Syntactically Motivated Task Definition for Temporal Relation Identification

Georgiana Marşic

Research Institute in Information and Language Processing
University of Wolverhampton
MB Building, Stafford Street, Wolverhampton, WV1 1LY, UK
georgie@wlv.ac.uk

ABSTRACT. *The annotation of temporal relations remains a challenge, being a very difficult task for humans, not to mention machines, to reliably and consistently annotate temporal relations in natural language texts. This paper advocates a change in the definition of the problem itself, by proposing a staged divide-and-conquer approach guided by syntax, that offers a more principled way of selecting temporal entities involved in a temporal relation. The decomposition of the problem into smaller syntactically motivated tasks, and the identification of accurate and linguistically grounded solutions to solve them, promote a sound understanding of the phenomena involved in establishing temporal relations. We illustrate the potential of linguistically informed solutions in the area of temporal relation identification by proposing and evaluating an initial set of syntactically motivated tasks.*

RÉSUMÉ. *L'annotation de relations temporelles demeure encore aujourd'hui un défi : annoter manuellement de façon fiable et cohérente les relations temporelles dans des textes reste difficile et l'est bien plus encore lorsqu'il s'agit d'annotation automatique. Cet article préconise un changement dans la définition du problème en proposant une approche qui, en s'appuyant sur la syntaxe et sur une stratégie de type « diviser pour conquérir », offre une manière plus élaborée de sélectionner les entités impliquées dans une relation temporelle. La décomposition du problème en de plus petites questions se concentrant sur la syntaxe et l'identification de solutions précises et linguistiquement fondées pour les résoudre favorisent une meilleure compréhension des phénomènes impliqués dans l'établissement de relations temporelles. Nous illustrons le potentiel des solutions linguistiquement fondées dans le cadre de l'identification de relations temporelles en proposant et évaluant une première série de tâches se concentrant sur la syntaxe.*

KEYWORDS: *temporal relations, task decomposition, syntactic approach.*

MOTS-CLÉS : *relations temporelles, décomposition des tâches, approche syntaxique.*

1. Introduction

The temporal dimension of information is fundamental for reasoning about how the world changes. The ability to capture this temporal dimension of a natural language text has become more and more relevant to many natural language processing (NLP) applications such as Question Answering (Moldovan *et al.*, 2005), Automatic Summarization (Mani and Shiffman, 2005), Information Retrieval (Alonso *et al.*, 2007), and Information Extraction (Surdeanu *et al.*, 2003). The need for temporally aware systems has become increasingly popular. This need is justified by the fact that most of the information available electronically is temporally sensitive, in the sense that something that was true at some point in time could be false at another (e.g. example [1.1] presents a fact that is currently true, but at some point of time in the future its truth value will change to false). Despite its omnipresence, agreeing on how time can be formalized has historically been a difficult task, as well as incorporating it into automatic systems that can access the temporal dimension and extract the temporal meaning of a text, known as **temporal processing systems**.

[1.1] *David Cameron is the Prime Minister of the UK.*

Temporal processing systems aim to derive a precise temporal representation of a natural language text by extracting time expressions, events and temporal relations, and then representing them according to a chosen knowledge framework. The most difficult challenge for researchers working in the area of temporal processing is finding methodologies for annotating the temporal relations between time expressions and events or between events and other events. This difficulty is justified by the variety of mechanisms used by language to express temporal relations, as well as by the fact that temporal relations are not always stated explicitly, being often implicit and requiring interpretations or inferences derived from world knowledge. For example, the sentences in [1.2] and [1.3] have similar syntax, but the events they describe are not in the same temporal order. The temporal information in these examples is implicit, as the events described are neither anchored to precise points in time, nor specifically ordered with respect to neighbouring events. To derive the correct temporal interpretation for these examples, one must rely on semantic content, knowledge of causation and knowledge of language use. Despite their structure and syntax being so similar, in [1.2] the event of *falling* is temporally after the event of *pushing*, while in [1.3] the event of *falling* precedes the event of *asking*.

[1.2] *John **fell**. Mary **pushed** him.*

[1.3] *John **fell**. Mary **asked** for help.*

Temporal relations remain a challenge even for human annotators. Previous work in this area (Katz and Arosio, 2001; Setzer and Gaizauskas, 2002; Pustejovsky *et al.*, 2006; Mani and Shiffman, 2005; Verhagen *et al.*, 2009) has shown that it is very difficult to ask humans, let alone computers, to reliably and consistently annotate temporal relations. This calls for a more comprehensive study that would provide a sound understanding of temporal relations. To this end, it is advisable to decompose the annotation task into more linguistically motivated and structured subtasks, as this would enable a thorough analysis of the temporal phenomena that interact in establishing a temporal relation, an informed revision of the annotation guidelines, and therefore a more effective data creation process and system evaluation.

This paper is structured as follows. Section 2 looks at the mechanisms used by language to express temporal relations. This is followed by an overview of computational approaches that address the annotation of temporal relations in Section 3. Section 4 presents the main problems we are confronted with when performing temporal relation annotation (either manually, automatically or in the context of evaluation exercises). Section 5 elaborates on the syntactically motivated staged approach to temporal relation identification that we propose for automatic temporal relation identification at intra-sentential level. Sections 6 and 7 describe experiments involving various syntactically constrained subtasks defined at intra-clausal and inter-clausal level, respectively. Finally, in Section 8, conclusions are drawn.

2. Expressing Temporal Relations in Natural Language

An investigation of how language is used to convey temporal relations is necessary to gain computational insights into the mechanisms that can be exploited in the development of automatic systems targeting the identification of temporal relations.

Mechanisms like tense and grammatical aspect encode temporal relations between the three time points defined by Reichenbach (1947): the speech time (S), the event time (E), and the reference time (R). The speech time is the time at which the utterance is produced. The event time is the time at which the described event occurred. The reference time is the time from which the speaker is viewing the event on a timeline. Three temporal relations can hold between these time points: *at* (=), *before* (<), or *after* (>). In terms of Reichenbach's theory, the relation between the reference time R and the speech time S is established by tense, while the relation between the event time E and the reference time R is provided by grammatical aspect. In example [2.1], the event time (i.e. the time of reading the book) is situated before the reference time (i.e. the time when John told Mary the plot), as suggested by the presence of the perfective aspect. The past tense of the verb *told* locates the reference time before the speech time. A simple temporal illustration of this example would be $\mathbf{E} < \mathbf{R} < \mathbf{S}$.

[2.1] *<S>Mary had [read the book]<E> [when John told her about the plot]<R>.*

Temporal relations are especially dependent for their expression upon time adverbials. Time adverbials convey temporal relations between the time they denote and the verbal event they syntactically depend on. Time adverbials are syntactically realized by means of adverbs, noun phrases, prepositional phrases and temporal clauses. Most adverbs (e.g. *yesterday*), noun phrases (e.g. *last week*) and prepositional phrases (e.g. *on Monday*) that express the semantic role of time are considered temporal expressions (TEs), and it should be noted that temporal expressions form the largest subclass of time adverbials. However, there is a particular class of time relationship adverbials (adverbs that signal temporal sequence, such as: *afterwards*, *then*, *before*, *later*, *next*, *subsequently*) that are not considered temporal expressions. They are used to indicate the temporal relation that holds between the event expressed by the verb they syntactically depend on, and the reference time point or the event that was last introduced in the preceding discourse. Temporal clauses (e.g. *John came home* ***after Mary left****.*) are another realization of time adjuncts, a temporal clause being able to relate the time of the event it mentions to the time of the event described in the clause it syntactically depends on. While temporal expressions relate an event to a time, temporal clauses establish temporal relations between two events.

Besides the above mechanisms, one can encounter other ways in which language expresses temporal relations, and they are enumerated below.

At the syntactic level, temporal relations can be inferred from dependency relations. For example, in syntactic constructions where a temporal expression is included in a noun phrase to qualify a noun event (e.g. *the Sunday election*), there is a temporal relation of overlap between the noun event and the time indicated by the temporal expression.

At the semantic level, an important role is played by world knowledge. Without world knowledge it is often impossible to know that an event represents an integrating part of another event, or that an event causes another event. **Subevents** trigger temporal inclusion, as illustrated by example [2.2] where the event of painting the walls is part of redecorating the house, thus leading to the interpretation that the temporal relation between the two events is one of temporal inclusion.

[2.2] *John* ***redecorated*** *his house. He first* ***painted*** *the walls.*

Causality is another factor that intervenes at the semantic level, and indicates temporal precedence, as the cause always comes before the effect (as in [1.2]).

Temporal relations can also be expressed by **narrative sequence**. In this case the sequence in which the events appear in text reflects the order in which they happened.

In a given text, events can be mentioned several times, and this leads to the phenomenon of **event co-reference**. The referential instance of an event takes place at the same time as the referred event, and one can infer that all the temporal relations holding for one instance of the event also hold for the other instance. This brings about an important source of temporal relations: **inference**. Temporal relations can be inferred using simple rules, such as the transitivity rule: if an event A happens before an event B, and B happens before C, then one can infer that A happens before C.

This wide variety of mechanisms used by language to express temporal relations gives an idea of how complex the entire process of annotating temporal relations in text can be. While temporal relations made explicit in text via mechanisms such as tense, grammatical aspect or temporal adverbials can be automatically identified, the types of semantically implicit temporal relations mentioned above pose real challenges to automatic systems due to the world knowledge required for their identification.

3. Temporal Relation Identification

Research in this area has matured from the initial approaches to time stamp events and to annotate corpora using various representations of temporal relations to more complex approaches targeting the automatic identification of temporal relations. All these efforts will be described in detail below.

3.1. *Time Stamping Events*

Approaches for time stamping events typically attempt to associate a calendar time or time interval with some or all events present in a text.

