# Temporal information processing of a new language: fast porting with minimal resources

**Francisco Costa** and **António Branco**
Universidade de Lisboa

## Abstract

We describe the semi-automatic adaptation of a TimeML annotated corpus from English to Portuguese, a language for which TimeML annotated data was not available yet. In order to validate this adaptation, we use the obtained data to replicate some results in the literature that used the original English data. The fact that comparable results are obtained indicates that our approach can be used successfully to rapidly create semantically annotated resources for new languages.

## 1 Introduction

Temporal information processing is a topic of natural language processing boosted by recent evaluation campaigns like TERN2004,[1] TempEval-1 (Verhagen et al., 2007) and the forthcoming TempEval-2[2] (Pustejovsky and Verhagen, 2009). For instance, in the TempEval-1 competition, three tasks were proposed: a) identifying the temporal relation (such as *overlap*, *before* or *after*) holding between events and temporal entities such as dates, times and temporal durations denoted by expressions (i.e. temporal expressions) occurring in the same sentence; b) identifying the temporal relation holding between events expressed in a document and its creation time; c) identifying the temporal relation between the main events expressed by two adjacent sentences.

Supervised machine learning approaches are pervasive in the tasks of temporal information processing. Even when the best performing systems in these competitions are symbolic, there are machine learning solutions with results close to their performance. In TempEval-1, where there were statistical and rule-based systems, almost all systems achieved quite similar results. In the TERN2004 competition (aimed at identifying and normalizing temporal expressions), a symbolic system performed best, but since then machine learning solutions, such as (Ahn et al., 2007), have appeared that obtain similar results.

These evaluations made available sets of annotated data for English and other languages, used for training and evaluation. One natural question to ask is whether it is feasible to adapt the training and test data made available in these competitions to other languages, for which no such data still exist. Since the annotations are largely of a semantic nature, not many changes need to be done in the annotations once the textual material is translated. In essence, this would be a fast way to create temporal information processing systems for languages for which there are no annotated data yet.

In this paper, we report on an experiment that consisted in adapting the English data of TempEval-1 to Portuguese. The results of machine learning algorithms over the data thus obtained are compared to those reported for the English TempEval-1 competition. Since the results are quite similar, this permits to conclude that such an approach can rapidly generate relevant and comparable data and is useful when porting temporal information processing solutions to new languages.

The advantages of adapting an existing corpus instead of annotating text from scratch are: i) potentially less time consuming, if it is faster to translate the original text than it is to annotate new text (this can be the case if the annotations are semantic and complex); b) the annotations can be transposed without substantial modifications, which is the case if they are semantic in nature; c) less man power required: text annotation requires multiple annotators in order to guarantee the quality of the annotation tags, translation of the markables and transposition of the annotations

[1] http://timex2.mitre.org

[2] http://www.timeml.org/tempeval2

*Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 671–677,
Uppsala, Sweden, 11-16 July 2010. ©2010 Association for Computational Linguistics

in principle do not; d) the data obtained are comparable to the original data in all respects except for language: genre, domain, size, style, annotation decisions, etc., which allows for research to be conducted with a derived corpus that is comparable to research using the original corpus. There is of course the caveat that the adaptation process can introduce errors.

This paper proceeds as follows. In Section 2, we provide a quick overview of the TimeML annotations in the TempEval-1 data. In Section 3, it is described how the data were adapted to Portuguese. Section 4 contains a brief quantitative comparison of the two corpora. In Section 5, the results of replicating one of the approaches present in the TempEval-1 challenge with the Portuguese data are presented. We conclude this paper in Section 6.

## 2 Brief Description of the Annotations

Figure 1 contains an example of a document from the TempEval-1 corpus, which is similar to the TimeBank corpus (Pustejovsky et al., 2003).

In this corpus, event terms are tagged with `<EVENT>`. The relevant attributes are `tense`, `aspect`, `class`, `polarity`, `pos`, `stem`. The `stem` is the term's lemma, and `pos` is its part-of-speech. Grammatical tense and aspect are encoded in the features `tense` and `aspect`. The attribute `polarity` takes the value `NEG` if the event term is in a negative syntactic context, and `POS` otherwise. The attribute `class` contains several levels of information. It makes a distinction between terms that denote actions of speaking, which take the value `REPORTING` and those that do not. For these, it distinguishes between states (value `STATE`) and non-states (value `OCCURRENCE`), and it also encodes whether they create an intensional context (value `I_STATE` for states and value `I_ACTION` for non-states).

