Hindi Causal TimeBank: an Annotated Causal Event Corpus

Tanvi Kamble and **Manish Shrivastava**
International Institute of Information Technology, Hyderabad
tanvi.kamble@research.iiit.ac.in, m.shrivastava@iiit.ac.in

Abstract

Events and states have gained importance in NLP and information retrieval for being semantically rich temporal and spatial information indicators. Event causality helps us identify which events are necessary for another event to occur. The cause-effect event pairs can be relevant for multiple NLP tasks like question answering, summarization, etc. Multiple efforts have been made to identify causal events in documents but very little work has been done in this field in the Hindi language. We create an annotated corpus for detecting and classifying causal event relations on top of the Hindi Timebank (Goel et al., 2020), the 'Hindi Causal Timebank' (Hindi CTB). We introduce semantic causal relations like Purpose, Reason, and Enablement inspired from Bejan and Harabagiu (2008)'s annotation scheme and add some special cases particular to Hindi language.

1 Introduction

Events and the relations between events form an important part of textual and verbal communication. A Dynamic event (or *event*, by TimeML Guidelines (Saurí et al., 2006)) as defined by Goel et al. (2020) is a cover term for situations that happen, occur, hold, or take place. Stative events (or *States*) are the predicates describing states or circumstances in which something obtains or holds. Detecting events, states, temporal expressions, and their relations provides a rich source of information and represents real-world information in the text. Causal relations between events is a comparatively recent task that detects the presence of causality between two events- In other words, detecting whether an event needs to occur for another event to take place or if an event 'causes' another event. Event causality plays a significant role in NLP applications. Question Answering, Summarization, Information Extraction, and Knowledge graphs are some of the many downstream tasks that can be solved better with the help of causality.

Sentence 1:	पुलिस	को देखकर	चोर	भाग	गए
	Police	*after seeing*	*thieves*	*ran*	*away*
Sentence 2:	दूध	पीकर	वरुण	सो	गया
	Milk	*after drinking*	*Varun*	*sleep*	*happened*

Figure 1: Examples of causality. Sentence 1 shows causality but Sentence 2 does not.

We extend the idea of event causality to the Hindi language by building upon the initial seed dataset of events and states (Goel et al., 2020) and annotate causal event and state relations in an augmented dataset of 1,000 Hindi news articles. We mark the causal event pairs and classify them into different relation types. We provide a comprehensive set of guidelines for identifying and classifying event relations based on the previous work (Bejan and Harabagiu, 2008). To our knowledge, this is the first work in the field of event causality in the Hindi language.

Event causality is a difficult task as causality is a psychological concept. It involves a lot of implicit connotations and real world context. One cannot simply look at the structure of a sentence and determine whether causality exists or not. For example, in figure 1 the two sentences follow the same structure, have two events but only one sentence exhibits causality. The second sentence simply shows a temporal relationship. This is because in the first sentence, the real world context tells us that a thief would probably want to stay away from the police and hence the first event of seeing the police would have caused the second event of running away to occur. Thus, manually identifying causal relations by two annotators can vary as they might not share the same real world context as they have different experiences. Therefore, while creating this dataset, multiple annotators annotated the data and we only marked the causal relations where majority of the annotators agreed upon the existence of causality between two events.

The paper is divided into the following sections; Section 2 talks about the previous work and gives brief overview of the existing datasets with annotated causal event relations. We introduce the dataset and present the annotation guidelines in Section 3. In Section 4, we conduct baseline experiments on our dataset, which is followed by conclusion in Section 5.

2 Related Work

The first work on events dates back to 1978 (Mourelatos, 1978). Ever since, there have been multiple datasets with events and event extraction guidelines. The TimeML annotation guidelines in 2004 (Saurí et al., 2006) introduced a rule-based approach for the event and state extraction. The existing large-scale datasets are all in the English Language. Following are some of the widely used datasets for event causality based on news events. :

1. EventStoryLine (ESC)(Caselli and Vossen, 2017): This dataset has both explicit and implicit causal relations marked for the Event Coreference Bank+ (ECB+) (Cybulska and Vossen, 2014) and supports both intra- and inter sentence causal relations. Both temporal and causal relations are marked in this dataset and a total of 2,265 causal relations are identified. These causal relations are categorised broadly into three categories out of which two categories are implicit causal relations that further have sub categories and the explicit causal relations look for markers like 'cause' and 'caused by' to assign them that label.

