

Enriching the Ita-TimeBank with Narrative Containers

Alice Bracchi

Università degli Studi di Pavia
C.so Strada Nuova 65
27100 Pavia

alice.bracchi@gmail.com

Tommaso Caselli

Vrije Universiteit Amsterdam
De Boelelaan 1105
1081 HV Amsterdam

t.caselli@vu.nl

Irina Prodanof

Università degli Studi di Pavia
C.so Strada Nuova 65
27100 Pavia

irina.prodanof@gmail.com

Abstract

English. This paper reports on an annotation experiment to enrich an existing temporally annotated corpus of Italian news articles with Narrative Containers, annotation devices representing temporal windows in text and marking up very informative temporal relations between temporal entities. The annotation has shown that the distribution of Narrative Containers is sensitive to the text genre and may be used to facilitate the creation of informative timelines.

Italiano. *Questo lavoro illustra i risultati di un esperimento di annotazione per l'identificazione di Contenitori Narrativi, ovvero marcatori di "finestre" temporali in un testo, come strategia per arricchire un corpus di articoli di quotidiano in lingua italiana, già annotato con informazioni temporali. L'annotazione ha mostrato che la distribuzione dei Contenitori Narrativi è legata al genere testuale e può essere usata per facilitare la creazione di linee temporali di eventi più informative.*

1 Introduction

Research in Temporal Processing has seen an increasing interest thanks to the availability of annotation schemes and corpora in multiple languages (Pustejovsky et al., 2003; Bittar et al., 2011; Caselli et al., 2011; Sauri and Badia, 2012), and the organization of evaluation campaigns (TempEval (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013), Clinical TempEval (Bethard et al., 2015; Bethard et al., 2016), Cross-Document TimeLine (Minaud et al., 2015), Temporal QA (Llorens et al.,

2015)), and EVENTI (Caselli et al., 2014)). This has established best practices, common evaluation frameworks, international standards (e.g. ISO-TimeML (Pustejovsky et al., 2010)), and approaches to solve such a complex task. However, the expression of time in text/discourse is by no means obvious and the automatic extraction of timelines is not a solved task yet. One of the limits of current annotation frameworks and corpora relies mainly in the sparseness of the available temporal relations and in the fine-grained values used to classify the temporal links. For instance, in the TempEval-3 corpus the ratio between temporal relations and event plus temporal expressions is 0.8 (Bethard et al., 2014) for 13 temporal values. In the EVENTI corpus, the ratio is even smaller, only 0.19 for 13 temporal values.¹ Furthermore, in some cases annotation guidelines are not informative enough concerning what types of temporal links to annotate, or they force the annotation of temporal relations between pairs of events when they should not be annotated. Attempts to overcome these limits have focused on three main strategies: i.) annotating particular sets of temporal relations (Kolomiyets et al., 2012); ii.) elaborating detailed annotation guidelines for each kind of temporal relations (event-temporal expression pairs, event-event pairs, and temporal expression-temporal expressions pairs); and iii.) developing densely connected temporal graphs, where all valid relations among the temporal entities (events and temporal expressions) are marked up, including inferred relations based on transitive properties of the temporal relations (e.g. if event A is BEFORE event B and event B IS_INCLUDED in event C, then event A is BEFORE event C) (Bethard et al., 2014). We

¹The smaller ratio for the Italian data is also due to specific restrictions on the annotation of the temporal relations as reported in the EVENTI Annotation Guidelines and explained in Section 2.

consider these solution as partial as they are not able to address the issue of identifying and extracting *informative* timelines, i.e. a set of maximally informative temporal links where relevant events in a text/discourse are correctly anchored to time, and then chronologically ordered. This paper reports on the first annotation effort to enrich existing resources for Temporal Processing in Italian by adopting a document-level approach rather than a sentence-level one. Following the proposal of **Narrative Containers** (NCs) (Pustejovsky and Stubbs, 2011), as embedding intervals where events occur, we developed an annotation scheme for their identification on the EVENTI corpus (Caselli et al., 2014)², as a strategy to increase the informativeness of the existing annotations and, possibly, improve systems’s temporal awareness.

