Projection-based Acquisition of a Temporal Labeller

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
`email@domain`

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
`email@domain`

Abstract

We present a cross-lingual projection framework for temporal annotations. Automatically obtained TimeML annotations in the English portion of a parallel corpus are transferred to the German translation along a word alignment. Direct projection augmented with shallow heuristic knowledge outperforms the uninformed baseline by 6.64% F_1-measure for events, and by 17.93% for time expressions. Subsequent training of statistical classifiers on the (imperfect) projected annotations significantly boosts precision by up to 31% to 83.95% and 89.52%, respectively.

1 Introduction

In recent years, supervised machine learning has become the standard approach to obtain robust and wide-coverage NLP tools. But manually annotated training data is a scarce and expensive resource. *Annotation projection* (Yarowsky and Ngai, 2001) aims at overcoming this resource bottleneck by scaling conceptually monolingual resources and tools to a multilingual level: annotations in existing monolingual corpora are transferred to a different language along the word alignment to a parallel corpus.

In this paper, we present a projection framework for *temporal annotations.* The TimeML specification language (Pustejovsky et al., 2003a) defines an annotation scheme for time expressions (*timex* for short) and events, and there are tools for the automatic TimeML annotation of English text (Verhagen et al., 2005). Similar rule-based systems exist for Spanish and Italian (Saquete et al., 2006). However, such resources are restricted to a handful of languages.

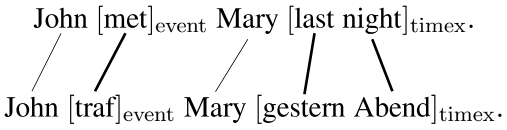


Figure 1: Annotation projection.

We employ the existing TimeML labellers to annotate the English portion of a parallel corpus, and automatically project the annotations to the word-aligned German translation. Fig. 1 shows a simple example. The English sentence contains an event and a timex annotation. The event-denoting verb *met* is aligned with the German *traf*, hence the latter also receives the event tag. Likewise, the components of the multi-word timex *last night* align with German *gestern* and *abend*, respectively, and the timex tag is transferred to the expression *gestern abend*.

Projection-based approaches to multilingual annotation have proven adequate in various domains, including part-of-speech tagging (Yarowsky and Ngai, 2001), NP-bracketing (Yarowsky et al., 2001), dependency analysis (Hwa et al., 2005), and role semantic analysis (Padó and Lapata, 2006). To our knowledge, the present proposal is the first to apply projection algorithms to temporal annotations.

Cross-lingually projected information is typically noisy, due to errors in the source annotations as well as in the word alignment. Moreover, successful projection relies on the *direct correspondence assumption* (*DCA*, Hwa et al. (2002)) which demands that the annotations in the source text be homomorphous with those in its (literal) translation. The DCA has been found to hold, to a substantial degree, for the above mentioned domains. The results we report here show that it can also be confirmed for temporal annotations in English and German. Yet, we cannot preclude *divergence* from translational correspondence; on the contrary, it occurs routinely and to a certain extent systematically (Dorr, 1994). We inquire two different techniques to filter noise. Firstly, the projection process is equipped with (partly language-specific) knowledge to account for alignment errors and cross-language discrepancies in the realisation of events and timexes. Secondly, we apply aggressive data engineering techniques to the noisy projections and use them to train statistical classifiers which generalise beyond the noise.

The paper is structured as follows. Sec. 2 gives an overview of the TimeML specification language and compatible annotation tools. Sec. 3 presents our projection models for temporal annotations, which are evaluated in sec. 4. Sec. 5 describes how we induce temporal labellers for German from the projected annotations; sec. 6 concludes.

