

TRIOS-TimeBank Corpus: Extended TimeBank corpus with help of Deep Understanding of Text

Naushad UzZaman, James Allen

Department of Computer Science

University of Rochester, Rochester, NY 14627

E-mail: naushad@cs.rochester.edu, james@cs.rochester.edu

Abstract

TimeBank (Pustejovsky et al, 2003a), a reference for TimeML (Pustejovsky et al, 2003b) compliant annotation, is widely used temporally annotated corpus in the community. It captures time expressions, events, and relations between events and event and temporal expression; but there is room for improvements in this hand-annotated widely used TimeBank corpus. This work is one such effort to extend the TimeBank corpus. Our first goal is to suggest missing TimeBank events and temporal expressions, i.e. events and temporal expressions that were missed by TimeBank annotators. Along with that this paper also suggests some additions to TimeML language by adding new event features (ontology type), some more SLINKs and also relations between events with their arguments, which we call RLINK (relation link). With our new suggestions we present the TRIOS-TimeBank corpus, an extended TimeBank corpus. We conclude by suggesting our future work to clean the TimeBank corpus even more and automatically generating larger temporally annotated corpus for the community.

1. Introduction

TimeML (Pustejovsky et al 2003b) is a formalism for annotating time expressions, events, and relations between events and temporal expressions in text. The TimeBank (Pustejovsky et al 2003a) corpus, based on TimeML, has been widely used in the community for training and evaluating systems that extract temporal information from text. But TimeBank contains some human errors and does not cover all temporal information present in text. This work is an effort to extend the TimeBank corpus and in some cases extending TimeML language too.

Our first goal is to identify and include missing events and temporal expression, i.e. events and temporal expressions that were missed by TimeBank annotators. Along with that this paper also suggests some additions to TimeML language by adding new event features (ontology type), few more SLINKs and also relations between events with words, which we call RLINK (relation link). With our new suggestions we present the TRIOS-TimeBank corpus, by extending TimeBank corpus.

We will start by describing TRIPS parser, a semantic parser, which we use in combination with other machine learning techniques for extracting events and temporal expressions. Next, we describe our system, TRIOS system to extract events and temporal expressions. Finally, we propose our extensions to TimeBank and TimeML.

2. Our System

2.1 TRIPS Parser

We use the existing TRIPS¹ parser² (Allen et al, 2008) to produce deep logical forms from text. The system is

¹ TRIPS: The Rochester Interactive Planning System:

<http://www.cs.rochester.edu/research/trips/>

² TRIPS parser demo:

<http://www.cs.rochester.edu/research/cisd/projects/trips/parser/cgi/web-parser-xml.cgi>

generic and no grammatical rules or lexical entries were added specifically for this task. The TRIPS grammar is lexicalized context-free grammar, augmented with feature structures and feature unification. The grammar is motivated from X-bar theory, and draws on principles from GPSG (e.g., head and foot features) and HPSG. The parser uses a packed-forest chart representation and builds constituents bottom-up using a best-first search strategy similar to A*, based on rule and lexical weights and the influences of the statistical preprocessing. The search terminates when a pre-specified number of spanning constituents have been found or a pre-specified maximum chart size is reached. The chart is then searched using a dynamic programming algorithm to find the least cost sequence of logical forms according to a cost table that can be varied by genre.

The TRIPS system uses a wide range of statistically driven preprocessing, including part of speech tagging, constituent bracketing, interpretation of unknown words using WordNet, and named-entity recognition (Allen et al, 2008). All these are generic off-the-shelf resources that extend and help guide the deep parsing process.

