REVIEW ARTICLE

MACHINE TRANSLATION WITH REFERENCE TO 'CASE' SYSTEM IN ENGLISH AND TAMIL

Dr. P.ARUNA DEVI

Associate Professor of English
Institute of Distance Education, University of Madras, Chennai

Dr. P.ARUNA DEVI

ABSTRACT

Technology has rendered positive avenues in research in the field of language .Machine translation is one such avenue that deploys technology for many purposes like translation, pedagogy, cross linguistic communication and so on. Using the application of natural language processing this machine translation is made possible .This paper makes an attempt to focus the role of linguistic knowledge in both the source language and the target language to feed the machine to solve the issues that arise out of the difference in the grammar and structure of Tamil and English. Both the languages hail from different families. English is from Indo European and Tamil is from Dravidian family. English is head first language having SVO sentence structure and syntactically a rich language. Tamil is head final, having SOV structure and morphologically rich. Case system in both the language is taken for analysis and dealt with in detail. In Tamil the noun inflects for case and is manifested overtly. But in English the overt inflection is only for a case but mostly the case is shown either through prepositions or word order. These differences have to be sorted for the translation and the language tools necessary for the process to be developed in the machine system is also discussed in the paper. But to give a detailed analysis of the development of such tools a deep knowledge of the computer language and programming are mandatory. Hence the paper is limited with an analysis at the linguistic level in the process. The grammatical nuances of both language is the base for machine translation and a linguistic survey is made with special reference to 'case' in English and Tamil.

Key words: Machine translation ,syntax, morphological parser, case marker , inflection, Linguistic knowledge

With the proliferation of language technology machine translation has come on handy for various purposes along with the academic one. It is the utilization of computers to translate one language to the other for various purposes. Machine translation plays a key role in bilingual and cross- lingual communication at various domains(Allen James,1995). While learning or teaching a foreign language the machine translation aids

to collect information about the language . Research in Cross –Language Information Retrieval (CLIR) based on machine translations is a boon to pedagogy and translation. It is to be noted that translation is not word level translation but is at a higher plane involving syntax, semantics and pragmatics. Hence in the process the machine should be able to solve the issues and the database should cover the sufficient data in both the languages. Knowledge based approach and rule based approach can be considered as a best methods for machine translation and the grammar of both the language, the source language and the target language is mandatory. In the rule based system the rules applied will provide a smooth handling of translation in the transfer module.

 Take a simple sentence in English as below

He spoke to me

If translated into Tamil it will be

Avan pecinaan ennidam

Avan   ennidam pecinaan

In the above translation the following details can be noted. There is equal lexical representation only in one category, the pronoun that carries the gender and number representation. The verb "spoke' in English carries the time of action, the tense whereas the verb 'pecinaan' in Tamil gives an additional information of gender and number. Gender is not an inflectional category in English. The   infinitive is manifested through an inflection in noun in tamil -'ennidem' but in English language it is marked lexically by two words –' to'an infinitive marker and 'me'. Again the word order also changes. English has SVO order and Tamil has the SOV order. While giving the input for the machine to translate i n the target language the grammatical knowledge aids to process the translation in an accurate manner. This paper tries to focus on an important grammatical category ,the case system at the morphological and syntactic level for the rule and knowledge based translation from English to Tamil and vice versa. . But the paper attempts to explain only the language part involved in the process and not the technical aspects involved like the computer language or the algorithms involved in the process. The essential features of this translation process can be summarized as follows.

Input

Tokenization

Morphological parsing

POS tagging

Transfer from SL to TL

The input

The iput is a sentence with any case manifested either by a suffix or a postpostion in tami/English. Look at the following sentence

Naan kamalaip paarteen ('ai',accusative case)

Tokenization

This identification involves individual words. The task of the tokenizer is to identify the words and the punctuation marks are not considered and would get deleted. Individual words include a lexeme with the affixes and suffixes as given in the input.

