

TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles

Hector Llorens, Estela Saquete, Borja Navarro-Colorado

Natural Language Processing Group

University of Alicante

{hlllorens, stela, borja}@dlsi.ua.es

Abstract

This paper analyzes the contribution of semantic roles to TimeML event recognition and classification. For that purpose, an approach using conditional random fields with a variety of morphosyntactic features plus semantic roles features is developed and evaluated. Our system achieves an F1 of 81.4% in recognition and a 64.2% in classification. We demonstrate that the application of semantic roles improves the performance of the presented system, especially for nominal events.

1 Introduction

Event recognition and classification has been pointed out to be very important to improve complex natural language processing (NLP) applications such as automatic summarization (Daniel et al., 2003) and question answering (QA) (Pustejovsky, 2002). Natural language (NL) texts often describe sequences of events in a time line. In the context of summarization, extracting such events may aid in obtaining better summaries when these have to be focused on specific happenings. In the same manner, the access to such information is crucial for QA systems attempting to address questions about events.

The analysis of events as well as the classification of the different forms they adopt in NL text is not a new issue (Vendler, 1967). It relates not only to linguistics but different scientific areas such as philosophy, psychology, etc.

In NLP, different definitions of event can be found regarding the target application.

On the one hand, in topic detection and tracking (Allan, 2002), event is defined as an instance of a topic identified at document level describing something that happen (e.g., “wars”). The aim of

this task is to cluster documents on the same topic, that is to say, the same event.

On the other hand, information extraction (IE) provides finer granularity event definitions. IE proposes standard schemes to annotate the individual events within the scope of a document. STAG scheme (2000) was aimed to identify events in news and their relationship with points in a temporal line. More recently, TimeML (Pustejovsky et al., 2003a) presented a rich specification for annotating events in NL text extending the features of the previous one.

This paper is focused on the TimeML view of events. TimeML defines events as situations that *happen or occur*, or elements describing *states or circumstances* in which something obtains or holds the truth. These events are generally expressed by tensed or untensed verbs, nominalizations, adjectives, predicative clauses or prepositional phrases. TimeML guidelines define seven classes of events:

- *Reporting*. Action of a person or organization declaring or narrating an event (e.g., “say”)
- *Perception*. Physical perception of another event (e.g., “see”, “hear”)
- *Aspectual*. Aspectual predication of another event (e.g., “start”, “continue”)
- *IAction*. Intensional action (e.g., “try”)
- *IState*. Intensional state (e.g., “feel”, “hope”)
- *State*. Circumstance in which something holds the truth (e.g., “war”, “in danger”)
- *Occurrence*. Events that describe things that happen (e.g., “erupt”, “arrive”)

The following sentence shows an example of an occurrence event and a state event.

```
It's <EVENT class="OCCURRENCE">turning</EVENT>  
out to be another <EVENT class="STATE">bad</EVENT>  
financial week.
```

The automatic annotation of events has been addressed with different data-driven approaches. Current approaches are mainly based on morphosyntactic information. Our hypothesis is that semantic roles, as higher language level analysis information, may be useful as additional feature to improve the performance of such approaches.

Within this setting, the main objective of this paper is to analyze (1) the contribution of semantic roles, as additional feature, and (2) the influence of conditional random fields (CRFs), as machine learning (ML) technique, in the events automatic recognition and classification task.

This paper is structured as follows. Firstly, related work in the task is reviewed in Section 2. The next section provides a detailed description of our proposal to address event recognition and classification. After that, Section 4 includes an evaluation of the proposal, and a comparative analysis of the results. Finally, conclusions are drawn in Section 5.

2 Related Work

There is only one corpus available annotated with TimeML events: TimeBank (Pustejovsky et al., 2003b). Hence, all the approaches regarding TimeML events extraction have been evaluated using this corpus.

EVITA system (Saurí et al., 2005) recognizes events by combining linguistic and statistical techniques. The main features used to manually encode event recognition rules are the following: part-of-speech (PoS) tagging, lemmatizing, chunking, lexical lookup and contextual parsing. Furthermore, WordNet information combined with Bayesian learned disambiguation was used to identify nominal events. EVITA obtained 74.03% precision, 87.31% recall, and 80.12% $F_{\beta=1}$ in event recognition over TimeBank.

