

The Semantic Representation of Temporal Expressions in Text

Robert Dale and Paweł Mazur

Centre for Language Technology
Macquarie University
North Ryde, NSW 2109
rdale,mpawel@ics.mq.edu.edu

Abstract. Temporal expressions—references to points in time or periods of time—are widespread in text, and their proper interpretation is essential for any natural language processing task that requires the extraction of temporal information. Work on the interpretation of temporal expressions in text has generally been pursued in one of two paradigms: the formal semantics approach, where an attempt is made to provide a well-grounded theoretical basis for the interpretation of these expressions, and the more pragmatically-focused approach represented by the development of the TIMEX2 standard, with its origins in work in information extraction. The former emphasises formal elegance and consistency; the latter emphasises broad coverage for practical applications. In this paper, we report on the development of a framework that attempts to integrate insights from both perspectives, with the aim of achieving broad coverage of the domain in a well-grounded manner from a formal perspective. We focus in particular on the development of a compact notation for representing the semantics of underspecified temporal expressions that enables the component-level evaluation of systems.

1 Introduction

Obtaining a precise semantic representation for utterances that contain references to time is interesting both from a theoretical point of view, as there are many complex phenomena to be addressed, and for purely practical applications such as information extraction, question answering, or the ordering of events on a timeline. In the literature, work on the interpretation of temporal expressions comes from two directions. On the one hand, work in formal semantics (see, for example, [1]) aims to provide a formally well-grounded approach to the representation of the semantics of these expressions, but such approaches are difficult to scale up to the broad coverage required for practical applications. On the other hand, work that has its roots in information extraction, while it emphasises broad coverage, often results in the use of ad hoc representations. The most developed work in this direction is focused around the TimeML markup language (described, for example, in [2] and in the collection edited by Mani et al. [3]). Some work attempts to bring these two traditions together: notable in

this respect is Schilder's [4] work on temporal expressions in German newswire text, and Hobbs and Pan's [5] work on axiomatisation in terms of OWL-Time.

Our contribution towards a truly broad coverage yet semantically well-founded approach is to recognize a principled distinction between (a) the interpretation of the semantics of a temporal expression devoid of its context of use, and (b) the fuller interpretation of that expression when the context is taken into account. The first of these, which we refer to as the **local semantics** of a temporal expression, should be derivable compositionally from the components of the expression; determining the value of the second, which we refer to as the **global semantics** of the expression, may require arbitrary inference and reasoning. Such a distinction is implicit in other accounts: Schilder's [4] use of lambda expressions allows representation of partially specified temporal entities, and the temporary variables that Negri and Marseglia [6] construct during the interpretation of a given temporal expression capture something of the same notion.

Our proposal is to reify this level of intermediate representation based on a formalization in terms of recursive attribute-value matrices, as traditionally used in other areas of natural language processing (see, for example, [7]). This has two distinct advantages: (a) it provides a convenient representation of underspecification, and (b) it leads naturally to a compositional approach to the construction of the semantics of temporal expressions via unification.

We also provide a compact encoding of this representation that is an extension of the existing TIMEX2 representation for temporal expressions. This brings the advantages that (a) existing tools and machinery for evaluation can be used to determine how well a given implementation derives these local semantic values; and (b) performance in the determination of local semantics and global semantics can be tested independently. To ensure breadth of coverage, we initially developed our representation on the basis of all 256 examples of temporal expressions provided in the TIMEX2 guidelines [8]; this has been subsequently refined on the basis of examination of the evaluation data sets provided under the Automatic Context Extraction (ACE) program.¹

The rest of this paper is structured as follows. In Section 2, we describe the architecture of DANTE, a system which embodies our approach to the interpretation of temporal expressions; in particular, we focus on the architecture we employ, and on the particular levels of representation that it makes use of. In Section 3, we argue for an intermediate representational level that captures the semantics of temporal expressions independent of the context of their interpretation, and introduce the idea of using recursive attribute-value matrices to represent the semantics of temporal expressions. In Section 4, we provide an encoding of these attribute-value matrices in a compact string-based representation that is effectively an extension of the ISO-based date-time format representations used in the TIMEX2 standard, thus enabling easy evaluation of system performance using existing tools. In Section 5, we report on the performance results of DANTE using the ACE 2007 evaluation data. Finally, in Section 6 we draw some conclusions and point to future work.