MUC (Message Understanding Conferences) campaigns (Sundheim and Chinchor, 1993) required in their scenario template task the assignment of a temporal expression to the slot LAUNCH_DATE of the predefined scenario template. All systems achieved poor results on this particular slot (approximately 40% F-measure), reflecting the difficulty of the task.

As part of the MUC task, times were only assigned to certain scenario events. A more general approach aimed to assign a time point or time interval to absolutely every event in a text. Filatova and Hovy (2001) experimented with both manually and automatically decomposing news stories into their constituent event clauses and assigning time stamps to each event clause following an analysis of the temporal adverbials and of the verbal tense information characterizing each event clause. A time-stamp assignment was considered to be correct whenever the event clause's real time-stamp was included in the time interval provided by the system for that event clause. The time-stamper was evaluated both on manually annotated event clauses achieving 77.85% accuracy on a set of 158 clauses, and on correctly identified clauses extracted from the output of the syntactic parser: in this case the accuracy was 82.29% on 96 clauses.

Schilder and Habel (2001) made the transition between time-stamping events and proper temporal relation identification by defining a set of temporal relations and assigning them to time-event pairs. Their approach relied on the assumption that relations between times and events should be marked only when they are explicitly signalled by prepositions, or whenever they are syntactically implicit. The authors designed a temporal annotation system for German that assigned a default temporal relation to each pair of event-TE connected via a preposition, while the inclusion relation was assumed for cases when no preposition was present. Their system marking only temporal relations between events and temporal expressions involved in a direct syntactic relation was evaluated on a small corpus of 10 German news articles and achieved an overall precision and recall of 84.49%.

The ultimate goal of event time-stamping approaches was to anchor all events in a text on a time line. However, natural language texts rarely specify the exact position an event should have on a time line, and mostly provide a partial ordering between events. This calls for a different representation that does not require a total event ordering, and that does not leave out temporal relations that are explicit in text. Efforts to define such representations and to demonstrate their suitability by applying them to natural language texts are presented in the following sections.

3.2. *Annotation of Corpora with Temporal Relations*

In an effort towards a better understanding of temporality from a computational perspective, interest has shifted towards designing annotation schemes and applying them to natural language texts.

Katz and Arosio (2001) proposed a semantic formalism suitable for the annotation of intra-sentential temporal relations, and then applied it to syntactically annotated sentences with the aim of creating a treebank annotated with temporal relations and morpho-syntactic information. This resource was designed for examining the influence of lexical and syntactic structure on temporal ordering. The manual annotation process targeted pairs consisting of a verb and the speech time, as well as pairs of verbs expressing states or events that were situated in the same sentence. Each verb was associated with a temporal interval and the relations among these intervals were reduced to either precedence or inclusion. The authors reported an inter-annotator agreement of 70% on a set of 50 sentences.

Setzer and Gaizauskas (2002) promoted a novel temporal representation for a text in the form of a time-event graph, the nodes of the graph being either times or events, and the arcs representing event-event and time-event temporal relations. This representation is preferred to the "time-stamping" paradigm because, in many cases, texts position events in time only by relation to other events, and any attempt to place these events on a timeline must either lose information or invent information. At the same time, both event-event and time-event temporal relations are required to accurately position events on a timeline. To assess the utility of this representation, the authors annotated a trial corpus of 6 newswire articles using the STAG annotation scheme (Setzer, 2001), an annotation scheme which enabled TEs, events and temporal relations to be marked up in newswire texts. This scheme adheres to the time-event graph representation. The annotation took place in two stages. The first stage covered the annotation of events, time expressions, signals and temporal relations that were either explicitly expressed or syntactically implicit. The second stage relied on the information annotated at the first stage, and automatically derived all possible inferences, thus enriching the annotation with new temporal relations. At this stage, the human annotator was only prompted to manually specify a temporal relation when the system was unable to infer a relation for a given pair of entities. The process then continued cyclically until every event-event and event-time pair in the text were temporally related. This annotation experiment showed that the task proved very difficult for human annotators, and that the low inter-annotator agreement was due to several causes: imprecision/incompleteness of the guidelines, imperfect annotator understanding of the task, difficulty of establishing a temporal relation in some cases, annotator fatigue, and annotator carelessness.

The efforts of Setzer and Gaizauskas to define an annotation scheme have proved essential for the development of the generally adopted standard for temporal annotation: TimeML (Pustejovsky *et al.*, 2003; Saurí *et al.*, 2006), and of the TimeML proof of concept: the TimeBank corpus (Pustejovsky *et al.*, 2006).

TimeML is a formal specification language for events, temporal expressions, time anchoring of events (i.e. the temporal relations between events and TEs), and relative orderings of events with respect to one another, that was developed as a result of a wide interest in temporal analysis and event-based reasoning. This interest was manifested in a number of important specialized workshops and satellite events organized at major conferences including ACL 2001 (ACL-2001, 2001), LREC 2002 (LREC-2002, 2002), TERQAS 2002 (TERQAS, 2002), TANGO 2003 (TANGO, 2003), Dagstuhl 2005 (Dagstuhl, 2005), TIME 2006 (TIME-2006, 2006), ARTE 2006 (ARTE, 2006). Significant progress was made during these events, leading to the design, refinement and standardization of TimeML to an ISO international standard for temporal information markup, ISO-TimeML (ISO-TimeML, 2007). Both the TimeML and the ISO-TimeML annotation standards define the following basic XML tags: <EVENT> for the annotation of events, <TIMEX3> for the annotation of time expressions, <SIGNAL> for capturing the textual elements that indicate a temporal relation, and the tags <TLINK>, <SLINK> and <ALINK> that capture different types of relations: temporal, subordination, and aspectual relations, respectively. The tag <EVENT> has attributes that capture morpho-syntactic features, such as the event's part of speech, its tense, aspect, modality and polarity. Each event is also characterized by an attribute **class** that indicates the event's semantic class and has one of the following values: OCCURRENCE (something that happens), STATE (circumstance in which something holds true), PERCEPTION (events involving the physical perception of another event), REPORTING (events that narrate or inform about another event), ASPECTUAL (events that capture the aspectual predication on different facets of another event's history: initiation, culmination, termination), I_STATE (intensional state) and I_ACTION (intensional action). The tag <TIMEX3> has attributes that capture the normalized value of the time expression, modifiers, and the ID of the TE which was used during its normalization. The tags <TLINK>, <SLINK> and <ALINK> are characterized by attributes which indicate the temporal entities linked by these tags, together with the type of relation that holds between them.

TimeBank (Pustejovsky *et al.*, 2006) is the human-annotated corpus marked up for temporal expressions, events, and temporal relations as a proof of concept for the TimeML standard. Its current version, TimeBank 1.2 contains 183 documents with just over 61,000 words. A low inter-annotator agreement score is observed for temporal relations (two annotators agree on which two entities to link in 55% of the cases), which is due to the large number of temporal entity pairs that can be selected for specifying temporal links, and any two annotators working on the same text are very likely to select different pairs of entities to be linked via temporal relations. Therefore, one important problem is that annotators do not create the same temporal links between temporal entities, and when they do, they only agree 77% of the time on the temporal relation that holds between the two entities. These official inter-annotator figures reveal low agreement for annotating temporal relations. This illustrates the difficulty of annotating temporal relations in accordance with TimeML, but also the fact that underlying temporal phenomena are not very well understood by humans. This lack of a clear picture over how time in general, and temporal relations in particular, are expressed in text and over what is the best way to represent them formally is easily inferred from the ambiguous TimeML annotation guidelines that leave many aspects of the annotation underspecified. The existing problems concerning TimeML and TimeBank are also highlighted by other authors (Boguraev and Ando, 2005; Boguraev and Ando, 2006; Derczynski and Gaizauskas, 2010a), who bring additional evidence of inconsistency in the annotation of TimeBank.

3.3. *Automatically Identifying Temporal Relations*

As annotated data started to become available, research efforts were dedicated to devising complex approaches targeting the automatic identification of temporal relations.

Mani *et al.* (2003) proposed a machine learning approach for temporally anchoring and ordering events in news. Events were associated with clauses, and several heuristics were used to assign each clause a reference time value **tval**, the concept of reference time being the one proposed by Reichenbach (1947). Then they trained a statistical classifier to order the event denoted by a clause with respect to the **tval** associated with that clause (equivalent to classifying the temporal relation between them into one of the following classes: AT, BEF, AFT, or undefined). Based on the predictions made by the classifier, the authors inferred a partial ordering of the events belonging to the same document. This approach achieved 59% accuracy in assigning a **tval** to a clause, 84.6% accuracy in finding the temporal relation between an event and its associated **tval**, and 75.4% F-measure in partially ordering events whenever the temporal relations between them and their associated **tvals**, as well as the temporal order of the two **tvals** allowed a relation to be inferred.

In a later publication (Mani and Shiffman, 2005), the authors described their efforts of simplifying the task of manually and automatically annotating temporal relations by focusing on ordering pairs of consecutive clauses where either both clauses were in Past Tense, or the first clause was in Past Perfect and the second one in Past Tense. In their experiment, the temporal relations between the central events of the clause pairs exemplifying the two tense sequences were annotated by 8 subjects with one of the following six relations: *Entirely Before*, *Entirely After*, *Upto*, *Since*, *Equal*, *Unclear*. The inter-annotator kappa agreement (Cohen, 1960) for this annotation task was 0.5, the conclusion drawn by the authors being that such fine-grained distinctions are hard for people to make. The task was then simplified by collapsing the categories

Entirely Before and *Upto* into *BEF*, and *Entirely After* and *Since* into *AFT*, observing an increase in inter-annotator agreement to a kappa of 0.61. The annotated data was then used to train/test a classifier that achieved 58.07% in ordering two successive Past Tense clauses using the coarse-grained set of relations, and 70.38% in finding the temporal relation between a Past Perfect and a Past Tense clause.

