Temporal Signals Help Label Temporal Relations

Leon Derczynski and Robert Gaizauskas
Natural Language Processing Group
Department of Computer Science
University of Sheffield
211 Portobello, S1 4DP, Sheffield, UK
{leon,robertg}@dcs.shef.ac.uk

Abstract

Automatically determining the temporal order of events and times in a text is difficult, though humans can readily perform this task. Sometimes events and times are related through use of an explicit co-ordination which gives information about the temporal relation: expressions like *"before"* and *"as soon as"*. We investigate the rôle that these co-ordinating temporal signals have in determining the type of temporal relations in discourse. Using machine learning, we improve upon prior approaches to the problem, achieving over 80% accuracy at labelling the types of temporal relation between events and times that are related by temporal signals.

1 Introduction

It is important to understand time in language. The ability to express and comprehend expressions of time enables us to plan, to tell stories, and to discuss change in the world around us.

When we automatically extract temporal information, we are often concerned with events and times – referred to collectively as temporal **intervals**. We might ask, for example, *"Who is the current President of the USA?."* In order to extract an answer to this question from a document collection, we need to identify events related to persons becoming president and the times of those events. Crucially, however, we also need to identify the temporal relations between these events and times, perhaps, for example, by recognizing a temporal relation type from a set such as that of Allen (1983). This last task, **temporal relation typing**, is challenging, and is the focus of this paper.

Temporal **signals** are words or phrases that act as discourse markers that co-ordinate a pair of events or times and explicitly state the nature of the temporal relation that holds between them. For example, in *"The parade reached the town hall before noon"*, the word *before* is a temporal signal, co-ordinating the event *reached* with the time *noon*. Intuitively, these signal words act as discourse contain temporal ordering information that human readers can readily access, and indeed this hypothesis is borne out empirically (Bestgen and Vonk, 1999). In this paper, we present an in-depth examination into the role temporal signals can play in machine learning for temporal relation typing, within the framework of TimeML (Pustejovsky et al., 2005).

2 Related Work

Temporal relation typing is not a new problem. Classical work using TimeML is that of Boguraev and Ando (2005), Mani et al. (2007) and Yoshikawa et al. (2009). The TempEval challenge series features relation typing as a key task (Verhagen et al., 2009). The take-home message from all this work is that temporal relation typing is a hard problem, even using advanced techniques and extensive engineering – approaches rarely achieve over 60% on typing relations between two events or over 75% accuracy for those between an event and a time. Recent attempts to include more linguistically sophisticated features representing discourse, syntactic and semantic role information have yielded but marginal improvements, e.g. Llorens et al. (2010); Mirroshandel et al. (2011).

Although we focus solely on determining the *types* of temporal relations, one must also identify which pairs of temporal intervals should be temporally related. Previous work has covered the tasks of identifying and typing temporal relations jointly with some success (Denis and Muller, 2011; Do et al., 2012). The TempEval3 challenge addresses exactly this task (UzZaman et al., 2013).

Investigations into using signals for temporal relation typing have had promising results. Lapata and Lascarides (2006) learn temporal structure according to these explicit signals, then predict temporal orderings in sentences without signals. As part of an early TempEval system, Min et al. (2007) automatically annotate signals and associate them with temporal relations. They then include the signal text as a feature for a relation type classifier. Their definition of signals varies somewhat from the traditional TimeML sig-

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 645–650,
Sofia, Bulgaria, August 4-9 2013. ©2013 Association for Computational Linguistics

	Event-event relations			Event-time relations		
	Non-signalled	Signalled	Overall	Non-signalled	Signalled	Overall
Baseline most-common-class	41.4%	57.4%	43.0%	49.2%	51.6%	49.6%
Maxent classifier	57.7%	58.6%	57.8%	81.4%	59.6%	77.3%
Error reduction	*27.8%*	*2.74%*	*25.4%*	*64.5%*	*16.4%*	*55.5%*
Sample size (number of relations)	3 179	343	3 522	2 299	529	2 828

Table 1: Relation typing performance using the base feature set, for relations with and without a temporal signal.

nal definition, as they include words such as *reporting* which would otherwise be annotated as an event. The system achieves a 22% error reduction on a simplified set of temporal relation types.

