Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAEREST 2016)

Annotations of Connectives and Arguments in Malayalam Language

Kumari Sheeja S[a,b], Sobha Lalitha Devi[a*]

[a] AU-KBC Research Centre, Anna University, Chrompet, Chennai ,
[b] KCG College of Technology, Karapakkam,Chennai,
sobha@au-kbc.org

Abstract

Discourse relations in natural languages link clauses in text and compose overall text structure. Discourse connectives are an important part of modeling the Malayalam discourse structure. We followed the annotation procedure of Penn Discourse Tree Bank and worked on tagging of discourse connectives and arguments of Malayalam text and also report the senses of relation. We present our work on annotations of Malayalam discourse connectives and arguments which helps to know more about the discourse connectives and their appearance in case of semantic rules in Malayalam discourse. Discourse connectives may or may not be explicitly present in the relation. In our work, we focus on the annotation of both explicit and implicit connectives and arguments in Malayalam text and showed encouraging results.

1. Introduction

Natural Language Processing is a field of Computational Linguistic   which is mainly used for various text processing in natural languages such as part-of-speech tagging ,word segmentation, word sense disambiguation , named entity recognition, parsing etc. Discourse analysis is one of the most important task in Natural Language Processing that has many applications in natural languages such as text summarization,information extraction ,opinion mining etc. Discourse connectives are extensively used in NLP applications to describe how two sentences or clauses are semantically connected in large text. Discourse connectives may or may not explicitly exist in the text. Explicit connectives signal the presence of discourse connectives between sentences or clauses .

Ramu is an intelligent boy **but** he cannot score good marks in the examination.

---

[*] Corresponding author. Tel.: +0-91984-013-3444;
E-mail address: sobha@au-kbc.org

In this example the connective "but" makes a relation between two clauses or sentences and making the text coherent.In implicit discourse relation ,there is no connective in between the arguments of text which results a very difficult task.We describe an approach to annotate a large scale corpus in terms of more basic characterization of discourse structure in malayalam text in terms of discourse connectives and their arguments.

Discourse connectives have multiple senses like verbs depends on the semantic structure of the text in Malayalam discourse .In this paper our effort to annotate Malayalam discourse connectives and arguments based on the guidelines of PDTB. The content of the paper started with introduction of the work. Section 2 describes the literature survey of our paper. Section 3 gives an overview of discourse relations in Malayalam , section 4 describes Malayalam text corpus and annotation process and section 5 describes the sense annotations of connectives. Finally the paper ends with the conclusion of the work.

2. Related Work

Annotation process of discourse connectives and their arguments have been explored in various languages such as Hindi[1],Tamil[3], Arabic [10], etc.They worked on tagging of discourse connectives and their arguments .They developed discourse relation bank of Hindi,Tamil and Arabic that will be useful as a resource for further researches in discourse.Versley[5] worked on tagging German discourse connectives and arguments using English training data and a German_ English parallel corpus. Versely's approaches were to transfer a tagger for English discourse connectives. They have done this work using a freely accessible list of connectives by annotation projection. Annotating the discourse connectives in Turkish language is done by Zeyrek, D., and Webber, B. L. [7]. Faiz, S. I., and Mercer, R. E. [9] worked on explicit connectives and their arguments in discourse with examples. Wang et al.,[6] used sub-trees as features and achieved a significant improvement in identifying arguments, explicit and implicit discourse relation .Yuping Zhou and Nianwen Xue [11] worked on tagging the discourse relations and their senses in Chinese language. In this paper we have explored the annotation results of various discourse connectives and their arguments in Malayalam language.

3. Discourse Connectives in Malayalam

Malayalam, a South Indian or Dravidian language is a free word order language but maintains the verb in final position. Discourse connectives in Malayalam text are important for interpreting or producing text.Malayalam has a special place in the classification of languages.There are different spoken forms in Malayalam eventhough the literary dialect throughout kerala is almost same.Malayalam has its own distinct script and most of the syntactic information is embedded as morphological structure.This causes the analysis of word forms of Malayalam to cross the limits of morphology and it reaches to syntactic and semantic level.Malayalam is a free-word order language and words are seen agglutinated, hence most of the connectives are seen in agglutinated form. The discourse relation in Malayalam language can be syntactic (a suffix) or lexical[8]. Discourse relation can be within a clause, inter-clausal or inter-sentential. Discourse connectives are an important part of modeling discourse structure. In this paper, we now describe various connectives and arguments present in Malayalam language and the sense of relations. The two major category of relations are Explicit and Implicit relations. We also observed the other types of relations such as Entity relation,Alternative lexicalized relation and No relation between the arguments.

3.1 Explicit connectives

The explicit connectives are morphemes or free words that trigger discourse relations in Malayalam language .The connectives can occur at the initial, medial or final position of arguments in Malayalam language [9]. Here shows the example for explicit connective in malayalam language.

(2)[prakqthi sundarangngaLAya pradESangngaL kEraLaththil dhArALamuNTu.]
    </ARG1>
  Natural beauty places kerala+in many.
  AthukoNTu <CON>
  therefore
  <ARG2>[vinOdasaFcaArikaL kEraLaththe ishTapetunnu.]</ARG2>
    tourist+people kerala like.
(There are many natural beauty places in kerala .**Therefore** many tourist
people like to visit kerala.)

The connective "therefore" in the above example occurs inter- sententially by connecting the two sentences. Connective occur at the initial position in the second argument. We see that the connectives are explicitly realizing the relations between two arguments.