# Temporal Relations with Signals: the Case of Italian Temporal Prepositions

Tommaso Caselli, Felice Dell'Orletta, Irina Prodanof
*Istituto di Linguistica Computazionale*
*CNR Pisa, Italy*
*firstName.secondName@ilc.cnr.it*

***Abstract*—This paper presents a Maximum Entropy tagger for the identification of intra-sentential temporal relations between temporal expressions and eventualities mediated by temporal signals in constructions of the kind "eventuality + signal + temporal relation". The tagger reports an accuracy rate of 90.8%, outperforming the baseline (81.8%). One of the main results of this work is represented by the identification of a set of robust features which may be automatically obtained with a relative computational effort.**

***Keywords*-temporal relations; signals; Italian**

## I. Introduction

In recent years a renewed interest in temporal processing of real texts has spread in the NLP community, boosted by the presence of specific mark-up languages and efforts to create an international standard (ISO-TimeML, SemAF/Time Project), and by a growing number of initiatives (CLEF[1], TERN[2], TempEval[3]). Extracting temporal information is both a challenging and an extremely useful task to improve semantic access to information from which many NLP applications, such as Information Extraction, (I.E.), and Question-Answering (both Domain-Specific and Open-Domain), can improve their performance.

The identification of temporal relations in text/discourse is not a trivial task. A lot of previous research, ([1], [2], [3], [4], [5], [6]), has explored and analyzed the different sources of information which are at play when inferring the temporal orders of eventualities, such as tense, temporal adverbials, signals, viewpoint and lexical aspect, discourse rhetorical relations, commonsense and pragmatic knowledge. For instance, in absence of more informative elements such as temporal expressions or signals of temporal relations, a sequence of two eventualities with same tense forms can be ordered by making reference either to pragmatic conventions, or commonsense knowledge, or by means of lexical aspectual information. Thus, in example 1. the precedence relation ($\prec$) between *e1* and *e2* is obtained by exploiting our commonsense knowledge which suggests that before eating an ice-cream, one has to buy it. On the other hand, in example 2. the overlap/inclusion relation ($\subset$) is obtained by inferencing mechanisms based on the knowledge of the lexical aspect, namely to the fact that *e4* is a state and *e3* is a point-like event.

1.) Marco *ha comprato*$_{e1}$ un gelato. L'*ha mangiato*$_{e2}$ con calma.
*Marco bought an ice-cream. He ate it slowly.*
$e_1 \prec e_2 \longrightarrow$ `commonsense knowledge`.

2.) Marco *è uscito*$_{e3}$. Giovanni *è rimasto*$_{e4}$ nella stanza.
*Marco left. Giovanni remained in the room*
$e_4 \subset e_3 \longrightarrow$ `lexical aspect`.

Substantial linguistic processing is required for a system to perform such inferences and commonsense knowledge can hardly be encoded in a domain independent program. Most sources of information for inferring a temporal relation rarely code in an explict and clear-cut way the specific temporal relation holding between two entities (i.e. eventualities and temporal expressions) which may lead to biases and incorrect tagging. This calls for the development of procedures and techniques to maximize the role of the various sources of information and the conditions under which they are necessary and sufficient to determine the current temporal relation.

This paper presents a Maximum Entropy tagger for the identification of temporal relations between temporal expressions and eventualities in presence of temporal signals, which represents a specific module of a more complex empirically based model for temporal processing of Italian text/discourse. Signals, together with a restricted set of other elements like temporal expressions, represent a relative explict means to code temporal relations, thus the identification of their semantics is a relevant task in order to improve the extraction of temporal information from a text/discourse. In section II we will present the theoretical background at the basis of this work and a description of the methodology followed. Section III is devoted to the presentation of the features used to train the tagger and their role for the accomplishment of our task. In section IV we will present the evaluation of our algorithm. Provided the fact that, to our knowledge, no other algorithm has been developed for this task in Italian, no cross systems comparison can be performed. Finally in section V we will present our conclusions and extensions for future works.