Later, Derczynski and Gaizauskas (2010) saw a 50% error reduction in assignment of relation types on signalled relation instances from introducing simple features describing a temporal signal's interaction with the events or times that it co-ordinates. The features for describing signals included the signal text itself and the signal's position in the document relative to the intervals it co-ordinated. This led to a large increase in relation typing accuracy to 82.19% for signalled event-event relations, using a maximum entropy classifier.

Previous work has attempted to linguistically characterise temporal signals (Brée et al., 1993; Derczynski and Gaizauskas, 2011). Signal phrases typically fall into one of three categories: monosemous as temporal signals (e.g. *"during"*, *"when"*); bisemous as temporal or spatial signals (e.g. *"before"*); or polysemous with the temporal sense a minority class (e.g. *"in"*, *"following"*). Further, a signal phrase may take two arguments, though its arguments need not be in the immediate content and may be anaphoric. We leave the task of automatic signal annotation to future work, instead focusing on the impact that signals have on temporal relation typing.

Our work builds on previous work by expanding the study to include relations other than just event-event relations, by extending the feature set, by doing temporal relation labelling over a more carefully curated version of the TimeBank corpus (see below), and by providing detailed analysis of the performance of a set of labelling techniques when using temporal signals.

3 Experimental Setup

We only approach the relation typing task, and we use existing signal annotations – that is, we do not attempt to automatically identify temporal signals.

The corpus used is the signal-curated version of TimeBank (Pustejovsky et al., 2003). This corpus, TB-sig,[1] adds extra events, times and relations to TimeBank, in an effort to correct signal under-annotation in the original corpus (Derczynski and Gaizauskas, 2011). Like the original TimeBank corpus, it comprises 183 documents. In these, we are interested only in the temporal relations that use a signal. There are 851 signals annotated in the corpus, co-ordinating 886 temporal relations (13.7% of all). For comparison, TimeBank has 688 signal annotations which co-ordinate 718 temporal relations (11.2%).

[1] See http://derczynski.com/sheffield/resources/tb_sig.tar.bz2

When evaluating classifiers, we performed 10-fold cross-validation, keeping splits at document level. There are only 14 signalled time-time relations in this corpus, which is not enough to support any generalizations, and so we disregard this interval type pairing.

As is common with statistical approaches to temporal relation typing, we also perform relation folding; that is, to reduce the number of possible classes, we sometimes invert argument order and relation type. For example, A BEFORE B and B AFTER A convey the same temporal relation, and so we can remove all AFTER-type relations by swapping their argument order and converting them to BEFORE relations. This lossless process condenses the labels that our classifier has to distinguish between, though classification remains a multi-class problem.

We adopt the base feature set of Mani et al. (2007), which consists mainly of TimeML event and time annotation surface attributes. These are, for events: `class`, `aspect`, `modality`, `tense`, `polarity`, `part of speech`; and, for times: `value`, `type`, `function in document`, `mod`, `quant`. To these are added `same-tense` and `same-aspect` features, as well as the string values of events/times.

The feature groups we use here are:

- **Base** – The attributes of TimeML annotations involved (includes tense, aspect, polarity and so on as above), as with previous approaches.
- **Argument Ordering** – Two features: a boolean set if both arguments are in the same sentence (as in Chambers et al. (2007)), and the text order of argument intervals (as in Hepple et al. (2007)).
- **Signal Ordering** – Textual ordering is important with temporal signals; compare *"You walk before you run"* and *"Before you walk you run"*. We add features accounting for relative textual position of signal and arguments as per Derczynski and Gaizauskas (2010). To these we add a feature reporting whether the signal occurs in first, last, or mid-sentence position, and features to indicate whether each interval is in the same sentence as the signal.
- **Syntactic** – We add syntactic features: following Bethard et al. (2007), the lowest common constituent label between each argument and