2. Causal TimeBank (CTB) (Mirza et al., 2014) Another widely used dataset, CTB identifies a total of 318 Causal Links and 117 Causal Signals from the TempEval-3 dataset (UzZaman et al., 2013). The Causal signals help identifying the relations explicitly. The causal links are divided into three categories: Cause, Prevent and Enable.

3. Causal News Corpus (CNC) (Tan et al., 2022a) This is an annotated news corpus with 1,957 causal and 1,602 non causal events. It uses five tests of causality to classify relations as causal or not. The causal relations are further classified as Purpose, Cause, Condition and Negative Condition based on linguistic cues present in the test.

4. Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008): Causal relations are one of the many discourse relations. It does not take into account all the possible causal relations as it ignores the possibility of intra clausal causal relations. (Tan et al., 2022b)

Since English and Hindi are typographically and semantically dissimilar languages, we cannot directly use the annotation guidelines from these datasets or translate them to Hindi to create a Hindi language dataset. There have been attempts to create event extraction models in Hindi (Ahmad et al., 2020; Kuila et al., 2018), but there hasn't been any significant work in ECI for this language.

3 Hindi Causal Timebank

We build the Hindi Causal TimeBank (HindiCTB) on top of the Hindi TimeBank by annotating causal relations between events. The further sub-sections talk about the annotation guidelines for the same.

3.1 Background

Goel et al. (2020) divides events into five categories; Perception, Aspectual, Reporting, I_Action, and Occurrence and states into Descriptors and Declarative states. This is in accordance with the TimeML annotation guidelines. Along with this, Time expressions of Time, Date, Duration, and Set have been identified. Furthermore, temporal relations are annotated according to the TimeML annotations (Saurí et al., 2006). We extend the dataset and add causal relations. These relations adapt to the guidelines provided by Bejan and Harabagiu (2008) in the Hindi language. The different types of relations are elaborated in further sections.

3.2 Causal Relations

A causality relationship exists between two events, if one event causes another event to take place. Based on their temporality and type of cause, causal relations can be defined in various ways. Based on our definition of events and states, causal relations can occur between two events or an event and a state. For example[1]:

kevala	***pAsa***	***hue***	*CAwra*	*hI*
ONLY	PASS	IS	STUDENTS	WILL
aMwima	*parIkSA*	*meM*	***bETeMge***	
FINAL	EXAM	IN	APPEARED	

[1]All examples are written in the WX notation (Bharati et al., 2002)

Here *pAsa hue* is a state whereas *bETeMge* is an event.
We propose the following types of Causal relation for the Hindi Language:

1. REASON If event E_2 is the direct consequence of event E_1, they have a REASON causal relation. When multiple reason events cause one consequence event, this relation is applied recursively. Discourse markers like *{ke kArana, -ne para, kI vajaha se}* often occur between events connected by the REASON relation. For example:

pEse	*xene*	*se*	***inkAra***	***karne***
MONEY	GIVING	ON	DENYING	IS
para	*coro*	*ne*	*rAma*	*ko* ***mArA***
ON	THIEVES	THE	RAM	BEAT

2. PURPOSE: Event E_1 is said to be the PURPOSE of event E_2 if the intention of event E_1 is to achieve a goal event E_2. E_1 must necessarily occur before E_2 for this relation to exist. In multiple cases, the presence of the signal word between two consecutive events in the same sentence implies the existence of a PURPOSE relation between the two events. Discourse markers like *{ke liye}* often occur between events connected by the PURPOSE relation. For example:

avnI	*ne*	***nAcane***	*ke liye*	*jUwe*
AVNI	DID	DANCE	FOR	SHOES
uwAre.				
REMOVED				

3. ENABLEMENT: When the occurence of event E_1 is necessary for an event E_2 to happen such that E_1 does not directly cause E_2, the relation between the two events is that of ENABLEMENT. If three events may be related to one another, A enables B, and A enables C does not imply that B enables C. For example:

rAma	*ne*	*pulIsa*	*ko*	***xeKkara***
RAM	DID	POLICE	TO	AFTER SEEING
rAswA	***baxlA***	*Ora*	***ravAnA ho gayA***	
PATH	CHANGED	AND	WENT AWAY	

We see that A(*xeKkara*) enables both B(*rAswA baxlA*) and C(*ravAnA ho gayA*), but there is no causal relations between B and C. However, a temporal relation still exists.