[1]http://clef-qa.itc.it/2008/important-dates.html
[2]http://www.nist.gov/speech/tests/ace/2007/index.html
[3]http://www.timeml.org/tempeval/

978-0-7695-3727-6/09 $25.00 © 2009 IEEE
DOI 10.1109/TIME.2009.23

## II. Theoretical Background on Italian Signals and Methodology

The term "signals" is used in this work as a cover term for a relative homogeneous set of parts of speech which are used to express, either explicitly or implicitly, different types of relations between textual entities. We claim, following [7] that the semantics of signals in general can be expressed by the formula *Rel (X, Y)*, where *Rel* represents the associated relation(s) of the signal, and *X* and *Y* the textual entities connected together by the signal. Signals can be divided into two macrocategories on the basis of their semantic transparency, namely **explicit** and **implicit** signals. Explicit signals are a set of signals whose meaning is self-evident and stable and can be directly represented by the *Rel* without taking into account the contribution of the elements it connects. On the contrary, implicit signals are realized by a set of parts of speech whose meaning is highly abstract and gets specialized according to the semantic properties of the elements which precede and follow the signal itself. The semantics of this set of signals needs more fine-grained information and thus can be represented as $Rel(\lambda(X), \lambda(Y))$.

Temporal signals are a restricted set of signals whose *Rel* value corresponds to a temporal relation(s), and the *X* and *Y* to temporal entities, namely eventualities and temporal expressions. The possible realizations of temporal signals comprehend:

- (temporal) prepositions;
- (temporal) conjunctions;
- (temporal) adverbs and adverbial constructions.

Temporal signals, both explicit and implicit, can occur in three different types of constructions involving text/discourse temporal entities, namely:

- they can temporally relate two temporal expressions;
- they can temporally relate a temporal expression and an eventuality (or viceversa); and finally
- they can temporally relate two adjacent eventualities.

As already stated, in this work we have taken into account the second type of constructions, i.e. signals which relates an eventuality and a temporal expression. The choice of this kind of constructions has been dictated by two observations which have emerged as the results of a corpus exploration: first, most temporal expressions, when present in a sentence and not realized by time or date patterns, are introduced by a signal and, secondly, the relative vast majority of signals introducing a temporal expressions are implicit signals realized by temporal prepositions like *in* [in], *a* [at/on], *da* [since/for], *per* [for]. Moreover, this type of constructions are very important since they provide, either directly or indirectly by means of reduced temporal reasoning mechanisms, information on the temporal location of the eventuality, i.e. they may be used to provide answers to questions like *When did X happen?*.

Table I
Implicit signals and their frequencies.

| Implicit Signals | Absolute freq. | Temporal freq. |
|---|---|---|
| in | 76457 | 5265 |
| per | 30200 | 1658 |
| da | 42498 | 2318 |
| a | 80673 | 3587 |
| di | 110721 | 3378 |
| entro | 637 | 637 |
| tra | 5138 | 242 |

### A. Identifying Temporal Signals

One of the main results of the corpus exploration has been the identification of an extensive set of temporal signals to start with. Nevertheless, in order to obtain the most comprehensive list of temporal signals we have performed a corpus study by using a 5 million word shallow parsed corpus of contemporary Italian, drawn from a subset of the PAROLE corpus ([8]). We have then automatically extracted only those parts of speech which may assume the status of temporal signals. As far as conjunctions and adverbs are concerned we have manually checked their immediate context (a reduced window of the five preceding and following chunks) and we have observed that none of them occurs in the type of construction we have taken into account in this work. As for prepositions, in addition to the local context, i.e. the five chunk windows, we have also extracted the noun heads and matched them to their ontological types by associating the head noun of each prepositional chunk to its ontological type by means of a database query with the SIMPLE Ontology ([9]). Only those noun heads which have a positive match with the ontological type of "TIME", which is defined in SIMPLE as all nouns referring to temporal expressions, have been extracted and all instances of false positives have been eliminated, i.e. we have excluded words like "*incubazione*" [incubation], "scuola" [school], "assessorato" [chancellorship] and suchlike, since they may have an associated temporal reading but do not correspond to the *de facto* standard definition of temporal expression used by the NLP community, i.e. "those temporal expressions which contained a reserved time word, called **lexical trigger**" ([10]: 2). Temporal signals, thus identified, have been classified on the basis of their semantic transparency, i.e. whether they are explict signals or implicit ones. The working hypothesis we have applied to state the semantic transparency is based on their morphological form: polysyllabic signals should be semantically explicit while monosyllabic ones should be semantically implicit. The working hypothesis has been confirmed by this first study, though not as a rule but as a general tendency. In Table I we report the set of implicit temporal signals we have identified, their absolute frequency in the 5 million-word corpus and the relative frequency when they assume a temporal value.