Morphological parser:

Here a word has to be segmented as   individual morphemes. In Tamil the root word or stem is segregated and other grammatical categories attached is identified as suffixes from the verb and noun. Since Tamil is an agglutinative language after the root word parsing the whole of the suffixes also should be parsed. In a word like *'pattittaan'* the machine segments it as 'pati (root stem),'tt' and 'aan'. The morphological parser should therefore have a root word database and a suffix database. The root word database should have a store of the root words of the source language along with that of the target language. Similarly the suffix database of each language should have   a repertoire of the inflectional suffixes of both the languages and the common categories are noun suffixes like gender (Tamil), number(Tamil, English), case (Tamil English), postpositions   (Tamil),   verbal   suffixes   like   tense(Tamil ,English),aspects (Tamil,   English),PNG (Tamil),modal(Tamil).

Dr. P.ARUNA DEVI

POS tagging

After the word is parsed the tagging has to be done to identify whether the segment is an inflected form and if it is what is the category .For example the case inflection attached with the noun gets identified in Tamil language here. The segment 'ootu' in Tamil can either be a case marker as in '*Kamalottu vanteen' I came with Kamal'* or it can be a lexeme as in *inda vottu enna vilai? what is the price of this roof tile*? To identify the word as either a case marker or a noun we need a tagger.

Transfer Module

The transfer module gives the output. This module utilizes the programme and the rules and translates the sentence given as input .The initial step is to get the lexemes at the root level ,then equate the suffixes either with suffixes or words and finally takes care of the syntax. It is to be noted that mostly what is manifested at the morphological level in Tamil is expressed syntactically with words and postpositions in English as for as case is concerned. The sentence structure is reordered as per the syntax of the target language.

Case System:

It is a system of indicating the relationship the noun/pronoun bear with the heads. It can be realized either by a case marker through inflectional suffixes or through case forms as individual lexemes like prepositions in English. Case in a language signifies both grammatical relations as well as semantic roles which ultimately gives meaning. (Marudur Arangarasan ,p.71) Fillmore also in his paper "The case for case"(1968) talked about the case indicating the syntactic semantic relation Consider the sentence *she made the bowl from clay*(Barry J.Blake,2001) where the case speaks of the object and the material out of which it is made. Agesthialingom (2002) has categorized this as material case and has distinguished it from the instrumental case with a valid test saying that in a material case the noun phrase that inflects for case cannot become the subject whereas that can happen in the instrumental case. However this paper does not intend to go deep into such analysis but stops with quoting the most commonly addressed cases in the two languages for the purpose of machine translation. The case marking are realized in different layers in different languages and Marudur Arangarasan has given six layers namely

1.overt case markers

2.postpositions

3.prepositions

4.word order

5.without / zero markers

6.variations in tone

In traditional Tamil the case relationship is manifested through suffixes and postpositions. There are eight cases in Tamil and the manifestations are tabulated as follows

case	Name	Suffix	Postposition
1st case	Nominative	Unmarked	-
2nd Case	Accusative	ai	-
3rd case	Sociative	ootu,aan,aal	Kontu,vaittu,utan
4th case	Dative	ku	
5th case	Ablative	In,il	ninru
6th case	Genitive	Atu,aatu,a	utaiya
7th case	Locative	Kan,il	-
8th case	Vocative	Change in the noun endings	

But in modern Tamil twelve cases are found as given below

1. Nominative case
2. Accusative case
3. Instrument case

Dr. P.ARUNA DEVI

4. Material case
5. Causal Case
6. Associative case
7. Dative case
8. Purposive case
9. Ablative case
10. Genitive case
11. Locative case
12. Vocative case

1.Nominative Case

Here the case is unmarked. The subject is inflected. Since the finite verb in Tamil is in agreement with the subject of the sentence for PNG the identification of the subject is easy even if the sentence has two names.

Kamal peecinaan

Kamal spoke

where the subject is Kamal Look at the next example.

Kamal ramanodu vantaan

Kamal came with Raman.

In the above example the subject is 'Kamal' as it is in agreement with the finite verb that manifests the PNG suffix 'aan' in 'vantaan'. Hence the noun inflected but unmarked is 'Kamal'.

2.Accusative Case

This second case, irrandaan verrumai as called in Tamil is manifested with the suffix 'ai'.

Kaml raamanaip paarttaan

Kamal saw Raman

3.Instrumental Case

The third case is marked with a suffix 'aal'.

Kamal kattiy**aal** avanaik konraan

Kamal killed him with the knife

Kamal tan ceikayaal avanai mahizvittaan

Kamal made him happy with his action.