The remainder of this paper is structured as follows: the EVENTI corpus will be shortly introduced in Section 2, with a particular emphasis on the available temporal relations. Section 3 will present the notion of Narrative Container and the proposed annotation scheme. In Section 4 the results of a pilot annotation on the EVENTI dataset will be reported. Finally, conclusion, future work, and a pointer to the annotated data and guidelines will be reported in Section 5.

2 Temporal Relations in the EVENTI Corpus

The EVENTI corpus, released in the context of the EVALITA 2014³ workshop, consists of 3 datasets: the Main task training data, the Main task test data, and the Pilot task test data. The corpus has been annotated with a simplified version of the It-TimeML Annotation Guidelines (Caselli et al., 2011), an adapted version to Italian of the TimeML Guidelines. Four tags have been used to annotate the data: `EVENT`, `TIMEX3`, `SIGNAL`, and `TLINK`.

The `EVENT` tag is used to annotate all lexical items which may realize an event mention. It includes verbs, nouns, adjectives, and prepositional phrases. The tag is enriched with 8 attributes expressing tense, (grammatical) aspect, part-of-speech, mood, modality, verb form, TimeML class, and polarity.

²<https://sites.google.com/site/eventievalita2014/>

³<http://www.evalita.it/2014/tasks/eventi>

The `TIMEX3` tag is used for the annotation of temporal expressions (timexes), expressing the type, the value and whether the timex is absolute or relative (e.g. “2015-05-18” vs. “yesterday”[ieri]).

The `SIGNAL` tag is employed to mark any linguistic elements, such as prepositions (e.g. *in* [in]), adverbs (e.g. *before* [prima]), or conjunctions (e.g. *when* [quando]), which support the identification and classification of a temporal relation between target entities (e.g. events and timexes).

Finally, the `TLINK` tag is used to annotate temporal relations. In the EVENTI task, the subset of possible temporal relations has been restricted to three subtypes of intra-sentence relations, namely: i.) pairs of syntactic main events in the same sentence; ii.) pairs of syntactic main event and subordinate event in the same sentence; and iii.) pairs of event and timexes. All 13 temporal relation values from It-TimeML (`BEFORE`, `AFTER`, `IS_INCLUDED`, `INCLUDES`, `SIMULTANEOUS`, `I(MMEDIATELY)_AFTER`, `I(MMEDIATELY)_BEFORE`, `IDENTITY`, `MEASURE`, `BEGINS`, `ENDS`, `BEGUN_BY` and `ENDED_BY`) have been used.

The Main task datasets, which have been enriched with Narrative Containers, add up to 130,279 tokens, divided into 103,593 tokens for training and 26,686 for test. They contain 21,633 `EVENTS` (17,835 in training and 3,798 in test), 3,359 `TIMEX3` (2,753 in training and 624 in test), 1,163 `SIGNALS` (923 in training and 231 in test), and 4,561 `TLINKS` (3,500 in training and 1,061 in test).

3 Adding Narrative Containers to News Articles

The notion of Narrative Container (NC) was first introduced by Pustejovsky and Stubbs (2011) to deal with some aspects of Temporal Processing, such as sensitivity to the text genre and interaction with discourse relations, not addressed in the TimeML Guidelines nor in the TimeBank corpus. NCs were proposed as a temporal window, providing left and right boundaries, to when events not anchored to timexes could have happened, thus overcoming issues related to linking of events with the Document Creation Time (DCT), i.e. when a text was written or published. In particular, standard TimeML markup imposes that all events have

a link with the DCT but fail to specify that each event should also be annotating to its actual temporal anchor, i.e. to its moment of occurrence. As reported in Pustejovsky and Stubbs (2011), in example 1, TimeML guidelines will order both event mentions, e_1 and e_2 , to the DCT with a BEFORE relation, anchor e_1 to the timex “yesterday” (t) but will fail to provide the anchoring of e_2 :

1. The bomb **exploded** _{e_1} *yesterday* _{$t_{2011-09-09}$} and **killed** _{e_2} three people. [DCT=2011-09-10]

A further justification to the introduction of NCs is related to the different informational status of temporal relations. Assuming the informativeness of a temporal link as a function of the information contained in the individual links and their closure, an anchoring relation, that is a relation between a timex and an event explicitly stating when the event occurred as the one between e_1 and timex “yesterday” in example 1 (i.e. a temporal value of INCLUDES or IS_INCLUDED), is assumed to be more informative than an ordering relations, i.e. a precedence relation between two events.