Features	Classifier	Event-event accuracy	Event-time accuracy
N/A	Baseline most-common-class	57.4%	51.6%
Base	Baseline maximum entropy	58.6%	59.6%
DG2010	Maximum entropy	72.6%	72.4%
	Random forest	76.7%	78.6%
All	Adaptive boosting	70.4%	73.0%
	Naïve Bayes	73.8%	71.5%
	Maximum entropy	75.5%	78.1%
	Linear SVC / Crammer-Singer	79.3%	75.6%
	Linear SVC	80.7%	77.1%
	Random forest	**80.8%**	**80.3%**

Table 2: Results at temporal relation typing over TB-sig, for relations that use a temporal signal

the signal; following Swampillai and Stevenson (2011), the syntactic path from each argument to the signal, using a top-level `ROOT` node for cross-sentence paths; and three features indicating whether there is a temporal function tag (`-TMP` between each of the intervals or the signal to the root note. These features are generated using the Stanford parser (Klein and Manning, 2003) and a function tagger (Blaheta and Charniak, 2000).

- **Signal Text** – We add the signal's raw string, as well as its lower-case version and its lemma.
- **DCT** – For event-time relations, whether the time expression also functions as the document's creation timestamp.

Collectively, these feature groups comprise the **All** feature set. For comparison, the feature set we reported in previous work (Derczynski and Gaizauskas, 2010) is also included, labeled **DG2010**. This set contains the base and the signal ordering feature groups only, plus a single signal feature for the signal raw string.

Using these feature representations we trained multinomial naïve Bayes (Rennie et al., 2003), maximum entropy (Daumé III, 2008), adaptive boosting (Freund and Schapire, 1997; Zhu et al., 2009), multi-class SVM (Crammer and Singer, 2002; Chang and Lin, 2011) and random forest[2] (Breiman, 2001) classifiers via Scikit-learn (Pedregosa et al., 2011).

We use two baselines: most-common-class and a model trained with no signal features. We also introduce two measures replicating earlier work: one using the DG2010 features and the classifier used in that work (maximum entropy), and another using the DG2010 features with the best-performing classifier under our All feature set, in order to see if performance changes are due to features or classifier.

Classifiers were evaluated by determining if the class they output matched the relation type in TB-sig. Results are given in Table 2. For comparison with the general case, i.e. for both signalled and non-signalled temporal relation instances, we list performance with a maximum entropy classifier and the base feature set on TB-sig's temporal relations. Results are in Table 1. These are split into those that use a signal and those that do not, though no features relaying signal information are included.

In order to assess the adequacy of the dataset in terms of size, we also examined performance using a maximum entropy classifier learned from varying subproportions of the training data. This was measured over event-event relations, using all features. Results are given in Figure 1. That performance appears to stabilise and level off indicates that the training set is of sufficient size for these experiments.

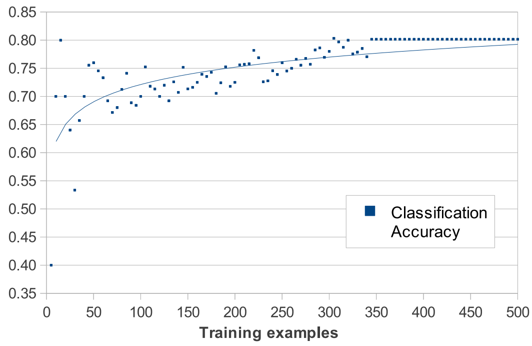

Figure 1: Effect of training data size on relation typing performance.

4 Analysis

The results in Table 2 echo earlier findings and intuition: temporal signals are useful in temporal relation typing. Results support that signals are not only helpful in event-event relation typing but also event-time typing. For comparison, inter-annotator agreement across all temporal relation labels, i.e. signalled and non-signalled relations, in TimeBank is 77%.

Using the maximum entropy classifier, our approach gives a 2.9% absolute performance increase over the DG2010 feature set for event-event relations (10.6% error reduction) and a 5.7% absolute increase for event-time relations (20.7% error reduction). Random forests

[2] With $n_{estimators} = 200$, a minimum of one sample per node, and no maximum depth.