Lapata and Lascarides (2004) associated the task of identifying sentence-internal temporal relations among clause pairs with the task of identifying the marker that had the highest probability of linking the two clauses. Several classification models were trained to predict one of the following temporal markers: *after*, *before*, *while*, *when*, *as*, *once*, *until*, *since*, and their accuracy was 70.7%. With this approach one is not able to directly derive the temporal relation between the two clauses, as some of the markers predicted by the models are ambiguous. In their later work, Lapata and Lascarides (2006) use these classification models to predict a reduced set of TimeML temporal relations comprising *BEFORE*, *INCLUDES*, *ENDS*, *BEGINS*, *SIMULTANEOUS*. Their evaluation results (45.8% accuracy) show that one can infer temporal information from corpora that is not semantically annotated in any way.

Chambers *et al.* (2007) described a two-stage machine learning architecture for the identification of event-event temporal relations. The first stage involved identifying the event attributes tense, aspect, modality, polarity and event class by training a Naïve Bayes classifier on TimeBank. The event attributes obtained at the first stage were then used together with other features in a second stage to classify the temporal relation between two events with an SVM classifier (Chang and Lin, 2011). The novelty of this work is the use at the second stage of imperfect feature values such as the ones obtained during the first stage, which still produces a small improvement (approx. 3%) over methods that used human-annotated feature values.

Other authors like Boguraev and Ando (2005), Vasilakopoulos and Black (2005), Li *et al.* (2004) have also explored the use of machine learning for temporal relation identification. Most of the work dedicated to automatic temporal relation identification has been stimulated by the two evaluation exercises TempEval and TempEval-2 organized as part of the SemEval 2007 and SemEval 2010 semantic evaluation campaigns.

3.3.1. *TempEval*

TempEval (Verhagen *et al.*, 2007) was the first evaluation exercise that focused on temporal relation identification. It was organized in the context of SemEval 2007, and consisted of three tasks that tested the capability of participating systems to relate an event and a TE located in the same sentence, an event and the TE representing the Document Creation Time (DCT), and two events located in consecutive sentences. The data used for this exercise consisted of a simplified version of TimeBank, in the sense that only certain events and event attributes were preserved, and a simplified set of temporal relations was used (consisting of: *BEFORE*, *AFTER*, *OVERLAP*, *BEFORE-OR-OVERLAP*, *OVERLAP-OR-AFTER*, and *VAGUE*). The test data included the events and temporal expressions together with their TimeML annotations, as well as pairs of temporal entities for which the temporal relation was supposed to be identified automatically. Six participating systems have approached the three TempEval tasks using different methods.

Four of the participating systems adopted a machine learning approach for solving the three tasks (Hepple *et al.*, 2007; Min *et al.*, 2007; Cheng *et al.*, 2007; Bethard and Martin, 2007), with the most popular classifier being Support Vector Machines. The feature engineering process employed in these approaches involved rules of varying complexities to derive values for their syntactic and semantic features not explicitly annotated in the data.

The other two participating systems took a rule-based approach, by relying on a deep syntactic analysis of the texts. XRCE-T, the system developed at XEROX (Hagege and Tannier, 2007), relied on an initial stage which performed the automatic detection and normalization of TEs and events, and on a syntactic analyzer that was extended to associate TEs with the events they modify just as thematic roles are attached to predicates. These associations were then used to order events in certain syntactic configurations. A more complex approach was adopted by the author of the present work in the system that achieved the best results at TempEval (Puşcaşu, 2007). The approach is described in detail in Section 5.

The lessons learned from TempEval were that some of the tasks were not well defined (Verhagen *et al.*, 2007; Verhagen *et al.*, 2009) and they proved difficult to carry out, as illustrated by the relatively low inter-annotator agreement (69%, 74% and 65% for each of the three tasks), and that the disjunctive and VAGUE labels should have not been included in the target set of temporal relations (Lee and Katz, 2009). The TempEval organizers considered that the definition of the first task should be changed from linking all events in a sentence with all TEs situated in the same sentence to more syntactically motivated subtasks that would link temporal entities according to syntactic considerations, such as syntactic dominance, argument structure and discourse structure. The results of the TempEval competition were also analyzed by Lee and Katz (2009) who concluded that only three labels should be used as target temporal relations: BEFORE, AFTER and OVERLAP. Some of these issues were addressed in TempEval-2.

3.3.2. *TempEval-2*

TempEval-2 (Verhagen *et al.*, 2010) was organized in the context of SemEval 2010 and consisted of six tasks. Unlike the first TempEval, the tasks now targeted not only temporal relations, but also the identification of TIMEX3 time expressions and of TimeML events. Four of the six tasks involved determining the temporal relations holding between an event and a TE situated within the same sentence with the restriction that the event syntactically dominates the TE, events and the DCT, main events in consecutive sentences, and two events where one syntactically dominates the other. Eight teams participated in this competition, but only three teams attempted all tasks, and it is surprising that their results did not demonstrate improvement over the results obtained in the first TempEval, despite certain tasks being simplified.

All the systems participating in TempEval-2 adopted a machine learning approach, either using Markov Logic (UzZaman and Allen, 2010; Ha *et al.*, 2010), Conditional Random Fields (Llorens *et al.*, 2010; Kolya *et al.*, 2010), or Maximum Entropy classifiers (Derczynski and Gaizauskas, 2010b). Adding more semantic-based features had a beneficial impact on system performance, but still the results were not encouraging. For example, the first task included in TempEval-1 that required systems to identify the temporal relation between events and TEs located in the same sentence was simplified in TempEval-2 by adding the restriction that the event and time expression had to be syntactically adjacent, with the event syntactically dominating the TE. The system performance was however very similar: the best score for this task was 0.62 in TempEval-1, and 0.65 for the simplified counterpart in TempEval-2.

In the following sections we argue that research in this area will not advance significantly without a sound understanding of the phenomena that are involved in establishing temporal relations. To this end we propose a shift in the way the temporal relation annotation problem is currently addressed to a more structured and linguistically motivated approach that looks at syntactically constrained subtasks.

4. Why Redefine the Problem?

4.1. *Scope Too Broad*

The manual annotation of TimeBank has revealed very low agreement (i.e. 55%) between annotators when they were asked to select which pairs of temporal entities should be involved in a temporal relation. This low agreement is due to giving the annotators the freedom to insert a temporal relation between any two temporal entities located anywhere in the document, without offering them appropriate guidance as to how those entities should be selected. A detailed analysis of their resulted annotation has shown that most temporal relations are established between entities located in the same sentence, as illustrated below.

The investigation of all the temporal relations annotated in TimeBank 1.2 (6,418 TLINKs) has revealed that approximately 60% of these relations link two temporal entities (events or TEs) situated in the same sentence. Another approximately 20% of the TLINKs annotated in TimeBank are relations between temporal entities and the Document Creation Time (DCT). The remaining percentage of temporal relations (approximately 20%) hold between temporal entities situated in different sentences. Table 1 captures the different types of temporal relations classified according to the categories and the relative position in the text of the two connected temporal entities.

The fact that human annotators link most frequently temporal entities situated in the same sentence via temporal relations is perfectly understandable, as local context provides many explicit clues as to what temporal relation holds between two entities. By broadening the context and increasing the textual distance between two entities,

Table 1. *Distribution of temporal relations in TimeBank 1.2*

Type of TLINK	Number of TLINKs
TLINKs between two events situated in the same sentence	2,368
TLINKs between an event and a TE from the same sentence	1,339
TLINKs between two TEs situated in the same sentence	28
TLINKs between an event and the DCT	1,275
TLINKs between a TE and the DCT	71
TLINKs between events situated in consecutive sentences	573
TLINKs between events situated in different sentences (more than one sentence apart)	540
TLINKs between an event and a TE located in different sentences	183
TLINKs between two TEs situated in different sentences	41

not only does one require more inferences to decide upon the temporal relation, but at the same time the chances of the two entities being temporally unrelated increase.

Therefore, sentence-internal temporal relation identification represents a significant problem that deserves exploration, and will be the focus of the research presented in this paper.

4.2. *High Number of Temporal Relations*

Assuming prior knowledge of the temporal entities that should be involved in a temporal relation, there is still an important problem that needs to be addressed, i.e. finding the temporal order of the two entities. The identification of the exact relation that holds between the two entities is not a straightforward task, its difficulty being also proven by the relatively low inter-annotator agreement achieved for the manual annotation of temporal relations. Given the initial TimeML set of 14 temporal relations derived from the original set of 13 interval-interval temporal relations distinguished by Allen (1983), the human agreement measured in the annotation of TimeBank with temporal relation types was 0.71 in terms of kappa agreement, and 0.77 in terms of percentage of the cases on which the annotators agree [1]. Despite the fact that the TempEval efforts were directed towards simplifying the task by defining a reduced set of five temporal relations, it was surprising to see lower inter-annotator agreement figures for the task of intra-sentential temporal relation annotation: 0.54 in terms of kappa agreement, and 0.69 in terms of percentage agreement.