Boguraev and Ando (2005) present an evaluation on automatic TimeML events annotation. They set out the task as a classification problem and used a robust risk minimization (RRM) classifier to solve it. The $F_{\beta=1}$ results obtained by a 5-fold cross validation over TimeBank were 78.6% for recognition and 61.3% for classification. Moreover, they evaluated the impact of applying word-profiling techniques over their ap-

proach to exploit unannotated data. Using this additional information, the $F_{\beta=1}$ results improved to 80.3% and 64.0%. In this evaluation, neither precision nor recall were given.

STEP (Bethard and Martin, 2006) is a system for TimeML event recognition and classification. This approach uses a rich set of textual, morphological, dependency and WordNet hypernymy features to build a Support Vector Machine (SVM) model. The model was trained using 9/10 of the TimeBank. The test, carried out using the remaining 1/10 of the corpus, obtained a 82.0% precision, 70.6% recall and 75.9% $F_{\beta=1}$ for recognition and a 66.7% precision, 51.2% recall and 57.9% $F_{\beta=1}$ for classification.

Finally, March and Baldwin (2008) present an evaluation on event recognition using a multi-class classifier (BSVM). The main features used to train the classifier are word and PoS context window, stop words removal and feature generalization through words grouping (numbers, named entities, etc.). The result for the best feature combination in a 10-fold cross validation over TimeBank was 76.4% $F_{\beta=1}$.

It is worth mentioning that there are two versions of the TimeBank corpus, 1.1 and 1.2. The latest version is the current gold standard. Both versions consist of the same documents¹, mainly news articles and transcribed broadcast news from different domains. EVITA is the only reference which used TimeBank 1.2 while the rest of reviewed references used TimeBank 1.1.

3 Our proposal: semantic roles enhancing a CRF model

In this section, the motivation for our proposal, and our specific approach are presented.

3.1 Motivation

The next two subsections describe the main feature (semantic roles) and the ML algorithm (CRFs) we selected to address event recognition and classification; and the reasons why we think they could be useful in that task.

¹Except 3 documents removed in TimeBank 1.2

3.1.1 Semantic roles

Semantic role labeling (SRL) has achieved important results in the last years (Gildea and Jurafsky, 2002). For each predicate in a sentence, semantic roles identify all constituents, determining their arguments (agent, patient, etc.) and their adjuncts (locative, temporal, etc.). Currently, there exist different role sets aimed to cover opposed requirements. They range from more specific, such as FrameNet (Baker et al., 1998), to more general like PropBank (Palmer et al., 2005). Figure 1 illustrates a semantic role labeled sentence.

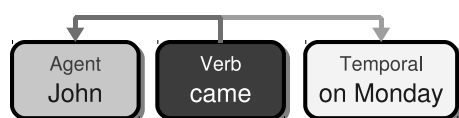


Figure 1: Semantic roles example

Many research efforts into the application of semantic roles demonstrated that this information is useful for different NLP purposes (Melli et al., 2006). Focusing on TimeML, semantic roles have been applied to temporal expressions recognition (Llorens et al., 2009), and temporal links classification (Hagège and Tannier, 2007). However, they have not been used to recognize and classify TimeML events.

Semantic roles provide structural relations of the predicates in which events may participate. Beyond syntactic relations expressed by means of the different types of phrases, semantic roles give further information about semantic relations between the arguments of a predicate. Therefore, as richer information, roles may better distinguish tokens to be candidate events. In addition, different semantic role settings may represent specific event classes.

Example 1 shows four sentences annotated with PropBank semantic roles (in square brackets) in which the noun “control” participates. In the sentences 1 and 2, “control” does not represent an event, while in the sentences 3 and 4, it represents an *state* event. It can be seen that the noun “control”, when it is contained by A1 role it may represent an event. However, it is not an event when contained by A0 or AM-MNR roles. The analysis may also take into account the governing

verb. In the example, we could specify that “control” represents an event when contained by A1 role of “seek” and “obtain” verbs; and the opposite for the A0 role of “emerge” and the AM-MNR of “had”.