2 Temporal Annotation

2.1 The TimeML Specification Language

The TimeML specification language (Pustejovsky et al., 2003a)[1] and annotation framework emerged from the TERQAS workshop[2] in the context of the ARDA AQUAINT programme. The goal of the programme is the development of question answering (QA) systems which index content rather than plain keywords. Semantic indexing based on the identification of named entities in free text is an established method in QA and related applications. Recent years have also seen advances in relation extraction, a variant of event identification, albeit restricted in terms of coverage: the majority of systems addressing the task use a pre-defined set of—typically domain-specific—templates. In contrast, TimeML models events in a domain-independent manner and provides principled definitions for various event classes. Besides the identification of *events*, it addresses their relative ordering and anchoring in time by integrating *timexes* in the annotation. The major contribution of TimeML is the explicit representation of dependencies (so-called *links*) between timexes and events.

Unlike traditional accounts of events (e.g., Vendler (1967)), TimeML adopts a very broad notion of eventualities as “situations that happen or occur” and “states or circumstances in which something obtains or holds true” (Pustejovsky et al., 2003a); besides verbs, this definition includes event nominals such as *accident*, and stative modifiers (*prepared*, *on board*). Events are annotated with `EVENT` tags. TimeML postulates seven event classes: REPORTING, PERCEPTION, ASPECTUAL, I-ACTION, I-STATE, STATE, and OCCURRENCE. For definitions of the individual classes, the reader is referred to Saurí et al. (2005b).

Explicit timexes are marked by the `TIMEX3` tag. It is modelled on the basis of Setzer’s (2001) `TIMEX` tag and the TIDES `TIMEX2` annotation (Ferro et al., 2005). Timexes are classified into four types: dates, times, durations, and sets.

Events and timexes are interrelated by three kinds of links: temporal, aspectual, and subordinating. Here, we consider only *subordinating links (slinks)*. Slinks explicate event modalities, which are of crucial importance when reasoning about the certainty and factuality of propositions conveyed by event-denoting expressions; they are thus directly relevant to QA and information extraction applications. Slinks relate events in modal, factive, counter-factive, evidential, negative evidential, or conditional relationships, and can be triggered by lexical or structural cues.

2.2 Automatic Labellers for English

The basis of any projection architecture are high-quality annotations of the source (English) portion of the parallel corpus. However, given that the projected annotations are to provide enough data for training a target language labeller (sec. 5), manual annotation is not an option. Instead, we use the TARSQI tools for automatic TimeML annotation of English text (Verhagen et al., 2005). They have been modelled and evaluated on the basis of the TimeBank (Pustejovsky et al., 2003b), yet for the most

---

[1] A standardised version ISO-TimeML is in preparation, cf. Schiffrin and Bunt (2006).

[2] See `http://www.timeml.org/site/terqas/index.html`

$e \in E$	temporal entity
$l \in E \times E$	(subordination) link
$w_s \in W_s, w_t \in W_t$	source/target words
$al \in Al : W_s \times W_t$	word alignment
$A_s \ni a_s : E \to 2^{W_s}$	source annotation
$A_t \ni a_t :$ $(E \times A_s \times Al) \to 2^{W_t}$	projected target annotation

Table 1: Notational conventions.

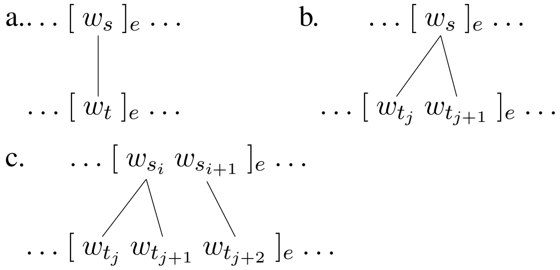


Figure 2: Projection scenarios: (a) single-word 1-to-1, (b) single-word 1-to-many, (c) multi-word.

