Full-fledged temporal processing: bridging the gap between deep linguistic processing and temporal extraction

Francisco Costa and António Branco
University of Lisbon

ABSTRACT

The full-fledged processing of temporal information presents specific challenges. These difficulties largely stem from the fact that the temporal meaning conveyed by grammatical means interacts with many extra-linguistic factors (world knowledge, causality, calendar systems, reasoning). This article proposes a novel approach to this problem, based on a hybrid strategy that explores the complementarity of the symbolic and probabilistic methods. A specialized temporal extraction system is combined with a deep linguistic processing grammar. The temporal extraction system extracts eventualities, times and dates mentioned in text, and also temporal relations between them, in line with the tasks of the recent TempEval challenges; and uses machine learning techniques to draw from different sources of information (grammatical and extra-grammatical) even if it is not explicitly known how these combine to produce the final temporal meaning being expressed. In turn, the deep computational grammar delivers richer truth-conditional meaning representations of input sentences, which include a principled representation of temporal information, on which higher level tasks, including reasoning, can be based. These deep semantic representations are extended and improved according to the output of the aforementioned temporal extraction module. The prototype implemented shows performance results that increase the quality of the temporal meaning representations and are better than the performance of each of the two components in isolation.

Keywords: temporal processing, temporal extraction, tense, aspect, hybrid approaches, deep linguistic processing, shallow linguistic processing

Journal of Language Modelling Vol 1, No 1 (2013), pp. 97–154

1 INTRODUCTION

Deep linguistic processing aims at providing grammatical representations of sentences, including their full-fledged semantic representations. This is undertaken by computational grammars whose handcrafted rules encode the regularities uncovered by theoretical linguistics. Deep natural language processing systems have been successfully employed in many applications, like machine translation (Müller and Kasper, 2000; Bond *et al.*, 2005), grammar checking (Bender *et al.*, 2004) and ontology acquisition (Nichols *et al.*, 2006), among others.

While these grammars typically deliver precise linguistic analyses and fine-grained semantic representations of given sentences, they perform less well when it comes to resolving ambiguity and getting at the appropriate representation of a sentence given its context of occurrence. The inverse tension is observed in shallow processing systems. Often resorting to statistical methods, these systems are very helpful at resolving ambiguity, but they perform much worse when it comes to getting at the sophistication of deep semantic representations.

The linguistic expression of time forms a highly intricate semantic subsystem that offers a particularly good illustration of the complementarity between the two approaches and the gap to bridge. Like in any other grammatical dimension, here too ambiguity is pervasive, and each sentence in isolation may bear different temporal readings.

Deep grammars typically handle such proliferation of readings by resorting to some underspecification formalism that allows for its packing. Although this makes it possible to address the efficiency problems associated with this ambiguity, rule-based grammars offer limited means to resolve this ambiguity and to support real-world applications that need to rely on the actual temporal information conveyed by sentences in their contexts.

The area of temporal information extraction, greatly fostered by the TempEval challenges (Verhagen *et al.*, 2007, 2010), has encouraged the development of systems able to extract from texts important pieces of information concerning time. But there is so far little or no exploration of how to combine them with the deep principled semantic representations of the sentences, so that they can help support higher-level temporal processing and reasoning systems. In the opposite direction, much of the sophisticated linguistic information that may be important to improve the accuracy of temporal information extraction is also waiting to be explored.

This paper explores the complementarity of the two approaches, drawing inspiration from other efforts of hybrid natural language processing, such as Crysmann *et al.* (2002) and Frank *et al.* (2003), among others. Our exercise is circumscribed here to the processing of temporal information. A proposal is presented that contributes to an enhanced processing of time by bridging the gap between temporal information extraction and deep linguistic processing.

More specifically, in this paper, we seek to incorporate the temporal information extracted by a system specialized in the processing of temporal information (and developed with machine learning methods) in the meaning representations produced by a deep processing grammar, resulting in semantic representations enriched with more information about time. Our motivation is partly due to the fact that the processing of the expression of time in natural language interacts with a number of extra-linguistic systems (such as calendar systems, or knowledge about the world) that are best handled outside a computational grammar. One example is the processing of temporal expressions, such as *today*, *the fifth of May*, or *two days later*. It may be preferable to compute the exact date that these expressions refer to outside the grammar. Their processing requires access to arithmetic operations (e.g. once the anchor date for the last of these example expressions is determined, it is necessary to add two days to it) and to a calendar system (e.g. so that we know that subtracting two days from March 1, 2012 gets us to February 28, 2012), which grammar formalisms are typically not designed to support.

Many different kinds of information are needed to accurately determine temporal relations conveyed in natural language text. Linguistic knowledge is obviously important, at various levels (lexical, morphological, syntactic, semantic). For instance, aspectual type, which distinguishes various types of eventuality descriptions, such as states, activities, accomplishments and achievements (Vendler, 1967; Dowty, 1979), is partly lexical but also interacts with syntax and semantics, and it features prominently in the semantics literature on the expression of time in natural language. But extra-linguistic knowledge of different sorts also comes into play, such as:

- Pragmatics and knowledge about the world. Relations like causation can override default constraints on interpretation. Like the examples of Lascarides and Asher (1993) show, the chronological order of events can be reflected in the order in which they are presented in text, (1a), but causality relations between the mentioned events can override that preference, (1b):

(1) a. Max stood up. John greeted him.
    b. Max fell. John pushed him.

- Calendar systems. Time expressions like *next Monday* or *two months earlier* must be interpreted relative to a calendar system, and furthermore there are implicit temporal relations between the referred dates and times that can be be made explicit and explored.
- Logical inference. For instance, temporal precedence is transitive, so sometimes the possible temporal relations conveyed in a piece of text are restricted by what has occurred before (e.g. if event A precedes B, a new event C that precedes A must also precede B).

All these factors are important and should be explored to leverage a temporal extraction system. However, they are difficult to handle in grammar formalisms and grammar development environments.

In this paper, a computational grammar that delivers detailed representations of the meaning of input sentences is extended with a representation of time and aspect, in order to enrich these meaning representations. This extension is based on the linguistic literature on tense and aspect, and it was also developed in such a way that the resulting meaning representations are straightforward to combine with the output of the temporal extractor.

Subsequently, the deep grammar and the temporal extractor are combined, with the purpose of extending and correcting the semantic representations delivered by the grammar as far as temporal meaning is concerned. This combination of the computational grammar with the dedicated temporal extraction system allows the meaning representations produced by the grammar to be improved in the following ways:

- Extending the representations
  It is possible to add further temporal information (that the gram-

mar does not have access to) to the meaning representations output by the grammar. One example is the normalization of temporal expressions (determining the exact date or time that they refer to), which in deep natural language processing systems are often processed separately, for instance by a pre-processing module. In our case, we use the temporal extractor, as it already deals with these expressions, thus avoiding the replication of this functionality in a pre-processing component.

- Specifying the representations
  The meaning representations are in many cases underspecified. When the temporal extractor produces more specific output, the grammar representations can also be made more specific, in accordance with this output.
- Correcting the representations
  Since the grammar only looks at grammatical information, while the temporal extraction system is sensitive to other kinds of information (as hinted at above), it is often more accurate than the grammar in resolving time-related ambiguity. Its output can thus be used to correct meaning representations.

As discussed in this paper, one obtains better and more detailed meaning representations with this combination.

The deep grammar we use is LXGram, presented in Section 2.1. The temporal extractor is LX-TimeAnalyzer, described in Section 2.2.3. The data we use to train LX-TimeAnalyzer, as well as for the evaluation reported here, is TimeBankPT (it is divided into a training set and a test set), which we introduce in Section 2.2.2. The particular systems and data that we used represent the Portuguese language, but the key issues stand for other languages as well: the expression of time and its interface with all these extra-linguistic kinds of knowledge and the meaning representations of time.

We evaluate the performance of this extractor system, the computational grammar, and a combined system that incorporates the two on a common data set. The combined system allows for full-fledged temporal processing and outperforms both the temporal extractor and the deep grammar.

This paper is structured as follows. The next section introduces the key topics that will be dealt with in the remainder of the paper:

deep linguistic processing, hybrid natural language processing, temporal information processing, and the semantics of tense and aspect. The particular systems that we used are also presented. The following sections explain how these elements will work together. Section 3 describes how a deep grammar can be extended to include information about time in the meaning representations that it produces. Section 4 describes and evaluates an approach to integrate the deep processing grammar and the temporal information extractor, and combine their contributions for the processing of the linguistic expression of time. Finally, in Section 5 this article closes with final remarks.

2 BACKGROUND

This section introduces the key elements that will be integrated, with the purpose of combining temporal information extraction and deep semantic representations: a deep grammar that produces such representations (Section 2.1), and temporal information extraction technology, which identifies and normalizes events, dates and times mentioned in a text and classifies temporal relations holding between these entities (Section 2.2). Additionally, we present previous work in the area of hybrid processing (Section 2.3) as well as in the area of the semantics of tense and aspect (Section 2.4).

2.1 *Deep linguistic processing*

Deep linguistic processing grammars associate each input sentence with its grammatical representation, including a representation of its meaning. For the sake of the research exercise reported in this article, LXGram was chosen as the working grammar. LXGram is a deep grammar for Portuguese (Costa and Branco, 2010a).

This grammar is based on the Head-Driven Phrase Structure Grammar (HPSG) grammatical framework (Pollard and Sag, 1994; Sag *et al.*, 2003). HPSG resorts to a unification-based grammatical representation formalism with a type system featuring multiple inheritance and recursive data structures called typed feature structures.

LXGram is implemented in the LKB (Copestake, 2002), an integrated development environment for typed feature structure grammars in general, popular within the HPSG community. The grammar runs on PET (Callmeier, 2000), an efficient parser for HPSG grammars, that allows several input methods, including interfacing with external morphological analyzers, which we make use of. These systems also allow the training and use of a statistical model to discriminate between competing analyses for each sentence (Oepen *et al.*, 2002; Toutanova *et al.*, 2005; Velldal, 2007). This facility is also used with LXGram to rank the parses produced for a given sentence. The grammar outputs all possible parses for a given input sentence, and this model selects the most probable one. The model is trained on CINTIL TreeBank, a treebank obtained by manually selecting the best parse from those produced by the grammar (Silva *et al.*, 2012). Around 2,000 sentences of newspaper text from this treebank were used to train the model.

When run over unrestricted newspaper text, LXGram produces a parse for about 30% of the input sentences, and the disambiguation model correctly identifies the preferred analysis for around 40% of these parsed sentences (Costa and Branco, 2010a). Despite this coverage, we will see below that it already produces very competitive results with temporal processing and results above the state of the art when combined with the shallow temporal extractor.

LXGram explores the core Grammar Matrix system (Bender *et al.*, 2002), which contains a set of implemented grammatical constraints relevant to many languages, following the HPSG framework. It employs Minimal Recursion Semantics (MRS; Copestake *et al.* 2005) as the formalism for the semantic representations it produces.

An important feature of MRS is that it supports underspecified semantic representation. An MRS representation is a tuple containing a global top, a bag of relations labeled with handles and a bag of constraints on handles. Relations labeled with handles are called *elementary predications*, but we will also refer to them as relations in this article. Conjunction is represented by shared labels. Handles can also appear as arguments of these relations, and they are used to represent scope. The main kind of constraint on handles is equality modulo quantifiers ($=_q$), which means that either the two handles are the same handle or one or more quantifier relations (but no relation of a different kind) intervene between the two. They enable the underspecification of the scope between the various relations. An example

MRS representation for the sentence *A black cat can fly* is:[1]

$$
\begin{array}{l}
<h1,\\
\quad \{h2 : \_a\_q(x3, h4, h5),\ h6 : \_black\_a(x3),\\
\quad\ h6 : \_cat\_n(x3),\ h7 : \_can\_v(h8),\\
\quad\ h9 : \_fly\_v(x3)\},\\
\quad \{h1 =_q h7, h4 =_q h6,\ h8 =_q h9\} >
\end{array}
$$

This representation corresponds to the two scoped formulas that can be obtained from it by scope resolution:

- $\_a\_q(x3,\ \_black\_n(x3) \wedge \_cat\_n(x3),\ \_can\_v(\_fly\_v(x3)))$
  (There is a black cat that possibly flies.)
- $\_can\_v(\_a\_q(x3,\ \_black\_n(x3) \wedge \_cat\_n(x3),\ \_fly\_v(x3)))$
  (It is possible that there is a black cat that flies.)

This is how the scope ambiguity between the existential quantifier and the modal operator is captured. The first reading is obtained when the constraints on the handles are resolved this way: $h1 = h2,\ h4 = h6,\ h5 = h7,\ h8 = h9$. The second one is when $h1 = h7,\ h8 = h2,\ h4 = h6,\ h5 = h9$.

MRS representations are straightforwardly encoded in the typed feature structures manipulated by HPSG grammars. For the sake of readability of this text, we abstain from presenting them in that format.

For the purpose of experimentation, a concrete grammar has to be used. As will be apparent, the solutions put forth are tested with this working grammar but their principles can be easily adapted or transferred to other deep computational grammars delivering an underspecified semantic representation, developed under other grammatical frameworks or for other languages.

Existing computational HPSG grammars typically do not include the meaning representation of tense and aspect in the semantic repre-

[1] We follow the convention of including part-of-speech-inspired labels in the names of the relations in an MRS representation: *n* for relations denoted by nouns, *a* for those related to adjectives and adverbs, *q* in quantifier relations, *v* in verbal relations, etc.

sentations they produce. But because MRSs are used by applications[2] and this sort of information is important even if provided in a very approximate way, a common approach is to enrich the output MRSs with information about grammatical tense and aspect. For instance, the MRS representation for our working sentence *A black cat can fly* often looks like:

$$\begin{array}{l} <h1, \\ \quad \{h2 : \_a\_q(x3, h4, h5),\ h6 : \_black\_a(x3), \\ \quad\ h6 : \_cat\_n(x3),\ h7 : \_can\_v(e10\{tense\ present\}, h8), \\ \quad\ h9 : \_fly\_v(e11, x3)\}, \\ \quad \{h1 =_q h7, h4 =_q h6,\ h8 =_q h9\} > \end{array}$$

Here, two event variables have been added to the relations for *can* and *fly*, an approach similar to that of Davidson (1967). These event variables can have features of their own. The one for *can* has a *tense* feature with the value *present*. This is an indication of the verb tense used in the verb form corresponding to this relation.

