Relative_Clause

March 18, 2021

1 Project Plan: Language Typology And Universals

Team : 1. Chayan Kochar - 2019114008 2. Tanishq Goel - 2019114015

1.1 Introduction

In this project, we will be covering the analysis of relative clause constructions in English, Hindi and Telugu.

Relative Clauses: A relative clause is a clause that usually modifies a noun or noun phrase and is introduced by a relative pronoun (which, that, who, whom, whose), a relative adverb (where, when, why), or a zero relative. In English, a relative clause is a postmodifier--that is, it follows the noun or noun phrase it modifies. In this project we will see how it behaves in languages like Hindi and Telugu, and also confirm whether this is always true for English or not.

1.2 Literature Review

1. Bernard Comrie Chapter 7
2. https://www.aclweb.org/anthology/W04-1502.pdf

1.3 Corpora

Languages Used: English, Hindi, Telugu

1. English: https://www.cfilt.iitb.ac.in/~parallelcorp/iitb_en_hi_parallel/

2. Hindi: LTRC / above parallel corpora/ https://github.com/joshua-decoder/indian-parallel-corpora/tree/master/

3. Telugu: Issues with the Corpora: Telugu Corpus not on LTRC website.

https://bitbucket.org/sivareddyg/telugu-part-of-speech-tagger/src/master/
https://github.com/ltrc/telugu_treebank

1.4 Methodology

1.4.1 Type of Processing Done/To-do:

1. Find and explore different types of relative clauses, manually find out how the relative clauses appear in the test.
2. Preprocess data if necessary.
3. Then we can code to find particular observation/statistics related to the same -> thus figure out any implicational rule

1.4.2 Proposed Tools

1. NLTK/IndicNLP for tokenizing.
2. Stanford core nlp with Stanza
3. Tools in LTRC like shallow parser, etc(if needed)
4. If using dataset from LTRC, SSF_API will be used(as all data are annotated)
5. On the other hand, we can also use the annotated data like the dependency relations, karaka relations, etc to find in details about the relative clause constructions.