociative, locative and Ablative.

Some prepositions are marked as case markers. But, these case markers give different meanings in different places. So, translation become difficult.

For example,

(i) “ HE ATE WITH THE SPOON”

Here , ‘with’ comes as instrumental case marker.

(2) “HE ATE WITH HIS FRIEND”

Here, the same preposition ‘with’ comes as Sociative case marker.

More than this, in the parser output⁵, for both the sentences, the subject and object are no where mentioned as either instrumental or sociative case markers. But mentioned as nominative in both the cases . And the word ‘with’ is simply mentioned as preposition .

Problems with Gender suffixes and nouns:

In English, generally, the names of male persons , are spelled excluding the last letter.

For example,

‘**Rama** went to Srilanka’, instead of ‘Raman went to Srilanka’.

Now, there is a problem, whether to translate this as

?

(or)

?

he ate with the spoon.

"<he>"

"he" <NonMod> PRON PERS MASC **NOM** SG3 SUBJ @SUBJ

"<ate>"

"eat" <SVO> <SV> V PAST VFIN @+FMAINV

"<with>"

"with" **PREP** @ADVL

"<the>"

"the" <Def> DET CENTRAL ART SG/PL @DN>

"<spoon>"

"spoon" N **NOM** SG @<P

"<\$.>"

he ate with his friend.

"<he>"

"he" <NonMod> PRON PERS MASC NOM SG3 SUBJ @SUBJ

"<ate>"

"eat" <SVO> <SV> V PAST VFIN @+FMAINV

"<with>"

"with" **PREP** @ADVL

"<his>"

"he" PRON PERS MASC GEN SG3 @GN>

"<friend>"

"friend" N **NOM** SG @<P

"<\$.>"

Therefore, by rule based method alone, translation from English to Tamil cannot be done. So, we have to train the system accordingly and then only we can translate the sentences.

Conclusions:

In this paper, the problems created by case markers and problems created when forming the Thamizh words by adding tense, gender, plural suffixes with root words are dealt in detail. And it is more important to solve these problems, since translation from 'English to Thamizh' is a very important and timely needed task for Tamilnadu state government, and countries like Singapore, Malaysia and Sri Lanka, where Tamil is accepted as one of the official language, in order to improve the communication and education.

Bibliography:

1. A grammar of modern Tamil - Thomas Lehmann (Pondicherry University)
2. - Dr. Crowl
3. [μ - M. Raagava Iyengar, 1958
4. Hidden problems and challenges in Tamil computing
 - i. -S.Srinivasan, A James (Tamil Virtual University)
 - ii. -S.AnanthaKrishnan, K.R.S.Narayanan (IARC Kalpakkam)