This approach, which is common to several existing computational HPSG grammars, has the disadvantage of mixing semantic information with grammatical information. This mixing is undesirable, because semantic representations are supposed to explicitly describe truth conditions, which grammatical categories fail to do. The motivation for our work is also to eliminate grammatical information from semantic representations, as far as tense and aspect are concerned.

2.2 *Temporal information extraction*

There is a long research tradition on extracting the information about time that is conveyed in natural language text. Some recent evaluation campaigns have given it more attention, as they focused precisely on this task. They include TempEval (Verhagen *et al.*, 2007), TempEval-2 (Verhagen *et al.*, 2010), and TempEval-3 (UzZaman *et al.*, 2013). Besides encouraging work on the topic, the TempEval campaigns have provided data that can be and has been explored to develop and evaluate systems that automatically annotate natural language text with the temporal information they convey.

[2] Machine translation is one application where MRS representations have been extensively used, in this case as the level to which transfer rules apply (Flickinger *et al.*, 2005; Nygaard *et al.*, 2006; Nichols *et al.*, 2007).

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

Figure 1: Simplified sample of the annotations in TempEval for the fragment: *In Washington today, the Federal Aviation Administration released air traffic control tapes from the night the TWA Flight eight hundred went down*

These data are annotated with an annotation scheme similar to TimeML (Pustejovsky *et al.*, 2003a). Figure 1 shows a small, simplified extract of the data from the first TempEval challenge, with TimeML-style annotations.

The words that denote events are annotated using `EVENT` tags. An example is the word referring to the event of the FAA's releasing of the tapes. `EVENT` tags are also employed to annotate words denoting states (such as the situations denoted by verbs like *love* or *want*). For this reason, in this context the terms *event*, *situation*, and *eventuality* are employed interchangeably in this paper, to refer to states and events. This use of the term *event* is common in the literature on temporal extraction.

The `TIMEX3` tags surround temporal expressions, such as *today*. In this working example, the temporal expression *today* denotes the date normalized as `1998-01-14`. The attribute `value` of `TIMEX3` elements holds this normalized representation.

The `TLINK` elements at the end describe temporal relations between events and dates, times or other events. For instance, the event of the plane going down is annotated as temporally preceding the date denoted by the temporal expression *today*.

The first two TempEval challenges had as their main tasks the automatic identification of the temporal relations. That is, the value of the `relType` attribute of the `TLINK` elements (such as the ones in Figure 1) had to be determined, and all other annotations were given. Temporal relation classification is also the most interesting problem in temporal information extraction. The other tasks that are necessary to automatically annotate text with TimeML (identifying and normalizing temporal expressions and events) show better evaluation results, and they also have a longer research history.

TempEval featured three tasks: A, B and C.[3] Task A was about classifying the temporal relation that holds between an event and a time mentioned in the same sentence (they could however be far apart in the sentence, as the temporal relation represented by the `TLINK` with the `lid` with the value `l1` in Figure 1). Task B focused on the temporal relation between events and the document's creation time, which is also annotated in TimeML (not shown in that figure). Task C was about classifying the temporal relation between the main events of two consecutive sentences. The goal of all these tasks was to determine the type of a given temporal relation. The possible values for the type of relations are `BEFORE`, `AFTER` and `OVERLAP`, as well as `BEFORE-OR-OVERLAP`, `OVERLAP-OR-AFTER` and `VAGUE`, but the last three values occur very infrequently in the annotated data that were made available for TempEval.

2.2.1 State of the art in temporal information extraction

Table 1 shows a synopsis of the results of the first two TempEval competitions, taken from Verhagen *et al.* (2009, 2010), for the main tasks of classifying temporal relations. The data used in these two competitions are similar but not identical, hence the different baselines.

This table does not show the results of TempEval-3, because they are so difficult to compare to previous work: (i) the training data set used is substantially larger (twice the size), (ii) the evaluation setup is different (in TempEval-3, the temporal relation classification tasks are performed from raw text; in the first two TempEval competitions, the remaining gold annotations were given to participants), (iii) the inventory of relation types is different, (iv) and the evaluation measure is also different – the temporal awareness score of UzZaman and Allen (2011) is used instead of classification accuracy.

[3] TempEval-2 had additional tasks, about identifying and normalizing events and temporal expressions. It also had an additional temporal relation classification task, about pairs of events mentioned in the same sentence. Furthermore, the names of the tasks in TempEval-2 are different. We use the names employed in TempEval.

Table 1:
Results for English in TempEval

		Task		
		A	B	C
TempEval	Best system	**0.62**	**0.80**	**0.55**
	Avg. of all participants	0.56	0.74	0.51
	Majority class baseline	0.57	0.56	0.47
TempEval-2	Best system	**0.65**	**0.82**	**0.58**
	Avg. of all participants	0.61	0.78	0.53
	Majority class baseline	0.55	0.59	0.49

Table 2:
Results for English in TempEval-2: temporal expressions and events (F-measure for extent recognition and accuracy for the attributes)

Temporal expressions					
	Extents	`type`	`value`		
Best system	**0.86**	**0.98**	**0.85**		
Avg. of all participants	0.78	0.86	0.57		
Median	0.82	0.91	0.59		
Events					
	Extents	`class`	`tense`	`aspect`	`polarity`
Best system	**0.83**	**0.79**	**0.92**	**0.98**	**0.99**
Avg. of all participants	0.74	0.72	0.75	0.97	0.99
Median	0.79	0.77	0.86	0.97	0.99

TempEval-2 also evaluated the recognition of temporal expressions and events (i.e. identifying their extents in text) and their normalization (filling in the various attributes of the `EVENT` and `TIMEX3` elements visible in Figure 1). A synopsis of the results is in Table 2. The averages of all participants reported in this table are affected by a few extremely low scores; therefore we also show the median values.

The several systems that participated in the first two TempEval challenges resorted to different methods. There were symbolic solutions as well as machine learning approaches. Different levels of linguistic analysis, ranging from shallow processing, such as POS-tagging, to full syntactic parsing, were explored as a means to provide information used in rules or as classifier features. This variety of approaches can also be seen amongst the best systems of TempEval-2. The TRIPS and TRIOS systems (UzZaman and Allen, 2010) used a combination of parsing and machine learning methods such as conditional random fields (Lafferty *et al.*, 2001) and Markov logic networks

(Richardson and Domingos, 2006). TIPSem (Llorens *et al.*, 2010a) also used conditional random fields trained using several kinds of features, including features extracted from the output of a syntactic parser, namely that of Charniak and Johnson (2005) for English. Like UzZaman and Allen (2010), the NCSU systems (Ha *et al.*, 2010) employed Markov Logic using features taken from different natural language processing tools. Ha *et al.* (2010) gave a bigger emphasis to features that capture lexical relations between the event terms involved (such as similarity relations between *producing* and *creating* events, antonymy relations between the terms *open* and *close*, etc.).

TimeML, the TimeBank (Pustejovsky *et al.*, 2003b) – a TimeML annotated corpus which served as the basis for the data used in TempEval – and the TempEval challenges have been very influential in the area of temporal information extraction. The work of Denis and Muller (2010) offers a comparison of the set of temporal relations considered in TimeML and other temporal algebras developed earlier, namely those of Allen (1983, 1984) and Bruce (1972).

Also, a lot of recent work has used the TimeBank and the data sets made available in the two TempEval challenges. Verhagen and Pustejovsky (2008) present a system that automatically annotates raw text with TimeML, including annotations for events, time expressions and temporal relations. Chambers *et al.* (2007) trained machine learning classifiers on the TimeBank, namely Naïve Bayes (John and Langley, 1995) classifiers. They were concerned with temporal relations between pairs of events, which could be in the same sentence or not. Their system's goal intersects Task C of the first TempEval challenge (relations between events in different sentences). Their algorithm operates on two stages. In the first stage, they try to learn some properties of the events in the temporal relation, such as tense, grammatical aspect and aspectual class. Here they use some morpho-syntactic features as well as features based on information provided by WordNet (Fellbaum, 1998). In the second stage, they classify the temporal relation between those events. They use as classifier features the information obtained in the first stage, as well as other kinds of features based on the syntactic structure of the sentences where the events are mentioned. Llorens *et al.* (2010b), similarly to Llorens *et al.* (2010a), explore the contribution of semantic role labeling to temporal information processing.

Machine learning methods have become dominant in addressing the problem of extracting the temporal ordering of what is described in a text. One major limitation of machine learning methods is that they are typically used to classify temporal relations in isolation, and therefore it is not guaranteed that the resulting temporal ordering is globally consistent. Yoshikawa *et al.* (2009) and Ling and Weld (2010) seek to overcome this limitation using Markov logic networks, which learn probabilities attached to first-order formulas. Some of the participants of the second TempEval used a similar approach (Ha *et al.*, 2010; UzZaman and Allen, 2010). Denis and Muller (2011) cast the problem of learning temporal orderings from texts as a constraint optimization problem. They search for a solution using Integer Linear Programming (ILP), similarly to Bramsen *et al.* (2006), and Chambers and Jurafsky (2008). Because ILP is costly (it is NP-hard), the latter two only consider *before* and *after* relations. Rather than classifying a temporal relation between two time intervals, Denis and Muller (2011) and Lee (2010) classify four relations between four instants (the endpoints of the two original time intervals). Symbolic or hybrid approaches have also been used. This was the case of the WVALI system (Puşcaşu, 2007), one of the participants of the first TempEval competition and the one with the best results for some of the tasks.

The logical properties of temporal relations make temporal information processing stand out from many of the other natural language processing tasks. UzZaman and Allen (2011) propose a new way to evaluate temporal information processing systems. Instead of the usual precision and recall metrics used in the first two TempEval competitions, they argue that it is better to compute the temporal closure of the reference annotations and confront the result with a system's output. This is because a system may identify temporal relations that are not part of the reference annotations but nevertheless are logical consequences of the ones that are in fact annotated.

Despite the prominence of the TimeML annotated data sets mentioned earlier (the TimeBank and the data sets of the TempEval challenges) and the plethora of work using them, there are further resources with temporal annotations. One is the WikiWars corpus (Mazur and Dale, 2010). Its scope is more limited than that of the TimeBank and the data used in the TempEval challenges, because it is annotated only for temporal expressions, leaving out events and temporal relations. The kind of task it supports is thus similar to the early efforts of the Temporal Expression Recognition and Normalization evaluation (Ferro *et al.*, 2004) and the previous Message Understanding Conferences (MUC-6, 1995; MUC-7, 1998), where a concerted effort for the annotation of time expressions first took place. Yet another corpus of English featuring temporal annotations (Bethard *et al.*, 2007) contains annotated temporal relations between events denoted by words in a specific syntactic relation (one heads the clause that is the complement of the other one).

Work on the topics of temporal expression recognition (identifying the boundaries of temporal expressions in text) and normalization (assigning each of them a normalized representation of the time or date that they refer to) has produced quite good results for some time now (Negri and Marseglia, 2004; Strötgen and Gertz, 2013; Angeli *et al.*, 2012; Llorens *et al.*, 2012). Still, in recent years, the topics of temporal expression recognition and normalization have not been abandoned. WikiWars, just mentioned, is a recent corpus where time expressions are annotated. Other recent work on this topic includes that of Zhao *et al.* (2010). Additionally, there has been interest in new problems related to temporal expressions. Kolomiyets *et al.* (2011) investigate the portability of time expression recognition to non-newswire domains, since most of the annotated data consist of news articles (the exception being WikiWars). Their idea is to generate additional training examples by replacing temporal expression words with potential synonyms, taken from WordNet and other similar resources. This technique potentially increases the number of word types seen in training as part of a time expression.

2.2.2 Data with annotations about time

The data released in the first TempEval challenge were for English only. The second TempEval challenge released data for Chinese, English, French, Italian, Korean and Spanish (although only English and Spanish attracted participants to the competition). Since then, efforts to manually annotate temporal phenomena have continued for several languages (Pustejovsky and Stubbs, 2011; Xue and Zhou, 2010; Zhou and Xue, 2011), and a number of corpora featuring similar temporal annotations have been developed for several languages: Chinese

Table 3:
Size of the data set

	Train	Test
Sentences	2,281	351
Word tokens	60,782	8,920
Annotated events	6,790	1,097
Annotated temporal expressions	1,244	165
Annotated temporal relations		
Task A	1,490	169
Task B	2,556	331
Task C	1,735	258
Total	5,781	758

(Cheng *et al.*, 2008), French (Bittar *et al.*, 2011), Italian (Caselli *et al.*, 2011), Korean (Im *et al.*, 2009), Romanian (Forăscu and Tufiş, 2012).

For Portuguese, there is the TimeBankPT corpus (Costa and Branco, 2010b, 2012d). This corpus is an adaptation of the original TempEval data to Portuguese, obtained by translating it and then adapting the annotations. The two corpora – TimeBankPT and the original English data set used in the first TempEval challenge – are quite similar (Costa and Branco, 2012d), but the languages are of course different.

TimeBankPT is used here to train and evaluate the temporal information extraction component. Just like the original English corpus for TempEval, it is divided into a training part and a testing part. The original English corpus is composed of news documents. Many of these documents are taken from the Wall Street Journal, and they belong to the domain of economics. TimeBankPT is thus also composed of documents of this genre and domain. Some figures pertaining to the size of this data set are presented in Table 3.