¹ See <http://www.nist.gov/speech/tests/ace>

2 The DANTE System

2.1 Processing Steps

In our work, our goal is very close to that for which the TIMEX2 standard was developed: we want to annotate each temporal expression in a document with an indication of its interpretation, in the form of an extended ISO-format date and time string, normalised to some time zone. So, for example, suppose we have the following italicised temporal expression in an email message that was sent from Sydney on Friday 30th November 2007:

- (1) We will be able to present this at the meeting on *Monday at 2pm*.

In our application, this temporal expression should be marked up as follows:

- (2) <TIMEX2 VAL="2007-12-03T03:00GMT">Monday at 2pm</TIMEX2>

We have to do three things to achieve the desired result:

- First, we have to detect the extent of the temporal expression in the text. We refer to this process as **temporal expression recognition**.
- Then, we have to use information from the document context to turn the recognized expression into a fully specified date and time. We refer to this as **temporal expression interpretation**.
- Finally, we have to normalise this fully specified date and time to a predefined time zone, which in the case of the present example is Greenwich Mean Time. We refer to this as **temporal expression normalisation**.²

We observe that, at the time the extent of a temporal expression within a text is determined, it is already possible to derive a semantic representation of that expression irrespective of the wider context required for a full interpretation: for example, having recognized an occurrence of the string *Friday* in a text, we already know that this is a reference to a specific day of the week. It is likely that most existing systems for the interpretation of temporal expressions make use of such a level of representation implicitly. Schilder's [4] approach captures the semantics here in terms of a lambda expression like $\lambda x \text{Friday}(x)$; Negri and Marseglia [6] capture information at this stage of processing via a collection of temporary attributes.

In our system, each of the three steps above corresponds to a distinct processing component in the DANTE system architecture. These components communicate in terms of a number of distinct representations, which we now describe.

Local Semantics: We use this term to refer to a level of representation that corresponds to the semantic content that is derivable directly from the text

² This third step is not required by the TIMEX guidelines, but is an additional requirement in the context of our particular application. This also means that our use of the term 'normalisation' here is not consistent with the standard usage in the TIMEX context; however, we would argue that our distinction between interpretation and normalisation describes more accurately the nature of the processes involved here.

representation; in the case of temporal expressions that are arguments to prepositions, this includes the interpretation of the preposition. Such representations are often incomplete, in that they do not in themselves denote a specific point or period on the time line; however, usually they do partially specify points or periods, and constrain the further interpretation of the string.

In-Document Semantics: This refers to the fully explicit interpretation of the text string, to the extent that this can be determined from the document itself, in conjunction with any metadata associated with the document (in particular, the time and date of publication). This level of representation corresponds to the information encoded in the attributes of the TIMEX2 tag as defined in the TIMEX guidelines.

Global Semantics: The TIMEX guidelines do not have anything to say beyond the representation just described. In our application, however, we are also required to normalise all temporal expressions to a specific time zone. This requires that some further temporal arithmetic be applied to the semantics of the found expressions. To calculate this, we have to determine the difference between the time zone of the document containing the temporal reference and the target time zone, here Greenwich Mean Time.

3 Representing Temporal Expressions

3.1 Temporal Entities

As is conventional in this area of research, we view the temporal world as consisting of two basic types of entities, these being **points in time** and **durations**; each of these has an internal hierarchical structure. In the present paper we focus on the representation of points in time; the representation of durations is a natural extension of this representation. Figure 1 shows the representation of a reference to a specific point in time; in the ISO date and time format used in the TIMEX2 standard, this would be written as follows:

(3) 2006-05-13T15:00:00Z

Each atomic feature in the attribute–value structure thus corresponds to a specific position in the ISO format date–time string.

3.2 Underspecification

Of course, very few temporal expressions in text are so fully specified. The attribute–value matrix representation makes it easy to represent the content of underspecified temporal expressions. For example, the content of the temporal expression *Thursday* in a sentence like *We met on Thursday* can be expressed as:

$$\left[\begin{array}{c} \textit{point} \\ \text{TIMEANDDATE} \left[\text{DATE} \left[\text{DAY} \left[\text{DAYNAME} \ D_4 \right] \right] \right] \right] \end{array} \right]$$