During data annotation for TempEval, the task organizers noticed a small number of cases tagged by humans using the disjunctive relation labels BEFORE-OR-OVERLAP and OVERLAP-OR-AFTER. This small number was surprising especially as these labels were added to facilitate annotation in the cases when annotators faced difficulties in deciding between two temporal relations. The TempEval organizers also noticed far more disagreement than agreement in the case of the disjunctive relation types, thus raising the question of whether these labels are truly useful in a temporal relation annotation scheme. The poor distribution of these disjunctive

1. http://timeml.org/site/timebank/documentation-1.2.html.

labels in the training data, as well as the observed low system performance on these labels due to unclear guidelines as to when these labels should be used, all suggest using only three labels (OVERLAP, BEFORE and AFTER) in the task of temporal relation identification. Several authors (Verhagen *et al.*, 2009; Lee and Katz, 2009; Bethard, 2007) indicate that such a simplification would help drive research forward. We adopt the same perspective in this paper, by using only this minimal set of labels in our research, thus allowing a simplification of the annotation task down to a level that eliminates much of the confusion created by the extended set of temporal relations. The temporal vagueness present in natural language make subtle distinctions between fine-grained relations hard even for human annotators. The following examples extracted by Boguraev and Ando (2005) from the TimeBank annotation illustrate this difficulty in the case of two TimeML relations IS_INCLUDED and DURING:

[4.1] *In <the nine months>, net income <rose> 4.3% to $525.8...*
TLINK(*rose*, *the nine months*) = IS_INCLUDED

[4.2] *...said that its net income <rose> 51% in <the third quarter>.*
TLINK(*rose*, *the third quarter*) = DURING

These examples show very similar contexts that contain very similar temporal entities which are linked by human annotators using different temporal relations. Unfortunately this is not an isolated case, and many such inconsistencies are present in TimeBank, which clearly indicates that the fine-grained distinctions that would discriminate between the 14 TimeML relations are hard for people to make. If we move to less fine-grained relations by collapsing for example the relations IS_INCLUDED and DURING into a more general relation OVERLAP, the agreement between annotators would go up considerably, fact which was demonstrated by Mani and Shiffman (2005) in an experiment that involved linking pairs of adjacent Past Tense clauses. The authors report an increase of 0.11 in kappa agreement when switching from the fine-grained set of relations to the minimal set of only three relations. Such choices are commonly encountered in the annotation of complex phenomena. In many cases, theoretically important linguistic distinctions that are hard for people to make have been eliminated to improve annotation consistency. For example, the distinction between arguments and adjuncts has been avoided in the annotation of the Penn TreeBank. This simplification, on the one hand, had the advantage that it simplified the annotation task, the result being a more consistent annotation. But, on the other hand, certain information is left out and needs to be recovered later. The annotation of the Penn TreeBank also showed that by reducing the size of the tag set, one reduces the chance of tagging inconsistencies.

4.3. *Problematic Task Definition*

Another important problem posed by temporal relation identification was highlighted during the TempEval evaluation exercises, and involved the definition of the tasks. The low inter-annotator agreement observed during data annotation not only showed that humans cannot agree on the temporal relation to be assigned to a pair of temporal entities, but it was also an indicator of the performance level that can be expected from an automatic system that tries to solve the tasks at hand. It was proved once again that it is very complex to ask humans, let alone machines, to annotate temporal relations without imposing any constraints or predefined structure to the tasks, or without creating detailed guidelines. The tasks of identifying temporal relations, in the manner that they have been defined so far, give too much freedom and too little guidance to the annotators. Therefore, an important lesson learned from TempEval is that task decomposition is extremely advisable. Not only will clearer and focused task definitions facilitate a more reliable data annotation process, but it will also allow better system evaluation and error analysis in order to identify task-specific problems and solutions.

By focusing on different tasks with different scopes, TempEval-1 and TempEval-2 have tried to overcome the low agreement observed during the annotation of TimeBank for the problem of selecting which event pairs should be involved in a temporal relation (i.e. 55%). The only task that was successfully solved in both TempEvals (with an accuracy of around 80%) required the identification of temporal relations between any event and the DCT. Since the DCT can be associated to Reichenbach's Speech Time, tense and aspect are important cues in predicting the temporal relation between events and the DCT, offering a linguistically grounded solution to solving this task. This indicates that a better definition of the temporal relation identification problem should involve a more principled way of selecting and linking temporal entities. It is therefore desirable to divide the temporal relation identification problem into smaller tasks that have the potential of being resolved using linguistically motivated solutions. We suggest that the decomposition follows syntactic structure, given that syntactic information is relatively reliable, in comparison with deeper semantic information. After finding accurate and linguistically grounded solutions to these smaller tasks, one can then proceed to a higher level by composing more complex tasks once gaining knowledge as to how different temporal phenomena interact when they are grouped in complex utterances.

The present work aims at defining and investigating such syntactically motivated subtasks at intra-sentential level. Our focus is on finding computational solutions to the tasks/subtasks involved in solving the temporal relation identification problem. While the solutions presented in the remainder of this paper can prove useful in the temporal relation manual annotation process, they are mainly aimed at automatically solving the problems at hand.

5. Syntactically Motivated Staged Approach to Temporal Relation Identification

Section 4 has shown that the problem of temporal relation identification requires a better definition that should tackle the three main problems identified above:

1) **scope too broad**: to address this problem, this research focuses on intra-sentential temporal relations, thus narrowing the scope of the problem;

2) **high number of temporal relations**: this is addressed by focusing only the minimal set of temporal relations: BEFORE, OVERLAP and AFTER. However, the approach presented below can be further refined to deal with extended sets of relations;

3) **problematic task definition**: this aspect is tackled by dividing the temporal relation identification problem into smaller tasks that can be solved with linguistically informed solutions.

This section describes the syntactically motivated staged approach we use for automatic temporal relation identification at intra-sentential level. We employ a knowledge-intensive methodology that relies on a complex syntactic analysis of the text, and uses as an important source of information the evidence encoded in the syntax of a given sentence to derive temporal relations between co-sentential temporal entities. This methodology was devised following careful examination of the mechanisms used by language to express temporal relations (see Section 2 for more details), and it involves the following stages:

Stage 1: intra-clausal temporal ordering: each clause is individually processed to obtain a temporal ordering of the clause constituents by relying on the clause's syntactic tree. A recursive bottom-up process of finding the temporal order between directly linked syntactic constituents is employed at this stage;

Stage 2: inter-clausal temporal ordering: each pair of clauses involved in a syntactic dependency relation is involved in a recursive bottom-up process of finding the temporal order between directly linked clauses. At the end of Stages 1 and 2, each branch of the syntactic tree connecting a non-root node with its ancestor is labelled with a temporal relation;

Stage 3: ordering two co-sentential temporal entities: eligible pairs of temporal entities situated in the same sentence processed as above are temporally ordered by relying on an inference mechanism that links the two targeted temporal entities to their closest ancestor in the syntactic tree, and then to each other.

To be able to proceed with the processing involved in any of these stages, the following pre-processing steps are mandatory:

Preprocessing step 1: the sentence is first annotated with morpho-syntactic and functional dependency information by employing Connexor's FDG parser (Tapanainen and Jarvinen, 1997). For newspaper articles this parser reports a success rate of 96.4% at morpho-syntactic level and an F-measure of 91.45% when attaching heads in a dependency relation;

Preprocessing step 2: a clause splitting module is employed to detect clause boundaries and to establish the dependencies between the resulted clauses by

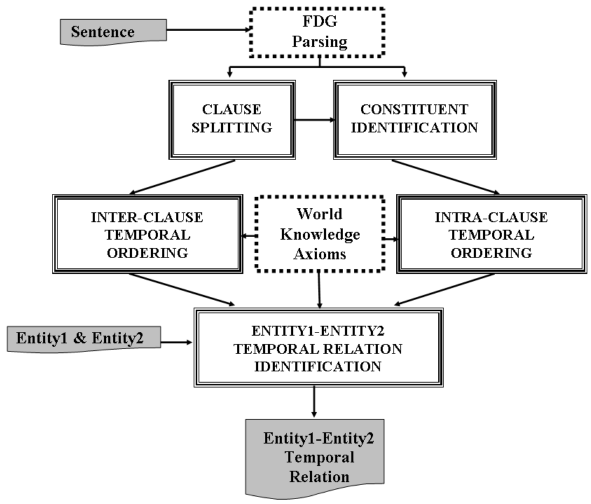

Figure 1. *Processing stages for the intra-sentential temporal relation identifier*

relying on formal indicators of coordination and subordination and, in their absence, on the functional dependency relation predicted by the FDG parser.

Figure 1 depicts the two preprocessing steps, and the three main stages involved in the identification of the temporal relation between any two co-sentential temporal entities.

The following sections describe each of the three stages involved in finding the intra-sentential temporal relations between any two temporal entities, and, wherever applicable, identifies the syntactically grounded tasks relevant for each stage.

5.1. *Stage 1: Intra-Clausal Temporal Ordering*

This stage begins by identifying the set of temporally relevant constituents present in each clause by examining the morpho-syntactic information provided by Connexor's FDG parser. The temporally relevant clause constituents are considered to be: the verb phrase VP, the noun phrases NPs, the prepositional phrases PPs, the non-finite verbs and the adverbial temporal expressions present in the analyzed clause. Given that events are mainly expressed using verbs, from this point forward we will use the term **central event of a clause** to refer to the event expressed by the verb that represents the head of a clause's VP (e.g. in the clause *John walked to the park*, the central event is *walked*). Events can also be expressed using nouns (e.g. *lecture*), and they can appear in clause constituents such as NPs or PPs. It is worth mentioning that there are NPs and PPs that do not include either noun events, or temporal expressions. Such constituents are still considered temporally relevant, as in many cases they are instrumental in linking for example temporal entities that they syntactically dominate with temporal entities syntactically dominating them. It is true that they can be interpreted as having no temporal value, but at the same time they can be associated at an abstract level with a temporal interval covering their existence. If we eliminate them from the syntactic path connecting other events and/or temporal expressions, we would lose important information that in certain cases can help relate other temporal entities. However, in most cases, their presence does not influence the process of identifying temporal relations, as they are typically leaf nodes in the syntactic tree.

The identified constituents and the syntactic tree of the corresponding clause are afterwards employed in a recursive bottom-up process of finding the temporal order between directly linked constituents. The leaf nodes are first linked to their immediate syntactic ancestors, then by going up the syntactic tree each non-leaf and non-root node is linked to its ancestor until there is a path of temporal relations from each leaf node up to the root of each clause's syntactic tree – the central verb phrase. Each constituent is linked only with the constituent it syntactically depends on using one of the predefined temporal relations.