2.2.3 LX-TimeAnalyzer

For the experiments reported in the present paper, an independent temporal extractor is used. It is called LX-TimeAnalyzer (Costa and Branco, 2012b,c) and annotates raw text with temporal annotations. These annotations are similar to the ones used in the first two TempEval challenges, based on TimeML (Pustejovsky *et al.*, 2003a), and illustrated in Figure 1 above. LX-TimeAnalyzer annotates raw text with events, temporal expressions, and temporal relations. This system runs on Portuguese input text, and it was trained with the data just presented above in Section 2.2.2.

In order to produce these annotations, several tasks are performed: (i) identifying temporal expressions and events mentioned in the text; (ii) normalizing these time expressions (annotating the `value` attribute of `TIMEX3` elements, where the date or time referred to by the temporal expression is recorded in a standardized format); (iii) filling in the values of the remaining attributes of the `EVENT` and `TIMEX3` elements that were recognized; (iv) identifying temporal relations, i.e. which pairs of entities (events and times) should be linked with temporal relations; and (v) classifying these temporal relations (overlap, precedence, etc.).

Most of these tasks are performed with machine learning classifiers trained on the training data of TimeBankPT. The tasks of normalizing temporal expressions and identifying temporal relations are performed by handcrafted rules, and most of the annotated attributes of `EVENT` elements are directly based on the output of other natural language processing tools, namely a part-of-speech tagger and morphological analyzer. The classifiers used to identify event terms and temporal expressions use features based on information that also comes from these tools (part-of-speech, lemma, inflectional features) and a context window of two words on each side of the target word. There is a dedicated machine learning classifier for the attribute `class` of `EVENT` terms.

The normalization of temporal expressions makes use of Joda-Time 2.0,[4] which implements calendar systems as well as many operations between dates (e.g. it can calculate that two days after February 28, 2013 is March 2, 2013).

The models that classify temporal relations are produced with machine learning classifiers that use several features that capture many types of information. These features are numerous, and for this reason it is not possible to provide a full account of them, which is presented in (Costa, 2013). Briefly, there are:

- Superficial features based on information from a part-of-speech tagger (e.g. the conjunction nearest the event that enters the temporal relation under classification);

[4] `http://joda-time.sourceforge.net`

- Features that encode information about logical inferences. For instance, we solve task B before the other two tasks, and sometimes information about task B temporal relations as well as the implicit temporal relations between the times and dates mentioned in the text can constrain the temporal relations in the other subsequent tasks;
- Fine-grained information about aspectual type. TimeML makes a distinction between states and non-states in the attribute `class` of `EVENT` elements. We explore a more fine-grained distinction, as we make use of four aspectual types, following the work of Vendler (1967) and Dowty (1979) (as well as the large body of literature that follows them) more closely;
- Information about the world (e.g. a verb like *predict* typically precedes in time what is predicted, but a verb like *report* typically follows in time what is reported).

Crucially, LX-TimeAnalyzer makes use of several pieces of extra-linguistic information, such as the logical constraints between temporal relations (when classifying temporal relations) or calendar systems (when normalizing temporal expressions), that are typically not available to a deep natural language processing system. Depending on the formalism employed in the implementation of a deep grammar, it may not even be feasible or practical to implement this kind of knowledge in such a system. It certainly is not possible in the LKB, where LXGram is implemented, but even if it were, there is still the question of whether it would be appropriate to encode extra-linguistic information in a deep grammar.

Evaluation results show that LX-TimeAnalyzer performance is at the level of the state-of-the-art for English (Costa and Branco, 2012b,c), except for the task of event detection (determining whether a given word token denotes an event). This problem is somewhat hard for nouns. The best system to identify events in the second TempEval resorted to, among other things, WordNet (Llorens *et al.*, 2010a), an approach that is not available for Portuguese currently, as there is no WordNet for this language with the breath and maturity of the English WordNet. This makes event identification harder for Portuguese (Costa and Branco, 2012c). Table 4 presents the evaluation results for LX-TimeAnalyzer, using the test data of TimeBankPT. The evaluation

Temporal expressions	Score	Events	Score	Temporal relations	Score
Extents	0.85	Extents	0.72	Task A	0.67
`type`	0.91	`class`	0.74	Task B	0.80
`value`	0.81	`tense`	0.95	Task C	0.55
		`aspect`	0.96		
		`polarity`	0.99		

Table 4: Performance of LX-TimeAnalyzer on the test data of TimeBankPT

measures reported in that table are the F-measure for the problems of identifying the extents of event terms and temporal expressions and accuracy for the remaining tasks. These results are very similar to the state of the art for English (cf. Table 1 and Table 2).

2.3 *Hybrid natural language processing*

The present paper follows a hybrid approach to natural language processing.

Within the HPSG community, we find, among others, the work of Adolphs *et al.* (2008), which allows the grammars developed in the LKB (presented above in Section 2.1) to see the output of shallow tools as AVMs (Attribute-Value Matrices, the data structures that HPSG grammars manipulate). This work builds on previous efforts to combine shallow and deep processing with HPSG, like the work of Crysmann *et al.* (2002) and Frank *et al.* (2003). Frank *et al.* (2003) combines a deep grammar with a shallower parser, resulting in efficiency gains of a factor of 2.25. Crysmann *et al.* (2002) additionally use shallow morphological analysis, part-of-speech tagging and named entity recognition to guess information about unknown words (words not in the lexicon of the deep grammar). This results in an increase in grammar coverage from 12.5% to 22.1%, on a corpus of 20,000 newspaper sentences.

Similar work is that of Schäfer (2006), who develops a software architecture designed to combine shallow and deep systems, with the purpose of making the deep systems more robust. The author shows that this approach increases the efficiency and the coverage of the deep system by a factor of more than two. Since then, hybrid techniques such as these have become popular within deep processing. LXGram uses a similar approach, where morphological information output by shallow tools is used to enable the grammar to process unknown words (though we do not use a shallow parser to improve efficiency).

Grover and Lascarides (2001) is an earlier work that also uses the morphological information coming from shallow tools to increase the robustness of a computational grammar, namely when it comes to dealing with out-of-vocabulary words.

In the Verbmobil project (Wahlster, 2000) on speech-to-speech translation, multiple parsers are used to aid machine translation. Several of them are run in parallel (a symbolic HPSG grammar, a statistical parser, and a chunker). They produce meaning representations in a common format. When the parsers fail to provide analyses that fully span an utterance, the fragments that they produce are combined, resulting in an analysis for the entire utterance (Rupp *et al.*, 2000).

Also in the context of the Verbmobil project, particularly relevant to our work is that reported in Alexandersson *et al.* (2000) and Stede *et al.* (1998). They extract mentions of times and dates from the semantic representations produced by the parsers and employ a specialized module to map these semantic representations to a canonical representation of these dates and times. Their work shows that recognizing temporal expressions can be done with a parser. However, like us, they consider that other problems, like this problem of temporal expression normalization, are best handled with external technology. In our work, where an existing and stand-alone temporal extraction system is available, it is not necessary to have the grammar recognize temporal expressions, since the extraction system (which must be used to normalize them anyway) already performs this task.

Within the Lexical-Functional Grammar (LFG) framework (Kaplan and Bresnan, 1982), Brun (1998) describes a pre-processing step where nominal multiword expressions as well as time expressions are recognized in the input that is to be subsequently parsed by a grammar. Named entity recognition has also been integrated in this pre-processing stage in several computational LFG grammars (Kaplan *et al.*, 2004; Butt *et al.*, 1999).

The approach we present in this paper is also a hybrid approach, where a deep grammar is combined with shallower tools. But in our case we combine information of a different kind. We are interested in putting together different methods to extract temporal relations from text: with the deep processing grammar, which looks exclusively at grammatical information, and with a dedicated temporal extraction system, which has access to extra-linguistic knowledge. Instead of using external tools to pre-process the grammar's input, we use the output of a tool specialized in temporal extraction to refine the grammar's output in a post-processing step. In our scenario, post-processing is preferred to pre-processing because: (i) the additional information that is being brought to the grammar is directly about meaning (i.e. it is about temporal relations and representations of the times denoted by time expressions); and (ii) the time expressions recognized and annotated by the temporal extraction system do not necessarily correspond to syntactic constituents.[5]

2.4 *The semantics of tense and aspect*

There is a vast body of linguistic literature on the semantics of tense and aspect. Our implementation of tense and aspect in the deep grammar, described below in Section 3, is inspired by previous work that we briefly describe in this section.

Davidson (1967) is the first author to reify events. In HPSG, this approach has been popularized in a number of analyses, including Sag *et al.* (2003), as well as in several HPSG implementations, like the English Resource Grammar (Flickinger, 2000) and the Grammar Matrix (Bender *et al.*, 2002). A survey of the advantages over the alternatives can be found in Kamp and Reyle (1993, pp. 504–10).

Reichenbach (1947) described tenses as temporal relations between several pairs of times, not just an event time and an utterance time (or speech time). In particular, he introduced the concept of a reference time that mediates the relation between those two times. This idea has been maintained in subsequent work by other authors.

[5] The system recognizes time expressions according to the TIMEX3 specification (Saurí *et al.*, 2006). Many TIMEX3 elements are syntactic constituents (for instance, many are noun phrases), but some elements of noun phrases (such as relative clauses) are left out of the annotated extents of these elements annotated with TIMEX3 tags, as the inclusion of such elements would make a parser necessary to determine these extents. If the annotated time expressions always corresponded to syntactic constituents, this information could be exploited in order to contrain the parser's search space. As they do not, there is no benefit in detecting them in a pre-processing step.

Some influential ideas originating in Discourse Representation Theory (DRT), of Kamp and Reyle (1993), have also crept into many analyses of tense. This is the case in the observation that past tense denotes overlap of the event time with a past time in the case of stative situations but inclusion in the case of non-stative situations.

Intricately related to tense is aspect. A large body of literature exists on this topic, with the work of Vendler (1967) and Dowty (1979) being seminal.

Pustejovsky (1991) posits a separate level of representation for the event structure associated with predicates and their arguments and advocates the decomposition of events into sub-events. For instance, a sentence like *the door closed* is analyzed as a process (*the door closing*) followed by a state (*the door is closed*). This is similar in spirit to the work of Moens and Steedman (1988).

In the framework of HPSG, Van Eynde (2000) develops an analysis for the Dutch tenses and temporal auxiliaries inspired by DRT in its semantic aspects. The work of Yoshimoto and Mori (2002) combines HPSG with a DRT analysis of tense. Bonami (2002) is an HPSG analysis of aspect shift inspired by the work of de Swart (1998, 2000). This phenomenon is treated by positing implicit aspectual operators, which we also resort to. Flouraki (2006) focuses on aspectual constraints on the various tenses of Modern Greek, modeling them with HPSG. Relevant to our work is also that of Goss-Grubbs (2005), which develops an analysis of tense and aspect for English using MRS. This work encodes aspectual type by typing event variables, and it also resorts to positing explicit aspectual operators in the semantic representations. It does not make use of explicit temporal relations or the various Reichenbachian times (reference time, speech time, etc.); instead it encodes tense as a feature of time variables.

Bobrow *et al.* (2007) is also similar work, inasmuch as it is about a computational system that produces meaning representations of its input which contain non-trivial information about time. In its representations, the system includes explicit temporal relations between events and the speech time. It does not, however, include information about aspect or make use of reference times.

3 DEEP PROCESSING OF TENSE AND ASPECT

A semantic representation for tense and aspect was implemented in the grammar that was presented above in Section 2.1, taking into account the possibility of it being extended with additional information relevant to time coming from temporal information extraction systems.

The grammar was extended with an implementation of tense and aspect inspired by much of the literature just referred to above. The following running example illustrates the various aspects of the implementation:

(2) *A atriz mudou-se de França para os Estados Unidos em*
the actress moved from France to the United States in
fevereiro de 1947.
February of 1947
The actress moved from France to the United States in February 1947.

The MRS representation for this sentence, as produced by the grammar, is shown in Figure 2. Temporal information can be seen in the *is-before* and at relations, that relate the event time $t9$ with the utterance time $t10$, and aspectual information can be seen in the *aspectual-operator* relations as well as the feature *culmination*, which indicates that the associated eventuality (the *moving* event) contains a culmination as one of its sub-events (i.e. it is a culmination or a culminated process).

The remainder of this section provides more details on the implementation of tense and aspect in the working grammar, and how they are reflected in the meaning representations such as the one in Figure 2.

3.1 *Tense*

It is important to distinguish between grammatical tense and semantic tense: we will use the first expression to refer to inflectional morphology alone, and the second one to refer to the temporal and aspectual meaning they convey.

Each predicate denoted by a verb, adjective, preposition or adverb receives a Davidsonian semantic representation, with an event

$$
\begin{aligned}
&<h1,\\
&\quad \{h3 : \_o\_q(x4, h5, h6),\\
&\quad\ h7 : \_atriz\_n(x4),\\
&\quad\ h8 : at(e2\ \{culmination\ +\}, t9),\\
&\quad\ h8 : is\text{-}before(t9, t10\ \{t\text{-}value\ utterance\text{-}time\}),\\
&\quad\ h8 : aspectual\text{-}operator(e2, e12, h11),\\
&\quad\ h11 : \_mudar\_v(e12, x4),\\
&\quad\ h11 : \_de\_p(e14, e12, x13),\\
&\quad\ h15 : proper\_q(x13, h16, h17),\\
&\quad\ h18 : named(x13,\ \text{“França”}),\\
&\quad\ h11 : \_para\_p(e20, e12, x19),\\
&\quad\ h21 : \_o\_q(x19, h23, h22),\\
&\quad\ h24 : named(x19,\ \text{“Estados Unidos”}),\\
&\quad\ h11 : \_em\_p(e26, e12, x25),\\
&\quad\ h27 : udef\_q(x25, h28, h29),\\
&\quad\ h30 : \_fevereiro\_n(x25),\\
&\quad\ h30 : \_de\_p(e31, x25, x32),\\
&\quad\ h33 : proper\_q(x32, h34, h35),\\
&\quad\ h36 : named(x32,\ \text{“1947”})\},\\
&\quad \{h1 =_q h8,\ h5 =_q h7,\ h16 =_q h18,\ h23 =_q h24,\ h28 =_q h30,\\
&\quad\ h34 =_q h36\} >
\end{aligned}
$$

Figure 2: MRS for *A atriz mudou-se de França para os Estados Unidos em fevereiro de 1947* “The actress moved from France to the United States in February 1947”

variable as its first argument. This variable is not explicitly quantified, but assumed to be bound by an existential quantifier. This is in line with a substantial amount of the HPSG literature, including computational implementations such as the English Resource Grammar (Flickinger, 2000) and the Grammar Matrix (Bender *et al.*, 2002). An example is the predicate *_mudar_v* (for the verb form corresponding to English “move”) in Figure 2: its first argument ($e12$) is an event variable.