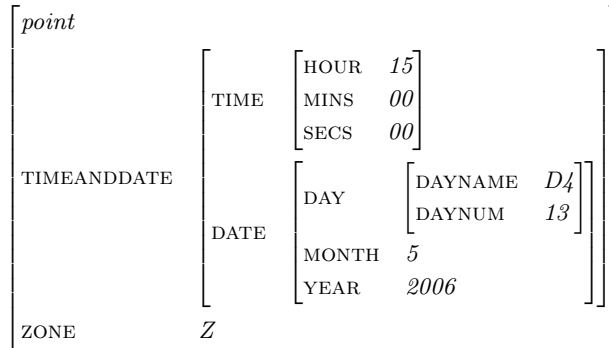
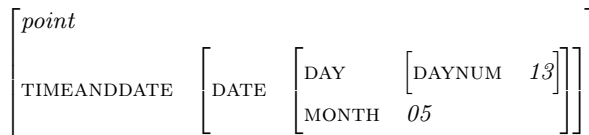
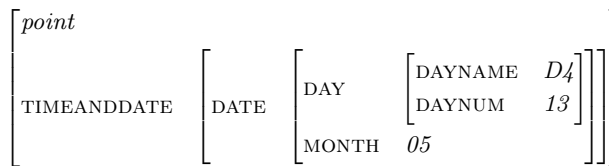


Fig. 1. The semantics of the expression *3pm Thursday 13th May 2006 GMT*

On the other hand, a reference to *13th May* in a sentence like *We will meet on 13th May* has this representation:



In the cases just described, the semantic representation corresponds to the entire temporal noun phrase in each case. The same form of representation is easy to use in a compositional semantic framework: each constituent in a larger temporal expression provides a structure that can be unified with the structures corresponding to the other constituents of the expression to provide a semantics for the expression as a whole. The values of the atomic elements in such an expression come from the lexicon; multiword sequences that are best considered atomic (such as, for example, idioms) can also be assigned semantic representations in the same way. The value of a composite structure is produced by unifying the values of its constituents. Unifying the two structures above, for example, gives us the following representation for *Thursday 13th May*:



So, these structures provide a convenient representation for what we have referred to above as the local semantics of a temporal expression, and correspond to the output of the recognition stage of our processing architecture.

3.3 Interpretation

We can now define the task of interpretation in terms of the content of these structures. We assume a **granularity ordering** over what we might think of as the **defining attributes** in a temporal representation:

- (4) year > month > daynum > hour > minute > second

These are, of course, precisely the elements that are represented explicitly in an ISO date–time expression.

Interpretation of a partially specified temporal expression then requires ensuring that there is a value for every defining attribute that is of greater granularity than the smallest granularity present in the partially specified representation. We refer to this as the **granularity rule** in interpretation.

In the case of the example in the previous section, the granularity rule tells us that in order to compute the full semantic value of the expression we have to determine a value for YEAR, but not for HOUR, MINS or SECS. This interpretation process may require various forms of reasoning and inference, and is qualitatively different from the computation of the local semantics.

In the context of our application, a third stage, the normalisation process, then requires taking the further step of adding a ZONE attribute with a specific value, and translating the rest of the construction into this time zone if it represents a time in another time zone.

4 A Compact Encoding

The structures described in the previous section are relatively unwieldy in comparison to the simple string structures used as values in the TIMEX standard. To enable easy evaluation of a system’s ability to construct these intermediate semantic representations, we would like to use a representation that is immediately usable by existing evaluation tools. To achieve this goal, we define a number of extensions to the standard TIMEX2 string representation for values of the VAL attribute; these extensions allow us to capture the range of distinctions we need. To save space, we also use these representations here in preference to using attribute–value matrices to show the coverage of the annotation scheme that results.

4.1 Partially Specified Dates and Times

As noted above, many references to dates or times are not fully specified in a text, with the result that some parts will have to be computed from the context during the interpretation stage. Typical examples are as follows:

- (5) a. We’ll see you in *November*.
b. I expect to see you at *half past eight*.

Table 1. Underspecified Dates and Times

#	String	Representation	#	String	Representation
1	9 pm	xxxx-xx-xxT21	6	the nineteenth	xxxx-xx-19
2	11:59 pm	xxxx-xx-xxT23:59	7	January 3	xxxx-01-03
3	eleven in the morning	xxxx-xx-xxT11:00	8	November	xxxx-11
4	ten minutes to 3	xxxx-xx-xxt02:50	9	summer	xxxx-SU
5	15 minutes after the hour	xxxx-xx-xxttx:15	10	the '60s	xx6

In the recursive attribute–value notation introduced above, the missing information in each case corresponds to those features that are absent in the structure as determined by the granularity rule introduced in Section 3.3.