It is to be noted that Agesthialingom(2002) has indicated that the case can also be marked with the suffix 'kondu' replacing 'aal'. The same sentence can be rendered as

Kamal kattiyaikkondu avaniak konraan.

4.Material Case

There is a controversy to be addressed as for as this case is concerned. Some grammarians like Agesthialingom (2002) and Arangarasan (2000)consider this as a separate case .Some others like Kothandaraman,P.(1997) and Paramasivam,K (2011) place this under the instrumental case. This talks about the material out of which the object represented by the noun phrase is made.

Inda chilai thangattaal ceyyappattatu

This statue is made up of gold.

As in the previous case there are more suffixes like 'kontu', 'il', that is used to show the case and the examples are

Inda chilai thangathil ceytattu

Inda chilai thangattai kondu ceytatu

5 Causative case

This case is included in the instrumental case itself by grammarians like Paramasivam, K(20110) but grammarians like Kodantaraman,P(1997),Agesthialingom(2002) consider this as a separate case. This is shown by the suffix 'aal'.

Kamal nanraaha pattitataal parisu perraan

Dr. P.ARUNA DEVI

Since Kamal read well he got the prize

6.Associative Case

This case is shown by the suffixes 'kuta',ootu,utan. In modern Tamil this is a separate case but in old Tamil it is placed under the instrumental case.

Kamal ramanootu vantaan

Kamal came with Raman

Kamal Ramankuta vantaan

Kamal came with Raman

Kamal Ramanutan vantaan

Kamal came with Raman

The suffix 'ootu' also gives another meaning as in

Naan inrodu unnai velai vittu niruttukireen

I am stopping you from job from today(onwards)

7.Dative case

There are many meanings conveyed through this case with suffixes 'ku',irku ,ukku, arku

Time

Pattu manikku vaa

Come at teno' clock

Inta kaykarikal oru vaarathirku varum

These vegetables will suffice for a week

Space

Avan vittukku munnal nirkkiraan

He is in front of the house.

En munnaal nirkatee

Don't stand in front of me

Reason

Dengu kaaycchalukku neraya per paliyaanarkal

Many people became victims/died of Dengu fever

Experience

Enakku talai churrukiratu

My head reels

Relation

Avan enakku tambi

He is my younger brother

Goa:

Avan pallikku cenraan

He went to school

8.Purposive case

The case suffixes here are three namely 'ku',kkaka','inporuttu' and this case is specified in modern tamil .by Agesthialingom,S (2002).

Kamal tanatu mahizcikkaka velai ceytaan

Kamal did the work for his happiness.

Avan pannat tin poruttu itaic ceytaan

He did this for money.

Avan pannathirku velai ceytaan

He worked   for money.

9.Ablative case

The ablative case talks about the source and origin of the action mentioned and the suffixes are 'iliruntu','itamiruntu,' 'mutal'. The suffix 'iruntu' occurs with inanimate noun phrases and 'itamiruntu' with human and other animate noun phrases.

Kamal dubailiruntu nerru vantatan

Kamal came from Dubai yesterday.

En tantaidamiruntu naan itaik karrukonteen

I learnt this from my father.

Naalai mutal enakku vidumuai

I have holiday from tomorrow.

10.Genetive case

This case indicates the relationship between the possessor and the possessed with the suffixes 'utaiya ',-atu', 'in' and zero allomorphs.

Itu ennutaiya pai

This is my bag

Itu avanatu puttagam

This is his book

Itu kamalin appa

This is kamal's father.

Kamal en makan

Kamal is my son

11.Locative case

As the term indicates, the case talks about the place of action with two suffixes namely 'il', itam'.

Avan poongaavil irukkiraan

He is in/inside the park

Avan kamalidam pecukiraan

He speaks with kamal

There are postpositions also like '-miitu', 'mel', 'torum' etc that show the case relationship.

Meccaiyin melulla caaviai edu

Take the key that is on the table

Viithi torum pookkatai ullatu

There is flower shop in every street

12. Vocative case

This case is to call someone and stands out of the sentence

Kamal inke vaa

Kamal come here

Or the noun takes a phoneme like 'a' as in

Ram 'a' inke vaa

Postpositions

Postpositions are word like elements that functions as suffix. There are five types of postpositions namely genitive, accusative, dative, locative and plain postpositions.