Temporal expressions (timexes) are inside `<TIMEX3>` elements. The most important features for these elements are `value`, `type` and `mod`. The timex's `value` encodes a normalized representation of this temporal entity, its `type` can be e.g. `DATE`, `TIME` or `DURATION`. The `mod` attribute is optional. It is used for expressions like *early this year*, which are annotated with `mod="START"`. As can be seen in Figure 1 there are other attributes for timexes that encode whether it is the document's creation time (`functionInDocument`) and whether its value can be determined from the expression alone or requires other sources of information (`temporalFunction` and `anchorTimeID`).

The `<TLINK>` elements encode temporal relations. The attribute `relType` represents the type of relation, the feature `eventID` is a reference to the first argument of the relation. The second argument is given by the attribute `relatedToTime` (if it is a time interval or duration) or `relatedToEvent` (if it is another event; this is for task C). The `task` feature is the name of the TempEval-1 task to which this temporal relation pertains.

## 3 Data Adaptation

We cleaned all TimeML markup in the TempEval-1 data and the result was fed to the Google Translator Toolkit.[3] This tool combines machine translation with a translation memory. A human translator corrected the proposed translations manually.

After that, we had the three collections of documents (the TimeML data, the English unannotated data and the Portuguese unannotated data) aligned by paragraphs (we just kept the line breaks from the original collection in the other collections). In this way, for each paragraph in the Portuguese data we know all the corresponding TimeML tags in the original English paragraph.

We tried using machine translation software (we used GIZA++ (Och and Ney, 2003)) to perform word alignment on the unannotated texts, which would have enabled us to transpose the TimeML annotations automatically. However, word alignment algorithms have suboptimal accuracy, so the results would have to be checked manually. Therefore we abandoned this idea, and instead we simply placed the different TimeML markup in the correct positions manually. This is possible since the TempEval-1 corpus is not very large. A small script was developed to place all relevant TimeML markup at the end of each paragraph in the Portuguese text, and then each tag was manually repositioned. Note that the `<TLINK>` elements always occur at the end of each document, each in a separate line: therefore they do not need to be repositioned.

During this manual repositioning of the annotations, some attributes were also changed man-

[3] http://translate.google.com/toolkit

```
<?xml version="1.0" ?>
<TempEval>

ABC<TIMEX3 tid="t52" type="DATE" value="1998-01-14" temporalFunction="false"
functionInDocument="CREATION_TIME">19980114</TIMEX3>.1830.0611
NEWS STORY

<s>In Washington <TIMEX3 tid="t53" type="DATE" value="1998-01-14" temporalFunction="true"
functionInDocument="NONE" anchorTimeID="t52">today</TIMEX3>, the Federal Aviation Administration <EVENT
eid="e1" class="OCCURRENCE" stem="release" aspect="NONE" tense="PAST" polarity="POS" pos="VERB">released
</EVENT> air traffic control tapes from <TIMEX3 tid="t54" type="TIME" value="1998-XX-XXTNI"
temporalFunction="true" functionInDocument="NONE" anchorTimeID="t52">the night</TIMEX3> the TWA Flight
eight hundred <EVENT eid="e2" class="OCCURRENCE" stem="go" aspect="NONE" tense="PAST" polarity="POS"
pos="VERB">went</EVENT>down.</s>
...
<TLINK lid="l1" relType="BEFORE" eventID="e2" relatedToTime="t53" task="A"/>
<TLINK lid="l2" relType="OVERLAP" eventID="e2" relatedToTime="t54" task="A"/>
<TLINK lid="l4" relType="BEFORE" eventID="e2" relatedToTime="t52" task="B"/>
...
</TempEval>
```

Figure 1: Extract of a document contained in the training data of the first TempEval-1

ually. In particular, the attributes `stem`, `tense` and `aspect` of `<EVENT>` elements are language specific and needed to be adapted. Sometimes, the `pos` attribute also needs to be changed, since e.g. a verb in English can be translated as a noun in Portuguese. The attribute `class` of the same kind of elements can be different, too, because natural sounding translations are sometimes not literal.

### 3.1 Annotation Decisions

When porting the TimeML annotations from English to Portuguese, a few decisions had to be made. For illustration purposes, Figure 2 contains the Portuguese equivalent of the extract presented in Figure 1.