- (1) 1. “[Control procedures A0] will emerge”
2. “[Iraq A0] had [thousands of Americans A1] [under its control AM-MNR]”
3. “[Crane Co. A0] may obtain [control of Milton Roy Corp. A1]”
4. “[Pattison’s A0] decided to seek [control A1]”

Our hypothesis is that semantic roles, as additional information, may help in the recognition and classification of events. The information about the role of a token and the verb it depends on, or the set of roles of the sentence, could be useful for determining whether a token or a sequence of tokens is an event or not. Due to the fact that roles represent high level information in NL text, they are more independent from word tokens. Hence, roles may aid in learning more general models that could improve the results of approaches focused on lower level information.

3.1.2 CRF probabilistic model

Conditional Random Fields is a popular and efficient ML technique for supervised sequence labeling (Lafferty et al., 2001). CRFs are undirected graphical models, a special case of conditionally-trained finite state machines. A key advantage of CRFs is their flexibility to include a wide variety of arbitrary, non-independent features of the input.

We see the task set out in this paper as a sequence labeling problem. Assume X is a random variable over data sequences to be labeled, and Y is a random variable over the corresponding label sequences (hidden), being all Y components (Y_i) members of a finite label alphabet γ . X might range over NL sentences and Y range over event annotations of those sentences, with γ the set of possible event IOB2² labels. The following example illustrates the event recognition problem.

(2)	X	Y	
	was	?	
	another	?	B-EVENT
	bad	?	? = I-EVENT
	week	?	O

²IOB2 format: (B)egin, (I)nside, and (O)utside

The variables X and Y are jointly distributed over both label and observation sequences. However, unlike Hidden Markov Models (generative) in which $p(X, Y)$, CRFs (discriminative) construct a conditional model from paired observation and label sequences: $p(Y|X)$. Graphically, CRFs are represented by undirected graphs, $G = (V, E)$ such that $Y = (Y_v)$, $v \in V$, so that Y is indexed by the vertices of G . Then (X, Y) is a *conditional random field* if Y_v variables obey the *Markov property* with respect to the graph when conditioned on X :

$$P(Y_v|X, Y_w, v \neq w) = P(Y_v|X, Y_w, v \sim w),$$

where $v \sim w$ means that Y_v and Y_w are connected neighbors in G .

To extend the problem to event classification, the alphabet γ must be extended with the event classes (state, aspectual, etc.).

CRFs have been successfully applied to many sequence labeling tasks (Sha and Pereira, 2003; McCallum and Li, 2003).

From our point of view, the task addressed in this paper is well suited for this ML technique. Events may depend on structural properties of NL sentences. Not only the word sequence, but morphological, syntactic and semantic information is related with the event structure (Tenny and Pustejovsky, 2000).

For example, sequences of verbs may represent *i.action+occurrence* or *aspectual+occurrence* events (see Example 3).

(3) “The president will <EVENT class="i.action"> try </EVENT> to <EVENT class="occurrence"> assist </EVENT> to the <EVENT class="occurrence"> conference </EVENT>”

“Saddam will <EVENT class="aspectual"> begin </EVENT> <EVENT class="occurrence"> withdrawing </EVENT> troops from Iranian territory on Friday”

In addition, for instance, many *state* event instances are represented by “to be” plus a variable quality (see Example 4).

(4) “It is <EVENT class="occurrence"> turning </EVENT> out to be another <EVENT class="state"> bad </EVENT> financial week.”

Given this analysis, our hypothesis is that CRFs will be useful in the recognition of events in which the sequential and structural properties are relevant.

3.2 Approach description

This paper proposes CRFs as learning method to infer an event recognition and classification model. Our system includes CRF++ toolkit³ for training and testing our approach. The learning process was done using *CRF-L2* algorithm and hyper-parameter $C=1$.

The definition of the features is crucial for the architecture of the system. The features used in our approach are grouped in two feature sets. On the one hand, general features, which comprise morphosyntactic and ontological information. On the other hand, semantic roles features, which are the main focus of this paper.

The general features used to train our CRF model are described regarding each language analysis level.

- **Morphological:** The lemma and PoS context, in a 5-window (-2,+2), was employed. This basic linguistic feature showed good results in different NLP tasks, as well as in event recognition and classification (March and Baldwin, 2008). Tokenization, PoS and lemmatization were obtained using TreeTagger (Schmid, 1994).
- **Syntactic:** Different events are contained in particular types of phrases and syntactic dependencies. This feature tries to tackle this by considering syntactic information. Charniak parser (Charniak and Johnson, 2005) was used to obtain the syntactic tree.
- **Lexical semantics:** WordNet (Fellbaum, 1998) top ontology classes have been widely used to represent word meaning at ontological level, and demonstrated its worth in many tasks. We obtained the four top classes for each word.