### B. Temporal relations: resolving the Rel

In order to identify the possible temporal relations coded by the implicit signals, i.e. the *Rel* value, we have used a sub-corpus of 499 randomly extracted occurrences of the constructions "eventuality + signal + temporal expression", which has been manually annotated by one investigator with temporal relations in a bottom-up approach. The tagset used was partially inspired by the TimeML TLINK ([11]) values. The main differences are:

- the relations of *includes* and *during* have been collapsed to the coarse-grained value of *overlap*;
- two new values, namely *before_ending* and *equals*, have been introduced to account for the temporal relations of the signals *entro* [by] and *a* [at] whose semantics could not be represented by using the TimML tagset;

The final tagset used is composed by 9 relations, namely *overlap*, *simultaneous*, *before*, *after*, *no tlink*, *begin*, *end*, *before_ending*, *equals*.

The assignment of the *Rel* values has been performed by means of paraphrase tests. All the paraphrases have the same scheme, based on the general formula *Rel* ($\lambda$(e), $\lambda$(t)), illustrated in 3.

3.) *The event/state of X R t.*

where X stands for the eventuality, R is the set of temporal relations available from the tagset, and *t* represents the temporal expression introduced by the implicit signal. Thus, a sequence like the following:

4.) Sono stato sposato per quattro anni.
*I have been married for four years.*

can be paraphrased and represented as in 5.

5.) The state of "being married" *equals* four years.
$\lambda$(e) = *Perf* (BE_MARRIED (x) $\wedge$ x $\subseteq I \wedge$ ($i_1 \prec I \prec i_2$))
$\lambda$(t) = (FOUR_YEARS (y) $\wedge$ y $\subseteq I_1 \wedge$ ($i_3 \preceq I_1 \preceq i_4$))
($i_1 \equiv i_3$) $\wedge$ ($i_2 \equiv i_4$) $\rightarrow$ EQUALS ($I$, $I_1$)

The only exception is represented by the simultaneous relations which is paraphrased as in 6.

6.) *The event/state X OCCURS (-ED)/HOLDS AT t.*

The inverse relations of binary temporal relations have been excluded, so that all temporal relations have a unique temporal directionality going from the eventuality towards the temporal expressions.

Interesting data have emerged for the class of Temporal Movement Events i.e that subset of verbs or nominal events which semantically express a temporal relation, e.g. *prevedere* [to schedule/to forsee], *anticipare* [to anticipate], *postporre* [to postpone], *ritardare* [to delay], *aggiornare* [to postpone/to update] and similar. In these cases, the implicit signal which co-occurs with them qualifies as a particle pre-selected by the verb. For instance, the verb *prevedere* [to schedule/to forsee] alway co-occurs with the signal *per* which does not have a temporal meaning. This class of eventualities has been excluded from the creation of the training data.