Additionally, an *at* relation pairs this event variable with a temporal index: in Figure 2 this relation is labeled with $h8$ and relates the event variable $e2$ with the temporal index $t9$. This temporal index represents the event time. In the existing literature on tense, some authors use quantified time variables, while other authors use free time variables. Partee (1973) presents arguments for a free variable approach. Our temporal indices are compatible with this approach. Temporal indices have their own type in the grammar, and a feature T-VALUE is appropriate for this type. This feature locates the index in the time line.

Depending on the grammatical tense, there are then temporal relations between temporal indices, in the spirit of Reichenbach, who also describes tense as temporal relations between various times.

In our example, the Portuguese verb is in the *pretérito perfeito* tense. The semantics of this tense is ambiguous between a simple perfective past (i.e. the situation occurred in the past and is culminated) and a present perfect (the situation has a resulting state that holds and is relevant at the present). The event time is before the utterance time and, accordingly, there is a temporal relation *is-before* with the event time as its first argument.

This particular example is an adaptation of a Reichenbachian representation, where one would expect two time relations (the event time is simultaneous with a reference time and this reference time precedes the utterance time). Our option to diverge in this particular case is motivated by the ambiguity of grammatical tenses like the *pretérito perfeito*. This grammatical tense is ambiguous with respect to semantic tense, viz. the simple past (which has the Reichenbachian analysis just mentioned) and the present perfect (where the event time precedes the reference time, and the reference time is simultaneous with the utterance time). Since it is not possible to underspecify this distinction in the semantic representations, there are two options: duplicate the number of analyses provided by the grammar for each verb with this tense in the input (this is the approach of Van Eynde 2000 for Dutch, but it is computationally costly and does not seem justifiable as both representations essentially describe a past event); or use a simplified representation that covers both interpretations. We chose the second route, arriving at what has just been described.

With other tenses, the grammar delivers representations resorting to reference times.

The second argument of the temporal relation *is-before* is another temporal index, *t*10, with a T-VALUE specified to have the value *utterance-time*. This is how the speech time is represented. According to what has been presented so far, the relevant representation fragment is thus:

$$at(e2,\ t9) \wedge \textit{is-before}(t9,\ t10\ \{\textit{t-value utterance-time}\}) \wedge \_mudar\_v(e2,\ x4)$$

That is, the event described by the form of the verb *mudar* "move" occurred in a time that precedes the utterance time.

It is thus worth noting that grammatical tense presents two levels of ambiguity that must be resolved:

- The same form can correspond to more than one grammatical tense. An English example is the verb form *put*, which can, for instance, be present tense or past tense. Portuguese also contains similar ambiguities, e.g. forms like *corremos* ("we run" or "we ran").
- The same grammatical tense can cover more than one meaning when it comes to locating a situation in time. An English sentence like *I leave tomorrow* shows that present tense can refer to the future. Usually this tense locates an event in the present. Portuguese has similar cases.

This two-fold ambiguity is accounted for by a two-layer analysis in the working grammar. The first layer consists of a set of rules that map surface form to grammatical tense. The second layer consists of a set of rules that map grammatical tense to semantic representations of tense. Both are implemented as lexical rules, i.e. unary rules that apply to single lexical items (verb forms in this case).[6]

[6] One example is the following. In Portuguese, present tense can be frequently used with a future meaning, although of course it can also be used to refer to a present situation. This possibility exists in English, too (e.g. *The train leaves tomorrow*). With this organization in two layers, a present tense verb form is analysed in the following fashion. A rule in the first layer is responsible for the morphology: it maps between the lemma of the verb, which is what is encoded in the grammar's lexicon, and the actual surface form. It also produces a morphological representation in which the grammatical tense of this verb form is encoded, in a dedicated feature. In the second layer, two rules can apply. One of them associates present tense morphology with present semantics. It adds to the meaning representation for the sentence where the verb form occurs that the situation denoted by the verb holds at a time that overlaps the speech time. The second rule

In the case of the rules in the first layer, the orthographic form of their output is different from that of their input (one is the dictionary form of the word, as that is what is listed in the grammar's lexicon, and the other one is an inflected form). The rules in the second layer do not change the spelling of their input. When we combine the grammar with an external morphological analyzer, the second layer of rules is still applied in the grammar, but the application of the rules in the first layer is dependent on the annotations coming from the external analyzer.

3.2 *Aspect*

Aspectual type is accounted for with the help of three Boolean features: *culmination* (positive for culminations and culminated processes), *process* (positive for processes and culminated processes) and *state* (positive for states). This representation is intended to capture the proposal of Moens and Steedman (1988), who decompose a culminated process into a process followed by a culmination. In our representation the two features *process* and *culmination* would be positive, which indicates that this culminated process is composed of two sub-events: a process and a culmination (although the order in which they occur is not made explicit in our representation). These three features are appropriate for event variables.

Even though aspectual type is also a lexical property, it is difficult to annotate it (Pustejovsky *et al.*, 2006). In our implementation, we abstain from recording aspectual type in the lexicon. This would require the annotation of a large part of the existing lexicon, which already contains several thousands of lexical entries. Another difficulty is that aspectual type depends on word sense, which is typically not dealt with by deep grammars, including LXGram.

However, contextual (i.e. syntactic) constraints on aspect are indeed implemented. These are represented by aspectual operators, which are functions from situation descriptions to situation descriptions, and they appear as relations in the MRS representations.

that can possibly apply to a morphological present is one that encodes future semantics, expanding the meaning representation with a temporal precedence relation between the speech time and the time at which the situation denoted by the verb holds.

For instance, we represent a function from state descriptions to culmination descriptions as *aspectual-operator*(e_2 {*culmination* +}, e_1 {*state* +}, X). Here, e_1 is a state, e_2 is a culmination, and X is the MRS representation for the state e_1. The event variable of the resulting situation (e_2 in this example) is included in the representation. We also make use of an extra argument, which is just a pointer for the event variable of the argument (e_1 in this example), as this is useful when post-processing MRS representations.

We follow Bonami (2002) in assuming that all aspectually sensitive relations allow for at most one implicit aspectual operator. These implicit aspectual operators account for aspectual coercion. Therefore every context that allows aspectual coercion must introduce either zero or one aspectual operators in the semantic representation: zero if no aspectual coercion actually occurs, or one otherwise.

Because it is not possible to underspecify the number of relations in an MRS, one *aspectual-operator* is introduced in every aspectually sensitive context, although in general it is not specified which operator it is (in line with Bonami 2002). That is, one underspecified operator is always introduced. We assume that sometimes it stands for a dummy relation (i.e. the identity function), in the cases when no aspectual shift occurs.

Several elements are sensitive to aspectual type. Tense is one of them. Consider the two example sentences below. They correspond to the English sentence *Samuel liked that wine.*

(3) a. O Samuel gostou desse vinho.
    b. O Samuel gostava desse vinho.

The difference between the two is grammatical tense, but they also convey different temporal and aspectual meanings. In the first one the verb is in the *pretérito perfeito*, discussed above. In the second one the verb is in the *pretérito imperfeito*. Both are past tenses, but the first is perfective whereas the second one is imperfective.

Perfective aspect constrains the whole event to be telic (a culmination or a culminated process). Imperfective aspect constrains it to be a state in Portuguese. The first sentence means that Samuel liked the wine at some point in the past, but he no longer does. It may suggest a particular wine tasting episode that has ended (i.e. he liked the wine that he drank at some specific time in the past, as in the English sentence *Samuel enjoyed that wine*), or it may mean that for some time Samuel liked that (kind of) wine, but he no longer does. The second one cannot be about a particular episode. It says that Samuel used to like that kind of wine, and he may still like it.

The grammar assigns to the first sentence a semantic representation expressing this:

$$\begin{aligned}&at(e\ \{culmination\ +\},\ t)\ \wedge\\ &is\text{-}before(t,\ t2\ \{t\text{-}value\ utterance\text{-}time\})\ \wedge\\ &aspectual\text{-}operator(e,\ e2,\ gostar(e2,\ X)),\end{aligned}$$

where X is the representation for the verb's arguments.

This representation is similar to the one presented above in the discussion about tense, but it includes information about aspect as well. In particular, an *aspectual-operator* was added scoping over the relation for the main verb in this sentence. This operator is introduced in the semantics by the lexical rule responsible for semantic tense (together with the temporal relations seen in this MRS fragment), as tenses impose aspectual constraints at the clausal level (Bonami, 2002). The constraint that the event variable e be telic (its feature *culmination* has the value +) also comes from the *pretérito perfeito* tense.

By contrast, the second sentence receives a representation like:

$$\begin{aligned}&at(e\ \{state\ +\}, t)\ \wedge\ overlaps(t,\ t2)\ \wedge\\ &is\text{-}before(t2,\ t3\ \{t\text{-}value\ utterance\text{-}time\})\ \wedge\\ &aspectual\text{-}operator(e,\ e2,\ gostar(e2,\ X)),\end{aligned}$$

where X is the representation for the verb's arguments.

The *pretérito imperfeito* conveys a different temporal meaning, and therefore the temporal relations in the semantic representation are different. This tense does not indicate that the associated situation no longer holds at present, and accordingly the associated temporal relations are more vague with respect to the relation between the event time t and the utterance time $t3$. Unlike the *pretérito perfeito* tense, which introduces an aspectual operator that produces telic situations, the *pretérito imperfeito* constrains the whole clause to be a state. In this example, this is encoded in the event variable e, with its feature *state* constrained to have the + value.

The verb *gostar* "like", instantiating the third argument of the *aspectual-operator* relation, is a state. Even though lexical aspect is not encoded in the grammar (and therefore there is no restriction on the aspectual features of $e2$) for the reasons mentioned above, our encoding of aspect at the syntactic level, as was just illustrated, is important because it can capture distinctions such as the one illustrated by this pair of sentences.

Additionally, it can be straightforwardly extended with lexical aspect: if we knew that "like" is lexically a state, then the *aspectual-operator* in the second sentence is a function from states to states (i.e. it is the identity function, and does not change the basic meaning of the verb). The aspectual operator in the first sentence would be a function from states to telic situations. This causes a shift in meaning, as a culmination is added, corresponding to the end of the underlying state. As mentioned above, there can be two results: Samuel's liking of that kind of wine ended in the past, or the situation is associated with a specific episode that similarly ended in the past.

The implementation of aspect in the grammar interacts with many elements that are sensitive to aspect: many verbs, which impose aspectual constraints on their complements (some examples are the progressive auxiliary, which combines with processes, but also verbs like *stop* and *finish*); durational adverbials (*for* adverbials, which combine with processes, and *in* adverbials, which combine with culminated processes, are widely studied with respect to this phenomenon); tenses (as just briefly illustrated); etc.

A full description of the semantics of all tenses implemented in the grammar is outside the scope of this paper and would be tedious, but an example with the present tense can also be presented. A sentence like *O Samuel gosta desse vinho* "Samuel likes that wine" receives an MRS representation along the following lines:

$$\begin{aligned}&\textit{at}(e\ \{\textit{state}\ +\}, t)\ \wedge\\ &\textit{includes}(t,\ t2\ \{\textit{t-value utterance-time}\})\ \wedge\\ &\textit{aspectual-operator}(e,\ e2,\ gostar(e2,\ X)),\end{aligned}$$

where X is the representation for the verb's arguments.

Here t is the event time, and $t2$ is the utterance time. The present tense is assumed to be an imperfective tense, similar to the past imperfective tense mentioned above: it is associated with an overlap relation, and constrains the clause where it occurs to describe a state. We follow DRT in further assuming that semantic present is special in that this overlap relation is more specific than just overlap, and it is an inclusion relation: the event time includes the utterance time. Because the verb *gostar* "like" is a state lexically, this is another example where the aspectual operator involved is the identity function.

3.3 *Backshift*

There is also an implementation of backshift, or sequence of tense, in this grammar. The pairs of English sentences in (4), adapted from Michaelis (2006), illustrate this issue, which is visible in indirect speech. Each sentence in parentheses is the direct speech counterpart of the embedded clause in the same line, and yet they often (but not always) show different tenses. For instance, the example in (4b) shows an embedded past tense that corresponds to a present tense form in the direct speech utterance.

(4) a. Debra said she **likes** wine. ("I like wine")
  b. Debra said she **liked** wine. ("I like wine")
  c. Debra said she **brought** the wine. ("I brought the wine")
  d. Debra said she **had brought** the wine. ("I brought the wine")
  e. Debra said she **will bring** some wine. ("I will bring some wine")
  f. Debra said she **would bring** some wine. ("I will bring some wine")

The example in (5), from Rodríguez (2004), clearly shows that in syntactic contexts such as the one exemplified by these sentences, tense can be interpreted relatively. In (5) the past tense that occurs in the embedded clause (i.e. in *drank*) is associated with a verb that describes a situation that in the most natural reading for this sentence will occur in the future. Here, past tense merely indicates precedence with respect to the situation mentioned in the matrix clause, through the use of a future construction (*will tell*). In other words, this past tense form is interpreted relative to another mentioned event rather than with respect to the speech time.