The specific semantic roles features used to enhance the training framework of the CRF model were developed considering PropBank role set. PropBank was applied in our system due to the high coverage it offers in contrast to FrameNet. In order to get PropBank semantic roles, the CCG

³<http://crfpp.sourceforge.net/>

SRL tool (Punyakanok et al., 2004) was used for labeling the corpus.

- **Role:** For each token, we considered the role regarding the verb the token depends on. Semantic roles information may be useful for distinguish particular lemmas that are events only when appearing under a precise role.
- **Governing verb:** The verb to which the current token holds a particular role. This may distinguish tokens appearing under the influence of different verbs.
- **Role+verb combination:** The previous two features were combined to capture the relation between them. This introduces new classification information by distinguishing roles depending on different verbs. The importance of this falls especially on the numbered roles of PropBank (A0, A1, ...) holding different meanings when depending on different verbs.
- **Role configuration:** This consists of the set of roles depending on the verb the token depends on. This may be particularly useful for distinguish different sentence settings and thus, whether a token denotes an event in a particular sentence type.

The system consists of two main processes. Firstly, given TimeML annotated text, it obtains the defined features plus the IOB2 tags of the annotated events. Then, using this data the system learns (trains) a model for event recognition and a model for event classification. Secondly, given plain text, it automatically gets the defined features using the described tools. With this data, the system applies the learned models to recognize and classify TimeML events.

4 Evaluation

In this section, firstly, the corpus, criteria and measures are defined. Secondly, the results obtained by our approach are presented. After that, the contribution of our approach is measured through different experiments: (1) general contribution, (2) semantic roles contribution, and (3) CRFs contribution. And finally, our approach is compared to the state of the art systems.

4.1 Corpus, criteria and measures

For the evaluation, the TimeBank 1.2 corpus (7881 events) was used without modification. All the results reported in this evaluation were obtained using a 5-fold cross validation. The n-fold train-test sets were built sorting the corpus files alphabetically and then sequentially select each set regarding the documents size. It is important to highlight the latter because if the n-folds were made regarding the number of documents, the sets had not been homogeneous due to the differences in TimeBank document sizes.

Only annotations matching the exact event span were considered as *correct* in recognition and classification, requiring also the class matching in the second case.

The following measures were used to score the evaluated approaches.

- Precision $\frac{\text{correct_annotations}}{\text{total_approach_annotations}}$
- Recall $\frac{\text{correct_annotation}}{\text{total_corpus_annotations}}$
- $F_{\beta=1} \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$

4.2 Our approach results

Table 1 shows the results obtained by our approach for both recognition and classification of events. The last column (BF) indicates the best $F_{\beta=1}$ results obtained in the individual folds.

	Precision	Recall	$F_{\beta=1}$	BF
<i>Recognition</i>	83.43	79.54	81.40	82.43
<i>Classification</i>	68.84	60.15	64.20	69.68

Table 1: Our approach (CRF+Roles) results

The results show a high $F_{\beta=1}$ score in both recognition and classification, showing a good balance between precision and recall. This indicates that our approach is appropriate to address this task.

Focusing on classification task, Table 2 shows the detailed scores for each event class.

Looking at the specific class results, *reporting* obtained the best results. This is due to the fact that 80% of *reporting* events are represented by lemmas “say” and “report” with PoS “VBD” and “VBZ”. *Occurrence*, *perception*, *aspectual* and *i_state* obtained classification results over 50%.