In our string-based notation, we use lowercase *xs* to indicate those elements for which a value needs to be found, but which are not available at the time the local semantics are computed; and we capture the granularity requirement by omitting from the string representation those elements that do not require a value.³ Table 1 provides a range of examples that demonstrate various forms of underspecification.

A lowercase *x* thus corresponds to a variable. By analogy, we also use a lowercase *t* instead of the normal ISO date–time separator of *T* to indicate that the time needs further specification: consider the fourth and fifth examples in Table 1, where it is unclear whether the time specified is a.m. or p.m.

For partially-specified dates and times, the string-based encoding thus both captures the local semantic content of the temporal expression, and provides a specification of what information the interpretation process has to add. The interpretation process makes use of a notion of **temporal focus**, the temporal point of reference that generally shifts through the course of a text. If the temporal focus is encoded in the same form of representation, then producing the final interpretation is often a simple process of merging the two structures, with the values already specified in the intermediate representation taking precedence over those in the representation of the temporal focus.

4.2 Relative Dates and Times

A relative date or time reference is one that requires a calendar arithmetic operation to be carried out with respect to some temporal focus in the text. Typical examples are as follows:

- (6) a. We'll see him *tomorrow*.
 b. We saw him *last year*.
 c. We'll see him *next Thursday*.
 d. We saw him *last November*.

³ Note that this does not mean the same thing as the use of an uppercase *X* in the TIMEX2 guidelines: there, an uppercase *X* means that no value can be determined. In our approach, if no value can be found for a variable element during the interpretation process, then the corresponding *x* will be replaced by an *X*.

Table 2. Relative dates and times in ISO-like format

#	String	Representation	#	String	Representation
1	today	+0000-00-00	6	sixty seconds later	+0000-00-00T+00:00:60
2	tomorrow	+0000-00-01	7	five minutes ago	+0000-00-00T-00:05
3	yesterday	-0000-00-01	8	in six hours time	+0000-00-00T+06:00
4	last month	-0000-01	9	at 6 a.m. today	+0000-00-00T06:00
5	three years ago	-0003	10	last night	-0000-00-01TNI

To handle relative references, we extend the ISO format with a preceding ‘+’ or ‘-’ to indicate the direction from the current temporal focus. Some examples of dates and times are provided in Table 2. Note the both the date and time elements in a relative reference can be independently either absolute or relative: compare the representations for *in six hours time* and *at 6am today*.

This representation leads to a very intuitive coordinate-based arithmetic for computing the final semantic interpretation of a given expression: the interpreter merely adds the temporal focus and the intermediate value element-by-element from the smallest unit upwards, using carry arithmetic where appropriate.

5 Evaluation

The primary focus of the work described here has been to develop a representation for local semantics that covers as wide a range of linguistic phenomena as possible; for this reason, the representation has been developed using the TIMEX2 guidelines as a target data set, since the examples presented there are intended to cover the broad range of phenomena that a human annotator is likely to encounter.

However, it is also important, of course, that the representational mechanism then be tested for its performance in a real task evaluation. To this end, we subsequently fine-tuned the representations on the basis of the ACE 2005 development data, and in January 2007, DANTE participated in the ACE Time Expression Recognition and Normalization (TERN) task.

In the ACE evaluations a correctly recognized time expression is one which has a strictly accurate extent and correct values for all the TIMEX2 attributes. An annotation generated by the system is classified as matched with an annotation from the gold standard if there is minimum 30% text span overlap between them

The ACE 2007 evaluation data included 2028 time expressions to be recognized and interpreted. Across all domains we currently achieve 54.7, 57.6 and 56.1 for precision, recall and F-measure, respectively, for correct recognition of temporal expressions. After applying weights to particular elements which are subject to evaluation,⁴ the scores are 69.7, 69.2 and 69.4 respectively; the overall

⁴ In the ACE 2007 TERN evaluations the weights were as follows: 1.0 for type VAL, 0.5 for ANCHOR_VAL, 0.25 for ANCHOR_DIR, 0.1 for MOD, 0.1 for SET, 0.1 for extent (where there is at least a 30% overlap between matched elements; otherwise elements are not mapped at all). The cost for spurious TIMEX2 mentions was -0.75.