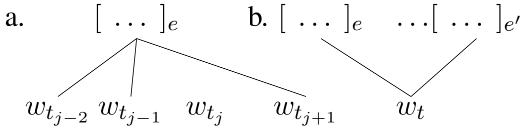


Figure 3: Problematic projection scenarios: (a) non-contiguous aligned span, (b) rivalling tags.

part rely on hand-crafted rules. To obtain a full temporal annotation, the modules are combined in a cascade. We are using the components for timex recognition and normalisation (Mani and Wilson, 2000), event extraction (Saurí et al., 2005a), and identification of modal contexts (Saurí et al., 2006).[3]

3 Informed Projection

3.1 The Core Algorithm

Recall that TimeML represents temporal entities with `EVENT` and `TIMEX3` tags which are anchored to words in the text. Slinks, on the other hand, are not anchored in the text directly, but rather relate temporal entities. The projection of links is therefore entirely determined by the projection of the entities they are defined on (see Table 1 for the notation used throughout this paper): a link $l = (e, e')$ in the source annotation a_s projects to the target annotation a_t iff both e and e' project to non-empty sequences of words. The projection of the entities e, e' themselves, however, is a non-trivial task. Given a temporal entity e covering a sequence $a_s(e)$ of tokens in the source annotation, the projection model needs to determine the extent $a_t(e, a_s, al)$ of e in the target annotation, based on the word alignment al. Possible projection scenarios are depicted in Fig. 2. In the simplest case (Fig. 2a), e spans a single word w_s which aligns with exactly one word w_t in the target sentence. In this case, the model predicts e to project to w_t. A single tagged word with 1-to-many alignments (as in Fig. 2b) requires a more thorough inspection of the aligned words. If they form a contiguous sequence, e can be projected onto the entire sequence as a multi-word unit. This is problematic in a scenario such as the one shown in Fig. 3a, where the aligned words do *not* form a contiguous sequence. There are various strategies, described in sec. 3.2, to deal with non-contiguous cases. For the moment, we can adopt a conservative approach which categorically blocks discontinuous projections. Finally, Fig. 2c illustrates the projection of an entity spanning multiple words. Here, the model composes the projection span of e from the alignment contribution of each individual word w_s covered by e. Again, the final extent of the projected entity is required to be contiguous.

With any of these scenarios, a problem arises when two distinct entities e and e' in the source annotation have conflicting projection extents, that is, when $a_t(e, a_s, al) \cap a_t(e', a_s, al) \neq \emptyset$. This is illustrated in Fig. 3b. The easiest strategy to resolve conflicts like these is to pick an arbitrary entity and privilege it for projection to the target word(s) w_t in question. All other rivalling entities e' project onto their remaining target words $a_t(e', a_s, al) \setminus \{w_t\}$.

Pseudocode for this word-based projection of

---

[3] TARSQI also comprises a component that introduces temporal links (Mani et al., 2003); we are not using it here because the output includes the entire tlink closure. Although Mani et al. (2006) use the links introduced by closure to boost the amount of training data for a tlink classifier, this technique is not suitable for our learning task (sec. 5), since the closure might easily propagate errors in the automatic annotations.

```
project($a_s$, $al$):
    $a_{t,\mathsf{C}} = \emptyset$
    for each entity $e$ defined by $a_s$:
        $a_{t,\mathsf{C}}(e, a_s, al) = \bigcup^{\mathsf{C}}_{w_s \in a_s(e)} \mathsf{proj}(w_s, e, a_s, al)$
    for each link $l = (e, e')$ defined over $a_s$:
        if $a_{t,\mathsf{C}}(e, a_s, al) \neq \emptyset$ and $a_{t,\mathsf{C}}(e', a_s, al) \neq \emptyset$
        then define $l$ to hold for $a_{t,\mathsf{C}}$
    return $a_{t,\mathsf{C}}$
```

where

$$\begin{aligned}\mathsf{proj}(w_s, e, a_s, al) = \{&w_t \in W_t \mid (w_s, w_t) \in al \wedge \\ &\forall e' \in a_s.\ e' \neq e \Rightarrow w_t \notin a_{t,\mathsf{C}}(e', a_s, al)\}\end{aligned}$$

and

$$\bigcup^{\mathsf{C}} S = \begin{cases} \bigcup S & : \bigcup S \text{ is convex} \\ \emptyset & : \text{otherwise} \end{cases}$$

Figure 4: The projection algorithm.

temporal annotations is provided in Fig. 4.