Class (instances)	Precision	Recall	$F_{\beta=1}$
<i>Reporting</i> (1021)	91.90	89.18	90.51
<i>Perception</i> (48)	65.93	66.83	66.37
<i>Aspectual</i> (258)	81.35	47.00	59.57
<i>I.Action</i> (673)	51.40	29.30	37.32
<i>I.State</i> (582)	68.44	43.70	53.34
<i>State</i> (1107)	50.01	24.84	33.19
<i>Occurrence</i> (4192)	66.73	72.07	69.29

Table 2: CRF+Roles 5-fold detailed results

Although *perception* and *aspectual* are quite restricted to some lemmas, they obtained results below *reporting*. This is due to the fact that TimeBank contains very few examples of these classes. *I.action* and *state* show poorer results. In the case of the former, this is because some non-intensional verbs (e.g., “look”) appear in the corpus as *i.action* under certain conditions, for example, when there is modality or these verbs appear in conditional sentences. This suggests the necessity of incorporating a word sense disambiguation (WSD) technique. Our approach did not take into account this information and thus the results are lower for this event class. In the case of *state*, the reasons for the low performance are the richness of this event class by means of lemmas, PoS, and phrases.

Finally, Table 3 shows the results of our approach by word class.

		Precision	Recall	$F_{\beta=1}$
Recognition	<i>Verb</i>	91.56	92.15	91.33
	<i>Noun</i>	72.67	48.26	58.42
	<i>Adj.</i>	66.78	38.09	48.35
Classification	<i>Verb</i>	73.86	74.21	73.51
	<i>Noun</i>	62.73	41.33	49.53
	<i>Adj.</i>	55.69	31.12	40.41

Table 3: CRF+Roles 5-fold word class results

It may be seen that the best results in both recognition and classification are obtained in verb events, followed by noun and adjective.

4.3 Contribution analysis

This subsection details the contribution of each aspect of our approach through three comparative experiments.

First experiment: general contribution

This experiment measures the general contribu-

tion of our approach by comparing its results with a baseline. TimeBank was analyzed to find a basic general rule to annotate events. The events are mainly denoted by verbs, pertaining to *occurrence* class. Hence, we propose a baseline that annotates all verbs as *occurrence* events. Table 4 shows results obtained by this baseline for both recognition and classification of events.

		Prec.	Recall	$F_{\beta=1}$
Our approach	<i>Recog.</i>	83.43	79.54	81.40
	<i>Class.</i>	68.84	60.15	64.20
Baseline	<i>Recog.</i>	72.50	65.20	68.60
	<i>Class.</i>	46.01	53.19	49.34

Table 4: Our approach vs Baseline results

Given the simplicity of the baseline, the results obtained are quite high. However, our approach $F_{\beta=1}$ significantly improves baseline by 19% for recognition and 30% for classification.

Second experiment: roles contribution

The main objective of this paper is to determine the impact of semantic roles in this task. To quantify it, a non-roles version of our approach was evaluated. This version only uses the general features described in section 3. Table 5 shows the results obtained.

		Precision	Recall	$F_{\beta=1}$
Our approach	<i>Recog.</i>	83.43	79.54	81.40
	<i>Class.</i>	68.84	60.15	64.20
Non-roles	<i>Recog.</i>	82.96	74.81	78.67
	<i>Class.</i>	67.53	54.80	60.50

Table 5: Our approach vs Non-roles results

Comparing these results with the ones obtained by our full featured approach, the application of roles improved especially the recall. Specifically, recall improved by 6% and 10% for recognition and classification respectively. The main improvement was achieved by *state* and *occurrence* classes (60% of the total improvement), especially, nominal events of that classes that concentrate around the 70% of the total contribution.

To illustrate corpus examples that have been improved by roles, Example 5 shows two sentences containing state events that were correctly tagged by the roles approach and missed by the

non-roles. In the examples, the TimeML events annotation and below the semantic roles annotation is reported.

- (5) “There are still few buyers and the mood is <EVENT class=STATE>gloomy</EVENT>”
 “[There A0] are [still AM-TMP] [few buyers A1] and [the mood A0] is [gloomy AM-MNR]”

“Security is now <EVENT>better</EVENT>”
 “[Security A0] is [now AM-TMP] [better AM-MNR]”

In these cases, AM-MNR role information lead to a correct *state* event recognition.

Third experiment: CRFs contribution

In order to measure the CRFs contribution to this task, an extra experiment was carried out. This consisted of comparing, under the same setting, CRFs with a popular learning technique: support vector machines (SVM). As in Bethard and Martin (2006), YamCha⁴ software was used (parameters: $C=1$ and $\text{polynomial_degree}=2$).

Table 6 shows the results obtained by the SVM-based approach in recognition and Table 7 reports the improvement (CRFs over SVM) distribution in the different word classes.