Case markers in English

In English case is realized by zero marker, prepositions and by word order. Look at the following examples.

Ram came (zero marker)

He has come from India (preposition)

I like you (word order)

There are five cases in English namely nominative, accusative, genitive, dative and vocative.

Nominative case

When the noun or pronoun acts as the subject of a sentence then it is named as nominative class

Dr. P.ARUNA DEVI

Ram is a good teacher

Accusative case

When the noun takes the role of object then it is the accusative case

Giri met Ram

Genitive case

When the noun inflects to show the possession or ownership it is said to be in genitive case.

This is Seetha's book

Dative case

It indicates a noun ,a receiver (indirect object), to whom something is given.

Kamal gave prize to Ram

Vocative case

When the noun is a person addressed to then it is the vocative case

Ram, come

Apart from the five cases discussed above there are cases like the instrumental, material, causal, associative, purposive and locative and the examples are as follows:

Instrumenta:

She hit the mosquito with the electric bat .

Material case

This chain is made with/of gold

Causal case

Due to ill health he could't attend the interview

Associative Case

He is with us for the past ten years.

Purposive case:

He glided into the office for searching important files.

Ablative case

I have to start my work from today.

The smell comes from the store room.

Locative case

The book is on the table.

English and Tamil case System: A comparison

S.No	Case	Enlish	Tamil
1.	Nominative	Unmarked	Unmarked
2.	Accusative	unmarked	'ai',
3.	Instrumental	By, with	-al','kondu,(ai)vaittu
4.	Material	By, with ,of	-'al','kondu',(ai) vaittu
5.	Causal	Because of, due to, thanks to	-'al'
6.	Associative	with	-out,utan,kuta,irku
7.	Dative	to	-'ku','-ukku' '-akku'
8.	Purposive	for	-'ku',kkaaka,inporuttu
9.	Ablative	from	-iliruntu',itamiruntu',-mutal'
10.	Genitive	-'s',of	-'utaiya',-'atu',-'in'
11.	Locative	In,at ,within	-'il',-'itam'
12.	Vocative	Unmarked	-'aa',-'ee'

Look at the postpositions   equivalences

S.No.	Tamil	English
1.	Itam, kontu	with ,to
2.	pal	towards, in the direction of
3.	mutal	from, starting with
4.	varai	till, until, upto
5.	utan, kuta	with
6.	vita,, vitavum, kaattilum	than,
7.	parri	about
8.	miitu,mela	on, above, top, over
9.	kile	under, down, below
10.	pakkattil,aruke, attuttu	near by, next ,close to,
11.	mun, munne, ethire	in front of, before, opposite to
12.	Peril,	on, on behalf of
13.	vaayilaaka,valiye,muulam	through
14.	Itaiye,natuve	between, in the middle of, in the midst of
15.	venti	for the sake of, on account of
16.	pin	after, behind
17.	ulle	inside, in, into
18.	veliye	out, of, outside
19.	curri	around
20.	patil	instead of
21.	takunta	suitable

Issues to be addressed in the Transfer Module

The case markers in Tamil are functional morphemes attached to nouns/pronouns and the parameter is framed in the native learners as noun+ case marker. This is reset in English as preposition + noun and noun takes no case marker except   the possessive case 's' which is 's attached to noun-proper and personal pronouns. In a prepositional phrase the preposition in English precedes the noun but in Tamil it is a case agglutinated to the noun and hence any morpheme comes only as suffix and exactly in the final position .

Naan chennaiy**iruntu** varukireen

I Chennai from coming

I am coming from Chennai.

Avan panth**ai** marattinmeli**runtu** erintaan

He ball tree top from threw

He threw the ball from the tree top.

In the above given example the verb's position changes in both the languages. The order of case also falls into particular order in English. The placement of the noun ' ball' cannot be moved either to the final position after the preposition as the preposition 'from' indicates location and is qualifying the noun ' tree'. If at all it has to be moved the entire phrase 'from the treetop' has to be moved to the initial place and then the sentence goes as

From the treetop he threw the ball

This is also the case with Tamil but instead of the lexical words it is the noun stem with the suffix that is positioned in a different place as in

Marrat**in**meel**iruntu** avan panth**ai** erintaan.