Table II
SIGNALS AND THEIR MEANINGS.

| Signals | Meaning |
|---|---|
| a | overlap — after — begin — end — equals — no tlink — simultaneous |
| da | begin — no tlink |
| di | equals — overlap — simultaneous — no tlink |
| durante | overlap |
| entro | before — before_ending |
| fino a | end |
| fino da | begin |
| in | overlap — after — simultaneous — no tlink |
| per | overlap — equals — no tlink |
| tra | begin — after |

The value *no tlink* has been introduced to account for the fact that the presence of a signal followed by a temporal expression does not always code a temporal relation. It may be the case that it offers information about the temporal distance between the moment of occurrence of an eventuality and the temporal expression. To clarify this statement, consider the following sentence:

7.) Marco *è uscito*$_{e1}$ **da**$_{Signal}$ *10 minuti*$_t$.
*Marco left 10 minutes ago.*
$\lambda$(e) = *Perf* (GO_OUT (x) $\wedge$ x $\subset I \wedge$ (($i_1 \equiv I$) $\vee$ ($i_2 \equiv I$))
$\lambda$(t) = (TEN_MINUTES (y) $\wedge$ y $\subseteq I_1 \wedge$ ($i_3 \preceq I_1 \preceq i_4$)))
($i_2 \equiv i_3$) $\rightarrow$ NO_TLINK ($I$, $I_1$)

In this case, the role of the signal is that of measuring the interval of time which has passed since the event of leaving of Marco, represented by the interval *I*, and the moment of utterance, which coincides with the ending point $i_5$ of the temporal expression *10 minuti* [ten minutes]. Thus, no temporal relation is expressed between the interval representation of the event and of the temporal expression, but a measure of how much time is passed.

### C. Corpus-study results

The combination of the corpus exploration and the corpus study has provided us with a list of signals, both implicit and explicit, and of their meaning(s), i.e. corresponding temporal relation(s). In Table II we report the list of signals which we have identified to occur in the particular construction in analysis together with their meaning(s). Notice that those signals with a unique temporal relation value are cases of explicit signals, while those with more than one are instances of implicit signals. The list in Table II illustrates also the signals for which we have built the training data and submitted to the learner.

## III. Feature identification

In order to identify a set of relevant features for the learner, we have adopted the following strategy which mixes together theoretical statements and corpus frequency data on temporal relations. We have assumed, in particular for implicit signals, that the most frequent temporal relation represents the prototypical meaning. Every other associated temporal relation is given by a different setting of the information coded both from the eventuality and the temporal expressions. For instances, for the signal *in* [in] we have identified four temporal relations, namely *overlap*, *simultaneous*, *after* and *no tlink*, where the overlap relation is the most frequent value. Thus, we have analyzed those sentences with different temporal relations and looked for differences with respect to those which coded the most frequent ones. Such a bottom-up approach has lead to the identification of 15 features, which, in the perspective of our work, provide all the necessary and sufficient information in order to perform the abstraction operation required to determine the semantics of signals. It is important to point out that the lambda abstraction process can be extended to explicit signals as well, since the formula used to represent their semantics, namely *Rel (e, t)*, can be thought as derived from the more general one used for implicit signals, namely *Rel* ($\lambda(e)$, $\lambda(t)$).

The signals we have examined are represented by means of the feature PREP, whose values correspond to the signal lemmas, e.g. *in*, *a*, *da*, *per*.... The set of remaining features can be divided into three main groups: a) features to capture information of the temporal expressions; b) features to capture information on the eventuality, and c) features to capture context information.

### A. Temporal Expression Features

As for temporal expressions we have identified 4 main features, namely:

- the ontological status of the temporal expression which is represented by the features INSTANT and INTERVAL;
- the type of temporal expression; this feature (TIMEX) has four main values, namely DATE, TIME, DURATION and SET. Theses values correspond to the TimeML TIMEX3 tag attribute `type` and offer a standard classification of temporal expressions. In particular, the value DATE applies to temporal expressions describing a calendar date; TIME is used for temporal expressions which describes times of the day, like *tramonto* [sunset], *sera* [evening], *mattina* [morning], and clock times as well; DURATION is used for temporal expressions describing a period of time; finally SET is used to describes set of times which regularly repeat themselves.
- the fact that the temporal expression presents an internal quantifier (QUANTIFIER).