The TRIPS LF (logical form) ontology³ is designed to be linguistically motivated and domain independent. The semantic types and selectional restrictions are driven by linguistic considerations rather than requirements from reasoning components in the system (Dzikovska et al. 2003). As much as possible the semantic types in the LF ontology are compatible with types found in FrameNet (Johnson & Fillmore 2000). FrameNet generally provides a good level of abstraction for applications since the frames are derived from corpus examples and can be reliably distinguished by human annotators. However TRIPS parser uses a smaller, more general set of semantic roles for linking the syntactic and semantic arguments

³ TRIPS ontology browser:

<http://www.cs.rochester.edu/research/trips/lexicon/browse-ont-lex.html>

rather than FrameNet's extensive set of specialized frame elements. The LF ontology defines approximately 3000 semantic types and 30 semantic roles. The TRIPS parser will produce LF representations in terms of this linguistically motivated ontology.

As an example, the result of parsing the sentence, *He fought in the war*, is expressed as set of expressions in an unscoped logical formalism with reified events and semantic roles.

```
(SPEECHACT V1 SA-TELL :CONTENT V2)
(F V2 (:* FIGHTING FIGHT) :AGENT V3 :MODS
(V4) :TMA ((TENSE PAST)))
(PRO V3 (:* PERSON HE) :CONTEXT-REL HE)
(F V4 (:* SITUATED-IN IN) :OF V2 :VAL V5)
(THE V5 (:* ACTION WAR))
```

The main event (V2) is of ontology type FIGHTING, which is a subclass of INTENTIONAL-ACTION, and which corresponds to the first WordNet sense of fight, and includes verbs such as *fight*, *defend*, *contend* and *struggle*. The AGENT role of this event is the referent of the pronoun “he”, and the event is situated in an event described by the word “war”. For words not in the TRIPS core lexicon, the system looks up the WordNet senses and maps them to the TRIPS ontology. The word *war* is not in the core lexicon, and via WordNet is classified into the TRIPS ontology as the abstract type ACTION.

2.2 TRIOS System

The TRIOS system (TRIPS Temporal Reasoning System), uses the deep language understanding of TRIPS in combination with machine learning techniques to extract events, temporal expressions, their linguistic features, and relations in the text.

For event extraction, we take the TRIPS Logical Form (LF) and apply around hundred of hand-coded extraction patterns to extract events and features, by matching semantic patterns of phrases.

For instance, given the logical form of the sentence *he fought in the war*, shown above, we then use a set of hand-built extraction patterns, which match logical form expressions in terms of the ontology types, to identify relevant events. Because of the ontology, we can usually express general rules that capture a wide range of events. For instance, all noun-phrases describing objects that fall under the TRIPS Ontology's top-level type SITUATION-ROOT are extracted as described events. This situation will be captured in our extraction rule as:

```
((THE ?x (? type SITUATION-ROOT))
-extract-noms>
(EVENT ?x (? type SITUATION-ROOT) :pos NOUN
:class OCCURRENCE ))
```

Since WAR is ontology type ACTION, which falls under SITUATION ROOT in TRIPS ontology, this extraction rule will semantically match with (THE V5 (:* ACTION WAR)) and will extract WAR as event. Beside matching WAR under SITUATION ROOT in ontology, it also

matches the specifier THE, which explains it is a definite form of a noun phrase.

The result of matching around hundred of such rules to the sentence above is:

```
<EVENT eid=V2 word=FIGHT
pos=VERBAL ont-type=FIGHTING
class=OCCURRENCE tense=PAST
voice=ACTIVE aspect=NONE
polarity=POSITIVE
nf-morph=NONE>
<RLINK eventInstanceID=V2
ref-word=HE
ref-ont-type=PERSON
relType=AGENT>
<SLINK signal=IN
eventInstanceID=V2
subordinatedEventInstance=V5
relType=SITUATED-IN>
<EVENT eid=V5 word=WAR pos=NOUN
ont-type=ACTION
class=OCCURRENCE
voice=ACTIVE
polarity=POSITIVE
aspect=NONE tense=NONE>
```

Readers can see some differences (in bold) from usual TimeML annotations; these differences are our new suggestions that we will describe in next sections. A brief description of these new additions will be described in later sections.