Table 3. The results of evaluating DANTE on the ACE 2007 evaluation data set

Domain	Entities in gold standard	Spurious	Missing	Error	Precision	Recall	F-score	ACE Value
Broadcast Conversation	142	33	29	43	47.9	49.3	48.6	46.5
Broadcast News	322	103	38	69	55.6	66.8	60.6	55.2
Newswire	894	128	110	273	56.0	57.2	56.6	58.8
Telephone Conversation	70	23	11	25	41.5	48.6	44.7	51.4
Usenet News- groups	167	20	22	43	61.8	61.1	61.4	65.3
Weblogs	433	68	58	139	53.3	54.5	53.9	57.3
Total	2028	375	268	592	54.7	57.6	56.1	57.2

ACE TERN score for DANTE is 57.2. These results indicate that DANTE's performance is already very close to state-of-the-art systems: the best performing system in the ACE evaluations received an overall score of 61.6.

Table 3 shows the results of the system on the ACE 2007 data broken down by text domain. An analysis of the errors indicates that the representation remains robust when faced with real data; the bulk of our errors are due to limitations in our grammatical coverage (often due to the long tail of low-frequency lexical items and phrases used in temporal expressions) and deficiencies in our focus tracking mechanism.

6 Conclusions

In this paper, we have argued that, in the context of interpreting temporal expressions, there is value in identifying a level of semantic representation that corresponds to the meaning of these expressions outside of any particular document context. Many existing systems appear to make use of such representations at least implicitly. However, we have proposed that this level of representation be made explicit; and by providing an encoding of this level of representation that is an extension of the existing TIMEX2 annotations in terms of element attributes and their values, we make it possible to assess the performance of systems with respect to intermediate values, final values, or both, using standard evaluation tools.

We have developed the representation described here on the basis of the set of 265 examples provided in the TIMEX2 guidelines [8], and tested the performance of DANTE, a system using the representation, on the ACE 2005 and 2007 gold standard data sets. We achieve results that are broadly comparable with the state of the art.

DANTE's temporal expression recognizer is implemented using GATE's JAPE grammar formalism [9]; about 80 macros and 250 rules are used in the current

implementation. Recognition is thus by means of finite-state patterns; however, an examination of those cases where we fail to determine the correct extents for temporal expressions indicates that we suffer when the expressions in question are quite complex syntactically, as in *four days after Americans first penetrated the Baghdad outskirts*. Here, DANTE only recognises the string *four days* as a temporal expression and fails to detect the attached prepositional phrase. A major focus for future work is to integrate a syntax-based mechanism for detecting the extents of arbitrarily complex temporal expressions. This will also allow us to more rigorously test the compositional properties of our intermediate representation.

Acknowledgements

We acknowledge the support of the Defence Science and Technology Organisation in carrying out the work described here.

References

1. Pratt, I., Francez, N.: Temporal prepositions and temporal generalized quantifiers. *Linguistics and Philosophy* 24, 187–222 (2001)
2. Pustejovsky, J., Castaño, J., Ingria, R., Saur, R.G., Setzer, A., Katz, G.: TimeML: Robust Specification of Event and Temporal Expressions in Text. In: IWCS-5. Fifth International Workshop on Computational Semantics (2003)
3. Mani, I., Pustejovsky, J., Gaizauskas, R. (eds.): *The Language of Time*. Oxford Univ. Press, Oxford (2005)
4. Schilder, F.: Extracting meaning from temporal nouns and temporal prepositions. *ACM Transactions on Asian Language Information Processing* 3(1), 33–50 (2004)
5. Hobbs, J.R., Pan, F.: An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing* 3(1), 66–85 (2004)
6. Negri, M., Marseglia, L.: Recognition and normalization of time expressions: ITC-IRST at TERN 2004. Technical Report WP3.7, Information Society Technologies (2005)
7. Shieber, S.M.: *An Introduction to Unification-Based Approaches to Grammar*. Center for the Study of Language and Information (1986)
8. Ferro, L., Gerber, L., Mani, I., Sundheim, B., Wilson, G.: TIDES 2005 Standard for the Annotation of Temporal Expressions. Technical report, MITRE (2005)
9. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A framework and graphical development environment for robust NLP tools and applications. In: *Proceedings of the 40th Anniversary Meeting of the ACL* (2002)