

Assignment 5

Language Typology and Universals

Tejasvi Chebrolu

2019114005

Preprocessing

The dataset was created from the link <https://preon.iiit.ac.in/~jerin/bhasha/>. The *hi-en* parallel corpora was used. Then, the sentences containing relative clauses and causatives were extracted from the English half of the dataset. The corresponding sentences were then extracted from the Hindi half of the dataset. For Telugu, the selected sentences were manually translated so that all the languages have the same sentences. These sentences can be found in the `sents` directory.

Relative Clauses

Manifestation

English

In English, relative clauses appear usually as pronouns using a *wh*- particle. Examples are *which*, *what*, *who*, etc.

Hindi

In Hindi, relative clauses usually contain the word *जो* or consist of the words which contain prefixes like *जिस* or *जिन*. Examples are *जिनमे*, *जिससे*, *जिन्हे*, *जिसको*, etc. There are many examples in the data which have the words. These words are mostly nominal modifiers, usually determiners. They can also be relative adjectives or relative adverbs as well, and not just relative pronouns.

Telugu

In Telugu, relative clauses usually have suffixes attached to the root word. This is because Telugu is an agglutinating language and there is a high level of inflection. Examples are దీనిలో, దానితో, ఎవరిని, ఎవరికీ, etc.

Extraction

English

The sentences with the relative clause extracted can be found in file - `final/en_rel.md`.

Hindi

The sentences with the relative clause extracted can be found in file - `final/hi_rel.md`.

Telugu

The sentences with the relative clause extracted can be found in file - `final/tel_rel.md`.

Position of Clause

English

The relative clause occurs either **before or after** the head noun depending on the construction of the sentence. Therefore, it cannot be generalised.

Hindi

The relative clause occurs either **before or after** the head noun depending on the construction of the sentence. Therefore, it cannot be generalised.

Telugu

The relative clause occurs either **before or after** the head noun depending on the construction of the sentence. Therefore, it cannot be generalised.

Causatives

Extraction

English

The sentences with the causative extracted and its type can be found in file - `final/en_cau.md`.

Hindi

The sentences with the causative extracted and its type can be found in file -

`final/hi_cau.md`.

Telugu

The sentences with the causative extracted and its type can be found in file -

`final/tel_cau.md`.

Manifestation

English

In English, causatives are usually expressed by certain verbs like *let*, *have*, *make*, etc. English also shows lexical causatives with certain kinds of verbs.

Hindi

In Hindi, causatives are usually expressed in the suffixes of the verbs that they are attached to like *करवा*, *वाकर*, *वायेगा*, *वाओगे*, etc.

Telugu

In Telugu, causatives are usually expressed by words that have suffixes attached to them like *ప్రోత్సహించారు*, *పదవీకాలం ఉంటుంది*, *పంపిణీ చేశారు*, etc.

Directory Structure

The directory structure is as follows -

```
.
├── data
│   ├── train_en.txt
│   └── train_hi.txt
├── final
│   ├── en_cau.md
│   ├── en_rel.md
│   ├── hi_cau.md
│   ├── hi_rel.md
│   ├── tel_cau.md
│   └── tel_rel.md
├── report.pdf
└── sents
```

- |— en_cau.txt
- |— en_rel.txt
- |— hi_cau.txt
- |— hi_rel.txt
- |— tel_cau.txt
- |— tel_rel.txt