TimeBank has an inter-annotator agreement (IAA)⁴ of 78% (average of precision and recall on subset of 10 documents) on event extraction. Our system performed with 57.85% precision and 84.68% recall (average 71.265%) on TimeBank. With an additional filtering step⁵, we can improve our precision to 83.32% with ~13% decrease in recall. This TRIPS parser and extraction rule with filtering system gives us a 77.22% average of precision and recall, which is comparable to TimeBank's inter annotator agreement and it is the state-of-the-art performance on event extraction in TimeBank corpus. Details of our system on Event Extraction can be found in UzZaman and Allen (2010a).

For temporally annotating new documents, this second level filtering might be useful. But for the task of extending TimeBank, we refer to TimeBank annotations anyway. So, we skip second level filtering to keep higher recall, so that we can add our additional features and relations for more event instances.

⁴ Inter-annotator agreement (IAA) on subset of 10 documents from TimeBank 1.2. TempEval annotation for EVENT and TIMEX3 were taken verbatim from TimeBank 1.2 (Verhagen et al 2007).

IAA source:
<http://www.timeml.org/site/timebank/documentation-1.2.html#iaa>

⁵ We implemented a MLN classifier to classify TRIOS events into TimeBank events and wrong extraction. This extra step was to remove generics and wrong extraction from TRIOS generated output.

On the other hand, for temporal expression extraction, we build a system by making a hybrid between traditional machine learning classifier and TRIPS parser extractor. We used a token-by-token classification for temporal expressions represented by B-I-O encoding with a set of lexical and syntactic features, using Conditional Random Field classifier⁶. Separately, TRIPS parser extracts temporal expressions the same way as we extract events. The performance of TRIPS parser's temporal extraction alone doesn't outperform state-of-the-art techniques on the evaluation measures. However, we have found that TRIPS extracts some temporal expressions that are missed by our CRF based system and even sometimes missed by TimeBank annotators. At the same time, the TRIPS based extractor is backed by a domain independent semantic parser, so including it will make it easier to port in other domains for future. So eventually we implemented a system by making a hybrid between CRF based system and TRIPS suggestion.

The temporal expressions that are suggested by the TRIPS parser but are missed by CRF based system are passed to a filtering step that tries to extract type (type can be TIME, DATE, DURATION or SET) and a normalized value (specific values of date or time or duration, or set) of the temporal expression. If we can find a normalized value and type, we accept these temporal expressions along with CRF based system's extracted temporal expressions.

Our system on temporal expression for strict match⁷ has an average of precision and recall 82.47%, compare to TimeBank's IAA 83%; for relaxed match⁸ we get 90.97% compared to TimeBank's IAA 96%. Details of our temporal expression extraction module can be found in UzZaman and Allen (2010b).

3. Extensions to TimeBank and TimeML

3.1 Suggesting new events in TimeBank

The low inter-annotator agreement of TimeBank suggests that there should be some effort to refine TimeBank events. It is hard to automatically suggest that some annotated event in TimeBank is wrong; so we focus on suggesting new events that are missing in TimeBank.

The TimeML (Pustejovsky et al, 2003b) specification says not to tag generic interpretations, even though capturing them could be of use in question answering. By generics, they mean, events that are not positioned in time, or in relation to other temporally located events in the document. For example, they won't annotate use and travel in the sentence:

Use of corporate jets for political travel is legal.

⁶ We used off the shelf CRF++ implementation. <http://crfpp.sourceforge.net/>

⁷ Strict match admits recognition when both strings are strictly matched.

⁸ Relaxed match admits recognition as long as there are any common words.

It also suggests not tagging subordinate verbs that express events which are clearly temporally located, but whose complements are generics. For example, *He said students are prohibited from fighting with each other*. Even though the verb said is temporally located, it isn't tagged because its complement, *students are prohibited from fighting with each other*, is generic.

And finally an event nominalization that doesn't provide any extra information than the supplied verbs, are also not tagged.