	Precision	Recall	$F_{\beta=1}$
Our approach (CRF)	83.43	79.54	81.40
SVM	80.00	75.10	77.40

Table 6: Our approach (CRF) vs SVM results

	Verb	Noun	Adj.	Adv.	Prep.
General	22%	71%	5%	1%	1%

Table 7: CRF improvement distribution among the word classes

These results verify that CRF improves SVM $F_{\beta=1}$ by 5% in this task. Furthermore, especially noun events take advantage of using CRF.

Finally, Figure 2 illustrates the results of our approach over the described experiments.

4.4 Comparison with the state of the art

Most systems found in the literature are data-driven approaches using morphosyntactic features. SVM based approaches (Bethard and Martin, 2006; March and Baldwin, 2008) achieved,

⁴<http://chasen.org/~taku/software/YamCha/>

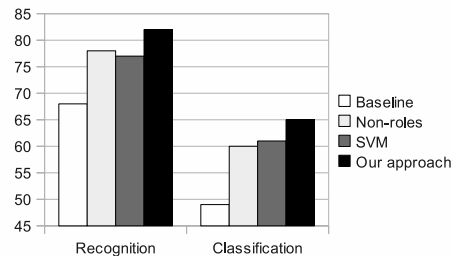


Figure 2: $F_{\beta=1}$ Results

approximately, 76% and 58% $F_{\beta=1}$ in event recognition and classification respectively. Boguraev and Ando (2005) used a robust risk minimization classifier to address this task and obtained 78.6% and 61% (without exploiting unannotated data). These results are very similar to the ones obtained by our non-roles approach. This suggests that using, apart from morphosyntactic features, additional features based on semantic roles could improve the approaches.

EVITA system (Saurí et al., 2005) combines linguistic and statistical techniques. On the one hand, it consists of a set of manually encoded rules based on morphosyntactic information. On the other hand, it includes a Bayesian learned disambiguation module to identify nominal events. The later was trained and tested using the whole corpus, therefore, the results could be inflated by this fact. For that reason, Bethard and Martin (2006) presented an EVITA implementation (Sim-Evita) to compare the results. Sim-Evita obtains an 73% and 51% $F_{\beta=1}$ in event recognition and classification respectively. These results suggest that data-driven improve rule-based approaches.

Only STEP evaluation showed detailed classification results. We agree that *state* events are the most complex and heterogeneous ones. Focusing on such events, our $F_{\beta=1}$ results (33%) improve Bethard’s (25%) by 32%. Regarding the results obtained for each word class. Bethard’s results presented good performance on classifying verb events (71%), but lower results in noun events (34%). Our approach results for noun events (49%) improve theirs by 44%. This suggests that the application of semantic roles enables our approach on making more general predictions. In this manner, our system may recog-

nize unseen nominal event instances as long as they share, with the seen instances, some semantic roles features.

5 Conclusions and Further Work

This paper presented an approach for the recognition and classification of TimeML events consisting of a CRF model learned using semantic roles as main feature. In addition to morphosyntactic features, the model was enhanced including extra semantic information, semantic role labeling, used for other applications with satisfactory results, but never employed before for this purpose. Our proposal was evaluated using the gold standard corpus, TimeBank 1.2, and the results obtained were analyzed and compared to measure the impact of both semantic roles and CRFs in the described task.

The obtained $F_{\beta=1}$ results demonstrated that semantic roles are useful to recognize (81.43%) and classify (64.20%) TimeML events, improving the presented baseline by 19% for recognition and 30% for classification. Specifically, Semantic roles employed as additional feature improved the recall of the non-roles version by 6% and 10% for recognition and classification respectively. This indicates that roles features led to more general models capable of better annotating unseen instances. The roles contribution was more significant in *state* and *occurrence* classes of noun events, concentrating around the 70% of the improvement.

Furthermore, it was verified that CRFs achieve higher results than models learned using other ML techniques such as SVM (5% improvement), contributing especially to nominal events. This demonstrated that CRF models are appropriate to face the task.

Finally, to the extent our results are comparable to state of the art evaluations, ours outperform the $F_{\beta=1}$ scores in both recognition and classification. Especially, our approach showed better performance than related works in *state* (32% improvement) and nominal events (44% improvement). Hence, the extension of the current approaches with semantic roles features could benefit their performance.