The temporal relation between two constituents is decided on the basis of generally applicable heuristics that involve parameters such as: the types of the two constituents, the syntactic relation holding between them, the presence of certain temporal signals (e.g. prepositions like *before*, *after*, *until*, *since*), the tense of the clause's verb phrase, the semantic properties of the two constituents' heads (whether their root forms denote reporting or aspectual events – this is decided using the resource described in Puşcaşu and Barbu-Mititelu (2008)), and the temporal relation between any of the clause's temporal expressions and the DCT. The default temporal relation holding between any constituent and its syntactic ancestor is OVERLAP, but this relation is overriden whenever any of the parameters enumerated above indicate a different relation. For example, given the clause *he likes the silence before the storm*, the relation between *the storm* and *the silence* is imposed by the preposition *before*, *the storm* being thus temporally located AFTER *the silence*. Sense ambiguity of prepositions like *after* and *before* is not currently addressed, all their occurrences being considered temporal.

Following the entire recursive process of linking any two syntactically related clause constituents via a temporal relation, there is a path of temporal relations from any clause constituent to the clause's central VP. After this process has been applied to each clause in a given sentence, the next stage is inter-clausal temporal ordering.

This stage can be divided into the following syntactically grounded tasks:

[1] identification of intra-clausal temporal relations between a TE and a governing nominal event: this subtask targets temporal relations holding between a temporal expression and a nominal event, given that the nominal event syntactically dominates the temporal expression;

[2] identification of intra-clausal temporal relations between a TE and a dominating verbal event: this subtask looks at temporal relations holding between a temporal expression and a verbal event, given that the verbal event governs the temporal expression;

[3] identification of intra-clausal temporal relations between two events involved in a syntactic dependency relation: this subtask investigates temporal relations that hold between two events that are involved in a syntactic dependency relation, given that the two events are located in the same clause. In this case, we propose a further decomposition of the task according to the syntactic dependency relation holding between the two events, as follows:

- **[3.1] identification of intra-clausal temporal relations between two events, where one event is the direct object of the other event;**
- **[3.2] identification of intra-clausal temporal relations between two events, where one event is the indirect object of the other event;**
- **[3.3] identification of intra-clausal temporal relations between two events, where one event is the subject of the other event;**
- **[3.4] identification of intra-clausal temporal relations between two events, where one event is the object complement of the other event;**
- **[3.5] identification of intra-clausal temporal relations between two events, where one event is the subject complement of the other event;**
- **[3.6] identification of intra-clausal temporal relations between two events, where one event is the adverbial subordinate of the other event;**
- **[3.7] identification of intra-clausal temporal relations between two events in a coordination relation.**

Tasks [1], [2], and a representative subtask of [3] (i.e. [3.1]) will be experimented with in more detail in Section 6.

5.2. *Stage 2: Inter-Clausal Temporal Ordering*

At this stage, each pair of clauses involved in a dependency relation are temporally ordered. The information provided by the tenses of their VPs and by the dependency relation holding between the two clauses is very important for this process. The underlying hypothesis is that the clause binding elements and the tenses of the two central VPs provide a natural way to establish temporal relations between two syntactically related clauses. The property of the superordinate clause's main verb of being a reporting, aspectual or perception event is also relevant at this stage. The object clause of a reporting event is typically situated prior to the reporting event on a timeline, except in the cases where the object clause talks about a future event indicated by tense or by mentioning TEs situated in the future with respect to the

DCT. Aspectual events refer to stages in the evolution of an event, thus overlapping temporally with the event they take as object. The same applies to perceptual events, as the perceived event happens roughly at the time when it is perceived. The information on whether a verbal event has the aspectual class REPORTING, or ASPECTUAL, or PERCEPTION, or OCCURRENCE, or STATE can be reliably identified automatically (85% accuracy) by using a resource we previously developed (Puşcaşu and Barbu-Mititelu, 2008) that maps each WordNet verb to its aspectual class. The temporal expressions modifying the verb phrases of the two clauses involved in a syntactic relation can also help in relating the two clauses temporally.

At the end of the intra-clausal and inter-clausal processing stages, each branch of the syntactic tree connecting a non-root node with its ancestor is labelled with a temporal relation, information which will be used at the following stage to infer the temporal relation between any two temporal entities belonging to the same sentence.

The temporal relation between two clauses involved in a syntactic dependency relation derived at this stage is in fact the temporal relation between the central events of the two related clauses, so this stage is reduced to the following task:

[4] identification of inter-clausal temporal relations between the central events of two clauses involved in a syntactic dependency relation: in this case, we propose a decomposition of this task according to the dependency relation that links the two clauses, targeting the following cases:

[4.1] one clause is the direct object of the other clause;

[4.2] one clause is the indirect object of the other clause;

[4.3] one clause is the subject of the other clause;

[4.4] one clause is the object complement of the other clause;

[4.5] one clause is the subject complement of the other clause;

[4.6] one clause is a relative clause of the other clause;

[4.7] one clause is the prepositional complement of the other clause;

[4.8] one clause functions as an adverbial wrt the other clause: this task can be further subdivided according to the type of adverbial clause: [4.8.1] adverbial clause of time; [4.8.2] adverbial clause of place; [4.8.3] adverbial clause of contingency; [4.8.4] adverbial clause of condition; [4.8.5] adverbial clause of concession; [4.8.6] adverbial clause of contrast; [4.8.7] adverbial clause of exception; [4.8.8] adverbial clause of reason; [4.8.9] adverbial clause of purpose; [4.8.10] adverbial clause of result; [4.8.11] adverbial clause of comparison; [4.8.12] adverbial clause of proportion; [4.8.13] adverbial clause of preference;

[4.9] the two clauses are involved in a coordination relation.

Task [4.8.1] is the target of a detailed investigation and evaluation in Section 7.

5.3. *Stage 3: Ordering Two Co-Sentential Temporal Entities*

This stage involves retrieving the temporal relation between pairs of temporal entities situated in the sentence processed as above. We believe that not every pair of temporal entities in a sentence should be involved in a temporal relation, and our intuition is that discourse theories could prove very useful in investigating if two temporal entities are eligible to be linked with a temporal relation. The process of checking for eligibility could resemble the one performed in the task of anaphora/co-reference resolution, where a domain of referential accessibility is defined for every co-referential expression, domain which includes only the locations where one should search for an antecedent. In a similar way, a domain of temporal accessibility could be defined so that for every temporal entity we would know the locations (for example the clauses) where to look for other temporal entities that should be temporally ordered with respect to the initial entity. Unfortunately this problem has not received the attention it deserves, and no solution is currently available.

Therefore, at this stage we presently link any pair of temporal entities located in the same sentence. Given a pair of entities, they are first verified to see if the temporal relation can not be directly inferred from the information associated to them. For example, if one entity is a TE indicating a date that is previous to the DCT, and the other entity is an event expressed via a future tensed verb, then the temporal relation between the event and the TE is AFTER. If no relation can be inferred, a temporal reasoning mechanism is employed to relate the two targeted temporal entities to their closest syntactic ancestor, and then to each other using the syntactic tree where each branch is labelled with a temporal relation.

The system implementing the staged approach described above for the identification of intra-sentential temporal relations took part in the TempEval evaluation exercise, being evaluated along with other systems in an independent setting for three different tasks, as described in Puşcaşu (2007). According to the TempEval evaluation results (Verhagen *et al.*, 2007), it achieved the highest strict (0.62) and relaxed scores (0.64) for the task of intra-sentential temporal ordering. This result is close to the upper bound of human agreement on this task, which was 0.69. However, given that the scope of this evaluation included only relations between a TE and an event, only the tasks [1] and [2] from Stage 1 received most attention during system development. Despite the fact that the entire framework presented in this section was implemented, certain tasks such as task [3] from Stage 1, or task [4] from Stage 2 were simplistically approached with solutions that did not cater for all the temporal phenomena involved. To address this issue, representative sub-tasks from [3] and [4] are the target of a thorough analysis in Sections 6 and 7, respectively. Section 6 also includes a detailed investigation of tasks [1] and [2] from Stage 1.

6. Experiments Involving Intra-Clausal Syntactically-Constrained Subtasks

6.1. *Identification of Intra-Clausal Temporal Relations Between a TE and a Governing Nominal Event*

This experiment evaluates the performance of the system implementing the methodology described in Section 5 on the task of identifying the temporal relation between a TE and a governing event expressed using a noun. If the nominal event is the immediate ancestor of the TE and no preposition intervenes between them, the relation is always OVERLAP. If a preposition links the TE to the event, the preposition indicates the temporal order of the two entities. In all other cases, the relation is propagated recursively in a bottom-up manner in the syntactic tree.

For this task, the accuracy of the system is measured using two different settings. The data used for evaluation is a simplified version of the TempEval data in the sense that each relation initially annotated with BEFORE-OR-OVERLAP or OVERLAP-OR-AFTER is now converted by a human annotator into one of the three core relations: OVERLAP, BEFORE or AFTER.

In the first evaluation setting, only the temporal expressions directly dependent on a noun event are looked at. Direct dependency includes the cases when the temporal expression modifies the noun either directly (*the Monday lecture*) or via a preposition (*the lecture on Monday*). The system identifies the correct temporal relation between the TE and the nominal event it directly modifies in 100% of the cases (45 out of 45 cases are correctly identified). This high accuracy is not surprising given the fact that, in the absence of a preposition, the TE that modifies the noun indicates the time when the noun event took place, thus always yielding the OVERLAP relation between the two temporal entities. Whenever a preposition intervenes between the TE and the noun it modifies, this preposition indicates the temporal order of the two entities.

The second evaluation setting allows any number of dependency links on the syntactic path between the TE and the noun it directly or indirectly modifies. This means that the TE is syntactically governed by the nominal event, any number of words (including 0) being allowed on the syntactic path linking the two entities. The entities are only restricted to being situated in the same clause. The TempEval data includes 73 such cases, and the system identifies the correct temporal relation for 68 of them with an accuracy of 93.15%. Errors are caused by the system's lack of semantic knowledge and by the syntactic parser in building the dependency tree.