Feature sets	Evt-evt	Evt-time
All	80.8%	80.3%
All-argument order	80.8%	78.3%
All-signal order	79.0%	77.5%
All-syntax	79.2%	79.6%
All-signal text	**70.8%**	**72.7%**
All-DCT	79.9%	79.4%
Base	54.2%	53.9%
Base+argument order	56.8%	60.1%
Base+signal order	59.7%	65.0%
Base+syntax	70.0%	**71.0%**
Base+signal text	**75.5%**	66.3%
Base+DCT	54.2%	53.9%
Base+signal text+signal order	80.4%	76.9%
Base+signal text+syntax	79.0%	74.1%
Base+arg order+signal order	77.8%	75.2%

Table 3: Relation typing accuracy based on various feature combinations, using random forests. Bold figures indicate the largest performance change.

offer better performance under both feature sets, with the extended features achieving notable error reduction over DG2010 – 17.6% for event-event, 7.9% for event-time relations. Linear support vector classification provided rapid labelling and comparable performance for event-event relations but was accuracy was not as good as random forests for event-time relation labelling.

Note, figures reported earlier in Derczynski and Gaizauskas (2010) are not directly comparable to the DG2010 figures reported here, as here we are using the better-annotated TB-sig corpus, which contains a larger and more varied set of temporal signal annotations.

Although we are only examining the 13.7% of temporal relations that are co-ordinated with a signal, it is important to note the performance of conventional classification approaches on this subset of temporal relations. Specifically, the error reduction relative to the baseline that is achieved without signal features is much lower on relations that use signals than on non-signalled relations (Table 1). Thus, temporal relations that use a signal appear to be more difficult to classify than other relations, unless signal information is present in the features. This may be due to differences in how signals are used by authors. One explanation is that signals may be used in the stead of temporal ordering information in surrounding discourse, such as modulations of dominant tense or aspect (Derczynski and Gaizauskas, 2013).

Unlike earlier work using maxent, we experiment with a variety of classifiers, and find a consistent improvement in temporal relation typing using signal features. With the notable exception of adaptive boosting, classifiers with preference bias (Liu et al., 2002) – AdaBoost, random trees and SVC – performed best in this task. Conversely, those tending toward the independence assumption (naïve Bayes and maxent) did not capitalise as effectively on the training data.

Features	Evt-evt	Evt-time
All	80.8%	80.3%
All-signal text	70.8%	72.7%
All-signal text-argument order	70.7%	72.2%
All-signal text-signal order	69.5%	71.2%
All-signal text-syntax	**59.5%**	**69.0%**
All-signal text-DCT	70.8%	72.8%

Table 4: Feature ablation without signal text features. Bold figures indicate largest performance change.

We also investigated the impact of each feature group on the best-performing classifier (random forests with $n = 200$) through feature ablation. Results are given in Table 3. Ablation suggested that the signal text features (signal string, lower case string, head word and lemma) had most impact in event-event relation typing, though were second to syntax features in event-time relations. Removing other feature groups gave only minor performance decreases.

We also experimented with adding feature groups to the base set one-by-one. All but DCT features gave above-baseline improvement, though argument ordering features were not very helpful for event-event relation typing. Signal text features gave the strongest improvement over baseline for event-event relations, but syntax gave a larger improvement for event-time relations. Accordingly, it may be useful to distinguish between event-event and event-time relations when extracting temporal information using syntax (c.f. the approach of Wang et al. (2010)).

A strong above-baseline performance was still obtained even when signal text features were removed, which included the signal text itself. This was interesting, as signal phrases can indicate quite different temporal orderings (e.g. *"Open the box while it rains"* vs. *"Open the box before it rains"*, and the words used are typically critical to correct interpretation of the temporal relation. Further, the model is able to generalise beyond particular signal phrase choices. To investigate further, we examined the performance impact of each group sans "signal text" features (Table 4). In this case, removing the syntactic features had the greatest (negative) impact on performance, though the absolute impact on event-event relations (a drop of 11.3%) was far lower than that on event-time relations (3.7%).