For `<TIMEX3>` elements, the issue is that if the temporal expression to be annotated is a prepositional phrase, the preposition should not be inside the `<TIMEX3>` tags according to the TimeML specification. In the case of Portuguese, this raises the question of whether to leave contractions of prepositions with determiners outside these tags (in the English data the preposition is outside and the determiner is inside).[4] We chose to leave them outside, as can be seen in that Figure. In this example the prepositional phrase *from the night*/*da noite* is annotated with the English noun phrase *the night* inside the `<TIMEX3>` element, but the Portuguese version only contains the noun *noite* inside those tags.

For `<EVENT>` elements, some of the attributes are adapted. The value of the attribute `stem` is obviously different in Portuguese. The attributes `aspect` and `tense` have a different set of possible values in the Portuguese data, simply because the morphology of the two languages is different. In the example in Figure 1 the value `PPI` for the attribute `tense` stands for *pretérito perfeito do indicativo*. We chose to include mood information in the `tense` attribute because the different tenses of the indicative and the subjunctive moods do not line up perfectly as there are more tenses for the indicative than for the subjunctive. For the `aspect` attribute, which encodes grammatical aspect, we only use the values `NONE` and `PROGRESSIVE`, leaving out the values `PERFECTIVE` and `PERFECTIVE_PROGRESSIVE`, as in Portuguese there is no easy match between perfective aspect and grammatical categories.

The attributes of `<TIMEX3>` elements carry over to the Portuguese corpus unchanged, and the `<TLINK>` elements are taken verbatim from the original documents.

## 4 Data Description

The original English data for TempEval-1 are based on the TimeBank data, and they are split into one dataset for training and development and another dataset for evaluation. The full data are organized in 182 documents (162 documents in the training data and another 20 in the test data). Each document is a news report from television broadcasts or newspapers. A large amount of the documents (123 in the training set and 12 in the test data) are taken from a 1989 issue of the Wall Street Journal.

The training data comprise 162 documents with

[4] The fact that prepositions are placed outside of temporal expressions seems odd at first, but this is because in the original TimeBank, from which the TempEval data were derived, they are tagged as `<SIGNAL>`s. The TempEval-1 data does not contain `<SIGNAL>` elements, however.

```
<?xml version="1.0" encoding="UTF-8" ?>
<TempEval>

ABC<TIMEX3 tid="t52" type="DATE" value="1998-01-14" temporalFunction="false"
functionInDocument="CREATION_TIME">19980114</TIMEX3>.1830.1611
REPORTAGEM

<s>Em Washington, <TIMEX3 tid="t53" type="DATE" value="1998-01-14" temporalFunction="true"
functionInDocument="NONE" anchorTimeID="t52">hoje</TIMEX3>, a Federal Aviation Administration <EVENT
eid="e1" class="OCCURRENCE" stem="publicar" aspect="NONE" tense="PPI" polarity="POS" pos="VERB">publicou
</EVENT> gravaoes do controlo de trfego areo da <TIMEX3 tid="t54" type="TIME" value="1998-XX-XXTNI"
temporalFunction="true" functionInDocument="NONE" anchorTimeID="t52">noite</TIMEX3> em que o voo TWA800
<EVENT eid="e2" class="OCCURRENCE" stem="cair" aspect="NONE" tense="PPI" polarity="POS" pos="VERB">caiu
</EVENT>
.</s>
...
<TLINK lid="l1" relType="BEFORE" eventID="e2" relatedToTime="t53" task="A"/>
<TLINK lid="l2" relType="OVERLAP" eventID="e2" relatedToTime="t54" task="A"/>
<TLINK lid="l4" relType="BEFORE" eventID="e2" relatedToTime="t52" task="B"/>
...
</TempEval>
```

Figure 2: Extract of a document contained in the Portuguese data

2,236 sentences (i.e. 2236 <s> elements) and 52,740 words. It contains 6799 <EVENT> elements, 1,244 <TIMEX3> elements and 5,790 <TLINK> elements. Note that not all the events are included here: the ones expressed by words that occur less than 20 times in TimeBank were removed from the TempEval-1 data.

The test dataset contains 376 sentences and 8,107 words. The number of <EVENT> elements is 1,103; there are 165 <TIMEX3>s and 758 <TLINK>s.

The Portuguese data of course contain the same (translated) documents. The training dataset has 2,280 sentences and 60,781 words. The test data contains 351 sentences and 8,920 words.