Inter-annotator agreement was measured for this task, and the annotators agree on the temporal relation in 98.6% of the cases, yielding a kappa agreement of 0.95.

6.2. *Identification of Intra-Clausal Temporal Relations Between a TE and a Dominating Verbal Event*

A similar methodology as in the case of nominal events is used here. The accuracy of the system is also measured using two different settings. The first setting evaluates only the assignment of temporal relations for TEs and verbal events in cases where the TE is directly linked to the verbal event. There are 330 such cases annotated in the TempEval data, the evaluation showing that the system is able to correctly identify the temporal relation holding between 304 verb-TE pairs. Therefore, the system performance in this setting is 92.12%. The largest source of errors (46.15%) arises from wrong PP-attachment in the cases where the TE is preceded by a preposition and the resulted PP is incorrectly linked to that verbal event. Another important source of disagreement is caused by wrong human annotations (26.92%). Vagueness and inaccessible semantic information account for the remaining errors.

The second setting relaxes the constraints imposed on the dependency between the TE and the event, allowing syntactic paths of variable lengths between the two temporal entities. Out of 520 cases, the system correctly assigns a temporal relation to 441, the accuracy being 84.80%. The errors produced by the system are mainly due to the lack of semantic information and world knowledge involving the words situated on the path between the TE and the verbal event. Some errors are introduced by the syntactic parser due to generating incorrect syntactic trees.

Inter-annotator agreement was measured for this task, and the annotators agree on the temporal relation in 94.23% of the cases, yielding a kappa agreement of 0.88.

6.3. *Identification of Intra-Clausal Temporal Relations Between Two Events Involved in a Syntactic Dependency Relation*

This experiment starts by investigating the event pairs located in the same clause and involved in a syntactic relation. The syntactic parser and the annotation present in TimeBank reveal 1,615 event pairs located in the same clause and involved in a syntactic relation. The TimeBank annotation shows that in 1,206 cases there is no temporal relation annotated for the syntactically related event pairs. An analysis of these cases reveals the following:

– **866 cases linked by SLINK:** the TimeML tag SLINK is used to annotate subordination relations between events (such as modal, evidential, factive, counter-factive, etc.). For example, given the context *John* ***said*** *that he* ***taught*** *on Monday*, the event *said* is a REPORTING event, and an SLINK of type EVIDENTIAL captures the subordination relation between *said* and *taught*, as evidence that the event of teaching really happened. However, the TimeML guidelines do not specify if events linked by an SLINK should also be annotated with a TLINK to indicate the temporal relation between them. Due to this under-specification of the guidelines, annotators either annotate an SLINK or a TLINK between a reporting event and its subordinate event.

We think that the temporal relations between such event pairs are important, and they should be annotated;

– 113 cases linked by ALINK: the TimeML tag ALINK is used to annotate the relation between an aspectual event and its subordinate event, as in *John* ***started*** *to* ***read***, an ALINK of type INITIATES connects the event *started* with the event *read*. As in the case of SLINKs, the TimeML guidelines do not specify if events linked by an ALINK should also be annotated with a TLINK. Therefore, human annotators either annotate an ALINK or a TLINK between an aspectual event and its argument event;

– 80 cases of coordination: a coordination syntactic relation links two events which are temporally unrelated, and for this reason the temporal relation is not annotated;

– 40 authentic cases of missing relations: these are cases where temporal relations should have been annotated between pairs of events, but were missed by the annotators;

– 38 cases which should be linked by SLINK: these cases include subordination relations of the type covered by SLINK, but neither an SLINK, nor a TLINK are annotated for these event pairs;

– 33 cases of wrong syntax the two events are not involved in any syntactic relation, but the parser wrongly links them;

– 18 cases which should be linked by ALINK: these cases include relations between an aspectual event and its argument (of the type covered by ALINK), but neither an ALINK, nor a TLINK are annotated for these event pairs;

– 18 cases where one of the events is odd: the annotation included in TimeBank marks certain prepositions, or numeric values as events. 18 such odd events are found as direct objects of other events, and no temporal relation is present between them, as it is the case of the event *$212* and the event *totaled* in the context *losses* ***totaled $212*** *million*.

There are only 409 event pairs that are linked by a temporal relation in TimeBank. Provided that this work does not focus on investigating which entities should be linked with a temporal relation, but its main aim is to identify the temporal relations between given pairs of temporal entities, only the pairs of syntactically related events involved in temporal relations according to the TimeBank annotation are further considered. A closer look at these pairs of events reveals that the most frequent syntactic relation linking them is the OBJ relation indicating the fact that one event is the direct object of the other. The OBJ relation is present in 55.50% of the cases (227 pairs out of the 409 annotated in TimeBank). Given its prominence among intra-clausal event-event syntactic relations, the OBJ relation is chosen as the focus of this experiment.

By analyzing the automatically identifiable sources of information available to us (e.g. the TimeML annotation of events and TEs, the morpho-syntactic information characterizing the sentence), the following rules are designed to identify the temporal relation between two events $EVENT_1$ and $EVENT_2$, where $EVENT_2$ is the direct object of $EVENT_1$:

[R1] if $EVENT_1$ is a REPORTING event, then $EVENT_2$ BEFORE $EVENT_1$ if no explicit TE indicating a future time is syntactically dependent on $EVENT_2$, otherwise the relation is $EVENT_2$ AFTER $EVENT_1$;

[R2] if $EVENT_1$ is an ASPECTUAL event, then $EVENT_2$ OVERLAP $EVENT_1$;

[R3] if $EVENT_1$ is a PERCEPTION event, then $EVENT_2$ OVERLAP $EVENT_1$;

[R4] if $EVENT_1$ is an OCCURRENCE event and $EVENT_2$ is an infinitive verb, then $EVENT_2$ AFTER $EVENT_1$;

[R5] if $EVENT_1$ is an OCCURRENCE event and $EVENT_2$ is a progressive verb, then $EVENT_2$ OVERLAP $EVENT_1$;

[R6] if $EVENT_1$ is an OCCURRENCE event and $EVENT_2$ is a noun and $EVENT_1$ is a synonym of the verb *to cause* [2], then $EVENT_2$ AFTER $EVENT_1$;

[R7] if $EVENT_1$ is an OCCURRENCE event and $EVENT_2$ is a noun and $EVENT_1$ is not a causal event, then $EVENT_2$ OVERLAP $EVENT_1$;

[R8] if $EVENT_1$ is a STATE event, then $EVENT_2$ OVERLAP $EVENT_1$, irrespective of $EVENT_2$ being a noun or an infinitive or progressive verb.

These rules are applied in the order they are presented, and they were evaluated on data extracted from the original TimeML annotation of TimeBank 1.2. For the event pairs linked by an OBJ relation, we extract the temporal relation annotated in TimeBank. The extracted temporal relations are mapped to the core set of only three temporal relations (OVERLAP, BEFORE, AFTER) by automatically mapping:

– SIMULTANEOUS, INCLUDES, IS_INCLUDED, DURING, DURING_INV, BEGINS, BEGUN_BY, ENDS, ENDED_BY and IDENTITY to OVERLAP;

– BEFORE and IBEFORE to BEFORE;

– AFTER and IAFTER to AFTER.

Inter-annotator agreement figures for this task show that the annotators agree on the temporal relation in 89.42% of the cases, with a kappa agreement of 0.74. The system that implements the rules described above correctly identifies the temporal relation between an event and its direct object subordinate event in 81.49% of the cases (185 out of 227 cases). The system encounters problems in the case of noun events being the direct object of OCCURRENCE events. These problems arise from the system's lack of semantic and world knowledge. In example [6.1], the event *calls* is the direct object of the event *return* and is temporally situated BEFORE it on a

2. Causality is dealt with in this work using a simplistic approach, by considering cause events to be expressed by the verb *to cause* or by any of its synonyms present in the *Roget's 21st Century Thesaurus*.

timeline. The system erroneously labels the temporal relation between the two events as OVERLAP. The system has access to very little semantic information, and this is one of its limitations.

[6.1] *Crane officials didn't <return> phone <calls> seeking comment.*

Another problem that appears in the case of two events linked by the OBJ dependency relation applies to aspectual events and their direct object dependents. The human annotation present in TimeBank is not consistent in such cases. For example in the case of the pair *<stop> <originating>* the temporal relation present in TimeBank is OVERLAP. However, in the similar case of the pair *<stopped> <providing>*, the annotated temporal relation between *providing* and *stopped* is BEFORE. Since the TimeML guidelines are rather unclear and lack a high level of detail, it is not surprising that many inconsistencies can be found in the annotation. This is one good reason to split the temporal relation annotation task into smaller subtasks, and include observations relevant for each subtask in the annotation guidelines, with a view towards achieving a high level of inter-annotator agreement, and at the same time facilitating the development of better systems.

7. Experimenting with an Inter-Clausal Syntactically Motivated Subtask

7.1. *Identification of Inter-Clausal Temporal Relations Between the Central Events of Two Clauses Involved in a Syntactic Dependency Relation*

The experiment described below focuses on pairs of clauses where one clause is a circumstance adverbial clause of time, or in other words a temporal clause, subordinated to the other. Temporal clauses bring into focus a novel temporal referent whose unique identifiability in the reader's memory is presupposed (Moens and Steedman, 1988), thus updating the current reference time (Reichenbach, 1947).

Temporal clauses are automatically identified in the context of this research using a machine learning approach that disambiguates the subordinating conjunctions used to introduce them (Puşcaşu *et al.*, 2006).

Temporal clauses are regularly marked by subordinators (e.g. *after*, *as*, *as/so long as*, *as soon as*, *before*, *once*, *since*, *until/till*, *when*, *whenever*, and *while/whilst*), many of which are ambiguous (e.g. *as*, *as/so long as*, *since*, *when*, and *while/whilst*), being able to introduce clauses of different semantic roles.

