Covering various Needs in Temporal Annotation: a Proposal of Extension of ISO TimeML that Preserves Upward Compatibility

Anaïs Lefeuvre-Halftermeyer, Jean-Yves Antoine, Alain Couillault, Emmanuel Schang, Lotfi Abouda, Agata Savary, Denis Maurel, Iris Eshkol-Taravella, Delphine Battistelli

Covering various Needs in Temporal Annotation: a Proposal of Extension of ISO TimeML that Preserves Upward Compatibility

Anaïs Lefeuvre-Halftermeyer[1,2], Jean-Yves Antoine[2], Alain Couillault[4], Emmanuel Schang[3], Lotfi Abouda[3], Agata Savary[2], Denis Maurel[2], Iris Eshkol-Taravella[3], Delphine Battistelli[5]

(1) Université Paris-Sorbonne, laboratoire STIH
1, rue Victor Cousin, 75005 Paris, France
(2) Université François Rabelais de Tours, laboratoire LI
3 place Jean Jaurès, 41000 Blois, France
(3) Université d'Orléans, laboratoire LLL, UMR 7270
10 rue de Tours, 45100 Orléans-la-Source, France
(4) Université de La Rochelle, laboratoire L3i
Avenue Michel Crépeau, 17000 La Rochelle, France
(5) Université Paris-Ouest Nanterre & MoDyCo UMR 7114
200, avenue de la République, 92 001 Nanterre Cedex, France
E-mail: anais.lefeuvre@paris-sorbonne.fr, jean-yves.antoine@univ-tours.fr, alain.couillault@univ-lr.fr, emmanuel.schang@univ-orleans.fr, lotfi.abouda@univ-orleans.fr, agata.savary@univ-tours.fr, denis.maurel@univ-tours.fr, iris.eshkol@univ-orleans.fr, delphine.battistelli@u-paris10.fr

Abstract

This paper reports a critical analysis of the ISO TimeML standard, in the light of several experiences of temporal annotation that were conducted on spoken French. It shows that the norm suffers from weaknesses that should be corrected to fit a larger variety of needs in NLP and in corpus linguistics. We present our proposition of some improvements of the norm before it will be revised by the ISO Committee in 2017. These modifications concern mainly (1) Enrichments of well identified features of the norm: temporal function of TIMEX time expressions, additional types for TLINK temporal relations; (2) Deeper modifications concerning the units or features annotated: clarification between time and tense for EVENT units, coherence of representation between temporal signals (the SIGNAL unit) and TIMEX modifiers (the MOD feature); (3) A recommendation to perform temporal annotation on top of a syntactic (rather than lexical) layer (temporal annotation on a treebank).

Keywords: ISO-TimeML, temporal annotation, eventualities, temporal relations, time expressions

1. Introduction

Corpus annotation is a time-consuming and a costly activity, thus impeding the development of large language resources. In order to favour the spread and the reuse of such valuable data, a crucial recommendation is to follow established annotation standards. Nonetheless, many areas of Natural Language Processing are still suffering from compliance problems between corpora following different annotation schemes.

Temporal annotation has avoided this pitfall. Most time-annotated corpora respect a common norm, TimeML, which was proposed as an international standard ISO DIS 24617-1 (ISO, 2009) now included in the ISO SemAF annotation framework. This standard has been applied to a large variety of languages (English, French, Italian, Portuguese, Korean, Romanian, Chinese...)[1] with a low number of language-specific adjustments which do not affect its coherence. This salutary standardisation effort shows, however, some drawbacks, since TimeML was originally meant for rather specific purposes. The main aim of the norm is to increase the performance of question answering systems, thus, few important problems are addressed by the norm, such as event temporal identification, time stamping of events, reasoning with contextually underspecified temporal expressions and reasoning over the persistence of events (Mani et al. 2005). These original specific motivations explain that TimeML results from some choices that do not necessarily cover all needs of current temporal annotation in NLP.

A call for revisions of this standard will be launched during year 2017. Our goal is to take this opportunity of standard improvement. We present the limitations of TimeML experienced in different temporal annotation projects (Temporal, TourInflux) and we propose some enhancements that aim at preserving coherence and upward compatibility with the current state of the norm.