To the best of our knowledge, the only corpus which extensively adopts the notion of NC and has available annotated data is the THYME corpus of clinical narratives (Styler IV et al., 2014). Our task is the first attempt at tackling temporal containment annotation over news articles in Italian.

A NC enables an accurate reproduction of the way events in text cluster around temporal reference points, explicitly or implicitly realized in the document, as the narration unfolds. NC relations are thus anchoring relations between pairs of events or events and temporal expressions. They are marked with an additional link tag, i.e. CONTAINS, to distinguish them from standard TLINKs. Each NC relation admits two components: i.) the narrative anchor, i.e. an element pointing to a specific temporal dimension shared by other events or timexes within the text; and ii.) the anchored element(s), i.e. events which satisfy the anchorability requirements (see Section 3.1 for details) and participate in an NC relation. Timex anchors are chosen on a transparency basis (i.e. granularity and nature of the timex), whereas Event anchors are chosen according to their relevance and salience for the timeline.

Two sub-types of NCs can be identified:

- **Temporal Containers (TCs):** they correspond to the timexes in the text which clearly anchor the events in analysis on a timeline; the relation can hold both at intra- and inter-sentence level. Example 2 from our annotated corpus shows a timex (2001) and the events it anchors (e_1 – e_4):

2. [...] *la Sonata* **composta** _{e_1} nel **2001**_[TCanchor], *il cui primo esecutore* **fu** _{e_2} *lo stesso Lucchesini. In questa* **esecuzione** _{e_3} *si* **ritrovavano** _{e_4} *già tutte le doti musicali di Lucchesini [...].*

- **Event Containers (ECs):** they correspond to event mentions which function as a temporal anchor for other event mentions. ECs can be useful in cases where no anchoring timex is available or to model event-subevent relations. Example 3 shows a sentence with no explicit temporal expression, where the anchoring of events (e_1 – e_3) is possible only with respect to the event (*ricognizione*).

3. [...] *Durante la* **ricognizione**_[ECanchor], *il tenente ha dato* **disposizioni** _{e_1} *per il presidio, e nella* **fase** _{e_2} *iniziale ha* **ordinato** _{e_3} *ai sottoposti di fare rapporto al campo base.*

Figure 1 serves as a visual representation of the NC as annotated in example 2. By means of NCs, a document timeline will result in an ordered succession of NCs rather than of isolated events. This is the NC resulting from the following sentence, taken from the annotated corpus.

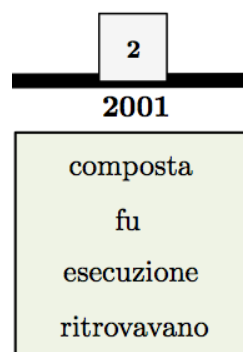


Figure 1: Visual representation of a NC for the sentence in Example 2.

Naturally, the NC represented here is only a visual aid picturing the conceptual outcome of

applying CONTAINS relations between the anchor (here, the `TIMEX3 2001`) and anchored elements (here, `EVENTs composta, fu, esecuzione,` and `ritrovavano`)

3.1 Event Anchorability Requirements

The set of events which can be anchored has been restricted to factual events. The identification of eligible anchorable events has been manually conducted at this stage of the annotation. We adopted the definition of **factuality** as proposed in the FactBank (Saurí, 2008) and which is based on the double axis of *polarity* (positive *vs.* negative) and *certainty*. For the sake of our annotation task, only positive and certain events can be anchored. Events in the future were generally not annotated as they normally do not have a certain status. However, those events with an established schedule (e.g. deadlines, meetings), or whose future temporal window is assumed to be certain, such as festivities, have been annotated in anchoring relations as well.