To examine helpful features, we trained a maxent classifier on the entire dataset and collected feature:value pairs. These were then ranked by their weight. The ten largest-weighted pairings for event-event relations (the hardest problem in overall temporal relation typing) are given in Table 5. Prefixes of 1- and 2- correspond to the two interval arguments (events). Negative values are those where the presence of a particular feature:value pair suggests the mentioned class is not applicable.

Weight	Feature	Value	Class
9.346	2-polarity	POS	ENDS
-8.713	1-2-same-sent	True	BEGINS
-7.861	2-aspect	NONE	BEGINS
-7.256	1-aspect	NONE	INCLUDES
6.564	2-sig-synt-path	NN-NP-IN	INCLUDES
6.519	signal-lower	before	ENDS
-6.294	2-tense	NONE	BEGINS
-5.908	2-modality	None	ENDS
5.643	2-text	took	BEGINS
-5.580	1-modality	None	ENDS

Table 5: Top ten largest-weighted feature:value pairs.

It can be seen that BEGINS and INCLUDES relationships are not indicated if the arguments have no TimeML aspect assigned; this is what one might expect, given how aspect is used in English, with these temporal relation types corresponding to event starts and the progressive. Also, notice how a particular syntactic path, connecting adjacent nominalised event and the word *in* acting as a signal, indicate a temporal inclusion relationship. Temporal polysemy, where a word has more than one possible temporal interpretation, is also observable here (Derczynski and Gaizauskas (2011) examine this polysemy in depth). This is visible in how the temporal signal phrase *"before"* is not, as one might expect, a strong indicator of a BEFORE or even AFTER relation, but of an ENDS relationship.

5 Conclusion

This paper set out to investigate the rôle of temporal signals in predicting the type of temporal relation between two intervals. The paper demonstrated the utility of temporal signals in this task, and identified approaches for using the information these signals contain, which performed consistently better than the state-of-the-art across a range of machine learning classifiers. Further, it identified the impact that signal text, signal order and syntax features had in temporal relation typing of signalled relations.

Two directions of future work are indicated. Firstly, the utility of signals prompts investigation into detecting which words in a given text occur as temporal signals. Secondly, it is intuitive that temporal signals explicitly indicate related pairs of intervals (i.e. events or times). So, the task of deciding which interval pair(s) a temporal signal co-ordinates must be approached.

Although we have found a method for achieving good temporal relation typing performance on a subset of temporal relations, the greater problem of general temporal relation typing remains. A better understanding of the semantics of events, times, signals and how they are related together through syntax may provide further insights into the temporal relation typing task.

Finally, Bethard et al. (2007) reached high temporal relation typing performance on one a subset of relations (events and times in the same sentence); we reach high temporal relation typing performance on another subset of relations – those using a temporal signal. Identifying further explicit sources of temporal information applicable to new sets of relations may reveal promising paths for investigation.

Acknowledgements

The first author was supported by UK EPSRC grant EP/K017896/1, uComp (http://www.ucomp.eu/).

References

J. Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Y. Bestgen and W. Vonk. 1999. Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language*, 42(1):74–87.

S. Bethard, J. Martin, and S. Klingenstein. 2007. Timelines from text: Identification of syntactic temporal relations. In *Proceedings of the International Conference on Semantic Computing*, pages 11–18.

D. Blaheta and E. Charniak. 2000. Assigning function tags to parsed text. In *Proceedings of the meeting of the North American chapter of the Association for Computational Linguistics*, pages 234–240. ACL.

B. Boguraev and R. K. Ando. 2005. TimeBank-Driven TimeML Analysis. In G. Katz, J. Pustejovsky, and F. Schilder, editors, *Annotating, Extracting and Reasoning about Time and Events*, number 05151 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

D. Brée, A. Feddag, and I. Pratt. 1993. Towards a formalization of the semantics of some temporal prepositions. *Time & Society*, 2(2):219.

L. Breiman. 2001. Random forests. *Machine Learning*, 45(1):5–32.

N. Chambers, S. Wang, and D. Jurafsky. 2007. Classifying temporal relations between events. In *Proceedings of the 45th meeting of the Association for Computational Linguistics*, pages 173–176. ACL.

C.-C. Chang and C.-J. Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27.