2. TimeML

The first step of temporal annotation consists in characterising all the linguistic items that refer to a constitutive element of the discourse temporality. These elements are usually called eventualities (Bach 1981, Mani et al. 2005). Eventualities encompass processes, punctual or protracted events and states. Temporal reasoning also requires the detection of relations between these eventualities. This is the aim of the second step of annotation, which usually follows the model of 13 temporal relations by (Allen, 1983).

Temporal annotation was originally motivated by applicative purposes, namely information extraction for question-answering tasks. Thus, the first evaluation

[1] See (Bittar 2011 ; Caselli 2009 ; Costa & Branco 2008 ; Im et al. 2009 ; Zhou & Xue 2011 and Cheng et al.2008).

campaigns (MUC-7, CoNLL 2002-2003) focused on time expressions (TIMEX) detection. Later campaigns investigated more complex tasks, from events characterization (ACE 2005-2007) to temporal relations, what led gradually to the elaboration of the ISO-TimeML standard.

ISO TimeML distinguishes several kinds of temporal units:

- EVENT corresponds to a linguistic expression denoting an eventuality.
- TIMEX identifies temporal expressions (dates, duration…).
- SIGNAL identifies linguistic markers which introduce temporal relations.

TimeML accounts for three kinds of temporal relations:

- TLINK characterises any temporal relation (identity or Allen relations) between EVENTs or TIMEXes.
- ALINK denotes an aspectual relation between two EVENTs. For instance, an ALINK relation is considered to select the initial phase of an EVENT on example (1).

(1) *Mary <u>started</u> her <u>lunch</u> at 2 p.m.*

- SLINK is a subordinate relation of influence between two EVENTs: The example (2) presents a *factive* SLINK relation between *forgot* and *was*.

(2) *I completely forgot I was in Paris last Tuesday.*

3. TimeML Limits and Improvements Propositions

The norm improvements proposed here have been elaborated through diverse projects using ISO-TimeML: first of all TEMPORAL[2], funded by the MSH Val de Loire and led by research teams in linguistics (LLL[3]) and computer science (LI and LIFO[4]). The second project, TourInflux[5], associated academic (L3I) and industrial teams (Syllabs, Proxem and APROGED association). First, section 3.1 presents extensions not implying any modification of the structure of the XML TimeML documents. These modifications are restricted indeed to the definition of values that instantiate some attributes of the norm. This preserves a structural compliance of the norm while minimizing backwards compatibility problems. Deeper adaptations will be detailed afterwards.

3.1 Level Enrichments to ISO-TimeML

TIMEX. The first annotation unit discussed here is TIMEX. Its attribute `temporalFunction` expresses whether its temporal reference needs to be calculated considering its linguistic expression and another reference. For instance, the TIMEX "*during the following hours*" will receive a positive value (`TRUE`). This feature is strictly binary in TimeML. It seems interesting to refine it to characterize the kind of temporal unit which it has to be calculated on. Studies on temporal references distinguish absolute references ("*in 1987*": `FALSE temporalFunction` in ISO-TimeML) from relative ones (`TRUE temporalFunction`), and add two sub-categories for the latter one: the ones depending on the enunciation time ("*now*" in example 3a below) and ones relative to another EVENT or TIMEX in the discourse ("*from then*" in example 3b):

(3a) **dependence on enunciation time**

Speaker 1 *when should we leave ?*
Speaker 2 *now*

(3b) **dependence on another eventuality**

The Mac Phersons settled in the wild Montana during the summer 1885 and stay in this area from then.

Instead of a binary value for the temporal function, we propose to use three values based on the seminal work of (Reichenbach 1947): `Null` for absolute references, `S` for enunciation-based relative references and `R` for discourse-based relative ones.

The details of the function themselves have never been described in details in the `temporalFunctionID` attribute of the norm. TourInflux (Drat 2014) explored different semantic function classes. For instance, the `tf5` class designates the calculus of the beginning and end of intervals with several values: "*at mealtimes*". We propose that the future version of ISO TimeML includes a detailed and comprehensive definition of the classes of temporal functions that should be accepted by the norm.