### B. Eventuality Features

A total of 6 features have been identified for eventualities, namely:

- the lemma of the eventuality (POTGOV_head);
- the POS of the eventuality which is represented by two features, namely VERB and NOUN;
- the diatesis, to express if we are dealing with an active or passive eventuality; this features (DIATESIS) is active only with verbal eventualities;
- the presence of negations (NEGATION);
- tense (TENSE) and viewpoint aspect (ASPECT) for verbal eventualities. This two features present a set of relative large values. The ASPECT feature has four values, namely PROGRESSIVE, for progressive eventualities when realized by specific periphrases like "*stare + gerund*", IMPERFECTIVE, corresponding both to the general imperfective value and to cases of progressive viewpoint not realized by periphrases, and PERFECTIVE, corresponding both to the general perfective eventualities and to the more specific perfect reading. The TENSE feature corresponds to the surface tense forms. It has five values, namely PRESENT, IMPERFECT, FUTURE, PAST, for all past tense forms, and INFINITIVE.
- the actual lexical aspectual value of the eventuality (AKTIONSAART). This features has three values, namely, TRANSITION, PROCESS and STATE. The value TRANSITION is used to represent the lexical aspectual value both for instances of telic eventualities, that is eventualities which have a natural culmination point to be reached in order to consider the eventuality as occurred, and also for all those cases of eventualities which cannot be considered a proper telic ones but which give rise to a resulting state which describes the new state of affairs resulting from the happening o the eventuality. This second case of transitions is represented by the subset of incremental eventualities, like *crescere* [to grow up], *aumentare* [to rise] and similar. The value PROCESS is used to describe durative events which do not give rise to resulting states, like *correre* [to run], *camminare* [to walk]; finally, STATE is used to describe stative eventualities.

### C. Contextual Features

The contextual features represent a way to keep under control the occurrence of constructions involving either a signal and a temporal expression, or a signal and an eventuality in the immediate left or right context of the "signal + temporal expression" part of the constructions we have examined. A total of 3 features of this kind as been identified, namely:

- the presence of a further "signal + temporal expression" which follows the current "signal + temporal expressions" (FOLLOWED_SIGNAL+TIMEX); e.g.:

8.) Marco ha corso dalle 3.00 *alle 4.00*$_{FOLLOWED\_SIGNAL+TIMEX}$.
*Marco ran from 3.00 to 4.00.*

- the presence of further "signal + temporal expression" which precedes the current "signal + temporal expressions" (PRECEDED_SIGNAL+TIMEX); e.g.:

9.) Marco ha corso *dalle 3.00*$_{PRECEDED\_SIGNAL+TIMEX}$ alle 4.00.
*Marco ran from 3.00 to 4.00.*

- the presence of a construction of the kind "signal + eventuality" which follows the constructions we have analyzed (FOLLOWED_SIGNAL+EVENT); e.g.:

10.) Il museo riaprirà a vent'anni *dalla chiusura*$_{FOLLOWED\_SIGNAL+EVENT}$.
*The museum will open again after 20 years since it was closed.*

## IV. Learning signal meanings

The feature annotation has been manually conducted on a subset of 1000 occurrences of sentences with the construction in analysis, i.e. instances of a signal relating an eventuality and a temporal expression, randomly selected from the 5 million-word corpus on the basis of two interlinked criteria: a) the level of semantic transparency of the signal, and, b) the relative frequency in the corpus. As a result of the application of these two criteria, the more transparent the semantics of the temporal signal being analyzed, the smaller the number of occurrences extracted (and therefore annotated). The annotation has been performed by one annotator and one of the authors who were allowed to interact with each other to resolve disagreements. In Table III we report the number of annotated instances of each signal together with the most frequent (or unique, if explicit) temporal relation they code.

The relative low values of some implicit signals - for instance, *tra* and *entro* - is due to the global sparseness both of the signals themselves with a temporal meaning and to its occurrence in the type of constructions analyzed.