We excluded all events which are presented as **subjective** (i.e. judgements, opinions). In example 4, *esplosione* is a factual event and was anchored as such, whereas *sbagliato* describes it through the grid of the writer’s judgement, who states that the explosion was a mistake, and thus not anchored.

4. *L’esplosione_{e1} è avvenuta a mezzanotte di lunedì [...]. Insomma, gli attentatori hanno sbagliato_{e2} obiettivo.*

Finally, **generic** events, i.e. events which acquire some kind of attributive value towards discourse participants, expressing persistent properties or reiterated, habitual activities, were not anchored.

4 The EVENTI-NC Corpus

The EVENTI-NC corpus includes documents from both the training and the test sections of the Main task of the EVENTI corpus. It includes 58 annotated articles, for a total of 24.259 tokens, covering roughly 11% of the EVENTI corpus; Table 1 shows the number of `EVENTs` and `TIMEX3` involved in our annotation.

Table 2 reports the number of annotated containers in our corpus, and their distribution according to their type. TCs make up for almost 63% of the total number of NCs, against the 37% of ECs.

General EVENTI-NC statistics	
Annotated tokens	24.259
Annotated articles	58
EVENT markables	3645
TIMEX3 markables	595

Table 1: Overview of corpus statistics.

Annotated NCs		
Type	Number	%
ECs		
Verbal anchors	61	19.5
Nominal anchors	55	17.6
Total EC n.	116	37.1
TCs		
Text-consuming TIMEX3s	160	51.1
Empty TIMEX3s	37	11.8
Total TC n.	197	62.9
Total NC n.	313	

Table 2: Distribution of Narrative Containers in the corpus.

It is interesting to notice that 11.8% of TCs is realized by empty `TIMEX3s`, i.e. temporal expressions which do not correspond to lexical items but can be inferred and which are necessary to for assigning a correct value to a `timex`.

4.1 Distribution of Narrative Containers anchors

We conducted an in-depth analysis of the NC anchors following two parameters: i.) the properties of NC anchors on their own; and ii.) the sensitivity to the document genre, i.e. the news domain, on the line of Pustejovsky and Stubbs (2011).

Concerning the first parameter, we first investigated the incidence of verbal anchors as opposed to nominal anchors. Whereas there appear to be no tendency towards verb or nouns being more likely to anchor other events, it is interesting to take a look within these categories. Out of all the verbal anchors, 42.9% are reporting verbs or verbs employed in a declarative context. We observed that there is a preference for ECs to correspond to the the event with the highest degree of topicality in the article, or the most important event (climax

event). For example, one article⁴ reports on President Clinton’s surgery in 2004: the largest EC in the document is anchored by *intervento* (*surgery*), with a total of 12 anchored items.

Sensitivity to text genre can be easily observed with TC anchors. 25% of them anchor events in a timespan that can be measured as ± 1 day with respect to the DCT. Anchors for these containers are mostly represented by non absolute temporal expressions, such as temporal adverbs (e.g. “ieri” [yesterday], “domani” [tomorrow], among others) and by the DCT itself, which represents 11% of the TC anchors.

Genre-sensitivity might also be the factor behind the average number of NCs in the corpus. The documents have an average of 5.17 NCs, and even for more lengthy articles, the textual anchors were rarely more than 7. The average of 5.17 NCs/article might be due to the fact that newspaper articles usually refer to a limited number of facts, whose core is usually made of a handful of recent happenings; whereas the fluctuating relationship between length and NC number usually depends both on the content of the article and on the granularity of the selected NC.

