PERMANENT REFERENCE

- Language Assistant -

Towards Automatically Translating Text from English to Tamil

in a Controlled Environment



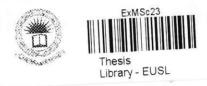
. K

NOW

By

S . Mohanarajah

(93 / M.Sc.Com.Sc / 32)



31405

Dissertation Submitted in Partial Fulfilment of the Requirement for the M.Sc in Computer Science at the

Department of Statistics and Computer Science University of Colombo, Sri Lanka

June – 1998

PROCESSE

PERMANENT REPERLNUE

Abstract

Automatic translation is the process of automatically converting a text in one natural language to another using computers. The resultant translated text should tally with the original text in meaning, effect and style. Since this automated translation process involves handling a huge amount of dictionary information, and complex grammar and word formation rules, the vocabulary and sentence structures of the input text may be restricted based on domain specific knowledge to get a better translation.

Sri Lanka is a multi-ethnic country. Communication between the ethnic groups is essential for co-operation in the interest of national development. Translation is the only viable solution for communication between communitier which use different languages. Sri Lanka requires a six-way translation tool between Tamil, English and Sinhala. The major objective of this study, as a first step for the overall motivation, is to build a model for a Machine Translation system that could translate text in English to Tamil in a controlled environment.

The proposed system is built on transformer based architecture. A code for information interchange in Tamil (SCIIT) is designed and used in this project. The system developed in this project works well for the English sentence structures found in elementary level textbooks. This system may be used readily for demonstration purposes, however, it needs further enhancements for it to be used in a real translation task.

Contents

Abstract

Acknowledgment Declaration Contents

CHAPTER I

PROJECT OVERVIEW	1
PROJECT OVERVIEW Introduction	2
Scope of Study	. 5
э.	
Deliverables Dissertation Overview	

CHAPTER II

MACHINE TRANSLATION BASICS	8
MACHINE TRANSLATION BASICS Introduction	8
o Dopio Eostures	
Decessing (NILP)	
Morphology: Syntax	16
Phrase Structure Analysis Parsing :	
Analysis (
Grammatical Relations Structure Analysis.	22
Generation (Synthesizing) Translation Engines	
Interlingua Engines. Dictionaries and MT systems	

CHAPTER III

3

LINGUISTIC ANALYSIS ON ENGLISH	
Introduction	
English Phonemes, Syllables and Stresses.	30
English Morphemes and Word formations.	32
English Word Classes.	
English Phrases, Clauses and Sentences	
Nouns and Noun phrases.	
Pronouns	
Adjective and Adjective Phrases	
Adverbs and Adverb Phrases:	
Prepositions and Prepositional Phrases:	
Verb and Verb Groups	
Negation	
Modal Verbs	
English Sentence Structure.	
Case I : Subject + Verb	
Case II: Subject + Verb + Complement	
Case III: Subject + Verb+ Object	
Case IV: Subject + Verb+ Object+ Ascriptive Complement	
Case V : Subject + Verb+ DirectObject + IndirectObject	
C another for a consistent interaction and another set of the set	

CHAPTER IV

INGUISTIC ANALYSIS ON TAMIL	
Introduction	. 45
Tamil Language Fundamentals	. 45
Tamil Alphabets, Phonemes, and Syllables	46
Roots of Tamil Words	49
Word Combination Rules in Tamil	51
Tamil Word Classes	53
Tamil Nouns	54
Caste	54
Gender	54
Number	55
Cases in Tamil Nouns	55
Pronouns and Cases	57
Tamil Verbs	58
Transitive, Intransitive and Causal Verbs	60
Passive Voice in Tamil	61
Negation in Tamil	61

,	
Modal Verbs	
Modal Verbs	
Modal Verbs Adjective Particles	
Adjective Particles Adverbial Particles	

CHAPTER V

DESIGN ISSUES	
DESIGN ISSUES Introduction	
Introduction Interface Design	
User Characteristics	65
Levels of Interfacing	66
Levels of Interfacing Form Link Diagram (FLD) Functional Design	
Functional Design	
Functional Design SCIIT - Simple Code for Information Interchange in Tamil	
SCIIT - Simple Code for Information Interchange III Furnition	
English Morphological Process Tamil Morphological Process	
Tamil Morphological Process	
Dictionary Dictionary Look-up (English to Tamil & Tamil to English)	
Dictionary Look-up (English to Tamil & Tamil to English)	
Translation Input Related Issues	
Controlled Environment	81
Tockenization	
Translating Engine English Language Parser.	83
English Language Parser.	
Verb Groubs	
Grammatical Relations Structure. Verb Çode	
Verb Çode Comparative Grammar Rules	
Comparative Grammar Rules Sentence and Document translator utilities	
Sentence and Document translator utilities.	

CHAPTER VI

RESULTS AND DISCUSSION	
RESULTS AND DISCUSSION Introduction	98
 Dictionary Look-Up Utilities 1.1 English to Tamil Dictionary 	
1.1 English to Tamil Dictionary1.2 Tamil to English Dictionary	
2.2 Document Translator Translation Mechanism	

CHAPTER VII

LIMITATIONS AND FUTURE EXTENSIONS	
Introduction	114
Limitations	114
Limitations on Implemented software	
Designed, but not Implemented portions	
Other Basic Requirements for an English to Tamil MT system	
Future Extensions	

CHAPTER VIII