(5) María will tell us after the party tomorrow that she drank too much.

The data are essentially identical for Portuguese as far as backshift is concerned. Further analysis can be found in Costa and Branco (2012a). The implementation of backshift in the grammar follows the analysis proposed in that paper.

The grammar makes use of the machinery of HPSG (unification, multiple inheritance and recursive data structures called typed feature structures) to implement constraints on the various tenses such that some of them are always interpreted relative to the speech time whereas others can be interpreted relative to the speech time or the event time of a higher verb, depending on the syntactic context where they occur.

The implementation accounts for cases like the examples in (4). An embedded present tense conveys an overlap temporal relation between the time of the eventuality described in the embedded clause and the speech time, as exemplified in (4a). An embedded future is similarly interpreted relative to the speech time, but conveying a precedence relation between the speech time and the time of the embedded eventuality (4e). An embedded past tense can be associated with an overlap relation or with a precedence relation between the time of the eventuality in the embedded clause and the time of the eventuality mentioned in the higher clause, as in (4b) and (4c). Constructions similar to the English past perfect, as in (4d), trigger a temporal precedence relation between the time of the eventuality mentioned in the embedded clause and the time of the eventuality in the main clause. Finally, sentences similar to the one in (4f) are associated with a precedence temporal relation between the time of the main event and the time of the embedded event.

4 FULL-FLEDGED TEMPORAL PROCESSING

This section describes how the information output by a temporal extraction system can be integrated with the deep semantic representations produced by the grammar.

4.1 *Integration of deep processing and temporal extraction*

The temporal extraction system outputs information that can be combined with the semantic representations delivered by the grammar, resulting in semantic representations enriched with more and better information about time. In some cases, it is preferable to compute these pieces of temporal information outside the grammar; in other cases it is not even possible to compute them in the grammar. One such example is the normalization of temporal expressions, which, as explained above in Section 1, requires access to arithmetic operations and to a calendar system. Deep grammars are implemented with specialized description formalisms and, in some cases, in platforms that do not even make arithmetic operations available.[7]

Typically, those specialized grammatical formalisms have a number of characteristics: they are developed exclusively with grammatical modeling in mind and often do not support operations that are not directly needed for this modeling; the formalisms used in handcrafted grammars are typically categorical (they let one say whether a sentence is either grammatical or ungrammatical, not whether it is better or worse than an alternative), thus making it difficult to represent gradient or statistical information; and, since computational efficiency is an important concern for these systems, many are very restrictive.[8] Another characteristic of computational grammars is that their context is limited, as they typically only look at one sentence at a time. Because of this, they do not have access to information present in other parts of the document, which temporal extraction systems can take advantage of.

The expression of time in natural language and its meaning representation make particularly strong cases where these limitations can be felt. These tasks deal with a number of aspects that require extralinguistic knowledge and as such are difficult or even impossible to implement in their full breadth in these specialized formalisms. Among

[7] This is the case of LXGram and all grammars implemented in the LKB. The LKB accepts a language called TDL – Type Description Language (Krieger and Schäfer, 1994) – which has no support for arithmetic. By contrast, modern programming languages make arithmetic operations available, and it is possible to find for them good implementations of calendar systems. For the implementation of the temporal extractor described above in Section 2.2.3, Joda-Time 2.0 (`http://joda-time.sourceforge.net`) was used, which provides many calendar operations as well as many operations on time intervals.

[8] For instance, the LKB, where LXGram is developed, is very fast, but, for efficiency reasons, does not allow the direct encoding of many kinds of constraints that are standard in the HPSG literature (Melnik, 2005).

```
<TIMEX3 tid="t0" functionInDocument="CREATION_TIME" value="2012-01-10T15:00:00"/>
<s>A atriz <EVENT eid="e5">mudou</EVENT>-se da França para os Estados Unidos em
<TIMEX3 value="1947-02" tid="t15">fevereiro de 1947</TIMEX3>.</s>
<TLINK lid="l2" eventID="e5" relType="BEFORE" relatedToTime="t0"/>
<TLINK lid="l3" eventID="e5" relType="OVERLAP" relatedToTime="t15"/>
```

Figure 3: Example text with (simplified) temporal annotations. The English translation is *The actress moved from France to the United States in February 1947.*

these aspects we find: (i) arithmetic and calendar systems (for the normalization of temporal expressions, as just mentioned); (ii) reasoning (temporal relations have several logical properties that can be exploited, such as the transitivity of temporal precedence); (iii) the modeling of world knowledge and pragmatics (where statistical information about what is usual or expected may constitute important heuristics to determining the chronological order of the described situations); etc.

In particular, it is possible to augment these semantic representations output by the grammar in the following ways:

- Extending the representations
  It is possible to add to the MRS representations output by the grammar further temporal information that the grammar does not have access to.
- Specifying the representations
  The MRS representations are in many cases underspecified, and in some such cases they can be made more specific.
- Correcting the specifications
  The temporal extraction system is sensitive to both grammatical and extra-grammatical information. It is often more accurate in resolving time-related ambiguity than the grammar, which considers grammatical features only. As such, the extractor's output can be used to correct the MRS representations produced by the grammar.

The following paragraphs provide details on how these three aspects are handled by our system that combines the deep grammar and the temporal extractor. To that end we return to our running example, presented above in (2) and repeated below for convenience:

(2) *A atriz mudou-se de França para os Estados Unidos em*
the actress moved from France to the United States in
fevereiro de 1947.
February of 1947

The actress moved from France to the United States in February 1947.

The temporal annotation obtained by the temporal extraction system for this running example is displayed in Figure 3. That example shows two annotated temporal relations, namely an overlap relation between the moving event and the month of February 1947, and a temporal precedence relation between this event and the document creation time.

The semantic representation obtained by the grammar for this example is shown in Figure 2 on page 120. The objective is thus to enrich the grammar-derived representation by exploring the temporal annotations shown in Figure 3.

4.1.1 Extending the MRS representations

The outcome of this combination is presented in Figure 4. As can be seen by comparing Figures 2, 3 and 4, there are several pieces of information that are incorporated into the resulting MRS representation. These additions are highlighted in bold in Figure 4.

The first one is the information about the document's creation time (the `TIMEX3` element in Figure 3). Temporal extraction systems register when a document was created (in our example this is `"2012-01-10T15:00:00"`), which can be determined from meta-data or with heuristics. This information can be incorporated in the MRS representations, specifying the utterance time. The normalized value for the document's creation time is used to fill in the `T-VALUE` of the temporal index for the utterance time. In Figure 4, this is the temporal index $t10$.

The second type of information to add is about temporal expressions. An argument is added to the relation for the head word of that expression that was identified as a temporal expression by the extraction system. This argument is instantiated with a temporal index whose *t-value* feature contains the normalized representation of the time expression. In our example, the temporal expression *fevereiro*

$<h1,$
$\{h3 : \_o\_q(x4, h5, h6),$
$h7 : \_atriz\_n(x4),$
$h8 : at(e2\ \{culmination\ +\}, t9),$
$h8 : is\text{-}before(t9, t10\ \{t\text{-}value\ \text{“}\mathbf{2012\text{-}01\text{-}10T15 : 00 : 00}\text{”}\}),$
$h8 : aspectual\text{-}operator(e2, e12, h11),$
$h11 : \_mudar\_v(e12, x4),$
$h11 : \_de\_p(e14, e12, x13),$
$h15 : proper\_q(x13, h16, h17),$
$h18 : named(x13,\ \text{“França”}),$
$h11 : \_para\_p(e20, e12, x19),$
$h21 : \_o\_q(x19, h23, h22),$
$h24 : named(x19,\ \text{“Estados Unidos”}),$
$h11 : \_em\_p(e26, e12, x25),$
$h27 : udef\_q(x25, h28, h29),$
$h30 : \_fevereiro\_n(x25, \mathbf{t69}\ \{t\text{-}value\ \text{“}\mathbf{1947\text{-}02}\text{”}\}),$
$\mathbf{h30 : overlaps(t9, t69)},$
$h30 : \_de\_p(e31, x25, x32),$
$h33 : proper\_q(x32, h34, h35),$
$h36 : named(x32,\ \text{“1947”}\},$
$\{h1 =_q h8,\ h5 =_q h7,\ h16 =_q h18,\ h23 =_q h24,\ h28 =_q h30,$
$h34 =_q h36\} >$

Figure 4: Final MRS for *A atriz mudou-se de França para os Estados Unidos em fevereiro de 1947* “The actress moved from France to the United States in February 1947”

de 1947 “February 1947” is originally given the MRS representation:

$$
\begin{array}{lll}
< h27, & \{ & h27 : udef\_q(x25,\ h28,\ h29), \\
 & & h30 : \_fevereiro\_n(x25), \\
 & & h30 : \_de\_p(e31, x25, x32), \\
 & & h33 : proper\_q(x32, h34, h35), \\
 & & h36 : named(x32,\ \text{“1947”})\ \}, \\
 & \{ & h28 =_q h30,\ h34 =_q h36\ \} >.
\end{array}
$$

An extra argument is added to the *_fevereiro_n* relation (with the label $h30$), filled with a temporal index containing the normalized value for the temporal expression, as shown in Figure 4: $< h30 :$ $\_fevereiro\_n(x25,\ t69\ \{t\text{-}value\ \text{“1947-02”}\}) >$.[9]

Finally, additional temporal relations detected by the temporal extraction system are incorporated in the MRS.

The only temporal relations originally present in the MRS representations are the ones directly related to verb tense, since the grammar only looks at grammatical information. These are always between an event and the utterance time or the event of the higher clause in the case of backshift phenomena (Costa and Branco, 2012a).

But temporal information systems can extract more temporal relations than those. These extra relations can be added to the MRS representations. In our example this is the *overlaps* relation between the event time *t9* of the moving event and the temporal index *t69* for the time conveyed by the temporal expression *fevereiro de 1947* “February 1947” : $< h30 : overlaps(t9, t69) >$.

[9] The resulting representation is somewhat redundant, and we believe it can be improved. However, this issue is far from trivial, although it may seem so at first. The intuitive alternative would be to replace the entire material in the original MRS for this temporal index. In this example, the five relations (and the two handle constraints) for the expression *fevereiro de 1947* “February 1947” would be completely eliminated from the MRS and replaced by a temporal index. This temporal index would occur as the second argument of the *_em_p* relation, for the preposition corresponding to English *in*: $\_em\_p(e26, e12, t69\{t\text{-}value\ \text{“1947-02”}\})$. This alternative has two problems that must be noted.

The first one is illustrated by a sentence like *2007 saw the birth of the iPhone*. Here, a temporal expression occurs as the subject of a verb. With the intuitive representation, the first argument of the predicate for the verb *to see* would end up being a temporal index. This seems wrong, as the first argument of that predicate would not be of the expected type.

The second problem is related to examples like *that awful year*. This is a time expression that includes material (namely the adjective *awful*) that is not present in the normalized value of the temporal expression (which would just consist of a number representing a calendar year). Replacing the entire MRS representation of this noun phrase for a temporal index would create a representation that does not include all the information present in the analyzed input sentence.

We believe that the problem of adequately modeling the semantic representation of temporal expressions is an interesting question for linguistics to further clarify, for these reasons. As such, an admittedly simplistic solution was chosen in our integrated representation.

To implement the integration of the original MRS produced by the grammar with the information coming from the temporal extraction system, all that is needed is an alignment between the word tokens in the original text and the semantic relations that correspond to those tokens. In our experimental setup, this is achieved quite straightforwardly since the PET parser, the parsing engine used with the grammar, allows the grammar to provide character spans next to each relation in the output MRS representations. These character spans describe the character positions of the linguistic material corresponding to that relation and are used for the alignment and merging of the deep temporal representations with the temporal relations extracted.

4.1.2 Increased semantic specification

The temporal relations identified by the grammar can be made more specific on the basis of the output of the temporal extractor. One example illustrating this is related to the following sentence, taken from the training data of TimeBankPT, with the original English sentence also presented below in italics:

(6) Esperava-se que Bush autorizasse os comandantes navais a usar "a mínima força necessária" para interditar os navios de carga para o Iraque e a partir do Iraque, disse um oficial americano.
Bush was expected to authorize naval commanders to use "the minimum force necessary" to interdict shipments to and from Iraq, a U.S. official said.

TimeBankPT (and the English data set used in TempEval) contains TimeML annotations for this sentence describing temporal relations between the document's creation time and several events, namely those represented by *esperava-se* "it was expected", *usar* "use", and *disse* "said". Similarly, the temporal extractor is capable of identifying these temporal relations.

The temporal semantics implemented in the grammar also encodes several temporal relations between situations described by finite verb forms and the speech time, which is similar to the document's creation time. However, in some cases, these semantic representations are less specific than the TimeML annotations.

A case in point is the imperfective past tense in indirect speech contexts, which is exemplified in this sentence with the verb form *esperava* "was expected". Here the semantics will encode that the event conveyed by the embedded *esperava* overlaps the one conveyed by *disse* "said". This is as expected, because this tense is associated with these kinds of readings in this context.[10] This semantic representation does not say anything about the relation between the embedded situation and the speech time or document's creation time. This is not a shortcoming of the implemented grammar; it is what is justified from the point of view of the linguistic analysis. But this information is readily available in the output of the temporal extractor, and therefore can be incorporated in the final MRS representation.

Another case that is not trivial to treat in the grammar alone is verb forms in the conditional mood. The grammar implementation assigns them a future of past interpretation: the described event occurs at a time that follows another time that precedes the speech time. Therefore, the direct relation between events introduced by verb forms in this tense and the speech time is not available in the MRS representation produced by the grammar, and in fact can be any one.