## 5 Comparing the two Datasets

One of the systems participating in the TempEval-1 competition, the USFD system (Hepple et al., 2007), implemented a very straightforward solution: it simply trained classifiers with Weka (Witten and Frank, 2005), using as attributes information that was readily available in the data and did not require any extra natural language processing (for all tasks, the attribute relType of <TLINK> elements is unknown and must be discovered, but all the other information is given).

The authors' objectives were to see "whether a 'lite' approach of this kind could yield reasonable performance, before pursuing possibilities that relied on 'deeper' NLP analysis methods", "which of the features would contribute positively to system performance" and "if any [machine learning] approach was better suited to the TempEval tasks than any other". In spite of its simplicity, they obtained results quite close to the best systems.

For us, the results of (Hepple et al., 2007) are interesting as they allow for a straightforward evaluation of our adaptation efforts, since the same machine learning implementations can be used with the Portuguese data, and then compared to their results.

The differences in the data are mostly due to language. Since the languages are different, the distribution of the values of several attributes are different. For instance, we included both tense and mood information in the tense attribute of <EVENT>s, as mentioned in Section 3.1, so instead of seven possible values for this attribute, the Portuguese data contains more values, which can cause more data sparseness. Other attributes affected by language differences are aspect, pos, and class, which were also possibly changed during the adaptation process.

One important difference between the English and the Portuguese data originates from the fact that events with a frequency lower than 20 were removed from the English TempEval-1 data. Since there is not a 1 to 1 relation between English event terms and Portuguese event terms, we do not have the guarantee that all event terms in the Portuguese data have a frequency of at least 20 occurrences in the entire corpus.[5]

The work of (Hepple et al., 2007) reports on both cross-validation results for various classifiers over the training data and evaluation results on the training data, for the English dataset. We we will

[5] In fact, out of 1,649 different stems for event terms in the Portuguese training data, only 45 occur at least 20 times.

|  | Task |  |  |
|---|---|---|---|
| Attribute | A | B | C |
| EVENT-aspect | ✓ | ✓ | ✓ |
| EVENT-polarity | ✓ | ✓ | × |
| EVENT-POS | ✓ | ✓ | ✓ |
| EVENT-stem | ✓ | × | × |
| EVENT-string | × | × | × |
| EVENT-class | × | ✓ | ✓ |
| EVENT-tense | × | ✓ | ✓ |
| ORDER-adjacent | ✓ | N/A | N/A |
| ORDER-event-first | ✓ | N/A | N/A |
| ORDER-event-between | × | N/A | N/A |
| ORDER-timex-between | × | N/A | N/A |
| TIMEX3-mod | ✓ | × | N/A |
| TIMEX3-type | ✓ | × | N/A |

Table 1: Features used for the English TempEval-1 tasks. N/A means the feature was not applicable to the task, ✓means the feature was used by the best performing classifier for the task, and × means it was not used by that classifier. From (Hepple et al., 2007).

be comparing their results to ours.

Our purpose with this comparison is to validate the corpus adaptation. Similar results would not necessarily indicate the quality of the adapted corpus. After all, a word-by-word translation would produce data that would yield similar results, but it would also be a very poor translation, and therefore the resulting corpus would not be very interesting. The quality of the translation is not at stake here, since it was manually revised. But similar results would indicate that the obtained data are comparable to the original data, and that they are similarly useful to tackle the problem for which the original data were collected. This would confirm our hypothesis that adapting an existing corpus can be an effective way to obtain new data for a different language.

### 5.1 Results for English

The attributes employed for English by (Hepple et al., 2007) are summarized in Table 1. The class is the attribute `relType` of `<TLINK>` elements.

The EVENT features are taken from `<EVENT>` elements. The EVENT-`string` attribute is the character data inside the element. The other attributes correspond to the feature of `<EVENT>` with the same name. The TIMEX3 features also correspond to attributes of the relevant `<TIMEX3>` element. The ORDER features are boolean and computed as follows:

- ORDER-`event-first` is whether the `<EVENT>` element occurs in the text before the `<TIMEX3>` element;
- ORDER-`event-between` is whether an `<EVENT>` element occurs in the text between the two temporal entities being ordered;
- ORDER-`timex-between` is the same, but for temporal expressions;
- ORDER-`adjacent` is whether both ORDER-`event-between` and ORDER-`timex-between` are false (but other textual data may occur between the two entities).

