

Rule Based Word Level Translation (Tamil - English) for Inflected and Non-inflected Tamil Words

Dr.R.Padmamala¹, R.Dhivya², V.Ranjitha³, P.Shobana⁴

1.Asst.Professor, Dept of MCA, Ethiraj College for Women, Chennai

2,3,4. Final year MCA students, Ethiraj College for Women, Chennai

Abstract

This paper deals with the technicalities and algorithms associated with word level translation of inflected and non-inflected Tamil words to English. A word can have more than one meaning or category associated with it. A word may give one meaning when it stands as a whole and a different meaning when it is parsed. For example, the meaning of the word *katalai* is 'groundnut'. But it can be parsed as *katal* + *ai*, where it gives the meaning 'Sea+ Accusative case'. This paper attempts at devising an algorithm, using rule based methods, which finds the meaning of a word and also finds out whether parsing is possible for the same. If so, the meaning for the word as such is given. Then the meaning corresponding to the morphologically parsed word is also given. With the help of the tool developed using this algorithm, a Tamil word and its all possible meanings can be listed out successfully.

Keywords: *Morphological parsing, word-level translation.*

1. Introduction

Language is a system of communication with certain grammar rules and pronunciation that are used to express thoughts, ideas, experiences etc. It becomes a challenge for a native speaker when introduced as a beginner to a foreign language. A translator allows bridging, the variation between these two languages using a universal language English. This paper attempts at developing an efficient tool kit for word level translation from Tamil to English.

2. Methodology:

The main functionality of this project is a word level translation of Tamil to English. If the given word is in non-inflected form, then its meaning is given straightly. If it is an inflected form, then morphological parsing is done to identify its stem and suffixes. With the obtained stem and its suffixes of a given word, its appropriate translation is done. Before we go into the parsing, we need to understand the structure of a word.

3. Structure of a Tamil Word

3.1 Noun

A word which is capable of taking a case suffix or postposition is a noun.
e.g. maram – 'tree'

Structure of a Noun

The following table describes the structure of a noun with various suffixes and particles.

Nst.	Numb	Filler	Case	Postpo.	Cl.1	Cl.2	Cl.3	Cl.4
<i>pen</i>	<i>kal</i>	<i>uk</i>	<i>ku</i>	<i>aka</i>	<i>mattum</i>	<i>tan</i>	<i>a</i>	<i>ata</i>
<i>aan</i>	<i>kal</i>	<i>in</i>	<i>al</i>	-----	<i>mattum</i>	<i>tan</i>	<i>a</i>	<i>ata</i>

3.2 Verb

A word which can take tense markers and which can be qualified by an adverb is a verb.
e.g. pati – ‘read’

Structure of a Verb

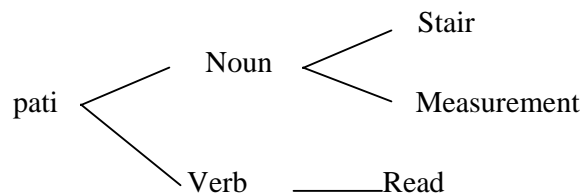
The finite form of a verb may be simple or complex. The longest finite form of a verb may have the following structure.

Vst.	Asp Aux	Voice Aux	Mod Aux	Tns	PNG	Cl.3	Cl.4
<i>ezuthik</i>	<i>katta</i>	<i>vaikkap</i>	<i>par</i>	<i>tt</i>	<i>an</i>	<i>a</i>	<i>ata</i>

where Vst. is verb stem, similarly Asp Aux is Aspectual Auxiliary, Mod Aux is Modal Auxiliary and PNG is Person-number-gender.

4. Translator

If a given word is in non-inflected form then its all possible meanings will be listed out.
e.g: ‘pati’ –



nei + *th* + *aaN* 'Wove (he)
Verb + TenseMarker + **PNG**



Thus, this tool gives all possible meanings of a given word, considering them as a root and also as inflected word wherever applicable. The database size is 2000 words which will be further expanded.

Conclusion

To achieve a Machine Translation system, many tools are required. This tool is one among them. This tool will play a prominent role in developing a full-fledged MT for Tamil – English.

Bibliography

- Knowledge sources for word level translation models, Philip Koehn, Kevin Knight, Information Sciences Institute, University of Southern California.
- Word-Level Confidence Estimation for Machine Translation, Nicola Ueffing, Herman Ney, RWTH Aachen University
- A Book of Contemporary Tamil, Pon.Kothandaraman