3.2 Incorporating Additional Knowledge

The projection model described so far is extremely susceptible to errors in the word alignment. Related efforts (Hwa et al., 2005; Padó and Lapata, 2006) have already suggested that additional linguistic information can have considerable impact on the quality of the projected annotations. We therefore augment the baseline model with several shallow heuristics encoding linguistic or else topological constraints for the choice of words to project to. Linguistically motivated filters refer to the part-of-speech (POS) tags of words in the target language sentence, whereas topological criteria investigate the alignment topology.

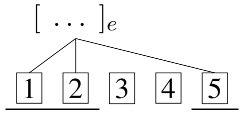
Linguistic constraints. Following Padó and Lapata (2006), we implement a filter which discards alignments to non-content words, for two reasons: (i) alignment algorithms are known to perform poorly on non-content words, and (ii) events as well as timexes are necessarily content-bearing and hence unlikely to be realised by non-content words. This *non-content (NC) filter* is defined in terms of POS tags and affects conjunctions, prepositions and punctuation. In the context of temporal annotations, we extend the scope of the filter such that it effectively applies to all word classes that we deem unlikely to occur as part of a temporal entity.

In order to adhere to the TimeML specification, a simple transformation ensures that articles[4] are included in the extent of timexes. Another heuristics is designed to remedy alignment errors involving auxiliary and modal verbs, which are not to be annotated as events. If an event aligns to more than one word, then this filter singles out the main verb or noun and discards auxiliaries.

Topological constraints. In sec. 3.1, we described a conservative projection principle which rejects the transfer of annotations to non-contiguous sequences. That model sets an unnecessarily modest upper bound on recall; but giving up the contiguity requirement entirely is not sensible either, since it is indeed highly unlikely for temporal entities to be realised discontinuously in either source or target language (Yarowsky and Ngai, 2001). Based on these observations, we propose two refined models which manipulate the projected annotation span so as to ensure contiguity. One model identifies and discards *outlier alignments*, which actively violate contiguity; the other one adds *missing alignments*, which form gaps. Technically, both models establish convexity in non-convex sets. Hence, we first have to come up with a backbone model which is less restrictive than the baseline, so that the convexation models will have a basis to operate on. A possible backbone model $a_{t,0}$ could simply gather all words aligned with any word covered by e in the source annotation, irrespective of contiguity. Discarding outlier alignments is then formalised as a reduction of $a_{t,0}$'s output to (one of) its greatest convex subset(s) (GCS). Let us call this model $a_{t,\mathrm{GCS}}$. In terms of a linear sequence of words, $a_{t,\mathrm{GCS}}$ chooses the longest contiguous subsequence.[5]

The second model, which fills gaps in the word alignment, constructs the *convex hull* of $a_{t,0}$ (cf. Padó and Lapata (2005)). We will refer to this model as $a_{t,\mathrm{CH}}$. The example in (1) illustrates both models.

(1)



$\bigcup^{C}$: $\emptyset$

GCS : $\{1, 2\}$

CH : $\{1, 2, 3, 4, 5\}$

---

[4] We also count contracted prepositions (*am, ins*) as articles.

[5] The GCS-model serves a filtering purpose similar to the NC filter. However, whereas the latter discards single alignment links on linguistic grounds, the former is motivated by topological properties of the alignment as a whole.

	events			slinks			time expressions		
model	prec	recall	F	prec	recall	F	prec	recall	F
timex-optimised	48.53	33.73	39.80	30.09	10.71	15.80	**71.01**	**52.76**	**60.54**
event-optimised	**50.94**	**44.23**	**47.34**	30.96	14.29	19.55	56.55	42.52	48.54
combined	**50.98**	**44.36**	**47.44**	**30.96**	**14.29**	**19.55**	**71.75**	**52.76**	**60.80**
baseline	52.26	33.46	40.80	26.98	10.71	15.34	49.53	37.80	42.87
full	51.10	40.42	45.14	29.95	13.57	18.68	73.74	54.33	62.56

Table 2: Performance of projection models over test data.