In the annotated data, however, there can be cases of temporal annotations between events introduced by verbs in the conditional and the document's creation time.

4.1.3 Corrections to the temporal representations

In some cases, the temporal extraction system can be used to correct the MRSs output by the grammar.

In cases of conflict between the initial temporal relations identified by the grammar and the ones given by the temporal extractor, the initial representations produced by the grammar can be corrected if the temporal relations identified by the extractor are considered more reliable than the ones that the grammar produces.

[10] "Past under past" constructions (Comrie, 1986; Declerck, 1990; Hornstein, 1991; Abusch, 1994; Michaelis, 2011) may be ambiguous in English. For example, in *John said he was ill* the two situations described can be simultaneous, but in *John said he fell down* the one described by the embedded verb precedes the one in the matrix clause. In Portuguese, the two interpretations are distinguished by the past tense used: the imperfective past is used in the former case, and the perfective past is used in the latter one (Costa and Branco, 2012a).

This is because the grammar only looks at grammatical tense, whereas the temporal information system takes other features into account, and can identify cases where grammatical tense is insufficient or misleading. An example of this is the case of the historical present, that is, the grammatical present being used to describe a past event, such as in the sentence *In 1939 Germany invades Poland*. This is an important property of our proposal.

Another example where corrections are fruitful is also connected to the use of present tense in Portuguese. English allows this tense to be used to describe future events, as in *The train leaves tomorrow*. In Portuguese this is much more pervasive, and because of that each occurrence of this tense is given this reading, as well as a present reading, by the grammar. The representations for the two different readings (present and future) are not underspecified (because they have different aspectual constraints, i.e. they constrain the three Boolean features that we use to encode aspect, as presented above, differently). Rather, each occurrence of this grammatical tense is ambiguous between present and future, triggering two distinct analyses. As mentioned before, the system uses a statistical model to discriminate between competing analyses for each sentence. By causing the analysis to branch out in these cases, the choice of present vs. future is determined by this parse selection model.

Not surprisingly, as far as this distinction goes, this parse selection model performs quite poorly when compared to a dedicated temporal annotation system, as shown in the next section. That is, there are several cases when the best interpretation given by the grammar erroneously assigns future semantics to present tense verb forms or vice versa. In these cases, the integration component corrects the final MRS representation by changing the temporal relations there so that said representation is in accordance with the output of the temporal extractor.

4.2 *Evaluation*

A test suite of sentences exemplifying the phenomena that the grammar should be able to deal with was created. It contains sentences in the various tenses, sentences with forms of the auxiliary *ter* "have" combining with a past participle, sentences with a progressive construction similar to the English construction composed of *be* and an

-ing form, sentences with forms of *ir* “go” with an infinitive (similar to English “going to” constructions), and sentences featuring adverbs like *hoje* “today”, *ontem* “yesterday”, and *amanhã* “tomorrow”, which feature different combinatorial possibilities with the different tenses. This test suite is used for regression tests during grammar development and contains 38 sentences. The grammar is able to correctly parse all of these sentences and provides correct temporal representations for them.

The test suite is useful to check for bugs in the implementation and ensure that the expected results are seen, but it might not be representative of what is seen in practical scenarios. So an evaluation with unseen data was conducted.

Evaluating this approach presents specific challenges. There is no gold-standard available with MRS annotations that contains temporal information similar to what is presented here. And in fact, it is quite difficult to produce MRS representations manually, as they contain many re-entrancies. For these reasons, we resort to manual evaluation. Since the temporal extractor was developed using the training set of TimeBankPT, the test part of this corpus is unseen and can be used for evaluation of the integrated solution as well.

To this end, the 20 documents comprising the test portion of TimeBankPT were parsed with the grammar. On large corpora of native Portuguese text taken from newspapers and the Wikipedia, the grammar is capable of analyzing around ⅓ of all sentences (Costa and Branco, 2010a), as already mentioned above in Section 2.1. In the present case, 24% of the sentences in the test set of TimeBankPT got a parse.[11] Since the integration of the grammar with the extractor is not meant to increase the coverage of the former, the sentences that receive no parse were left out of this evaluation exercise. There remained 84 sentences in the test set.

This section provides evaluation results for the several tasks directly involved in the integration of the grammar with the temporal extraction system. First, the recognition and normalization of tempo-

[11] We assume that this lower coverage is due to the fact that many of the documents composing this data set are taken from the Wall Street Journal (as TimeBankPT is a translation of the English corpus used in TempEval), and there was no effort to have the grammar deal with text from the financial and economic domains, which contain quite a number of syntactic idiosyncrasies.

ral expressions is discussed. This task is performed by the temporal extractor and then combined with the MRS representations output by the grammar, as discussed above. Here the results for the integrated output are thus the same as those for the temporal extractor.

After that, evaluation results are presented for two problems that are similar to the Tasks A and B of TempEval discussed above. Since the temporal extractor identifies events and temporal expressions and temporal relations between these, and these temporal relations are added to the MRS representations, the performance of the extractor and that of the integrated system are discussed. Finally, evaluation results are provided for the classification of temporal relations between events and the speech time or the document's creation time (i.e. Task B of TempEval). In this respect both the grammar and the temporal extractor are evaluated in isolation, since each can output these temporal relations. The integrated system, which corrects the MRS representations with the information coming from the extractor, is also evaluated.

The Task C of TempEval is not used by our integrated approach. Since Task C relates events mentioned in different sentences, a discourse representation is necessary to combine them in an informed way. This is not something that the typical deep linguistic technology does at the moment.[12]

Table 5 summarizes the results discussed in the rest of this section and obtained on the parsed sentences of the test data of TimeBankPT. In this table, n/a marks results that are not available, as the grammar is not intended to perform the corresponding tasks.

4.2.1 Evaluation of temporal expression recognition and normalization

Since the integrated system enriches the original MRS representations with representations for the temporal expressions that occur in the underlying text, this dimension was evaluated.

As mentioned above, we restricted our attention to the sentences for which there was a parse produced by the grammar. We looked at all temporal expressions that can be found in these sentences. The system was evaluated with respect to two factors. First, we want to know

[12] An exception is Boxer (Curran *et al.*, 2007), which can handle some cross-sentential phenomena, such as pronoun resolution and presupposition.

	Total	Grammar	Extractor	Combined system
Temporal expressions	32			
Recognition		n/a	28/32 (88%)	28/32 (88%)
Normalization		n/a	27/32 (84%)	27/32 (84%)
Event – time pairs	44			
Task A		n/a	25/44 (57%)	25/44 (57%)
Finite verbs	111			
Task B		83/111 (75%)	92/111 (83%)	104/111 (94%)

Table 5: Accuracy of the grammar, the temporal extraction system and the combined system for several tasks (% correct)

how many temporal expressions are recognized correctly. Second, we want to know how they are normalized, since these normalized values appear in the final representations.

Temporal expressions are somewhat infrequent and, in these 84 sentences, only 32 such expressions occur. Of these, 88% are recognized correctly. The remaining ones are either not recognized at all or their boundaries are not identified correctly. 84% are recognized correctly and also normalized correctly (or 96% of the ones that are recognized correctly). From the point of view of normalization, the difficult cases are very vague ones such as *the night*. These cases fail to be normalized and as such are not incorporated in the final MRS representations.

Although some of the temporal expressions occurring in this data set fail to be recognized and incorporated in the final MRS representations, the ones that are indeed inserted there are almost all correctly normalized (96%).

4.2.2 Evaluation of temporal relations between mentioned times and events

As mentioned above, the final MRS representations also include temporal relations between the times and dates and the events mentioned in the input sentences, since these relations are delivered by the temporal extractor (cf. Task A).

These temporal relations occurring in the semantic representations of the parsed sentences were checked for correctness. There are only 44 such relations, because only a few sentences contain multiple temporal expressions and multiple events. 57% of these relations are correctly encoded. A considerable number of the errors occur when the times and events being related are mentioned very far apart in the sentence or the syntactic relationship between the expressions denoting them is not direct. If we restrict our attention to pairs of events and times that are mentioned in the same clause, this score goes up to 68%.

Since the grammar provides us with this information, we are considering only adding these temporal relations to the MRS representations in these cases when the relevant expressions occur in the same clause. So even though temporal information processing technology still has a considerable amount of error, to some extent we can at least increase precision by sacrificing recall in a straightforward way if this is considered preferable.

4.2.3 Evaluation of temporal relations with the speech time

One final aspect to evaluate is how many of the temporal relations between events and the speech time or document's creation time, output by the final integrated temporal processing system, are correct. This is similar to the Task B of TempEval.

The grammar assigns temporal relations to events and states represented by finite forms of verbs only, for the reasons already mentioned. TimeBankPT includes annotations also for events denoted by words of other parts-of-speech, most importantly nouns. Even though the extractor can also identify these, it is not as accurate in doing so, as mentioned above. For this reason, the integrated system does not expand MRS representations with temporal information for events that are not given by verbs, and likewise we also ignore them in this evaluation.

For each sentence, only the preferred parse output by the grammar, as determined by the parse selection model, is considered. The grammar produced a correct output for 75% of all temporal relations between the situations described in these parsed sentences and the document's creation time/speech time.

As mentioned above, one difficulty is assigning the correct meaning to present tense verb forms. As they are ambiguous between future and present semantic values and this distinction is chosen by a general parse selection model, it is rarely the case that it is correctly resolved. The temporal extractor is much better at this particular problem, as it employs several features that are relevant to it. For instance, aspectual type is very relevant; depending on the language, the future interpretation of present tense is much harder or even impossible with stative verbs (Van Eynde, 1998, p. 249). The grammar has no information about lexical aspect, but the extractor has some, in the form of the aspectual indicators as well as the features `class` and even `stem` (since this is a lexical property). This problem accounts for 56% of the errors produced by the grammar for this task. Other errors were less interesting and had a smaller impact overall.

The temporal extractor gets 83% of these temporal relations between finite verb forms and the speech time/document's creation time right, better than the 75% of the grammar. The largest source of error has to do with identifying events: many of the verbs for which the grammar produces temporal relations are not recognized as events by the temporal extractor, and therefore no relation is posited for them. Note that TimeML does not annotate verbs used in generic statements (such as *Lions are mammals*) as events, and furthermore the annotations for event terms that occurred fewer than 20 times in the English data used in TempEval were removed. Therefore the training data of TimeBankPT, which is also used to train the event identification model used in LX-TimeAnalyzer, contains many examples of verbs that are not annotated as being event terms.[13]

The system combining the output of the grammar and that of the temporal extractor delivers temporal relations between finite verbs and the speech time/document's creation time with 94% accuracy. This is a better result than either the grammar (75%) or the temporal extractor (83%) in isolation.

Overall, these results show that integrating a specialized temporal extractor with a deep grammar can be fruitful in practice in increasing the quality of the temporal meaning representations and the accuracy of the resulting system.

[13] As a side note, if one removes these cases and looks only at those that were identified by both the grammar and the temporal extractor, the success rate of the latter in classifying the temporal relation with the document's creation time goes up to 97%. This is substantially better than the results presented above for the task B of TempEval because here we are looking exclusively at events denoted by verbs, which are easier to order with respect to the utterance time than those given by words with a different part-of-speech.

5 CONCLUSIONS

This article presents a novel contribution to the processing of the linguistic expression of time in deep natural language processing systems by combining them with data-driven methods. As interpreting the temporal ordering of the events mentioned in a text is indeed affected by phenomena that are difficult to model in a symbolic system, like knowledge of the world, machine learning methods can capture the contribution of factors whose impact is not well understood. To this end, it was discussed how to combine the outcome of temporal information extraction technology with the semantic representations produced by a deep processing grammar.

This combination helps to resolve the ambiguity preserved in the underspecified semantic representation. One very important point is that it also allows for the representations produced by deep grammars to encode extra-linguistic information – e.g. the normalized representation of the speech time – that is relevant to interpret these representations but hard to obtain with these grammars alone.

Finally, with the present contribution towards full-fledged temporal processing, this paper adds to the overall discussion and quest on how to make progress in natural language processing by means of hybrid systems that combine the complementarity of the symbolic and probabilistic approaches in a way that their strengths can be amplified and their shortcomings mitigated. The resulting system presents better performance than each of the two components in isolation, both quantitatively (as measured in terms of accuracy) and qualitatively (as it outputs truth-conditional representations of the meaning of sentences that includes but is not limited to information about time).

Future work is needed to address temporal relations between events mentioned in different sentences. In this respect, there is some work on the temporal structure of discourse, also using HPSG. One example is the work of Hitzeman *et al.* (1995), although in some cases this specific proposal leaves these temporal relations underspecified. It would be interesting to check how proposals such as this one compare with current temporal relation classification technology for the task C of the first TempEval challenge. Future work can check this by implementing a similar solution with the grammar.

Future work can address other ways to combine the two sub-systems (the extractor and the grammar). The integration can also work in the direction opposite of the one explored in this paper: for instance, the events recognized by the grammar can be proposed to the shallow temporal extraction system, as the latter failed to recognize some of them in our evaluation. Additional work could also investigate the use of a meta-learning component to detect correct and incorrect information in either sub-system. In this paper, we have shown, however, that even the simple approach that we explored already produces competitive results that improve the performance of the whole system.

REFERENCES

Dorit ABUSCH (1994), Sequence of Tense Revisited: Two Semantic Accounts of Tense in Intensional Contexts, in Hans KAMPF, editor, *Ellipsis, Tense and Questions*, pp. 87–139, University of Amsterdam, DYANA deliverable R.2.2.B.

Peter ADOLPHS, Stephan OEPEN, Ulrich CALLMEIER, Berthold CRYSMANN, Dan FLICKINGER, and Bernd KIEFER (2008), Some Fine Points of Hybrid Natural Language Parsing, in *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Genoa, Italy.