Many of the extra events generated by TRIPS parser with extraction rule that are not in TimeBank fall into these categories. We made a decision to only suggest verbal event, so we don't have to worry about the last instance. For verbal events, our task would be to keep the events that match with TimeML specification.

The extra TRIPS events that don't exist in TimeBank could be categorized as follows: (i) the result of wrong parse, (ii) a generic event and (iii) a legitimate event but missed by annotators.

Here are few examples that we think are legitimate events and also missed by TimeBank annotators:

1. *At least one of the sensitive sites was a barracks of the elite Republican Guard, a well-placed source **told** The Associated Press.*
2. *Net interest income for the third quarter **declined** to \$35.6 million from \$70.1 million a year ago.*
3. *About \$518 million of debt is **affected**.*
4. *If Iraq **chooses** a simple war of nerves and economic attrition, the Bush administration **knows** a long stalemate could try the patience of the American public and the West in general, and could **open** the possibility that moderate Arabs -- even including Saudi Arabia -- might **drop** out of the effort against Iraq and **accept** some deal from Saddam Hussein.*
5. *"It's the whole uncertainty about what's happening around us," **said** Valentin Von Korff, a trader at Credit Suisse First Boston in Frankfurt.*

Example 1, 2, 3 are obvious events that are missed by the annotators. There are some other cases where CONDITIONAL events are skipped by some annotators (example 4), however there is specific SLINK (relation between two events) in TimeML to handle CONDITIONAL relations. Some events with modality, which might not be specifically temporally located, because these are modal verbs, are also missed or neglected by some annotators. TimeML suggests for modality as event feature, so it is clear they want to include it as well.

On the other hand, we mentioned already, TimeML suggests not annotating subordinate verbs that express events which are clearly temporally located, but whose complements are generics. However, in this particular case we differ with TimeML, because we think these subordinate verbs being temporally located plays a role in overall temporal structure. For example, *He said the earth is round. People killed him*. Here **killing** was particularly

related to the event of **saying** something. If we don't keep **saying** as an event then the final temporal structure will be missing an important component of the story.

There are also events like **knows** in example 4, which is generic event according to guideline because it doesn't change throughout the document. However, many people might consider keeping these events as well.

We are interested in suggesting these legitimate events and filtering out wrong events. In case of generic events, we also want to identify which generic events are really generic and which of them are just discarded because they are true throughout the document. We first use the TRIOS system to extract events from text. We also get the POS tag for the events from Stanford POS tagger. We keep the TRIOS events for further processing that are suggested as Verbal events by both TRIOS and Stanford tagger. We only keep verbal events for new suggestions, because we have a higher accuracy for Verbal events (~92% accuracy) than other events (~75% accuracy). Now we take our filtered events and compare with TimeBank events to identify the possible extra events. We then classify these events as correct or not using an MLN classifier. The suggestions we get as correct events are then reviewed by humans to add legitimate events to the corpus. Flowchart for suggesting new TRIOS events is shown in Figure 1.

To train the system to identify the extra new events, we

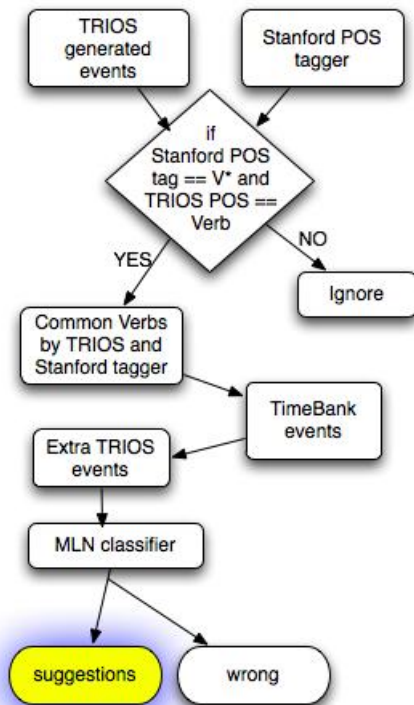


Figure 1: Flowchart for suggesting new events to TimeBank

train our system on TimeBank corpus for correct events

TimeML Spec	Our Analysis	Number	Performance
Correct	Correct	90	60%
	Generic complement	5	3.3%
Generic	Generic	28	18.8%
	True throughout	8	
	Others	2	1.3%
Wrong	Wrong	13	8.7%
	Not appropriate	3	2%