3.3 Other Discourse Relations

1. SUBEVENT: SUBEVENT is similar to the SUBFRAME relation from FrameNet (Baker et al., 1998). It holds between an event A that is part of a composite event B. A composite event can have multiple subevents, and a subevent can be a composite event for other events. The hierarchy of events resulting from using the SUBEVENT relation can encode complex semantic and temporal structures.

2. RELATED: RELATED refers to events between which there is a weak connection. These relations are very infrequent. For example, there's a RELATED relation between event A and event B if event A is a probable cause of event B, but the causality cannot be ensured. A RELATED relation can also exist between events if one is an irrelevant elaboration of the other event.

3.4 Special Causality

We define causality for some special cases as follows

- **Possible events:** Some events have a possibility of occurring in the future or they could have happened in the past. The most common way of detecting such cases is the use of *agara-wo* and *yaxi-wo* markers which translates to *if-then* markers in English. According to our definition of causality, we mark the causal relation between the possibly occurring events considering that the events do take place. An example of REASON relation, in this case, would be:

agara	*krUda*	*oIla*	*saswA*	***hogA***
IF	CRUDE	OIL	CHEAP(ER)	WILL HAPPEN
wo	*mahaMgAI*		*Wamne*	*kI*
THEN	INFLATION		STOP	OF
saMBAvanA	***baDegI***			
CHANCES	WILL INCREASE			

- **Negative events**: Events that do not occur also play a role in causality. An event that does not occur might still lead to another event. In such cases, we mark the causal relation that would exist had the event occurred. Non-occurring events are generally accompanied by the negation word *nahIM* in their Verb or Noun phrase. For example, the following sentence will have a PURPOSE relation.

yaSa	***ne***	*KAnA*	*KAne*	*ke liye*
YASH	DID	FOOD	TO EAT	FOR
hAWa	***nahIM***	***Xoye.***		
HANDS	NOT	WASHED		

3.5 Dataset Statistics

We use the same dataset as the base corpus as (Goel et al., 2020). It contains a total of 1000 news articles in the Hindi language. There are 292,517 tokens, 25,829 events, and 3,516 states. The corpus already has TLINKs or temporal relations marked which are 6069 in number. We have marked causal event relations in about 582 files which contain 14,739 events and 1,918 states. A total of 2,210 causal event relations pairs are manually annotated [2] as per the definitions stated in the above sections. Each pair contains the first event, ‘E1’, and the second event, ‘E2’ related by a causal relation type. Table 1 contains the distribution of the causal relations in the corpus.

Tag	Frequency
Subevent	122
Reason	815
Purpose	583
Enablement	533
Related	157
Total	**2210**

Table 1: Distribution of causal event relations according to the various tags

It is also evident from Table 2 that our dataset has similar number of annotated relations when compared to other benchmark datasets

Dataset	Number of Relations
Hindi CTB (Our)	2,210
EventStoryLine (Caselli and Vossen, 2017)	2,265
Causal TimeBank (Mirza et al., 2014)	318
Causal News Corpus (Tan et al., 2022a)	1,957

Table 2: Comparison of number of relations for different benchmark datasets

4 Baseline Experiments

We conduct Causality Identification for our dataset. In this task we aim to identify whether a sentence with annotated events has causality or not. The input to both models is a sentence with markers of <ARG0> and <ARG1> around tokens of the first and the second event respectively. The input is of the form: $tok_1 tok_2 ... tok_(k-1) < ARG0 > tok_k ... tok_(k+l) < /ARG0 > tok_(k+l+1) ... tok_(m-1) < ARG1 > tok_m ... tok_(m+n) < /ARG1 > tok_(m+n+1) ... tok_o$ where $tok_k ... tok_(k+l) \epsilon$ tokens of event 1 and $tok_m ... tok_(m+n) \epsilon$ tokens of event 2. We feed this sequence to a sequence classifier. For this task we have identified 14739 non causal event pairs whereas only 1918 causal event pairs. This number is different from the annotated causal relations as we are only considering intra-sentence causality for the sake of simplicity. We randomly sample a subset of 1200 non causal relations to avoid bias in the model.