Cross-validation over the training data produced the results in Table 2. The baseline used is the majority class baseline, as given by Weka's `rules.ZeroR` implementation. The `lazy.KStar` algorithm is a nearest-neighbor classifier that uses an entropy-based measure to compute instance similarity. Weka's `rules.DecisionTable` algorithm assigns to an unknown instance the majority class of the training examples that have the same attribute values as that instance that is being classified. `functions.SMO` is an implementation of Support Vector Machines (SVM), `rules.JRip` is the RIPPER algorithm, and `bayes.NaiveBayes` is a Naive Bayes classifier.

|  | Task |  |  |
|---|---|---|---|
| Algorithm | A | B | C |
| baseline | 49.8 | 62.1 | 42.0 |
| `lazy.KStar` | **58.2** | 76.7 | 54.0 |
| `rules.DecisionTable` | 53.3 | **79.0** | 52.9 |
| `functions.SMO` | 55.1 | 78.1 | **55.5** |
| `rules.JRip` | 50.7 | 78.6 | 53.4 |
| `bayes.NaiveBayes` | 56.3 | 76.2 | 50.7 |

Table 2: Performance of several machine learning algorithms on the English TempEval-1 training data, with cross-validation. The best result for each task is in boldface. From (Hepple et al., 2007).

|  | Task |  |  |
|---|---|---|---|
| Algorithm | A | B | C |
| baseline | 49.8 | 62.1 | 42.0 |
| `lazy.KStar` | **57.4** | 77.7 | 53.3 |
| `rules.DecisionTable` | 54.2 | 78.1 | 51.6 |
| `functions.SMO` | 55.5 | **79.3** | 56.8 |
| `rules.JRip` | 52.1 | 77.6 | 52.1 |
| `bayes.NaiveBayes` | 56.0 | 78.2 | 53.5 |
| `trees.J48` | 55.6 | 79.0 | **59.3** |

Table 3: Performance of several machine learning algorithms on the Portuguese data for the TempEval-1 tasks. The best result for each task is in boldface.

### 5.2 Attributes

We created a small script to convert the XML annotated files into CSV files, that can be read by Weka. In this process, we included the same attributes as the USFD authors used for English.

For task C, (Hepple et al., 2007) are not very clear whether the EVENT attributes used were related to just one of the two events being temporally related. In any case, we used two of each of the EVENT attributes, one for each event in the temporal relation to be determined. So, for instance, an extra attribute EVENT2-`tense` is where the tense of the second event in the temporal relation is kept.

### 5.3 Results

The majority class baselines produce the same results as for English. This was expected: the class distribution is the same in the two datasets, since the `<TLINK>` elements were copied to the adapted corpus without any changes.

For the sake of comparison, we used the same classifiers as (Hepple et al., 2007), and we used the attributes that they found to work best for English (presented above in Table 1). The results for the Portuguese dataset are in Table 3, using 10-fold cross-validation on the training data.

We also present the results for Weka's implementation of the C4.5 algorithm, to induce decision trees. The motivation to run this algorithm over these data is that decision trees are human readable and make it easy to inspect what decisions the classifier is making. This is also true of `rules.JRip`. The results for the decision trees are in this table, too.

The results obtained are almost identical to the results for the original dataset in English. The best performing classifier for task A is the same as for English. For task B, Weka's `functions.SMO` produced better results with the Portuguese data than `rules.DecisionTable`, the best performing classifier with the English data for this task. In task C, the SVM algorithm was also the best performing algorithm among those that were also tried on the English data, but decision trees produced even better results here.

For English, the best performing classifier for each task on the training data, according to Table 2, was used for evaluation on the test data: the results showed a 59% F-measure for task A, 73% for task B, and 54% for task C.

Similarly, we also evaluated the best algorithm for each task (according to Table 3) with the Portuguese test data, after training it on the entire training dataset. The results are: in task A the `lazy.KStar` classifier scored 58.6%, and the SVM classifier scored 75.5% in task B and 59.4% in task C, with `trees.J48` scoring 61% in this task.

The results on the test data are also fairly similar for the two languages/datasets.

We inspected the decision trees and rule sets produced by `trees.J48` and `rules.JRip`, in order to see what the classifiers are doing.

Task B is probably the easiest task to check this way, because we expect grammatical tense to be highly predictive of the temporal order between an event and the document's creation time.