K. Crammer and Y. Singer. 2002. On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research*, 2:265–292.

H. Daumé III. 2008. MegaM: Maximum entropy model optimization package. *ACL Data and Code Repository, ADCR2008C003*, 50.

P. Denis and P. Muller. 2011. Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1788–1793. AAAI Press.

L. Derczynski and R. Gaizauskas. 2010. Using Signals to Improve Automatic Classification of Temporal Relations. In *Proceedings of 15th Student Session of the European Summer School for Logic, Language and Information*, pages 224–231. FoLLI.

L. Derczynski and R. Gaizauskas. 2011. A Corpus-based Study of Temporal Signals. In *Proceedings of the Corpus Linguistics Conference*.

L. Derczynski and R. Gaizauskas. 2013. Empirical Validation of Reichenbach's Tense Framework. In *Proceedings of the 10th International Conference on Computational Semantics*, pages 71–82. ACL.

Q. X. Do, W. Lu, and D. Roth. 2012. Joint inference for event timeline construction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 677–687. ACL.

Y. Freund and R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

M. Hepple, A. Setzer, and R. Gaizauskas. 2007. USFD: preliminary exploration of features and classifiers for the TempEval-2007 tasks. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 438–441. ACL.

D. Klein and C. D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st meeting of the Association for Computational Linguistics*, pages 423–430. ACL.

M. Lapata and A. Lascarides. 2006. Learning sentence-internal temporal relations. *Journal of Artificial Intelligence Research*, 27(1):85–117.

Y. Liu, Y. Yang, and J. Carbonell. 2002. Boosting to correct inductive bias in text classification. In *Proceedings of the 11th international Conference on Information and Knowledge Management*, pages 348–355. ACM.

H. Llorens, E. Saquete, and B. Navarro. 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In *Proceedings of SemEval-2010*. ACL.

I. Mani, B. Wellner, M. Verhagen, and J. Pustejovsky. 2007. Three approaches to learning TLINKS in TimeML. Technical report, CS-07-268, Brandeis University.

C. Min, M. Srikanth, and A. Fowler. 2007. LCC-TE: A hybrid approach to temporal relation identification in news text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 219–222. ACL.

S. A. Mirroshandel, G. Ghassem-Sani, and M. Khayyamian. 2011. Using syntactic-based kernels for classifying temporal relations. *Journal of Computer Science and Technology*, 26(1):68–80.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830.

J. Pustejovsky, R. Sauri, R. Gaizauskas, A. Setzer, L. Ferro, et al. 2003. The TimeBank Corpus. In *Proceedings of the Corpus Linguistics Conference*, pages 647–656.

J. Pustejovsky, J. Castano, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, G. Katz, and D. Radev. 2005. TimeML: Robust specification of event and temporal expressions in text. In I. Mani, J. Pustejovsky, and R. Gaizauskas, editors, *The language of time: a reader*. Oxford University Press.

J. D. Rennie, L. Shih, J. Teevan, and D. Karger. 2003. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In *Proceedings of the International Conference on Machine Learning*. AAAI Press.

K. Swampillai and M. Stevenson. 2011. Extracting relations within and across sentences. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 25–32. ACL.

N. UzZaman, H. Llorens, L. Derczynski, M. Verhagen, J. F. Allen, and J. Pustejovsky. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. In *Proceedings of the 7th International Workshop on Semantic Evaluations*.

M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, and J. Pustejovsky. 2009. The TempEval challenge: identifying temporal relations in text. *Language Resources and Evaluation*, 43(2):161–179.

W. Wang, J. Su, and C. L. Tan. 2010. Kernel based discourse relation recognition with temporal ordering information. In *Proceedings of the 48th meeting of the Association for Computational Linguistics*, pages 710–719. ACL.

K. Yoshikawa, S. Riedel, M. Asahara, and Y. Matsumoto. 2009. Jointly identifying temporal relations with Markov logic. In *Proceedings of the International Joint Conference on Natural Language Processing*, pages 405–413. ACL.

J. Zhu, H. Zou, S. Rosset, and T. Hastie. 2009. Multi-class AdaBoost. *Statistics and Its Interface*, 2:349–360.