The main difficulties found in the task ad-

ressed in this paper are related to *i_action* and *state* events. In the former, we detected that modality and the word senses are important and must be treated to distinguish such events. In the later, although they were improved by our approach, *state* events are still the most complex class of events due to their richness in contrast to the reduced size of the training data. We agree with related literature that event classification results are still below other tasks performance, which indicates that this task is inherently complex and more training data may lead to significant improvements.

As further work we propose, firstly, improving the *i_action* results by taking into account the modality considering the *AM-MOD* role, and the word senses using a WSD technique. Secondly, the application of FrameNet role set (finer granularity) to determine which kind of roles are better to improve the current event annotation systems.

Acknowledgments

This paper has been supported by the Spanish Government, projects TIN-2006-15265-C06-01, TIN-2009-13391-C04-01 and PROMETEO/2009/119, where Hector Llorens is funded (BES-2007-16256).

References

- Allan, James. 2002. *Topic Detection and Tracking: Event-Based Information Organization*. Kluwer Academic Publishers, Norwell, MA, USA.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *COLING-ACL*, pages 86–90.
- Bethard, Steven and James H. Martin. 2006. Identification of event mentions and their semantic class. In *EMNLP: Proceedings of the Conference on Empirical Methods in NLP*, pages 146–154. ACL.
- Boguraev, Branimir and Rie Kubota Ando. 2005. Effective Use of TimeBank for TimeML Analysis. In *Annotating, Extracting and Reasoning about Time and Events 05151*.
- Charniak, Eugene and Mark Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *43rd Annual Meeting of the ACL*.
- Daniel, Naomi, Dragomir Radev, and Timothy Allison. 2003. Sub-event based multi-document summariza-

- tion. In *HLT-NAACL Text summarization workshop*, pages 9–16. ACL.
- Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. MIT Press.
- Gildea, Daniel and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- Hagège, Caroline and Xavier Tannier. 2007. XRCE-T: XIP temporal module for TempEval campaign. In *TempEval (SemEval)*, pages 492–495, Prague, Czech Republic. ACL.
- Lafferty, John D., Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th ICML*, pages 282–289. Morgan Kaufmann.
- Llorens, Hector, Borja Navarro, and Estela Saquete. 2009. Using Semantic Networks to Identify Temporal Expressions from Semantic Roles. In *VI RANLP*, pages 219–224.
- March, Olivia and Timothy Baldwin. 2008. Automatic event reference identification. In *ALTA 2008*, pages 79–87, Australia.
- McCallum, Andrew and Wei Li. 2003. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *HLT-NAACL*, pages 188–191.
- Melli, G., Y. Liu Z. Shi, Y. Wang, and F. Popowich. 2006. Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2006 Summarization Task. In *DUC*.
- Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31.
- Punyakanok, Vasin, Dan Roth, W. Yih, D. Zimak, and Y. Tu. 2004. Semantic role labeling via generalized inference over classifiers. In *HLT-NAACL (CoNLL)*, pages 130–133. ACL.
- Pustejovsky, James, José M. Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003a. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *IWCS-5*.
- Pustejovsky, James, Patrik Hanks, Roser Saurí, A. See, Robert Gaizauskas, Andrea Setzer, Dragomir R. Radev, Beth Sundheim, David Day, Lisa Ferro, and M. Lazo. 2003b. The TIMEBANK Corpus. In *Corpus Linguistics*, pages 647–656.
- Pustejovsky, James. 2002. TERQAS: Time and Event Recognition for Question Answering Systems. In *ARDA Workshop*.
- Saurí, Roser, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. Evita: A robust event recognizer for qa systems. In *HLT/EMNLP*. ACL.
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49.
- Setzer, Andrea and Robert Gaizauskas. 2000. Annotating Events and Temporal Information in Newswire Texts. In *LREC 2000*, pages 1287–1294.
- Sha, Fei and Fernando C. N. Pereira. 2003. Shallow parsing with conditional random fields. In *HLT-NAACL*.
- Tenny, Carol and James Pustejovsky. 2000. *Events as Grammatical Objects. The Converging Perspectives of Lexical Semantics and Syntax*. CSLI.
- Vendler, Zeno, 1967. *Linguistics and philosophy*, chapter Verbs and times, pages 97–121. Cornell University Press, Ithaca, NY.