But it is to be noted that in certain sentences like

Enakku unn**ai** pittikkum

I like you

even if the word order changes in Tamil the meaning doesn't change as in

Dr. P.ARUNA DEVI

Unn**ai** enakku pittikkum

Enakku unnai pittikkum

Both the sentences give the same meaning as the noun 'un' is moved along with the case marker 'ai' but if there is a change in the position of noun in English the meaning changes. *You like me* is different from *I like you*. Here the reason is that the case is not overtly marked and there is a change in the pronoun itself 'I 'is changed to 'me. 'The bound case morpheme in Tamil makes the word order flexible while in English the fixed word order plays a major role in case exhibition and the semantics based on it. But it can also be noted that even in Tamil sometimes the case marker is not overtly marked as in 'marakkilai'- 'marrattin kilai' and the word order here is maintained though even if the order is changed, damage will not be done to the meaning as it remains the same.

Valli kattiy**aal** palam narrukkinaal

Where the case marker 'attai'is not added to the stem 'palam'.

The sentence can also go as

Valli kattiy**aal** palam/palattai narrukkinaal

where the noun can either inflect for case marker 'attai' or need not but gives the same meaning..

The sentence can also go as

Valli palam katti**aal** narrukkinaal.

While developing a transfer module the structure of the two languages need to be considered to set the programming for translation using the tools tokenizer, morphological parser, word-class tagger and transfer module .The programmer have to be developed giving utmost care to the grammatical nuances of the languages involved .English and Tamil belong to two different families, and English is head-first and Tamil ,Head-final. Such language research in all categories at the morphological level renders proper anchoring for the programme development for machine translation. At the transfer level reordering as per the structure of the target language is mandatory. Syntax based rules of the language has to applied to render the output. Hence the database at the transfer module should have the syntax based rules of the target language.

REFERENCES:

Hutchins John 2001, Machine translation and human translation: in competition or in complementation?, *International Journal of Translation*, 13, 1-2, p. 5-20

Allen James (1995), "Natural Language Processing", 1-2.Redwood Benjamin/ Cummings.

Marudur Arangarasan(2000) Cases in Tamil [*Tamilil Veerrumaikal]*. Ariyalur: Balamurugan Press

Fillmore,C.J.1968 "the case for case".In E.Bach&R.T.Harms (eds.)Universals   in linguistic theory,1-88,London:Holt,Rinehart& Winston

Blake.Barry J 2001 *Case* , Cambridge: Cambridge University press p.66

Agesthialingom S, 2002 *Structure of Tamil Language* (in Tamil),Chidambaram: Meiyappan Publishers

Kothandaraman, (Porko)P 1997 *A Grammar of Contemporary Literary Tamil*. Chennai: International Institute Of Tamil Studies

Paramasivam,K.2011.An Introduction to Modern Linguistics(in Tamil).Tiruchi: Ataiyalam

A BRIEF BIO OF AUTHOR

Ponnammal Aruna devi was born on 15.01.1958   in Trivandrum, Capital of Kerela state. Ponnammal Aruna Devi received B.A.(1978)and M.A.(1980) in English from Annamalai university ,Tamil Nadu, India. From the university of Madras Aruna devi received her doctorate for the thesis   " A contrastive Analysis of Tense system in English and tamil". Ponnamal Aruna devi presently an associate professor of English, Institute of Distance Education, in the university of Madras is specializing in the field of applied linguistics and computational linguistics.

She has published papers in this field in the national and international level. She has produced three Ph.D on Stylistics, ELT and Machine translation. one of her candidates have submitted the thesis.   She has done a major UGC project on "Computer aided language teaching of English with special reference to speaking and reading skills" . She is now into a project as a honorary consultant on "The Development of Tamil - English Machine Translation". She has authored a book titled *Second Language Acquisition: An Exploration.* She is member of Board of Studies for English in her own institute, in Tamil Nadu Open university and two other colleges in Tamil Nadu. She is serving as subject experts in the recruitment committee for selection of college faculty. She is the subject coordinator for Soft Skills program in the University.