Table 1: Performance on suggested events

and train all extra TRIOS (TRIOS – TimeBank) events, which includes the generic and wrong events, as wrong events. In our MLN based classifier we used surface features capturing the current word, previous word, next word, contains a dollar sign, and combination of length range and penn tag (from Stanford tagger); also some grammatical features like pos, previous pos, next pos, tense, aspect; and some features involving semantic information like ont-type, RLINK related features (discussed later in detail), etc.

On complete TimeBank corpus, the system suggests around 150 new events. Both of the authors reviewed these suggestions individually and then jointly decided about these extra suggestions. According to TimeML guidelines we extracted extra 90 new events. But there are many other events that we think should be included, e.g. event whose complement is generics, but the main event is not (5 new events). TimeML also has a guideline that if some event doesn't change throughout the document should be counted as a generic. We also identify these, because we think these should be added as well and others might be interested in them too. There are also true generic events, which we also annotate. We removed the wrong suggestions and events that are not appropriate⁹ from text.

3.2 Adding temporal expressions in TimeBank

We mentioned before (Section 2.2), we extract temporal expressions from raw text by making a hybrid between CRF-based engine and TRIPS extraction. For suggesting extra temporal expressions, we consider both systems, but do it slightly differently. We get the temporal expressions suggested by both systems and compare with TimeBank. Then we process the extra temporal expressions in a filtering step, which tries to extract the normalized value for the temporal expressions. If the normalization module gets a normalized value then we consider that temporal

⁹ We parse all text and try to ignore titles, which are not annotated. But the system sometime parse the titles and extracts events. So these are correct events but not appropriate.

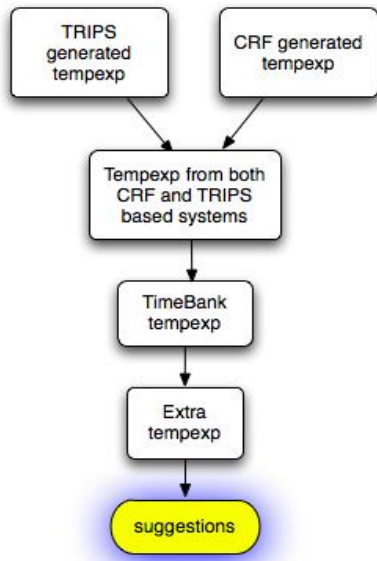


Figure 2: Suggesting new temporal expressions

expression as suggestion. The flowchart for suggesting new temporal expressions is shown in Figure 2.

The inter-annotator agreement for temporal expression identification is 96%, which means this is comparatively easy task for human annotators and we don't expect the annotators to miss many temporal expressions. On full TimeBank we suggested around 68 new temporal expressions, out of which, we found 50 (73.5%) temporal expressions to be legitimate. We added these new temporal expressions to the corpus. Some examples are shown below.

(1) *At the end of the broadcast this evening, one more trip around Havana to see what it's been like since **the last time**.*

(2) *And in **just a moment** Diane Sawyer will have some other news.*

(3) *And even terrorist groups that opposed Iraq in its war with Iran show signs of swinging behind Saddam Hussein **now** that he is in a confrontation with the U. S. And Iraq still has thousands of Americans and other Westerners under its control in Iraq and Kuwait.*

(4) *Turks feel they have special ties to the whole region, which they ruled for **hundreds of years** during the Ottoman Empire.*

(5) *In **the first days** after President Bush announced the dispatching of U.S. troops, they note, the Iraqi leader made several nationwide addresses indirectly -- having them read by a television announcer.*

(6) *Weisfield's, based in Seattle, Wash., **currently** operates 87 specialty jewelry stores in nine states.*

(7) ***Previously**, watch imports were denied such duty-free treatment.*

While the annotators didn't miss any obvious dates, they missed some temporal expressions like *now*, *currently*,

last time, *previously*, which are identified by our system and suggested as new temporal expressions. Such temporal expressions, although they have no specific temporal location as dates, helps to capture better temporal structure and are also annotated in TimeBank in general. Hence we want to suggest these new extra temporal expressions.

