

# Causative Constructions in Indo-Aryan and Dravidian Languages: Interim Progress Report

Prajneya Kumar - 2019114011

Jayant Panwar - 2019114013

April 5, 2021

## Introduction

This document serves to be an interim report for the work done by our team for the LTU course project. We try to identify and search for patterns in **causative constructions** in Hindi, English and Telugu languages.

Majority of the work carried out so far has been emphasized more on widely used English and other European languages. Indian Languages have been explored little because of the amount of information available in non-English language is relatively less. Even for English, not much research work has been conducted as far as generalizations about causative construction is concerned.

We approach this problem by first looking through a dataset of 500 sentences each for the three aforementioned languages: Hindi, English and Telugu; we annotate the data and extract sentences which we believe contain causative constructions.

## Literature Review

There are a lot of literature which exists for Causative Constructions, but a general pattern observed is that they are not transferable to Indian Languages. This is either because they are restricted to patterns observed in the language or due to lack of resources in IL. Some resources we reviewed which could be helpful in some way or the other include:

1. Bernard Comrie - Chapter 8: Causative Constructions: Bernard Comrie, in this chapter, explains typology of the syntax and semantics of causative constructions. Comrie explains the different types of causatives, which we will discuss further, and proposes different parameters that affect the existence of causatives in a sentence.

2. A Dependency Treebank for Telugu: This paper by Taraka Rama and Sowmya Vajjala does not talk about causatives in particular, but talks about annotations and development of Telugu treebank. This helped us in looking for patterns in the Telugu corpus.

## Causative Constructions

Comrie describes Causative Constructions as an expression that denotes a complex situation involving two component events:

1. the causing event, in which the causer does or initiates something and
2. the caused event, in which the causee carries out an action or undergoes some kind of change as a result of the causer's action

### Type I: Lexical Causatives

In this case, idea of causation is part of the semantics of the verb itself (Wikipedia). For example, die, kill etc.

### Type II: Morphological Causatives

These consist of words which have certain suffixes, prefixes or infixes attached to them to transform them into a causee/causer. For example, **सुलाना**, **सुलवाना** etc.

### Type III: Periphrastic Causatives

This case includes those sentences which make the use of more than one verbs or clauses to express causativity. For example, *I made him write*. The use of two verbs here, *made* and *him* denotes that this is a periphrastic causative.

## Methods

In order to find causative constructions in our corpus, we opted for different strategies. Generally speaking the methods were a combination of both manual and automated scanning. We scanned the entire corpus of a given language using python code at first and also opted to do manual scanning in cases where finding the causative constructions using code was a difficult task.

### Telugu

Since causative constructions in Telugu are constrained to a limited number of causative verbs and due to our non-native knowledge of Telugu, the causative constructions in Telugu corpus were scanned out using python code, while Ph.D. Scholar Hema Ala helped us in the manual annotation.

We first looked at the 'inchi' infix in each word, which implied the presence of causatives. Furthermore, we looked for words like:

చేయిస్తాను  
చేయించాను  
చేయిస్తానా?  
చేయించానా?  
చేయించను  
చేయించలేదు  
చేయించనా?  
చేయించలేదా?  
చేయిస్తాను  
చేయించలేదు?  
చేయించను?  
చేయించాను?  
చేయించాలని  
చేయించారు  
చేయించుకున్నాడు

In total, we were able to extract 370 sentences containing our desired construction.

### Hindi

Causative constructions in Hindi are easily identifiable owing to the morphological changes in the transitive and intransitive verbs. The suffixes such as 'वाया', 'वाना', 'वाओगे', etc. when attached to the verbs transform them into causative verbs.

However, causative constructions are not just limited to attaching suffixes to the main verb. For example, in some sentences, 'करवाना', 'करा दिया', and other variations of 'कर' denote presence of causative construction as well.

Finding these causative constructions in our corpus of Hindi sentences was partly a manual task and partly an automated one. Suffixes like 'वाना', 'वाता', 'वायेगा', etc. were searched for using code written in python meanwhile the other cases of causative constructions (like 'करवा दिया', 'करवाकर', etc.) were found out manually.

In total, we were able to extract 97 sentences containing our desired construction.

### English

In order to look for causative constructions in English, we listed certain verbs that occur in causative constructions with greater than a chance frequency. Once the sentences containing the verbs were filtered, we manually checked each one for a false positive, i.e., the sentences in which the verb did not take a causative form but was filtered solely due to its presence in the sentence. This process was repeated until we had scanned the entire English corpus for all the possible cases of causative constructions.

Here, we made a list of verbs which would help us classify causative sentences. These list of verbs included all three forms of: "let", "allow", "permit", "make", "made", "force", "require", "get", "got", "help".

In total, we were able to extract 1436 sentences containing our desired construction.

### Drawbacks and Challenges

The parallel corpora of English and Hindi did not ensure that if the sentence is of causative construction in English corpus it will be of causative construction in Hindi as well. We came across many such challenging translated cases where the construction became different. Therefore, in order to solve this we opted for a thorough individual investigation into both the English and Hindi corpus without worrying whether it's translation is of causative construction in the corresponding language or not.

For the Telugu corpus, it was kind of Ms. Hema to help us in finding causative constructions in the

Telugu corpora as neither of the team members have linguistic knowledge regarding Telugu.

As we can see, the number of Telugu and Hindi sentences we could obtain which contain causative construction is relatively less than those we could find in English. As far as Telugu is concerned, this was because we were not able to find a large enough dataset for Telugu language, and thus had to borrow a corpus from another team with a different construction. Our corpus consisted of 900 lines, and thus we could find only a limited number of sentences there. As far as Hindi is concerned, we had a corpus of 520 good complete sentences obtained from the IIT Bombay DevSet Corpus (consisting of 520 sentences), which yielded around 50% of the sentences we have obtained. To obtain even more sentences, we took the training set of the IIT Bombay Hindi Training corpus, whose majority of sentences were very short.

Keeping the above problem in mind, it is possible that we have problem in deriving patterns and rules for causative constructions in Telugu and Hindi. Should a problem like this arise, we

aim to translate some of the English sentences to Hindi and Telugu manually and then try to use the translated sentences as qualified causative constructions to continue our analysis work. This is why we kept the English dataset comparatively larger than the Hindi and Telugu datasets, as it was readily available.

## Conclusion & Future Work

Since our corpus has been completely scanned and analyzed with respect to causative constructions, we are looking forward to moving on with the final part of our project. The main motive now will be to find implicational rules that try to correlate the structure of the sentences to the presence of the causative constructions. This will be done for every language. Then we aim to generalize the form of causative constructions that exists across all the three languages that have been selected. If possible, we will try to analyze how typological classification can also help play a part in the presence or the absence of causative constructions, keeping in mind some more literature that we would read (apart from those already mentioned here).

\*\*\*\*\*