For the purpose of identifying temporal clauses by training classifiers capable of distinguishing between temporal and non-temporal usages of ambiguous subordinators, several classes of features have been designed: collocation features, features that characterize the VP of the subordinate clause (SubVP), and the VP of the main clause (MainVP), VPConnection features (capture regularities in the MainVP-SubVP tense pairs, as well as whether identical lemmas are present in the main and subordinate clauses), co-occurrence features (indicate whether certain subordinator-specific phrases appear in the scope of each feature to point to a certain semantic role, for example, in the case of *while/whilst*, a co-occurrence feature corresponding to the main clause span illustrates the presence within this span of *yet, besides, on the other hand, instead, nevertheless, moreover*, which point to a non-temporal usage), structural feature (captures the position of the subordinate clause wrt the main clause, as well as any punctuation signs separating the two clauses), FDG-relation (the relation predicted by the FDG parser for that particular clause pair). To identify the most appropriate model for the disambiguation of each subordinator, several feature combinations have been evaluated to compare the relevance of various feature classes to the classification of each temporal connective. The best model in the case of the subordinator *since* was the one including MainVP features, combined with VPConnection, structural and co-occurrence features, the correct distinction between temporal and non-temporal *since* being made in 98.33% of the cases. The best classifier for *when* combines the features corresponding to both VPs, VPConnection and co-occurrence features with an accuracy of 92.79%. Since each subordinator is disambiguated with its own classifier (the one offering the best performance for that subordinator), we call them **personalized classifiers**. The macro average accuracy of the personalized classifiers across all investigated connectives is 89.23%.

These classifiers are employed to identify temporal clauses present in TimeBank. To this end, the first step is extracting from TimeBank all potential temporal clauses introduced by any ambiguous or non-ambiguous temporal connective, together with their superordinate clauses. They are selected automatically so that the verb phrase of the subordinate clause is directly dependent on the verb phrase of the main clause, and so that the relation between the two clauses (relation provided by the syntactic parser) is of adverbial nature (e.g. a subordinate clause that is the direct object of the main clause is not considered).

The subordinate clauses corresponding to each ambiguous connective (e.g. *as*, *as/so long as*, *since*, *when*, and *while/whilst*) are manually annotated as temporal or non-temporal to be able to evaluate the performance of the machine learning algorithm mentioned above on this test data. A closer look at the TimeBank clause pairs reveals that there are 65 clauses introduced by ambiguous subordinators, out of which 33 are temporal clauses, and the rest non-temporal. One possible baseline for the task of distinguishing between temporal and non-temporal clauses would be to consider all clauses temporal, this yielding an accuracy of 50.76%. The system using personalized classifiers for each ambiguous subordinator correctly classifies 58 clauses as temporal or non-temporal, thus achieving a score of 89.23% and bringing a substantial improvement over the baseline.

The temporal clauses identified by the personalized classifiers, together with the temporal clauses introduced by non-ambiguous temporal connectives, form the focus of the current experiment targeting the identification of the temporal relation holding between the central events of two clauses where one clause is the temporal adjunct of the other clause.

The following rules were designed to infer the temporal relation between a temporal clause (CLAUSE_2) introduced by a subordinator (ambiguous or non-ambiguous) and its superordinate clause (CLAUSE_1):

[R1] the temporal relation between a temporal clause introduced by *after* and its main clause is BEFORE (CLAUSE_2 BEFORE CLAUSE_1);

[R2] according to Thompson (2005), temporal clauses introduced by *as* force the adjunct event time to be interpreted as simultaneous with the time of the matrix event (the event of the superordinate clause CLAUSE_1), therefore the temporal relation between a temporal *as*-clause and its superordinate clause is OVERLAP (CLAUSE_2 OVERLAP CLAUSE_1);

[R3] the temporal relation between a clause introduced by *as long as* or *so long as* and its main clause is OVERLAP (CLAUSE_2 OVERLAP CLAUSE_1);

[R4] in the case of *as soon as*, the action in the subordinate clause is temporally located BEFORE the action described by the main clause (CLAUSE_2 BEFORE CLAUSE_1);

[R5] a temporal clause introduced by *before* is always temporally AFTER its matrix clause (CLAUSE_2 AFTER CLAUSE_1);

[R6] *since* temporal clauses are temporally BEFORE their main clauses (CLAUSE_2 BEFORE CLAUSE_1);

[R7] the temporal relation between a temporal clause introduced by *until* or *till* and its superordinate clause is AFTER (CLAUSE_2 AFTER CLAUSE_1);

[R8] if the temporal clause CLAUSE_2 is introduced by *when*, the following parameters are important for deciding the temporal relation between the two clauses: the tense and aspect of the two verb phrases, the aspectual event types of the main events heading each clause, as well as the relative textual position of the subordinate clause with respect to the main clause;

[R8.1] if the aspect of the main verb phrase is Perfect, the presence of the aspectual morpheme *have* orders the Event Time of the main clause event as preceding the Reference Time that is normally modified by the temporal clause, thus situating the main event time before the subordinate event time. In this case the temporal relation between CLAUSE_2 and CLAUSE_1 is AFTER (CLAUSE_2 AFTER CLAUSE_1);

[R8.2] in the absence of the aspectual morpheme *have* from both clauses (i.e. both verb phrases are characterized by simple or progressive tenses), *when*-clauses are ambiguous in that they permit either a simultaneous or non-simultaneous reading. If the grammatical aspect of either verb phrase is Progressive, or the aspectual class of either head event is STATE, then the temporal relation between CLAUSE_2 and CLAUSE_1 is OVERLAP (CLAUSE_2 OVERLAP CLAUSE_1);

[R8.3] otherwise, in the absence of both the Perfective and the Progressive aspect from the two clauses, a *when*-clause preceding the main clause has only a non-simultaneous reading, according to Thompson (2005). In such cases the temporal relation between CLAUSE_2 (the *when*-clause) and CLAUSE_1 is BEFORE (CLAUSE_2 BEFORE CLAUSE_1);

[R8.4] for any remaining cases (simple tenses, non-stative events, and the *when*-clause is either embedded or after the main clause), it will be assumed that the temporal relation is also BEFORE, even if in reality this is not always the case, but due to the limitations of the system in accessing deeper semantic information, this will be the default behaviour (CLAUSE_2 BEFORE CLAUSE_1);

[R9] temporal clauses introduced by *whenever* receive the same treatment as *when*-clauses;

[R10] temporal *while*-clauses are contemporaneous with their matrix clauses, the temporal relation that applies to them being OVERLAP (CLAUSE_2 OVERLAP CLAUSE_1).

The temporal relation between the two clauses represents also the temporal relation between their central events.

These rules above are applied in the order they are presented. The data for evaluating the system performance on this task is obtained from TimeBank in a similar manner to the data for the previous task of intra-clausal event-event temporal relation annotation. Only the core set of three temporal relations is used.

When looking at the temporal relations between the temporal clauses and their superordinate clauses, one discovers that out of the 33 temporal clauses introduced by ambiguous subordinators, 12 are not linked via any temporal relation to the head of the main clause in the annotation present in TimeBank. Since this work does not focus on establishing which pairs of entities should be involved in a temporal relation, only the pairs of clauses linked by a temporal relation in TimeBank are further considered for system evaluation. Out of 21 clause pairs linked via a temporal relation in TimeBank, the system identifies the correct temporal relation using the methodology described in Section 5.2 in 17 cases, meaning that in 80.95% of the cases it identifies the correct temporal relation.

Besides clauses introduced by ambiguous subordinators, TimeBank also includes 29 temporal adverbial clauses introduced by non-ambiguous subordinators. In 12 cases, no temporal relation between the subordinate and main clause is annotated in TimeBank. Out of the 17 cases with a temporal relation associated in TimeBank, the system correctly specifies the temporal relation in 14 cases. The 3 errors appear in the case of *until*-clauses, as the system always labels a temporal relation between the *until*-clause and the main clause with AFTER, while the human annotators marked 3 cases with OVERLAP. The cases annotated with OVERLAP are most probably human annotation errors that could be avoided by specifying clearer and more detailed annotation guidelines for smaller and more specific tasks.

The baseline for this task involves assigning the most frequent relation encountered in the data (BEFORE) to all clause pairs. This baseline achieves a score of 55.26%. The overall system performance in identifying the temporal relation holding between a temporal clause and its matrix clause is 81.57% (31 correctly identified temporal relations for the 38 clause pairs linked via a temporal relation in TimeBank).

Inter-annotator agreement was measured for this task, and the annotators agree on the temporal relation in 89.47% of the cases, with a kappa agreement of 0.83.

8. Conclusion

This work was motivated by the belief that a more comprehensive study of temporal phenomena guided by syntactic structure would provide a sound understanding of temporal relations. Syntax and temporality are strongly intertwined, and while we believe that syntax alone cannot provide all the answers that would lead to a sound understanding of temporal relations, it can however prove instrumental in guiding the discovery process. Low human annotator agreement together with lessons learned from participating in TempEval have also suggested that it is advisable to decompose the annotation task into more linguistically motivated and structured subtasks, not only for a better understanding of the temporal phenomena that contribute to establishing a temporal relation, but also for improving the annotation process by adopting a staged approach. To this end, several syntactically motivated subtasks were investigated in this work. Evaluation results have shown that such subtasks can be reliably resolved by an automatic system which was able to identify the correct temporal relation in more than 80% of the investigated cases. The subtasks evaluated in Sections 6 and 7 are representative for the types of tasks one can encounter when dealing with intra-sentential temporal relation identification, and we are confident that similar results as those obtained for the investigated subtasks can be achieved for the other enumerated intra-clausal and inter-clausal tasks, except the tasks that involve coordination (i.e. tasks [3.7] and [4.9]). These two tasks can prove problematic without access to semantic information, since not all event pairs involved in a coordination relation are temporally related. In the future, an interesting research direction would be to investigate connections between syntax and temporality in contexts where the syntactically motivated subtasks outlined in this paper interact and form more complex tasks.