3.3 Adding ontology type as new event features in TimeML

TimeML comprises of event features, class, tense, aspect, nf-morph, pos, modality, and polarity. TRIOS system generates these features and also adds ontology type as event feature. The ontology type is the semantic type of word, particular word sense in the context, in TRIPS ontology¹⁰. TimeML tries to capture event information by very high-level class or pos. The ontology type feature captures more fine-grained information about the event, but in higher level than event word. Ontology type instances from our initial example are, FIGHTING for *fight* and ACTION for word *war*. Few other words with ontology type FIGHTING would be: *contend*, *defend*, and *struggle*, i.e. these words with similar meaning will get the same ontology type, in this case FIGHTING.

The TRIPS Ontology is available for public use, so people can use the ontology for their system. It also has mappings to WordNet, so can be connected to other lexical resources.

The accuracy of ont-type depends on parsing, and so will contain some errors. However, we have used our ont-type features in different classification sub-tasks in the next steps and it has improved our performance. To get the maximum benefit of this feature, we would need a much larger temporally annotated corpus, which is not available yet.

3.4 Adding improved relations in TimeML

Our next contribution is adding a richer set of relations to TimeML. TimeML captures the relations between different events with TLINK (temporal links), SLINK (subordinate link), and ALINK (aspectual link).

3.4.1. More SLINK instances

SLINKs or Subordinate Links are used for relations between two events. TimeML classify SLINKs into *Modal*, *Factive*, *Counter-Factive*, *Evidential*, *Negative evidential* and *Conditional*. This classification leaves out many instances where two events are related with each other, i.e. one event is argument of another event. We try to capture all possible relations when one event is related with another event. In following three examples from TimeBank corpus, we make one events in each sentence bold and another underlined. The bold event is the core

¹⁰ TRIPS ontology browser:

<http://www.cs.rochester.edu/research/trips/lexicon/browse-ont-lex.html>

event and the underlined is the reference event and the relType is noted in bracket afterwards.

(1) *Integra, which owns and operates hotels, **said** that Hallwood Group Inc. has agreed to exercise any rights that aren't exercised by other shareholders. (Theme)*

(2) *"They have to **continue** to tighten their belts," said Craig Kloner, an analyst at Goldman, Sachs amp Co. (Purpose)*

(3) *By mid afternoon, official Serb sources were saying the operation was over, but that has not yet been confirmed from Belgrade, the capital of Serbia, which is where the whole **attack** is thought to have been planned. (Theme)*

We try to capture all these relations as SLINK and the relation type will be the semantic role (or thematic roles). The most common semantic roles (relation types) found in the corpus are shown in Table 3 along with comparison with equivalent semantic roles from VerbNet and Lirics.

Our Role	VerbNet equivalents	Lirics equivalents	SLINK Count	RLINK Count
Agent	Agent, Actor	Agent	19	709
Theme	Stimulus, Theme	Theme	336	1137
Affected	Patient	Patient	13	92
Cause	Cause	Cause		49
Goal_as_Loc	Destination	finalLocation	47	
To_Loc	Recipient	Goal	46	
At_Loc	Location	Location	42	
In_Loc	Location	Location	28	
On	Location	Location	20	
Situated_in	Location?	Location?	39	
Purpose	--	Purpose	226	

Table 3: Most common relTypes used in SLINKs and RLINKs

There are other kind of SLINKs that we consider. Another instance from our initial example (*He fought in the war*) is:

```
<SLINK signal=IN
  eventInstanceID=V2
  subordinatedEventInstance=V5
  relType=SITUATED-IN>
```

We also try to capture the signal (connectives, that connects two events). The problem in these cases is identifying the relation type (relType). We decided to use

the ontology type of our signal (connective) as the relType for these kinds of extra SLINKs.