[2] https://anonymous.4open.science/r/HindiCTB-03F2

Dataset	Precision	Recall	F1
Hindi CTB (Ours)	88.99	55.52	68.38
Hindi CNC	85.36	72.09	781.16

Table 3: Results on different dataset when run on XLM-Roberta Model with Adapters

The experiment can be broken down into two model: The first one is ELECTRA based Hindi-bert by monsoon-nlp [3]. Since this model has been trained on Hindi Language, fine tuning it on our dataset might give favourable outcomes. We use a Classifier head on top of this model to identify the causality. However, this model overfits for the Hindi translated version of Causal Time Bank and Event StoryLine. For our dataset, we are able to achieve an accuracy of 71.9 percent and an F1 score of 59.3 percent.

The second experiment is run on a XLM-Roberta (Conneau et al., 2019) which is trained on a 100 different languages. We train it with Adapter on top and a causal LM head followed by a classification head to get the results. We get promising results on both our dataset as well as a hindi translated version of the Causal News Corpus as can be seen in Table 3. We use the HuggingFace models to perform all experiments.

This can be further extended to identify the type of causal relation that exists between the events but that is out of scope for this paper.

[3] https://huggingface.co/monsoon-nlp/hindi-tpu-electra

5 Conclusion

In this paper, we introduced the Hindi Causal Time Bank, by annotating the Hindi TimeBank with different causal relations and provided annotation guidelines for the same. Currently, Hindi CTB has news articles only. It can be used for pre-training models for other news datasets and the causal relations in our dataset can be used efficiently for downstream tasks like summarization, question answering. We suspect that this dataset might not help models for other topics as our dataset doesn't provide samples for them. However, our annotation guidelines can be used to prepare datasets for other topics as well.

We also provided baseline results for sequence and pair classification for this dataset. These models do need significant improvement and the results might improve as we add more data which will help the model learn the causal relation patterns. One major limitation of these models is that they require events marked in the sentences, which add another step in the pre-processing and might reduce the utility of the model as not every dataset has events marked in them. In the future, we aim to create a span detection algorithm such that there won't be a need to annotate a sentence with event pairs. It would also be feasible to make an end-to-end model which first annotates events in a sentence and then identifies the existence of causality between them. Overall, Hindi CTB has opened a lot of avenues for exploring causality in Indian languages and we hope that our work will inspire other authors to develop datasets ad model pertaining to Indian languages.

References

Zishan Ahmad, Deeksha Varshney, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Multi-linguality helps: Event-argument extraction for disaster domain in cross-lingual and multi-lingual setting.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Cosmin Bejan and Sanda Harabagiu. 2008. A linguistic resource for discovering event structures and resolving event coreference.

Akshar Bharati, Vineet Chaitanya, Rajeev Sangal, and Brendan Gillon. 2002. Natural language processing: A paninian perspective.

Tommaso Caselli and Piek Vossen. 2017. The event StoryLine corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86, Vancouver, Canada. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the 9th international conference on Language Resources and Evaluation (LREC2014)*.

Pranav Goel, Suhan Prabhu, Alok Debnath, Priyank Modi, and Manish Shrivastava. 2020. Hindi Time-Bank: An ISO-TimeML annotated reference corpus. In *16th Joint ACL - ISO Workshop on Interoperable Semantic Annotation PROCEEDINGS*, pages 13–21, Marseille. European Language Resources Association.

Alapan Kuila, Sarath chandra Bussa, and Sudeshna Sarkar. 2018. A neural network based event extraction system for indian languages. In *Fire*.

Paramita Mirza, Rachele Sprugnoli, Sara Tonelli, and Manuela Speranza. 2014. Annotating causality in the TempEval-3 corpus. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pages 10–19, Gothenburg, Sweden. Association for Computational Linguistics.

Alexander P. D. Mourelatos. 1978. Events, processes, and states. *Linguistics and Philosophy*, 2(3):415–434.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Roser Saurí, Jessica Moszkowicz, Bob Knippen, Rob Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. Timeml annotation guidelines version 1.2.1.

Fiona Anting Tan, Ali Hürriyetoğlu, Tommaso Caselli, Nelleke Oostdijk, Tadashi Nomoto, Hansi Hettiarachchi, Iqra Ameer, Onur Uca, Farhana Ferdousi Liza, and Tiancheng Hu. 2022a. The causal news corpus: Annotating causal relations in event sentences from news. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages

2298–2310, Marseille, France. European Language Resources Association.

Fiona Anting Tan, Xinyu Zuo, and See-Kiong Ng. 2022b. Unicausal: Unified benchmark and model for causal text mining. *ArXiv*, abs/2208.09163.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 1–9, Atlanta, Georgia, USA. Association for Computational Linguistics.