Acknowledgements

I would like to thank Sandra Kübler for her comments and suggestions on an earlier draft. My gratitude is also extended to the three anonymous reviewers, whose comments and recommendations have helped improve this article. I would also like to thank Iustina Ilisei and Lionel Nicolas for their help with the French translation.

9. References

ACL-2001, "Workshop on Temporal and Spatial Information Processing", 2001.

Allen J., "Maintaining Knowledge About Temporal Intervals", *Communications of the ACM*, vol. 26, p. 832-843, November, 1983.

Alonso O., Gertz M., Baeza-Yates R., "On the Value of Temporal Information in Information Retrieval", *ACM SIGIR Forum*, vol. 41, December, 2007.

ARTE, "ACL 2006 Workshop on Annotating and Reasoning About Time and Events", 2006.

Bethard S., "Finding Event, Temporal and Causal Structure in Text: A Machine Learning Approach", PhD thesis, University of Colorado at Boulder, December, 2007.

Bethard S., Martin J. H., "CU-TMP: Temporal Relation Classification Using Syntactic and Semantic Features", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 129-132, June, 2007.

Boguraev B., Ando R. K., "TimeBank-Driven TimeML Analysis", *in* G. Katz, J. Pustejovsky, F. Schilder (eds), *Dagstuhl International Workshop on Annotating, Extracting and Reasoning About Time and Events*, Dagstuhl Seminar Proceedings, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2005.

Boguraev B., Ando R. K., "Analysis of TimeBank as a Resource for TimeML Parsing", *Proceedings of LREC 2006: 5th International Conference on Language Resources and Evaluation*, Genova, Italy, p. 71-76, 2006.

Chambers N., Wang S., Jurafsky D., "Classifying Temporal Relations Between Events", *Proceedings of the ACL 2007 Demo and Poster Sessions*, Association for Computational Linguistics, Prague, p. 173-176, June, 2007.

Chang C.-C., Lin C.-J., "LIBSVM: a Library for Support Vector Machines", *ACM Transactions on Intelligent Systems and Technology*, 2011.

Cheng Y., Asahara M., Matsumoto Y., "NAIST.Japan: Temporal Relation Identification Using Dependency Parsed Tree", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 245-248, June, 2007.

Cohen J., "A Coefficient of Agreement for Nominal Scales", *Educational and Psychological Measurement*, vol. 20, p. 37-46, 1960.

Dagstuhl, "Dagstuhl International Seminar on Annotating, Extracting and Reasoning About Times and Events", 2005.

Derczynski L., Gaizauskas R., "Analysing Temporally Annotated Corpora with CAVaT", *in* N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (eds), *Proceedings of LREC 2010: 7th Conference on International Language Resources and Evaluation*, European Language Resources Association (ELRA), Valletta, Malta, 2010a.

Derczynski L., Gaizauskas R., "USFD2: Annotating Temporal Expresions and TLINKs for TempEval-2", *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 337-340, July, 2010b.

Filatova E., Hovy E., "Assigning Time-Stamps to Event-Clauses", *Proceedings of the ACL 2001 Workshop on Temporal and Spatial Reasoning*, Toulouse, France, 2001.

Ha E., Baikadi A., Licata C., Lester J., "NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology", *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 341-344, July, 2010.

Hagege C., Tannier X., "XRCE-T: XIP Temporal Module for TempEval Campaign", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 492-495, June, 2007.

Hepple M., Setzer A., Gaizauskas R., "USFD: Preliminary Exploration of Features and Classifiers for the TempEval-2007 Tasks", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 438-441, June, 2007.

ISO-TimeML, "Language Resource Management – Semantic Annotation Framework (SemAF) – Part 1: Time and Events", 2007.

Katz G., Arosio F., "The Annotation of Temporal Information in Natural Language Sentences", *Proceedings of the ACL 2001 Workshop on Temporal and Spatial Information Processing*, Association for Computational Linguistics, Toulouse, France, p. 104-111, 2001.

Kolya A.-K., Ekbal A., Bandyopadhyay S., "JU_CSE_TEMP: a First Step Towards Evaluating Events, Time Expressions and Temporal Relations", *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 345-350, July, 2010.

Lapata M., Lascarides A., "Inferring Sentence-Internal Temporal Relations", *Proceedings of the North American Chapter of the Association for Computational Linguistics*, p. 153-160, 2004.

Lapata M., Lascarides A., "Learning Sentence-Internal Temporal Relations", *Journal of Artificial Intelligence Research*, n° 27, p. 85-117, 2006.

Lee C. M., Katz G., "Error Analysis of the TempEval Temporal Relation Identification Task", *Proceedings of the NAACL HLT Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, Association for Computational Linguistics, Boulder, Colorado, p. 138-145, June, 2009.

Li W., Wong K.-F., Cao G., Yuan C., "Applying Machine Learning to Chinese Temporal Relation Resolution", *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, p. 582-588, 2004.

Llorens H., Saquete E., Navarro B., "TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2", *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 284-291, July, 2010.

LREC-2002, "Workshop on Annotation Standards for Temporal Information in Natural Language", 2002.

Mani I., Schiffman B., Zhang J., "Inferring Temporal Ordering of Events in News", *Proceedings of the Human Language Technology Conference (HLT-NAACL'03)*, 2003.

Mani I., Shiffman B., "Temporally Anchoring and Ordering Events in News", *in* J. Pustejovsky, R. Gaizauskas (eds), *Time and Event Recognition in Natural Language*, John Benjamins, Amsterdam, 2005.

Min C., Srikanth M., Fowler A., "LCC-TE: a Hybrid Approach to Temporal Relation Identification in News Text", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 219-222, June, 2007.

Moens M., Steedman M., "Temporal Ontology and Temporal Reference", *Computational Linguistics*, vol. 14, p. 15-28, June, 1988.

Moldovan D., Clark C., Harabagiu S., "Temporal Context Representation and Reasoning", *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2005)*, p. 1099-1104, 2005.

Puşcaşu G., "WVALI: Temporal Relation Identification by Syntactico-Semantic Analysis", *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 484-487, June, 2007.

Puşcaşu G., Barbu-Mititelu V., "Annotation of WordNet Verbs with TimeML Event Classes", *Proceedings of the 6th Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, p. 2793-2800, 2008.

Puşcaşu G., Martinez-Barco P., Saquete-Boro E., "On the Identification of Temporal Clauses", *Proceedings of the 5th Mexican International Conference on Artificial Intelligence (MICAI 2006)*, Apizaco, Mexico, p. 911-921, 2006.

Pustejovsky J., Castano J., Ingria R., Saurí R., Gaizauskas R., Setzer A., Katz G., "TimeML: Robust Specification of Event and Temporal Expressions in Text", *Proceedings of IWCS-5: 5th International Workshop on Computational Semantics*, 2003.

Pustejovsky J., Verhagen M., Saurí R., Littman J., Gaizauskas R., Katz G., Mani I., Knippen R., Setzer A., "TimeBank 1.2", , Linguistic Data Consortium LDC2006T08, 2006.

Reichenbach H., *Elements of Symbolic Logic*, The Macmillan Company, New York, 1947.

Saurí R., Littman J., Knippen B., Gaizauskas R., Setzer A., Pustejovsky J., "TimeML Annotation Guidelines Version 1.2.1", 2006.

Schilder F., Habel C., "From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages", *Proceedings of the ACL 2001 Workshop on Temporal and Spatial Information Processing*, Association for Computational Linguistics, Toulouse, France, p. 65-72, 2001.

Setzer A., "Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study", PhD thesis, University of Sheffield, 2001.

Setzer A., Gaizauskas R., "On the Importance of Annotating Temporal Event-Event Relations in Text", *Proceedings of the LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language*, p. 52-60, 2002.

Sundheim B., Chinchor N., "Survey of the Message Understanding Conferences", *Proceedings of the Workshop on Human Language Technology*, Association for Computational Linguistics, 1993.

Surdeanu M., Harabagiu S., Williams J., Aarseth P., "Using Predicate-Argument Structures for Information Extraction", *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, p. 8-15, July, 2003.

TANGO, “TimeML Annotation Graphical Organizer, ARDA Workshop on Advanced Question Answering Technology”, 2003.

Tapanainen P., Jarvinen T., “A Non-Projective Dependency Parser”, *Proceedings of the 5th Conference of Applied Natural Language Processing*, Washington DC, USA, p. 64-71, March, 1997.

TERQAS, “Time and Event Recognition for Question Answering Systems, ARDA Workshop on Advanced Question Answering Technology”, 2002.

Thompson E., *Time in Natural Language: Syntactic Interfaces with Semantics and Discourse*, Walter de Gruyter, 2005.

TIME-2006, “International Symposium on Temporal Representation and Reasoning”, 2006.

UzZaman N., Allen J., “TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text”, *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 276-283, July, 2010.

Vasilakopoulos A., Black W. J., “Temporally Ordering Event Instances in Natural Language Texts”, *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2005)*, 2005.

Verhagen M., Gaizauskas R., Schilder F., Hepple M., Katz G., Pustejovsky J., “SemEval-2007 Task 15: TempEval Temporal Relation Identification”, *Proceedings of the 4th International Workshop on Semantic Evaluation (SemEval-2007) at ACL 2007*, Association for Computational Linguistics, Prague, Czech Republic, p. 75-80, June, 2007.

Verhagen M., Gaizauskas R., Schilder F., Hepple M., Moszkowicz J., Pustejovsky J., “The TempEval Challenge: Identifying Temporal Relations in Text”, *Language Resources and Evaluation*, vol. 43, p. 161-179, 2009.

Verhagen M., Saurí R., Caselli T., Pustejovsky J., “SemEval-2010 Task 13: TempEval-2”, *Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval-2010) at ACL 2010*, Association for Computational Linguistics, Uppsala, Sweden, p. 57-62, July, 2010.