We suggest around 900 SLINKs to TimeBank corpus. Table 3 shows the statistics of most frequent SLINK types that we suggested.

3.4.2. New Relation Link, RLINK

Many researchers (Chambers et al, 2007), (Katsumasa et al, 2009) showed that having dependency information improves the performance for extracting temporal relations. They tried to capture the dependency relation with dependency parsers like Stanford dependency parser. This gives a hint that capturing how other dependent words are connected with the event will enrich the information about event.

We introduce new relation link, RLINK, to capture what other objects are related to the event (other than another event, which is captured with SLINKs), i.e. relation of event with its arguments.

In our initial example, for event FIGHT, we try to capture the information that the AGENT of that fighting event is HE, which is a PERSON. These relations give us information what are the arguments of an event and how they are connected.

```
<RLINK eventInstanceID=V2
  ref-word=HE
  ref-ont-type=PERSON
  relType=AGENT>
```

We considered the thematic/semantic roles that described in Table 3. In TRIOS-TimeBank corpus we suggest around 2000 RLINKs. We showed the distribution in Table 3.

Another example of RLINK's importance could be explained with (Chambers and Jurafsky, 2008). They learned narrative event chains considering the idea of protagonist (central actor). They are basically considering the events performed by same agent. We are trying to capture agent and other different thematic roles (or semantic roles) using RLINK, which would help many other applications like (Chambers and Jurafsky, 2008).

One argument against adding RLINK is, these information could be annotated with other layers like syntactic, semantic layers. We agree that this information could be added using other layers, however, for building a complete temporally aware system we would need these information. Hence to build a complete temporally annotated corpus it is better to have these information as well. This would benefit in advanced applications like question answering, summarization, etc, that would be using the temporal structure.

3.5 Building The TRIOS-TimeBank corpus

Our final task is to include all these new information in a new version of TimeBank. In case of event features, we keep TimeBank features for existing events and add additional ont-type features for events that we extract (in

84.68% cases) as well. For missing events (TimeBank – TRIOS events), we repeat *class* value as *ont-type* value. For new suggested events, we use features generated by TRIOS system. We distinguish these events with feature tag *source*, which has value *timebank-event* (TimeBank – TRIOS events), *trios-event* (TRIOS – TimeBank events, i.e. new suggestions) and *timebank-trios-event* (exists in both TimeBank and TRIOS, shares TRIOS generated ont-type feature). For *trios-event*, we also have another feature to distinguish legitimate events (according to TimeML guideline) with the *generic event*. Finally in *generic events*, we annotate if the event is true generic event or just discarded by TimeBank annotaters because it was true throughout the document.

We keep the existing temporal expressions as they are with additional feature tag *timebank-timex*. We add new temporal expressions with feature tag *trios-timex*. We also generate TimeBank features for new event and temporal expression suggestions. Finally we add extra SLINK and RLINKs in the document as well.

The TimeBank corpus (Pustejovsky et al 2003b) is annotated according to TimeML specification. Later in TempEval (Temporal Evaluation contest) (Verhagen et al 2007), the same corpus was released with modified event relations and minor modifications on some event features. Since TempEval contains the same documents as TimeBank, but more recently published with updated features, we used TempEval corpus, instead of old TimeBank corpus, as our base for our TRIOS-TimeBank corpus. Our newly developed corpus is available online¹¹.

4. Future Work

This is a work in progress and we plan modifications to create an even better temporal annotated corpus. We are already working on a few issues which we haven't finished yet, but will be available in our next release. All our release and updates will be available in TRIOS-TimeBank homepage¹¹.