EVENT. Concerning event instances, our attention has been drawn to the polarity feature. In TimeML polarity is a binary feature aiming at identifying when the EVENT is under the scope of a grammatical negation. The polarity feature covers only one kind of linguistic markers triggering event evidentiality. In this respect, we recommend to define an additional feature called `INQUISIT` to describe other markers implying evidentiality:

- `TQUEST` in a question about the time of the EVENT. *"When will you come?"*
- `QUEST` in a question about the actualization of the EVENT. *"Will you come?"*
- `TORDER` in an imperative sentence about the time. *"Do come tomorrow!"*
- `ORDER` in an imperative sentence. *"Do ask him!"*
- `DECL` other situations

TLINK. TLINK types available in the norm are the ones identified by (Allen 1983), in addition to a fourteenth one, `IDENTITY`, standing for coreference between two units. This `IDENTITY` value is highly questionable, since it suggests the existence of a referential identity between the

[2] http://tln.li.univ-tours.fr/Tln_Temporal.html. This project will be continued through the ODIL project starting in 2016.

[3] *Laboratoire Ligérien de Linguistique* (Orléans and Tours) and *Laboratoire d'Informatique Fondamentale d'Orléans* (Orléans).

[4] *Laboratoire d'Informatique* (Tours).

[5] http://tourinflux.univ-lr.fr/

related units which may concern objects of different nature (EVENTs, TIMEXes). We recommend not to mix referential and temporal relations between such objects (EVENTs, TIMEXes). The proper temporal relation for an exact matching of two pure temporal objects must be `SIMULTANEOUS`: the `IDENTITY` value must be dropped out. By the way, to the best of our knowledge, other projects have decided not to consider the TimeML `IDENTITY` relation (MERLOT 2015).

On the opposite, additional TLINK types should be considered and inserted in the norm. For instance, the TourInflux project has proposed to add a few more relations (Drat 2014) such as `IS_EXCLUDED` for utterances like: "open every day except Sunday".

Following this idea, a relation between a TIMEX and an EVENT giving information about its measure in the temporal dimension should be annotated with a special relation. Pustejovsky proposed to define an MLINK[6] in the same perspective.

3.2 Deeper Modifications

The enrichments listed previously can easily be integrated in the ISO-TimeML. We will now present some deeper modifications of the norm.

EVENT. The original norm uses a feature called `TENSE` representing the different standard tense values of verbs (`PAST`, `PRESENT`, `FUTURE`, `IMPERFECT` and `NONE`). This feature is defined by exhaustive tables of tensed verbal forms in English. (Bittar et al. 2011) followed the same approach for French. The combination of the two features tense and aspect addresses univocally every grammatical tense. Its only advantage in our opinion is to discriminate easily linguistic units that are already discriminated. It presents the serious drawback to make a confusion between grammatical tense and time. This is even more crucial when dealing with languages such as Creoles or Bantu languages, see Binnick (2012) and Winford (2012) amongst others. We propose to modify this feature. We suggest to call it `TIME` instead of `TENSE`, and to choose its value from: `PAST`, `PRESENT`, `FUTURE`, `NONE` (for infinitives not denoting actualized events for instance) and `OMNI` (for events that are always true). `IMPERFECT` being an aspect value, it has to be attached to the aspect feature.

SIGNAL. We have encountered some difficulties in classifying units regarding modification of temporal objects. Two units are considered completely different whereas in a discourse context they are really close:

- SIGNAL triggers a relation between EVENTs, TIMEXes or between an EVENT and a TIMEX: for instance before in "*I left before noon*" is annotated as a specific XML unit called SIGNAL : `<SIGNAL>before</SIGNAL>`
- Conversely, TIMEX has an attribute called `MOD` which specifies if the temporal unit is modified: a bit before noon in "*I left a bit before noon*" would be annotated as a specific XML unit, a TIMEX, whose `MOD` value would be `BEFORE`: `<TIMEX MOD=BEFORE VALUE=12:00>a bit before noon</TIMEX>.`

Two different objects (SIGNAL and `MOD`) convey the same kind of information which we recommend to annotate as a particular new unit we call MODIF. It can also modify an EVENT and make a TIMEX out of it: *the beginning of* in "*at the beginning of the war*", *at* is a signal, and *the war* is an EVENT. This modification of the norm aims at homogenizing it as argued in (Teissèdre et. al 2010).