And, indeed, the top of the tree induced by `trees.J48` is quite interesting:

```
eTense =  PI:  OVERLAP (388.0/95.0)
eTense =  PPI:  BEFORE (1051.0/41.0)
```

Here, `eTense` is the EVENT-`tense` attribute of `<EVENT>` elements, `PI` stands for present indicative, and `PPI` is past indicative (*pretérito perfeito do indicativo*). In general, one sees past tenses associated with the `BEFORE` class and future tenses associated with the `AFTER` class (including the conditional forms of verbs). Infinitives are mostly associated with the `AFTER` class, and present subjunctive forms with `AFTER` and `OVERLAP`. Figure 3 shows the rule set induced by the RIPPER algorithm.

The classifiers for the other tasks are more difficult to inspect. For instance, in task A, the event term and the temporal expression that denote the entities that are to be ordered may not even be directly syntactically related. Therefore, it is hard to

```
(eClass = OCCURRENCE) and ( eTense =  INF) and ( ePolarity =  POS) =>  lRelType= AFTER
                                                                            (183.0/77.0)
( eTense =  FI) =>  lRelType= AFTER (55.0/10.0)
(eClass = OCCURRENCE) and ( eTense =  IR-PI+INF) =>  lRelType= AFTER (26.0/4.0)
(eClass = OCCURRENCE) and ( eTense =  PC) =>  lRelType= AFTER (15.0/3.0)
(eClass = OCCURRENCE) and ( eTense =  C) =>  lRelType= AFTER (17.0/2.0)
( eTense =  PI) =>  lRelType= OVERLAP (388.0/95.0)
(eClass = ASPECTUAL) and ( eTense =  PC) =>  lRelType= OVERLAP (9.0/2.0)
 =>  lRelType= BEFORE (1863.0/373.0)
```

Figure 3: `rules.JRip` classifier induced for task B. `INF` stands for infinitive, `FI` is future indicative, `IR-PI+INF` is an infinitive form following a present indicative form of the verb *ir* (*to go*), `PC` is present subjunctive, `C` is conditional, `PI` is present indicative.

see how interesting the inferred rules are, because we do not know what would be interesting in this scenario. In any case, the top of the induced tree for task A is:

```
oAdjacent =  True: OVERLAP (554.0/128.0)
```

Here, `oAdjacent` is the ORDER-`adjacent` attribute. Assuming this attribute is an indication that the event term and the temporal expression are related syntactically, it is interesting to see that the typical temporal relation between the two entities in this case is an `OVERLAP` relation. The rest of the tree is much more *ad-hoc*, making frequent use of the `stem` attribute of `<EVENT>` elements, suggesting the classifier is memorizing the data.

Task C, where two events are to be ordered, produced more complicated classifiers. Generally the induced rules and the tree paths compare the tense and the class of the two event terms, showing some expected heuristics (such as, if the tense of the first event is future and the tense of the second event is past, assign `AFTER`). But there are also many several rules for which we do not have clear intuitions.

## 6 Discussion

In this paper, we described the semi-automatic adaptation of a TimeML annotated corpus from English to Portuguese, a language for which TimeML annotated data was not available yet.

Because most of the TimeML annotations are semantic in nature, they can be transposed to a translation of the original corpus, with few adaptations being required.

In order to validate this adaptation, we used the obtained data to replicate some results in the literature that used the original English data.

The results for the Portuguese data are very similar to the ones for English. This indicates that our approach to adapt existing annotated data to a different language is fruitful.

## References

David Ahn, Joris van Rantwijk, and Maarten de Rijke. 2007. A cascaded machine learning approach to interpreting temporal expressions. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 420–427, Rochester, New York, April. Association for Computational Linguistics.

Mark Hepple, Andrea Setzer, and Rob Gaizauskas. 2007. USFD: Preliminary exploration of features and classifiers for the TempEval-2007 tasks. In *Proceedings of SemEval-2007*, pages 484–487, Prague, Czech Republic. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

James Pustejovsky and Marc Verhagen. 2009. Semeval-2010 task 13: evaluating events, time expressions, and temporal relations (tempeval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 112–116, Boulder, Colorado. Association for Computational Linguistics.

James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003. The TIMEBANK corpus. In *Proceedings of Corpus Linguistics 2003*, pages 647–656.

M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, and J. Pustejovsky. 2007. SemEval-2007 Task 15: TempEval temporal relation identification. In *Proceedings of SemEval-2007*.

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco. second edition.