The task of assigning the right semantics to the temporal signal is in essence a tagging task, i.e. assign either the unique available value or one of the possible values. The use of a Maximum Entropy (M.E.) model perfectly fits for this task since given a linguistic context $c$ and an outcome $a \in A$, where $A$ corresponds to the set of possible values for each signal as illustrated in Table II, the conditional probability distribution $p(a|c)$ is established on the basis of the assumption that no *a priori* constraints must be met other than those related to a set of features $f_i(a,\ c)$ of $c$, whose distribution is derived from the training data. The fact that the signal itself is part of the list of features, preserves the M.E. framework even for the instances of explict signals we have identified.

The 1000 occurrences have been split in a test set (100 occurrences) and a training set data (900 occurrences). In order to perform a cross-validation we have randomized the 1000 instances and extracted the first 100 occurrences, which have been used to create the test data.

Table III
Annotated instances and baseline temporal value.

| Signals | Occurrences | Baseline value |
|---|---|---|
| a | 167 | overlap |
| da | 148 | begin |
| di | 81 | overlap |
| durante | 6 | overlap |
| entro | 33 | before_ending |
| fino a | 36 | end |
| fino da | 9 | begin |
| in | 346 | overlap |
| per | 151 | overlap |
| tra | 23 | after |

Table IV
Annotated instances and baseline temporal value.

| Models | Accuracy |
|---|---|
| Baseline | 81.8% |
| All-feature (15) | 90.8% |
| 10-feature | 90% |
| 9-feature | 89.8% |
| 8-feature | 89.8% |
| 8-feature (No QUANTIFIER) | 85% |
| 7-feature | 86.8% |
| 5-feature | 87.6% |
| 4-feature | 84% |

### A. Evaluation and results

All results are obtained from a 10-fold cross validation and they are reported in Table IV. All the measures illustrate the mean accuracy rate of the tagger, i.e. the percentages of correct tagging. The baseline has been computed by considering the most frequent temporal relation per signal as the correct one.

As the results in the Table IV show all feature configurations result in a net improvement of the baseline. Not surprisingly the best result is obtained when all features are selected, outperforming the baseline of more that 10%. For the All-feature model, it is interesting to notice that more than 60% of the errors occur with the *no tlink* value. A possible explanation for this bad performance of the tagger for this type of relation is mainly due to the relative low number of occurrences of this value. In fact, the wrong tagging of the *no tlink* value mainly occurs with the signals *in* [in], *per* [for], *da* [for/since] and *a* [at], where the occurrences of this values is always less than 10% of the occurrences of the other temporal relations. On the contrary, with the signal *di* [of], where the *no tlink* relation accounts for the 20% of the various possible temporal relations, the number of wrong assignments is almost null. The remaining 40% of errors are sparsed among the other signals and different types of temporal relations. Generalizing the observations, we have found two sources of errors: the first, is due to minor annotation errors in the features and the second to the very low occurrences of the specific temporal relations in the training data.

The 10-feature model is obtained by the following features: PREP, INTERVAL, INSTANT, FOLLOWED SIGNAL+TIMEX, PRECEDED SIGNAL+TIMEX, FOLLOWED SIGNAL+EVENT, TENSE, ASPECT, TIMEX, QUANTIFIER. The result obtained is still very good, though slightly worse than the All-feature model. However, with respect to the results obtained by the other models, the 10-features qualifies as the best. It is interesting to notice that the features are all relevant and, most importantly that they can be obtained with a relative surface based linguistic analysis, i.e. with no need of complicated systems. It is interesting to notice that the influence of the AKTIONSAART feature is almost null. This result is quite surprising since, theoretically, aktionsaart, or lexical aspect, is considered as one of the most distictive features for the identification of the temporal relations involving implicit temporal signals. The types of errors identified are very similar to those described for the All-feature model.