Jan ALEXANDERSSON, Ralf ENGEL, Michael KIPP, Stephan KOCH, Uwe KÜSSNER, Norbert REITHINGER, and Manfred STEDE (2000), Modeling Negotiation Dialogs, in Wolfgang WAHLSTER, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, pp. 441–451, Springer-Verlag, Berlin Heidelberg New York, Artificial Intelligence edition.

James ALLEN (1983), Maintaining Knowledge about Temporal Intervals, *Communications of the ACM*, 26(11):832–843.

James ALLEN (1984), Towards a General Theory of Action and Time, *Artificial Intelligence*, 23:123–154.

Gabor ANGELI, Christopher D. MANNING, and Daniel JURAFSKY (2012), Parsing Time: Learning to Interpret Time Expressions, in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 446–455, Association for Computational Linguistics, Montréal, Canada.

Emily M. BENDER, Dan FLICKINGER, and Stephan OEPEN (2002), The Grammar Matrix: An Open-Source Starter-Kit for the Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars, in John CARROLL, Nelleke OOSTDIJK, and Richard SUTCLIFFE, editors, *Procedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pp. 8–14, Taipei, Taiwan.

Emily M. BENDER, Dan FLICKINGER, Stephan OEPEN, Annemarie WALSH, and Timothy BALDWIN (2004), Arboretum. Using a precision grammar for grammar checking in CALL, in *Proceedings of the InSTIL Symposium on NLP and Speech Technologies in Advanced Language Learning Systems*, Venice, Italy.

Steven BETHARD, James H. MARTIN, and Sara KLINGENSTEIN (2007), Finding Temporal Structure in Text: Machine Learning of Syntactic Temporal Relations, *International Journal of Semantic Computing*, 1(4):441–457.

André BITTAR, Pascal AMSILI, Pascal DENIS, and Laurence DANLOS (2011), French TimeBank: An ISO-TimeML Annotated Reference Corpus, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 130–134, Association for Computational Linguistics, Portland, Oregon, USA.

Daniel G. BOBROW, Bob CHESLOW, Cleo CONDORAVDI, Lauri KARTTUNEN, Tracy Holloway KING, Rowan NAIRN, Valeria DE PAIVA, Charlotte PRICE, and Annie ZAENEN (2007), PARC's Bridge and Question Answering System, in Tracy Holloway KING and Emily M. BENDER, editors, *Proceedings of the GEAF07 Workshop*, pp. 46–66, CSLI Publications, Stanford, CA, http://csli-publications.stanford.edu/GEAF/2007/geaf07-toc.html.

Olivier BONAMI (2002), A syntax-semantics interface for tense and aspect in French, in Frank Van EYNDE, Lars HELLAN, and Dorothee BEERMANN, editors, *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pp. 31–50, CSLI Publications, Stanford.

Francis BOND, Stephan OEPEN, Melanie SIEGEL, Ann COPESTAKE, and Dan FLICKINGER (2005), Open Source Machine Translation with DELPH-IN, in *Proceedings of the Open-Source Machine Translation Workshop at the 10th Machine Translation Summit*, pp. 15–22, Phuket, Thailand.

Philip BRAMSEN, Pawan DESHPANDE, Yoong Keok LEE, and Regina BARZILAY (2006), Inducing Temporal Graphs, in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pp. 189–198, Sydney, Australia.

Bertram C. BRUCE (1972), A Model for Temporal References and Its Application in a Question Answering Program, *Artificial Intelligence*, 3(1–3):1–25.

Caroline BRUN (1998), Terminology Finite-State Preprocessing for Computational LFG, in *COLING '98 Proceedings of the 17th International Conference on Computational Linguistics*, volume 1, pp. 196–200, Association for Computational Linguistics, Stroudsburg, PA, USA.

Miriam BUTT, Tracy Holloway KING, María-Eugenia NIÑO, and Frédérique SEGOND (1999), *A Grammar Writer's Cookbook*, CSLI Publications, Stanford.

Ulrich CALLMEIER (2000), PET – A Platform for Experimentation with Efficient HPSG Processing Techniques, *Natural Language Engineering*, 6(1):99–108 (Special Issue on Efficient Processing with HPSG).

Tommaso CASELLI, Valentina BARTALESI LENZI, Rachele SPRUGNOLI, Emanuele PIANTA, and Irina PRODANOF (2011), Annotating Events, Temporal Expressions and Relations in Italian: the It-TimeML Experience for the Ita-TimeBank, in *Proceedings of the 5th Linguistic Annotation Workshop*, pp. 143–151, Association for Computational Linguistics, Portland, Oregon, USA, http://www.aclweb.org/anthology/W11-0418.

Nathanael CHAMBERS and Daniel JURAFSKY (2008), Jointly Combining Implicit Constraints Improves Temporal Ordering, in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 698–706, Association for Computational Linguistics, Honolulu, Hawaii.

Nathanael CHAMBERS, Shan WANG, and Dan JURAFSKY (2007), Classifying temporal relations between events, in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic.

Eugene CHARNIAK and Mark JOHNSON (2005), Coarse-to-fine n-best parsing and MaxEnt discriminative reranking, in *Proceedings of the 43rd Annual Meeting of the ACL*, pp. 173–180, Association for Computational Linguistics, Ann Arbor.

Yuchang CHENG, Masayuki ASAHARA, and Yuji MATSUMOTO (2008), Constructing a Temporal Relation Tagged Corpus of Chinese Based on Dependency Structure Analysis, *Computational Linguistics and Chinese Language Processing*, 13(2):171–196.

Bernard COMRIE (1986), Tense in Indirect Speech, *Folia Linguistica*, 20:265–296.

Ann COPESTAKE (2002), *Implementing Typed Feature Structure Grammars*, CSLI Publications, Stanford.

Ann COPESTAKE, Dan FLICKINGER, Ivan A. SAG, and Carl POLLARD (2005), Minimal Recursion Semantics: An Introduction, *Journal of Research on Language and Computation*, 3(2-3):281–332.

Francisco COSTA (2013), *Processing Temporal Information in Unstructured Documents*, Ph.D. thesis, Universidade de Lisboa, Lisbon, http://nlx.di.fc.ul.pt/~fcosta/papers/phdthesis.pdf.

Francisco COSTA and António BRANCO (2010a), LXGram: A Deep Linguistic Processing Grammar for Portuguese, in *Lecture Notes in Artificial Intelligence*, volume 6001, pp. 86–89, Springer, Berlin, http://nlx.di.fc.ul.pt/~fcosta/papers/propor2010.pdf.

Francisco COSTA and António BRANCO (2010b), Temporal Information Processing of a New Language: Fast Porting with Minimal Resources, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL2010)*, pp. 671–677, Association for Computational Linguistics, Uppsala, Sweden, http://nlx.di.fc.ul.pt/~fcosta/papers/acl2010.pdf.

Francisco COSTA and António BRANCO (2012a), Backshift and Tense Decomposition, in Stefan MÜLLER, editor, *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, pp. 86–106, ISSN 15351793,
`http://nlx.di.fc.ul.pt/~fcosta/papers/hpsg2012.pdf`.

Francisco COSTA and António BRANCO (2012b), Extracting Temporal Information from Portuguese Texts, in Helena CASELI, Aline VILLAVICENCIO, António TEIXEIRA, and Fernando PERDIGÃO, editors, *Computational Processing of the Portuguese Language – 10th International Conference, PROPOR 2012*, volume 7243 of *Lecture Notes in Artificial Intelligence*, pp. 99–105, Springer, Berlin, Germany,
`http://nlx.di.fc.ul.pt/~fcosta/papers/propor2012.pdf`.

Francisco COSTA and António BRANCO (2012c), LX-TimeAnalyzer: A Temporal Information Processing System for Portuguese, Technical Report DI-FCUL-TR-2012-01, Universidade de Lisboa, Faculdade de Ciências, Departamento de Informática,
`http://nlx.di.fc.ul.pt/~fcosta/papers/tr2012-1.pdf`.

Francisco COSTA and António BRANCO (2012d), TimeBankPT: A TimeML Annotated Corpus of Portuguese, in Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, Mehmet Uğur DOĞAN, Bente MAEGAARD, Joseph MARIANI, Jan ODIJK, and Stelios PIPERIDIS, editors, *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3727–3734, European Language Resources Association (ELRA), Istanbul, Turkey, ISBN 978-2-9517408-7-7,
`http://nlx.di.fc.ul.pt/~fcosta/papers/lrec2012.pdf`.

Berthold CRYSMANN, Anette FRANK, Bernd KIEFER, Stefan MÜLLER, Günter NEUMANN, Jakub PISKORSKI, Ulrich SCHÄFER, Melanie SIEGEL, Hans USZKOREIT, Feiyu XU, Markus BECKER, and Hans-Ulrich KRIEGER (2002), An Integrated Architecture for Shallow and Deep Processing, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02)*, pp. 441–448, Association for Computational Linguistics, Philadelphia, Pennsylvania.

James R. CURRAN, Stephen CLARK, and Johan BOS (2007), Linguistically Motivated Large-Scale NLP with C&C and Boxer, in *ACL '07 Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pp. 33–36, Association for Computational Linguistics, Stroudsburg, PA, USA.

Donald DAVIDSON (1967), The Logical Form of Action Sentences, in Nicholas RESCHER, editor, *The Logic of Decision and Action*, University of Pittsburgh Press.

Henriëtte DE SWART (1998), Aspect Shift and Coercion, *Natural Language and Linguistic Theory*, 16:347–385.

Henriëtte DE SWART (2000), Tense, aspect and coercion in a cross-linguistic perspective, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the Berkeley Formal Grammar Conference*, CSLI Publications, Stanford.

Rennat DECLERCK (1990), Sequence of Tenses in English, *Folia Linguistica*, 24:513–544.

Pascal DENIS and Philippe MULLER (2010), Comparison of different algebras for inducing the temporal structure of texts, in *Proceedings of COLING 2010*, pp. 250–258, Beijing, PRC.

Pascal DENIS and Philippe MULLER (2011), Predicting Globally-Coherent Temporal Structures from Texts via Endpoint Inference and Graph Decomposition, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 2011*.

David R. DOWTY (1979), *Word Meaning and Montague Grammar: the Semantics of Verbs and Times in Generative Semantics and Montague's PTQ*, Reidel, Dordrecht.

Christiane FELLBAUM, editor (1998), *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA.

Lisa FERRO, Laurie GERBER, Inderjeet MANI, Beth SUNDHEIM, and George WILSON (2004), TIDES 2003 standard for the annotation of temporal expressions, Technical report, The MITRE Corporation, McLean, Virginia.

Dan FLICKINGER (2000), On Building a More Efficient Grammar by Exploiting Types, *Natural Language Engineering*, 6(1):15–28 (Special Issue on Efficient Processing with HPSG).

Dan FLICKINGER, Jan Tore LØNNING, Helge DYVIK, Stephan OEPEN, and Francis BOND (2005), SEM-I Rational MT: Enriching Deep Grammars with a Semantic Interface for Scalable Machine Translation, in *Proceedings of the 10th Machine Translation Summit*, pp. 165–172, Phuket, Thailand.

Maria FLOURAKI (2006), Constraining Aspectual Composition, in Stefan MÜLLER, editor, *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pp. 140–157, CSLI Publications, Stanford.

Corina FORĂSCU and Dan TUFIŞ (2012), Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information, in Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, Mehmet Uğur DOĞAN, Bente MAEGAARD, Joseph MARIANI, Jan ODIJK, and Stelios PIPERIDIS, editors, *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, ISBN 978-2-9517408-7-7.

Anette FRANK, Marcus BECKER, Berthold CRYSMANN, Bernd KIEFER, and Ulrich SCHÄFER (2003), Integrated Shallow and Deep Parsing: TopP meets HPSG, in *Proceedings of the 41st Annual Meeting of the Association for*

Computational Linguistics (ACL 2003), pp. 104–111, Association for Computational Linguistics, Sapporo, Japan.

David GOSS-GRUBBS (2005), *An Approach to Tense and Aspect in Minimal Recursion Semantics*, Master's thesis, University of Washington, Seattle, Washington.

Claire GROVER and Alex LASCARIDES (2001), XML-Based Data Preparation for Robust Deep Parsing, in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pp. 252–259, Toulouse, France.

Eun Young HA, Alok BAIKADI, Carlyle LICATA, and James C. LESTER (2010), NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology, in Katrin ERK and Carlo STRAPPARAVA, editors, *SemEval 2010 – 5th International Workshop on Semantic Evaluation – Proceedings of the Workshop*, pp. 341–344, Uppsala University, Uppsala, Sweden.

Janet HITZEMAN, Marc MOENS, and Claire GROVER (1995), Algorithms for Analysing the Temporal Structure of Discourse, in *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 253–260, Dublin, Ireland.

Norbert HORNSTEIN (1991), *As Time Goes By*, MIT Press, Cambridge, USA.

Seohyun IM, Hyunjo YOU, Hayun JANG, Seungho NAM, and Hyopil SHIN (2009), KTimeML: Specification of Temporal and Event Expressions in Korean Text, in *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 115–122, Association for Computational Linguistics, Stroudsburg, PA, USA.

George H. JOHN and Pat LANGLEY (1995), Estimating Continuous Distributions in Bayesian Classifiers, in *Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338–345, San Mateo.

Hans KAMP and Uwe REYLE (1993), *From Discourse to Logic: An Introduction to Modeltheoretic Semantics, Formal Logic and Discourse Representation Theory*, Kluwer, Dordrecht.

Ronald M. KAPLAN and Joan BRESNAN (1982), Lexical-Functional Grammar: A Formal System for Grammatical Representation, in Joan BRESNAN, editor, *The Mental Representation of Grammatical Relations*, MIT Press Series on Cognitive Theory and Mental Representation, chapter 4, pp. 173–281, MIT Press, Cambridge, Massachusetts.