Here, entity e aligns to the non-contiguous token sequence $[1, 2, 5]$, or equivalently, the non-convex set $\{1, 2, 5\}(= a_{t,0}(e))$. The conservative baseline $a_{t,\mathrm{C}}$ rejects the projection altogether, whereas $a_{t,\mathrm{GCS}}$ projects to the tokens 1 and 2. The additional padding introduced by the convex hull ($a_{t,\mathrm{CH}}$) further extends the projected extent to $\{1, 2, 3, 4, 5\}$.

Alignment selection. Although bi-alignments are known to exhibit high precision (Koehn et al., 2003), we also use unidirectional alignments as a fallback in the face of sparse annotations. Furthermore, we follow Hwa et al. (2005) in imposing a limit on the maximum number of words that a single word may align to.

4 Experiments

Our evaluation setup consists of experiments conducted on the English-German portion of the Europarl corpus (Koehn, 2005); specifically, we work with the preprocessed and word-aligned version used in Padó and Lapata (2006). We put aside and manually annotated a development set of 101 and a test set of 236 bi-sentences.[6] All remaining data (approx. 960K bi-sentences) was used for training (sec. 5). We report the weighted macro average over all possible subclasses of timexes/events, and consider only exact matches. The TARSQI annotations exhibit an F_1-measure of 80.56% (timex), 84.64% (events), and 43.32% (slinks) when evaluated against the English gold standard.

In order to assess the usefulness of the linguistic and topological parameters presented in sec. 3.2, we determined the best performing combination of parameters on the development set. Not surprisingly, event and timex models benefit from the various heuristics to different degrees. While the projection of events can benefit from the NC filter, the projection of timexes is rather hampered by it. Instead, it exploits the flexibility of the GCS convexation model together with a conservative limit of 2 on per-word alignments. In the underlying data sample of 101 sentences, the English-to-German alignment direction appears to be most effective for timexes. Table 2 shows the results of evaluating the optimised models on the test set, along with the baseline from sec. 3.1 and a "full" model.[7] They confirm our initial assumption that linguistic and topological knowledge does indeed improve the quality of the projected annotations. The model which combines the optimal settings for timexes and events outperforms the uninformed baseline by 17.93% (timexes) and 6.64% (events) F_1-measure. However, exploration of the model space on the basis of the (larger and thus presumably more representative) test set shows that the optimised models do not generalise well. The *test set*-optimised model activates all linguistic heuristics, and employs $a_{t,\mathrm{CH}}$ convexation. For events, projection considers bi-alignments with a fallback to unidirectional alignments, preferably from English to German; timex projection considers all alignment links. This test set-optimised model, which we will use to project the training instances for the maximum entropy classifier, achieves an F_1-measure of 48.82% (53.15% precision) for events and 62.04% (73.74% precision) for timexes.[8]

With these settings, our projection model is capable of repairing alignment errors, as shown in Fig. 5, where the automatic word alignments are rep-

---

[6] The unconventional balance of test and development data is due to the fact that a large portion of the annotated data became available only after the parameter estimation phase.

[7] Full here means that all heuristics are activated.

[8] The model actually includes an additional strategy to adjust event and timex class labels on the basis of designated FrameNet frames.