5 Conclusion

This paper reports on a first proposal of an annotation scheme and accompanying annotated data for NCs in Italian news articles. The NC annotation is an additional layer on top of already available data for Temporal Processing in Italian. It addresses pending issues (e.g. the annotation of the temporal relations between event and the Document Creation Time) and increases the informativeness of the document timelines. Overall, we observed that there is a preference for NC to be realized by timexes in a limited time span. However, NCs may also be realised by events. In this case, nouns and verbs have a similar distribution with a preference for events which have a central role in the news or facilitate the clustering of the information (e.g. reporting events). Such a behaviour is different with respect to clinical narratives where nominal events are more frequently selected as NC (Bethard et al., 2016). This suggests that different text genres present different ways of organizing events on a timeline. The introduction of the factuality parameter to select the anchoring events is a strategy to clean timelines and to move Temporal Process-

ing from a single document to a cross-document task.

Future work will aim at assessing the reliability of the proposed scheme via an inter-annotator agreement study and at completing the annotation of the entire EVENTI corpus. Finally, the annotated data and guidelines are publicly available⁵ to encourage additional testing and experiments.

References

- Steven Bethard, Nathaniel Chambers, Bill McDowell, and Taylor Cassidy. 2014. An annotation framework for dense event ordering. In *52nd Annual Meeting of the Association for Computational Linguistics*, 52, Baltimore, MD, USA, June. Association for Computational Linguistics.
- Steven Bethard, Guergana Savova, James Pustejovsky, and Mark Verhagen. 2015. SemEval-2015 Task 6: Clinical TempEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 806–814. Association for Computational Linguistics.
- Steven Bethard, Savova Guergana, Leon Derczynski, Wei-Te Chen, James Pustejovsky, and Mark Verhagen. 2016. SemEval-2016 Task 13: TempEval-2. In *Proceedings of SemEval-2016*, SemEval ’13, pages 977–987, Stroudsburg, PA, USA. Association for Computational Linguistics.
- 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: Human Language Technologies: short papers-Volume 2*, pages 130–134. Association for Computational Linguistics.
- 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 Fifth Law Workshop (LAW V)*, pages pages 143–151, Portland, Oregon. Association for Computational Linguistics.
- Tommaso Caselli, Rachele Sprugnoli, Manuela Speranza, and Monica Monachini. 2014. EVENTI: EVALuation of Events and Temporal INFORMATION at Evalita 2014. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014*. Pisa University Press.

⁴adige20040709_id405401.txt

⁵<https://sites.google.com/site/ittimeml/documents/narrative-container-data.zip?attredirects=0&d=1>

- Oleksandr Kolomiyets, Steven Bethard, and Marie-Francine Moens. 2012. Annotating narrative timelines as temporal dependency structures. In *Proceedings of the International Conference on Linguistic Resources and Evaluation*.
- Hector Llorens, Nathanael Chambers, Naushad Uz-Zaman, Nasrin Mostafazadeh, James Allen, and James Pustejovsky. 2015. SemEval-2015 Task 5: QA TempEval—Evaluating temporal information understanding with question answering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 792–800.
- Anne-Lyse Minard, Manuela Speranza, Eneko Agirre, Itziar Aldabe, Marieke van Erp, Bernardo Magnini, German Rigau, Ruben Urizar, and Fondazione Bruno Kessler. 2015. Semeval-2015 task 4: Timeline: Cross-document event ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 778–786.
- James Pustejovsky and Amber Stubbs. 2011. Increasing informativeness in temporal annotation. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 152–160. Association for Computational Linguistics.
- James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. 2003. The TimeBank corpus. In *Corpus linguistics*, volume 2003, page 40.
- James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. ISO-TimeML: An international standard for semantic annotation. In *Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.
- Roser Sauri and Toni Badia. 2012. Spanish TimeBank 1.0. *LDC catalog ref. LDC2012T12*.
- Roser Saurí. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University.
- William F Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, et al. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.
- Naushad UzZaman, Hector Llorens, Leon Derczynski, James Pustejovsky, and James Allen. 2013. Semeval-2010 task 13: Tempeval-2. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, SemEval '13, pages 1–9, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. SemEval-2007 Task 15: TempEval Temporal Relation Identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07*, pages 75–80, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Marc Verhagen, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval '10*, pages 57–62, Stroudsburg, PA, USA. Association for Computational Linguistics.