

Discourse Translation from English to Telugu

Dr. T. Suryakanthi
Fellow, Faculty of Computing
Botho University
Gaborone, Botswana
+267 74777137

suryakanthi.tangirala@bothouniversity.ac.bw

Dr. Kamlesh Sharma
Associate Professor
Lingaya's University
Faridabad, India
91-9210474396

Kamlesh0581@gmail.com

ABSTRACT

Discourses are texts that are above the sentence level. Translating discourses requires special attention as the sentences should be translated keeping the context in mind. The first and second sentence should be interpreted as a whole and not as individual sentences. The present paper deals with translating two types of discourses, compound and complex sentences from English to Telugu language. The paper discusses the design of algorithms to translate compound and complex sentences from English to Telugu. The paper discusses the results obtained by implementing the two algorithms in a machine translation system on sample test suites.

Keywords

MT- Machine Translation SL: source language, TL: Target language, CC: coordinate conjunction, RB: adverb, POS: parts of speech

1. INTRODUCTION

Discourse is commonly described as a language above sentence level. Discourse comprises of sequence of sentences that must be interpreted as a whole keeping the context in mind. Majority of existing Machine Translation (MT) systems translate only single sentences. An input of two consecutive sentences, often gives disappointing results. In most of the cases, the second sentence is translated as a completely isolated piece of information, without taking into account its preceding text. Contextual knowledge is required to find the connection between two sentences, resolve ambiguities and infer new information to get a coherent view of information being communicated through the text [1]. The focus of discourse analysis is the understanding the structure of naturally occurring spoken and written language [2]. Some discourses are formed with connectives. The translation of discourses formed with connectives is described in this paper. Compound and complex sentences are discourses that are formed with connectives. Translating them from English to Telugu is a challenging task as the word order of source and target languages is not the same.

The examples in the paper show the source language (SL) sentences and its translation into target language (TL). The TL sentences are also shown in transliterated form. The machine translation system gives the transliterated outputs. The transliterations were generated based on Quillpad software [3]. The schema used for transliteration is given in Appendix 1.

1.1 Compound Sentences

Compound sentences are formed by connecting two independent clauses. Coordinate conjunctions are used to connect two independent clauses. These conjunctions are called discourse connectives which lay a relation between the two independent clauses [4].

1.2 Complex sentences

Complex sentences are formed by joining dependent and independent clauses using subordinate conjunctions (SC) [4].

2. COMPOUND SENTENCES TRANSLATION PERSPECTIVE

The two independent clauses of a compound sentence are joined by a coordinate conjunction followed a coma (.). Compound sentences are very rare in Telugu. When two sentences are connected by conjunctions like కావున, ఇంకనూ, అందువలన they can be considered as compound sentences [5]. List of English coordinate conjunctions and their corresponding Telugu conjunctions, their connection type and POS are shown in Table 1.

Table 1. Conjunctions of English and Telugu

English Connective	Connection type	POS Tag	Telugu Connective
For	connects explanation	IN	కావున
And	connects similar ideas	CC	-
But	connects conflicting ideas	CC	ఐనను, కానీ
Or	connects choices	CC	లేనిచో
Yet	connects explanation	CC	ఇంకనూ
So	connects explanation	CC	అందువలన

3. ALGORITHM FOR TRANSLTING COMPOUND SENTENCES

Compound sentences are formed by joining two independent clauses with a coordinate conjunction (CC) or conjunctive adverb (RB) and the two clauses are separated by a coma (,) or a semi colon (;).