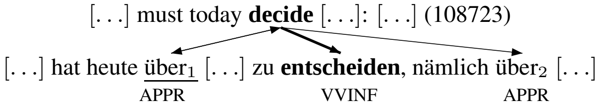


Figure 5: Amending alignment errors.

resented as arrows. The conservative baseline considering only bidirectional alignments discards all alignments but the (incorrect) one to *über*$_1$. The optimised model, on the other hand, does not exclude any alignments in the first place; the faulty alignments to *über*$_1$ and *über*$_2$ are discarded on linguistic grounds by the NC filter.

5 Robust Induction

The projected annotations, although noisy, can be exploited to train a temporal labeller for German. As Yarowsky and Ngai (2001) demonstrate for POS tagging, aggressive filtering techniques applied to vast amounts of (potentially noisy) training data are capable of distilling relatively high-quality data sets, which may then serve as input to machine learning algorithms. Yarowsky and Ngai (2001) use the Model-3 alignment score as an indicator for the quality of (i) the alignment, and therefore (ii) the projection. In the present study, discarding 25% of the sentences based on this criterion leads to gains in both recall and precision (Table 3). In accordance with the TimeML definition, we further restrict training instances on the basis of POS tags by basically re-applying the NC filter (sec. 3.2). But even so, the proportion of positive and negative instances remains heavily skewed—an issue which we will address below by formulating a 2-phase classification task.

	event		timex	
data	prec	recall	prec	recall
all	53.15	45.14	73.74	53.54
best 75%	54.81	47.06	74.61	62.82

Table 3: Correlation between alignment probability and projection quality.

The remaining instances[9] are converted to feature vectors encoding standard lexical and grammatical features such as (lower case) lemma, POS, governing prepositions, (X)COMP dependents, etc.[10] For slink instances, we further encode the syntactic subordination path (if any) between the two events.

	prec	recall	F	F
model		event		slink
1-step	83.48	32.58	46.87	17.01
1-step unk	83.88	32.19	46.53	16.87
2-step	**83.95**	**34.44**	**48.84**	**19.06**
2-step unk	84.21	34.30	48.75	19.06
		timex		
1-step	87.77	49.11	62.98	
1-step unk	87.22	49.55	63.20	
2-step	**89.52**	**51.79**	**65.62**	
2-step unk	88.68	50.89	64.67	

Table 4: Classifier performance over test data.

We trained 4 classifiers,[11] with and without smoothing with artificial unknowns (Collins, 2003), and as a 1-step versus a 2-step decision in which instances are first discriminated by a binary classifier, so that only positive instances are passed on to be classified for a subclass. The performance of the various classifiers is given in Table 4. Although the overall F_1-measure does not notably differ from that achieved by direct projection, we observe a drastic gain in precision, albeit at the cost of recall. With almost 84% and 90% precision, this is an ideal starting point for a bootstrapping procedure.

6 Discussion and Future Work

Clearly, the—essentially unsupervised—projection framework presented here does not produce state-of-the-art annotations. But it does provide an inexpensive and largely language-independent basis (a) for manual correction, and (b) for bootstrapping algorithms. Machine learning techniques like co-training (Blum and Mitchell, 1998) could further enhance projection, e.g. taking into account a third language.

---

[9] Note that slink instances are constructed for event *pairs*, as opposed to event and timex instances, which are constructed for individual words.

[10] The grammatical features have been extracted from analyses of the German ParGram LFG grammar (Rohrer and Forst, 2006).

[11] We used the `opennlp.maxent` package, `http://maxent.sourceforge.net/`.

References

A. Blum and T. Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-Training. In *Proceedings of the 1998 Conference on Computational Learning Theory*, pages 92–100, July.

M. Collins. 2003. Head-Driven Statistical Models for Natural Language Parsing. *Computational Linguistics*, 29(4):589–637, December.

B. J. Dorr. 1994. Machine Translation Divergences: A Formal Description and Proposed Solution. *Computational Linguistics*, 20(4):597–635.

L. Ferro, L. Gerber, I. Mani, B. Sundheim, and G. Wilson, 2005. *TIDES 2005 Standard for the Annotation of Temporal Expressions*, September.