4 Treebank Annotation for a Better Delimitation of Eventualities

TimeML annotation guidelines request <EVENT> tags to be restricted to the lexical head of the minimal event-denoting chunks (Bittar, 2008). This restrictive delimitation was questioned in (Pustejovsky et al. 2006). Unfortunately, their proposal to integrate the head with its arguments was not retained in the ISO norm, presumably to ease the annotation.

However, the experiments conducted during the TEMPORAL project did not demonstrate difficulties to delimit eventualities with a large span. We therefore advocate a broader annotation covering the whole event-denoting expression, to keep all the relevant information for temporal reasoning. Besides, this proposal enables a straightforward representation of nested eventualities. Consider the following example:

(4) *The Queen launched the yachting race*

(4a) **head annotation (TimeML)**

The Queen [launched] [the yachting race]

(4b) **large-span annotation**

[*The Queen launched [the yachting race]]*

A large-span annotation gracefully accounts for the dependence – directly expressed by syntax – between the two embedded EVENTs (the launch and the race). Conversely, the TimeML head annotation requires the definition of a spurious subordinate (SLINK) relation. The interest of a broader annotation is even more obvious with temporal abstract anaphora (Asher 1993), whose resolution often needs the consideration of a whole clause, like in the following example cited by (Dipper & Zinnmeister, 2010)

(5) Each fall, <u>penguins migrate to Fiji</u>. It happens just before the eggs hatch

[6] Oral seminar in Paris in 2014 (Pustejovsky 2014).

A large-span annotation presents however some drawbacks, in particular when the linguistic extent of an eventuality is discontinuous, as observed with speech disfluencies. Annotation tools have proposed formal solutions to model discontinuities. For instance, the Glozz or Analec platforms define a specific structure (schemata) gathering the different parts of discontinuous chunks (Widlöcher & Mathet 2012; Landragin et al. 2012). Their separate delimitation, and their explicit grouping may complicate however the annotation task.

In order to overcome these difficulties, we propose to adapt a solution that was investigated for multi-word expressions in the Prague Dependency Treebank (Bejček & Straňák, 2010). The idea is to characterize eventualities not at the text level, but on the syntactic structures of a treebank. The resulting benefits are twofold. Firstly, it eases the delimitation of the eventualities: practically, the annotation task boils down to the selection of the correct node, without going into details of the dependencies between the lexical head and its potential subcategorized arguments. In addition, discontinuities are directly merged on the same node in a treebank representation.

The preliminary experiments conducted in the TEMPORAL project suggest that this treebank annotation is workable, provided that a phrase-structure approach is considered: annotators encounter difficulties to characterize easily the span of time-denoting items with a dependency treebank. It is well known that phrase-structure approaches do not deal with non-projective structures that should result from discontinuities. Nevertheless, a previous study on two spontaneous spoken dialogue (Antoine & Goulian 2001) has shown that this phenomenon concerns less than 2,3% of the utterances, in addition, the quantitative impact of discontinuities on EVENTs expressions should be investigated.

The changing from a head annotation to a tree-based one modifies noticeably the overall philosophy of the ISO-TimeML norm. Indeed, it imposes a stand-off annotation in order to align tree nodes with the raw text, while TimeML implements an inline markup of time-denoting items. One should however consider that stand-off annotation is a valuable guarantee in terms of annotation adaptability. The ODIL project, starting in 2016, will investigate theoretically and practically the consequences of this modification. We will constantly endeavour to remain as compatible as possible with the current state of ISO-TimeML. One issue of the project will be to assess the reliability of the resulting tree annotation, and to investigate the impact of this change on annotation cost.

5 Conclusion

ISO-TimeML is a widespread standard which has guided the achievement of most time-annotated corpora. Despite this indisputable success, this paper has tried to demonstrate that TimeML suffers from various drawbacks that prevent a larger use of the norm. We have first proposed some modifications which enable an upward compatibility with the norm and with the existing resources. We expect these restricted but useful changes to be accepted by the scientific community as an answer to the next call for modification of the norm (2017). The question of the evolution of ISO-TimeML towards a stand-off treebank mark-up raises trickier issues that must be investigated carefully. This is precisely the aim of the ongoing ODIL project, which will start in 2016.

6 Bibliographical References

Allen J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), pp. 832–843.

