



## THE INDIAN LANGUAGES EVENTS PROJECT

### ABSTRACTS

The Indian Languages Events Project is a large effort aimed at creating event annotated datasets for Indian languages. Event analysis and extraction is very useful in Natural Language Processing and information retrieval, as events provide temporal information about the world.

Thus far, we have provided an in-depth analysis of the annotation and representation of events in Hindi (Goud et.al., 2019) and created a knowledge graph based representation for factoid question-answering from modifications to this annotated dataset (Goel et. Al., 2019). Furthermore, we have created neural systems for automatic detection of events (Prabhu et.al., 2019) Given the syntactic complexities of Hindi such as free word order, fragmentation of events and conjunct and light verb constructions, automatic detection is not a trivial task.

### DEFINITIONS AND METHODS

What is an event? An event, according to ISO-TimeML by Pustejovsky et. al.,(2003) is anything that "*occurs, takes place, obtains or holds true*", Event semantics has been a branch of linguistic philosophy for quite some time now, and it is ingrained into the syntax of the language we use, as was seen in Goud et. al. (2019), where directly defining events in Hindi as it was in English, did not prove to be semantically inconsistent with Hindi's representation of states and temporality. Event representation takes a pragmatic, real life view of the world and tried to represent it in text; concepts such as sequentiality, finiteness, and time. Events are not restricted to just the action verbs, as commonly misunderstood. Nouns and adjectives which have a notion of time are also events. "War", for example, is an event, and so are "election", "hunger strike" and so on. These nouns have a sense of time and finiteness associated with them .

We don't just annotate events. We also mark states, time expressions such as dates and months, participants in those events and relations between them. Each have their own guidelines on what should be annotated, what should be ignored and what is the nature of event representation in the syntax of that language. For example, in Hindi, we do not consider the postposition that follows time expressions such as "10 din mein", but we consider all auxiliaries in verbal predicates such as "bhaag gaya".

### PUBLICATIONS

For the list of relevant publications, please visit <https://www.tinyurl.com/events-iiit>

### PROJECT OBJECTIVES

This project has two aims. From the perspective of corpus linguistics, it aims to introduce the concept of events and temporal information extraction in Indian languages by creating TimeBanks and event-aware corpora. We also aim to use event information for downstream tasks such as inference and reasoning, ontology enrichment, question-answering and summarization tasks.

Development of general guidelines allows easier data creation in related languages, making temporal information representation a reality for Indian languages, Given the current work in Hindi, we have also begun working on event representation in Kannada. As the project expands, we hope that development of TimeBanks in all Indian languages is possible.

We also hope to create systems that analyze and retrieve temporal information from corpora and use that for creating a generalized, semantically rich knowledge representation across languages. This can be used in a variety of ways, from making downstream applications better to multilingual data enrichment.