R. Hwa, P. Resnik, A. Weinberg, and O. Kolak. 2002. Evaluating Translational Correspondence using Annotation Projection. In *Proceedings of ACL-2002*, Philadelphia, PA.

R. Hwa, P. Resnik, A. Weinberg, C. Cabezas, and O. Kolak. 2005. Bootstrapping Parsers via Syntactic Projection across Parallel Texts. *Natural Language Engineering*, 11(3):311–325.

P. Koehn, F. J. Och, and D. Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of HLT/NAACL 2003*, pages 127–133.

P. Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the MT Summit 2005*.

I. Mani and G. Wilson. 2000. Robust Temporal Processing of News. In *Proceedings of ACL-2000*, pages 69–76, Hong Kong.

I. Mani, B. Schiffman, and J. Zhang. 2003. Inferring Temporal Ordering of Events in News. In *Proceedings of HLT-NAACL 2003*. Short paper.

I. Mani, M. Verhagen, B. Wellner, C. M. Lee, and J. Pustejovsky. 2006. Machine Learning of Temporal Relations. In *Proceedings of ACL/COLING 2006*, pages 753–760, Sydney, Australia.

S. Padó and M. Lapata. 2005. Cross-lingual projection of role-semantic information. In *Proceedings of HLT/EMNLP 2005*, Vancouver, BC.

S. Padó and M. Lapata. 2006. Optimal constituent alignment with edge covers for semantic projection. In *Proceedings of COLING/ACL 2006*, Sydney, Australia.

J. Pustejovsky, J. Castaño, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, and G. Katz. 2003a. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of the 5th IWCS*.

J. Pustejovsky, P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003b. The TimeBank Corpus. In *Proceedings of Corpus Linguistics*, pages 647–656.

C. Rohrer and M. Forst. 2006. Improving coverage and parsing quality of a large-scale LFG for German. In *Proceedings of LREC 2006*, pages 2206–2211, Genoa, Italy, May.

E. Saquete, P. Martínez-Barco, R. Muñoz, M. Negri, M. Speranza, and R. Sprugnoli. 2006. Multilingual Extension of a Temporal Expression Normalizer using Annotated Corpora. In *Proceedings of the EACL 2006 Workshop on Cross-Language Knowledge Induction*, Trento, Italy, April.

R. Saurí, R. Knippen, M. Verhagen, and J. Pustejovsky. 2005a. Evita: A Robust Event Recognizer For QA Systems. In *Proceedings of HLT/EMNLP 2005*, pages 700–707.

R. Saurí, J. Littman, B. Knippen, R. Gaizauskas, A. Setzer, and J. Pustejovsky, 2005b. *TimeML Annotation Guidelines Version 1.2.1*, October.

R. Saurí, M. Verhagen, and J. Pustejovsky. 2006. SlinkET: A Partial Modal Parser for Events. In *Proceedings of LREC-2006*, Genova, Italy, May. To appear.

A. Schiffrin and H. Bunt. 2006. Defining a preliminary set of interoperable semantic descriptors. Technical Report D4.2, INRIA-Loria, Nancy, France, August.

A. Setzer. 2001. *Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study*. Ph.D. thesis, University of Sheffield, Sheffield, UK.

Z. Vendler, 1967. *Linguistics in Philosophy*, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, NY.

M. Verhagen, I. Mani, R. Saurí, R. Knippen, J. Littman, and J. Pustejovsky. 2005. Automating Temporal Annotation with TARSQI. In *Proceedings of ACL-2005*.

D. Yarowsky and G. Ngai. 2001. Inducing Multilingual POS Taggers and NP Bracketers via Robust Projection across Aligned Corpora. In *Proceedings of NAACL-2001*, pages 200–207.

D. Yarowsky, G. Ngai, and R. Wicentowski. 2001. Inducing Multilingual Text Analysis Tools via Robust Projection across Aligned Corpora. In *Proceedings of HLT 2001*.