Antoine J.-Y. & Goulian J. (2001). Etude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé : application à la CHM orale. In *Traitement automatique des langues*, 42(2), pp.413-440.

Asher N. (1993). *Reference to Abstract Objects in Discourse*, Kluwer Academic Publisher.

Bach E. W. (1981). Time, Tense, and Aspect: An Essay in English Metaphysics. In *Radical Pragmatics*, pp. 63–81.

Bejček E. & Straňák P. (2010). Annotation of multiword expressions in the Prague Dependency Treebank. *Language Resources and Evaluation*, 44(1–2), pp. 7–21.

Binnick, R. I. (2012). *The Oxford handbook of tense and aspect*. Oxford University Press.

Bittar A. (2008). Annotation des informations temporelles dans des textes en français. Actes *RECITAL'2008*.

Bittar A., Amsili P., Denis P., Danlos L. (2011). French TimeBank: An ISO-TimeML Annotated Reference Corpus. *Proc. ACL'2011*, Portland, Oregon: United States of America.

Caselli T. (2009). Time, Events and Temporal Relations: an Empirical Model for Temporal Processing of Italian Texts. Ph.D. Dissertation, University of Pisa, available at http://etd.adm.unipi.it/theses/available/etd04242009-113147/

Costa F. and Branco A. *(2013).* Full-fledged temporal processing: bridging the gap between deep linguistic processing and temporal extraction. *Journal of Language Modelling* Vol 1, No 1 (2013), pp. 97–154.

Cheng Y., Asahara M. and Matsumoto Y., 2008. In Proc. IJCLCLP, 13(2).

Dipper S., Zinsmeister H. (2010). Towards a standard for annotating abstract anaphora. In *Proc. LREC'10 Workshop on Language Resources and Language Technology Standards*, pp. 54-59.

Drat L. (2014). Projet TourInFlux. Annotation des expressions temporelles. Master's Thesis. Stendhal University, Grenoble, France. http://dumas.ccsd.cnrs.fr/dumas-01068476

Im S., You H., Jang H., Nam S., Shin H. (2009). KTimeM: Specification of Temporal and Event Expressions in Korean Text. In proc. of the 7th workshop on Asian Language Resources in conjunction with ACL-IJCNLP 2009, Suntec City, Singapore.

ISO (2009). ISO DIS 24617-1:2009 Language resource management - Semantic annotation framework - Part 1:

Time and Events. International Organization for Standardization, ISO Central Secretariat, Genève, Suisse.
Landragin F., Poibeau B., Victorri B. (2012). ANALEC: a New Tool for the Dynamic Annotation of Textual Data. In *Proc. LREC'2012*, pp. 357-362.
Mani I., Pustejovsky J., Gaizauskas R., Sauri R., Castano J., Littman J., Setzer A., & Katz G. (2005). The specification language TimeML. In Mani I., Pustejovsky J., Gaizauskas R. (Eds.), *The language of time: a reader*, chapter 27. Oxford.
MERLOT (2015) *Annotation scheme for the MERLOT French Clinical Corpus*. Tech. Report. LIMSI, Orsay, France.
Pustejovsky J. (2014). *Computational Models of Events*. Alpage Seminar. 24 th of March & 4th of April 2014. Paris : Université Paris Diderot, France.
Pustejovsky J., Littman J., Sauri R. (2006). Argument Structure in TimeML. In Katz G., Pustejovsky J., Schilder F. (Eds.) *Annotating, Extracting and Reasoning about Time and Events*, Dagstuhl Seminar Proc., Schloss Dagstuhl, Allemagne.
Reichenbach, H. (1947). *Elements of Symbolic Logic*. The Macmillan Company.
Teissèdre C., Battistelli D., Minel J.-L. (2010). Resources for Calendar Expressions Semantic Tagging and temporal Navigation through Texts. In *Proc. LREC'10*, pp. 3572-3577.
Widlöcher A., Mathet Y. (2012). The Glozz platform: a corpus annotation and mining tool. *Proc. of the 2012 ACM symposium on Document engineering*, pp. 171-180.
Winford, Donald. (2012). Creole Languages. *The Oxford Handbook of Tense and Aspect*, 428.
Zhou Y, Xue N.(2011). Discourse-constrained Temporal Annotation." In *Proc*. *LAW V*, Oregon, Portland.