Ronald M. KAPLAN, John T. MAXWELL III, Tracy Holloway KING, and Richard CROUCH (2004), Integrating Finite-state Technology with Deep LFG Grammars, in *Proceedings of the Workshop on Combining Shallow and Deep Processing for NLP (ESSLLI 2004)*, Nancy, France.

Oleksandr KOLOMIYETS, Steven BETHARD, and Marie-Francine MOENS (2011), Model-Portability Experiments for Textual Temporal Analysis, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human*

Language Technologies, pp. 271–276, Association for Computational Linguistics, Portland, Oregon, USA, `http://www.aclweb.org/anthology/P11-2047`.

Hans-Ulrich KRIEGER and Ulrich SCHÄFER (1994), TDL – A Type Description Language for Constraint-Based Grammars, in *Proceedings of the 15th International Conference on Computational Linguistics*, pp. 893–899, Kyoto, Japan.

John LAFFERTY, Andrew MCCALLUM, and Fernando C. N. PEREIRA (2001), Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proceedings of the 8th International Conference on Machine Learning 2001 (ICML 2001)*, pp. 282–289.

Alex LASCARIDES and Nicholas ASHER (1993), Temporal Interpretation, Discourse Relations, and Common Sense Entailment, *Linguistics and Philosophy*, 16:437–493.

Chong Min LEE (2010), Temporal Relation Identification with Endpoints, in *HLT-SRWS '10 Proceedings of the NAACL HLT 2010 Student Research Workshop*, pp. 40–45, Association for Computational Linguistics, Stroudsburg, PA, USA.

Xiao LING and Daniel S. WELD (2010), Temporal Information Extraction, in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*.

Hector LLORENS, Leon DERCZYNSKI, Robert GAIZAUSKAS, and Estela SAQUETE (2012), TIMEN: An Open Temporal Expression Normalization Resource, in Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, Mehmet Uğur DOĞAN, Bente MAEGAARD, Joseph MARIANI, Jan ODIJK, and Stelios PIPERIDIS, editors, *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3044–3051, European Language Resources Association (ELRA), Istanbul, Turkey.

Hector LLORENS, Estela SAQUETE, and Borja NAVARRO (2010a), TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2, in Katrin ERK and Carlo STRAPPARAVA, editors, *SemEval 2010 – 5th International Workshop on Semantic Evaluation – Proceedings of the Workshop*, pp. 284–291, Uppsala University, Uppsala, Sweden.

Hector LLORENS, Estela SAQUETE, and Borja NAVARRO-COLORADO (2010b), TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles, in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 725–733, Coling 2010 Organizing Committee, Beijing, PRC.

Paweł MAZUR and Robert DALE (2010), WikiWars: A New Corpus for Research on Temporal Expressions, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*, pp. 913–922.

Nurit MELNIK (2005), From "Hand-Written" to Computationally Implemented HPSG Theories, in Stefan MÜLLER, editor, *The Proceedings of the 12th*

International Conference on Head-Driven Phrase Structure Grammar, pp. 311–321, CSLI Publications, Stanford.

Laura MICHAELIS (2006), Tense in English, in Bas AARTS and April MCMAHON, editors, *The Handbook of English Linguistics*, Blackwell, Oxford.

Laura A. MICHAELIS (2011), Stative by Construction, *Linguistics*, 49:1359–1400.

Marc MOENS and Mark STEEDMAN (1988), Temporal ontology and temporal reference, *Computational Linguistics*, 14(2):15–28.

MUC-6 (1995), *Proceedings of the Sixth Message Understanging Conference (MUC-6)*, Defense Advanced Research Projects Agency.

MUC-7 (1998), *Proceedings of the Seventh Message Understanging Conference (MUC-7)*, Defense Advanced Research Projects Agency.

Stefan MÜLLER and Walter KASPER (2000), HPSG Analysis of German, in Wolfgang WAHLSTER, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, pp. 238–253, Springer-Verlag, Berlin Heidelberg New York, Artificial Intelligence edition.

Matteo NEGRI and Luca MARSEGLIA (2004), Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004, Technical report, Trento.

Eric NICHOLS, Francis BOND, Darren SCOTT APPLING, and Yuji MATSUMOTO (2007), Combining Resources for Open Source Machine Translation, in *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, pp. 1–10, Skövde, Sweden.

Eric NICHOLS, Francis BOND, Takaaki TANAKA, Sanae FUJITA, and Daniel FLICKINGER (2006), Robust Ontology Acquisition from Multiple Sources, in *Proceedings of the 2nd Workshop on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pp. 10–17, Sydney, Australia, `http://www.aclweb.org/anthology/W/W06/W06-0502`.

Lars NYGAARD, Jan Tore LØNNING, Torbjørn NORDGÅRD, and Stephan OEPEN (2006), Using a Bi-Lingual Dictionary in Lexical Transfer, in *Proceedings of the 11th Conference of the European Association for Machine Translation*, Oslo, Norway.

Stephan OEPEN, Kristina TOUTANOVA, Stuart SHIEBER, Christopher MANNING, Dan FLICKINGER, and Thorsten BRANTS (2002), The LinGO Redwoods Treebank: Motivation and Preliminary Applications, in *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, pp. 1253–1257, Taipei, Taiwan.

Barbara PARTEE (1973), Some Structural Analogies Between Tenses and Pronouns in English, *The Journal of Philosophy*, 70:601–609.

Carl POLLARD and Ivan SAG (1994), *Head-Driven Phrase Structure Grammar*, Chicago University Press and CSLI Publications, Stanford.

Georgiana PUŞCAŞU (2007), WVALI: Temporal Relation Identification by Syntactico-Semantic Analysis, in *Proceedings of SemEval-2007*, pp. 484–487, Association for Computational Linguistics, Prague, Czech Republic.

James PUSTEJOVSKY (1991), The Syntax of Event Structure, *Cognition*, 41:47–81.

James PUSTEJOVSKY, José CASTAÑO, Robert INGRIA, Roser SAURÍ, Robert GAIZAUSKAS, Andrea SETZER, and Graham KATZ (2003a), TimeML: Robust Specification of Event and Temporal Expressions in Text, in *IWCS-5, Fifth International Workshop on Computational Semantics*.

James PUSTEJOVSKY, Patrick HANKS, Roser SAURÍ, Andrew SEE, Robert GAIZAUSKAS, Andrea SETZER, Dragomir RADEV, Beth SUNDHEIM, David DAY, Lisa FERRO, and Marcia LAZO (2003b), The TIMEBANK Corpus, in *Proceedings of Corpus Linguistics 2003*, pp. 647–656.

James PUSTEJOVSKY, Jessica LITTMAN, Roser SAURÍ, and Marc VERHAGEN (2006), TimeBank 1.2 Documentation, `http://timeml.org/site/timebank/documentation-1.2.html`.

James PUSTEJOVSKY and Amber STUBBS (2011), Increasing Informativeness in Temporal Annotation, in *Proceedings of the Fifth Law Workshop (LAW V)*, pp. 152–160, Association for Computational Linguistics, Portland, Oregon, USA.

Hans REICHENBACH (1947), *Elements of Symbolic Logic*, University of California Press, Berkeley.

Matthew RICHARDSON and Pedro DOMINGOS (2006), Markov Logic Networks, *Machine Learning*, 62(1):107–136.

Joshua P. RODRÍGUEZ (2004), *Interpreting the Spanish* Imperfecto*: Issues of Aspect, Modality, Tense, and Sequence of Tense*, Ph.D. thesis, The Ohio State University, Columbus, Ohio.

C.J. RUPP, Jörg SPILKER, Martin KLARNER, and Karsten L. WORM (2000), Combining Analyses from Various Parsers, in *Verbmobil: Foundations of Speech-to-Speech Translation*, pp. 311–320, Springer, Berlin, Artificial Intelligence edition.

Ivan A. SAG, Thomas WASOW, and Emily M. BENDER (2003), *Syntactic Theory – A Formal Introduction*, CSLI Publications, Stanford.

Roser SAURÍ, Jessica LITTMAN, Bob KNIPPEN, Robert GAIZAUSKAS, Andrea SETZER, and James PUSTEJOVSKY (2006), TimeML Annotation Guidelines: Version 1.2.1, `http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf`, manuscript (retrieved October 19, 2012).

Ulrich SCHÄFER (2006), *Integrating Deep and Shallow Natural Language Processing Components – Representations and Hybrid Architectures*, Ph.D. thesis, Unversität des Saarlandes, Saarbrücken.

João SILVA, António BRANCO, Sérgio CASTRO, and Francisco COSTA (2012), Deep, Consistent and also Useful: Extracting Vistas from Deep Corpora for Shallower Tasks, in Jan HAJIČ, Koenraad De SMEDT, Marko TADIĆ, and António BRANCO, editors, *Proceedings of the Workshop on Advanced Treebanking at the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pp. 45–52, Istanbul, Turkey.

Manfred STEDE, Stefan HAAS, and Uwe KÜSSNER (1998), Understanding and tracking temporal descriptions in dialogue, in Bernhard SCHRÖDER, Winfried LENDERS, Wolfgang HESS, and Thomas PORTELE, editors, *Computers, Linguistics, and Phonetics between Language and Speech – Proceedings of the 4th Conference on Natural Language Processing KONVENS'98*, Peter Lang, Frankfurt, Germany.

Jannik STRÖTGEN and Michael GERTZ (2013), Multilingual and Cross-domain Temporal Tagging, *Language Resources and Evaluation*, 47(2):269–298.

Kristina TOUTANOVA, Christopher D. MANNING, Dan FLICKINGER, and Stephan OEPEN (2005), Stochastic HPSG Parse Selection using the Redwoods Corpus, *Journal of Research on Language and Computation*, 3(1):83–105.

Naushad UZZAMAN and James ALLEN (2011), Temporal Evaluation, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 351–356, Association for Computational Linguistics, Portland, Oregon, USA.

Naushad UZZAMAN and James F. ALLEN (2010), TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text, in Katrin ERK and Carlo STRAPPARAVA, editors, *SemEval 2010 – 5th International Workshop on Semantic Evaluation – Proceedings of the Workshop*, pp. 276–283, Uppsala University, Uppsala, Sweden.

Naushad UZZAMAN, Hector LLORENS, Leon DERCZYNSKI, Marc VERHAGEN, James ALLEN, and James PUSTEJOVSKY (2013), SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations, in *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*.

Frank VAN EYNDE (1998), Tense, Aspect and Negation, in Frank VAN EYNDE and Paul SCHMIDT, editors, *Linguistic Specifications for Typed Feature Structure Formalisms. Studies in machine Translation and Natural language Processing*, volume 10, pp. 209–280, Luxembourg.

Frank VAN EYNDE (2000), A constraint-based semantics for tenses and temporal auxiliaries, in Ronnie CANN, Claire GROVER, and Philip MILLER, editors, *Grammatical interfaces in HPSG*, pp. 231–249, CSLI Publications, Stanford University.

Eric VELLDAL (2007), *Empirical Realization Ranking*, Ph.D. thesis, University of Oslo, Oslo.

Zeno VENDLER (1967), Verbs and Times, in *Linguistics in Philosophy*, pp. 97–121, Cornell University Press, Ithaca, New York.

Marc VERHAGEN, Robert GAIZAUSKAS, Frank SCHILDER, Mark HEPPLE, Jessica MOSZKOWICZ, and James PUSTEJOVSKY (2009), The TempEval challenge: identifying temporal relations in text, *Language Resources and Evaluation*, 43(2):161–179, Special Issue: Computational Semantic Analysis of Language: SemEval-2007 and Beyond.

Marc VERHAGEN, Robert GAIZAUSKAS, Frank SCHILDER, Mark HEPPLE, and James PUSTEJOVSKY (2007), SemEval-2007 Task 15: TempEval temporal relation identification, in *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 75–80, Association for Computational Linguistics, Prague, Czech Republic.

Marc VERHAGEN and James PUSTEJOVSKY (2008), Temporal Processing with the TARSQI Toolkit, in *COLING 2008: Companion Volume: Demonstrations*, pp. 189–192.

Marc VERHAGEN, Roser SAURÍ, Tommaso CASELLI, and James PUSTEJOVSKY (2010), SemEval-2010 Task 13: TempEval-2, in Katrin ERK and Carlo STRAPPARAVA, editors, *SemEval 2010 – 5th International Workshop on Semantic Evaluation – Proceedings of the Workshop*, pp. 51–62, Uppsala University, Uppsala, Sweden.

Wolfgang WAHLSTER, editor (2000), *Verbmobil: Foundations of Speech-to-Speech Translation*, Springer, Berlin, Artificial Intelligence edition.

Nianwen XUE and Yuping ZHOU (2010), Applying Syntactic, Semantic and Discourse Constraints to Chinese Temporal Annotation, in *Proceedings of COLING 2010*, pp. 1363–1372, Beijing, PRC.

Katsumasa YOSHIKAWA, Sebastian RIEDEL, Masayuki ASAHARA, and Yuji MATSUMOTO (2009), Jointly Identifying Temporal Relations with Markov Logic, in *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*.

Kei YOSHIMOTO and Yoshiki MORI (2002), A compositional Semantics for Complex Tenses in Japanese, in Frank VAN EYNDE, Lars HELLAN, and Dorothee BEERMANN, editors, *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pp. 300–319, CSLI Publications, Stanford, `http://cslipublications.stanford.edu/HPSG/2/`.

Xujian ZHAO, Peiquan JIN, and Lihua YUE (2010), Automatic Temporal Expression Normalization with Reference Time Dynamic-Choosing, in *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 1498–1506, Association for Computational Linguistics, Stroudsburg, PA, USA.

Yuping ZHOU and Nianwen XUE (2011), Discourse-constrained Temporal Annotation, in *Proceedings of the 5th Linguistic Annotation Workshop (LAW V '11)*, pp. 161–169, Association for Computational Linguistics, Stroudsburg, PA, USA.