Suggesting Temporal Relations (TLINKs)

The TRIPS parser extracts the temporal relations between events and temporal expressions and events and events. TimeBank has TLINKs (TempEval 1 contest, Task A) between event and temporal expressions, but doesn't have intra-sentence event-event temporal relations. However, a new task in TempEval 2 includes intra-sentence event-event temporal realtions. Thus, such relations are in the documents of TempEval 2 contest, but are missing from TempEval 1 (the TimeBank corpus).

We plan to do two tasks in future:

- (1) Suggest new event-time TLINKs that are missed by TimeBank, as we did for events and temporal expressions.
- (2) Suggest intra-sentence event-event temporal relations (TLINKs), which asre ignored in TimeBank.

¹¹ TRIOS-TimeBank corpus is available online at: <http://www.cs.rochester.edu/u/naushad/trios-timebank-corpus>

Removing Wrong Events and Temporal Expressions

We suggested new events or temporal expressions in this paper, but ignored the task of removing incorrect events that are annotated by TimeBank annotaters. Our accuracy of event extraction is ~82% and temporal expression extraction is ~90%. It means that we are left with only small number of events and temporal expressions that we missed and TimeBank annotaters suggested. A preliminary examination reveals that some events should be removed to match the strict TimeML criterias. We plan to go over these and remove wrong events or temporal expressions to make it even cleaner corpus.

Building a Large TimeML-based Corpora

We plan to continue to improve the TRIPS parser and other machine learning tools in order to extract better events, temporal expressions, generate better event features and relations. With these new tools, we plan to automatically temporally annotate a significant corpus of texts (more newswire text and also other domains) and build a much larger temporally annotated corpus for the community.

5. Conclusion

In this paper, we presented the TRIOS-TimeBank corpus, an extended TimeBank corpus with suggestions for new events and temporal expressions. We also proposed an extension to TimeML language with richer event features, and event relations, all of which we generated with help of deep understanding of text using semantic parsing and some machine learning tools. This resource, the TRIOS-TimeBank corpus, with newly suggested events, event feature and relations and temporal expressions, is available to the community for further research on temporal reasoning.

6. Acknowledgements

This work was supported in part by the National Science Foundation, grant #0748942, and the Office of Naval Research (N000140510314). We thank Mary Swift and William DeBeaumont for help with different TRIPS parser related issues, Benjamin van Durme for many useful suggestions. and Sebestian Reidel for help on using the very useful MLN tool *TheBeast*. We are also very thankful to all three reviewers for useful reviews.

7. References

- James Allen, Mary Swift, and Will de Beaumont. 2008. Deep semantic analysis of text. Symposium on Semantics in Systems for Text Processing (STEP), Venice, Italy, 2008.
- Nathanael Chambers, Shan Wang, and Daniel Jurafsky. 2007. Classifying temporal relations between events. Proceedings of the 45th Annual Meeting on Association for Computational Linguistics. Association of Computational Linguistics, 2007.
- Nathanael Chambers and Daniel Jurafsky. 2008.

- Unsupervised Learning of Narrative Event Chains. Proceedings of the 46th Annual Meeting on Association for Computational Linguistics. Association of Computational Linguistics, 2008.
- Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto. 2009. Jointly Identifying Temporal Relations with Markov Logic. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009.
- James Pustejovsky, Jos M. Castao, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 2834. AAAI Press, 2003.
- J. Pustejovsky, P. Hanks, R. Saur, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003. The TIMEBANK corpus. *Corpus Linguistics*, 2003, Lancaster, March 2003.
- Naushad UzZaman and James Allen, 2010. Extracting Event and Event Features from Text, *In Progress*, 2010.
- Naushad UzZaman and James Allen, 2010. Extracting Temporal Information from Text, *In Progress*, 2010.
- Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz and James Pustejovsky, 2007. SemEval-2007 Task 15: TempEval Temporal Relation Identification, Proceedings of *4th International Workshop on Semantic Evaluations (SemEval 2007)*.