The 9-feature model and the 8-feature one obtain the same results. The 9-feature model is built by using the following features, namely PREP, INTERVAL, INSTANT, FOLLOWED SIGNAL+TIMEX, PRECEDED SIGNAL+TIMEX, FOLLOWED SIGNAL+EVENT, AKTIONSAART, TIMEX, QUANTIFIER. On the contrary the 8-feature differs with respect to the 9-feature for the absence of the AKTIONSAART feature. Comparing these two models is very important because it confirms that the AKTIONSAART feature is not very relevant for the good performance of our task. Nevertheless, a certain influence is still present. In fact, though the two models report the same accuracy, we have identified a different behaviour for the following couples of signal and associated temporal relation, namely *per* - *Rel* = *overlap* and *in* - *Rel* = *after*. We have noticed that with respect to the 9-feature model, in the 8-feature model the relation *per* - *Rel* = *overlap* is improved, i.e. less tagging errors are made, while the relation *in* - *Rel* = *after* worsen. Nevertheless, the results for these couples of signal and temporal relation are still better with the 10-feature model, where the AKTIONSAART is absent. Comparing the feature set of these last three models, namely 10-feature, 9-feature and 8-feature, we can observe that the role of the AKTIONSAART feature as a relevant feature for correct tagging can be accomplished by the TENSE and ASPECT features, which are much easier to be identified. As for the other errors reported by the 9-feature and the 8-feature models they are much the same the one as we have identified for the All-feature model.

A further relevant feature has been identified with the 8-feature (No QUANTIFIER) model. This model is very similar to the 8-feature one, but instead of the QUANTIFIER feature we have used the AKTIONSAART. The low accuracy reported is due to the absence of a crucial information on the internal structure of the temporal expressions. The absence of this features does not allow a fundamental distinction, namely that between quantified duration and non-specific ones. Such a distinction plays an essential role in assigning the correct value for the *in* signal, in particular to distinguish the *overlap* and the *after* temporal relation.

The 7-feature and the 5-feature models are both characterized by a complete lack of information on the eventuality side, in particular:

- the 7-feature model is obtained by the following features: PREP, INSTANT, INTERVAL, FOLLOWED SIGNAL+TIMEX, PRECEDED SIGNAL+TIMEX, TIMEX and QUANTIFIER; and
- the 5-feature model is obtained by the following features: PREP, INSTANT, INTERVAL, TIMEX, QUANTIFIER.

It is interesting to notice that the 5-feature model performs better than the 7-feature one, even if less features are present. The set of errors reported are relatively similar for both models. In particular, we have registered an increase of errors related to an over-extension of the baseline value for each signal. Nevertheless, the two models present also some differences. In particular, the 5-feature model is more accurate in the identification of the *after* relation for the *in* signal, but it always assigns the wrong value to the following couples of signal and temporal relation: *a* - *Rel* = *ending*; *tra* - *Rel* = *begin* and *a* - *Rel* = *after*. These values, in fact, can only be obtained by exploiting the information deriving from the features FOLLOWED SIGNAL+TIMEX, PRECEDED SIGNAL+TIMEX and FOLLOWED SIGNAL+EVENT. Similarly, the 7-feature model assigns the wrong value to the couple *a* - *Rel* = *after* due to the absence of the FOLLOWED SIGNAL+EVENT feature.

The last model, the 4-feature model, is the simplest possible model. It takes into account the most superficial features, namely PREP, VERB, NOUN and TIMEX. Though its performance is very far from the best models, i.e. the All-feature and the 10-feature models, it still outperforms the baseline. This means that there exists a very positive influence of the type of temporal expression. In particular, we have noticed this interesting phenomena:

- with implicit signals, the model does not apply the baseline value *tout court*, but it applies the most frequent values on the basis of the most frequent type of temporal expression. For instance, the *a* [at] signal presents as pure baseline the *overlap* relation. As for this value, the most frequent temporal expression is of type DATE. The model associates the *overlap* relation every time a temporal expression of type DATE is signaled. On the other hand, the second most frequent temporal relation is *simultaneous* which is mainly activated by temporal expressions of type TIME. Thus, every time a temporal

expression of type TIME is found in conjunction with the signal *a* [at], the model assign the *simultaneous* relation;
- with explicit signals it always assigns the correct value, since the the 4 features giving rise to the corresponding model are the basic feature needed for the identification of the semantics of the explicit signals, i.e. they corresponds to the simple *Rel (e, t)* formula.

## V. Conclusions and future work

This paper has illustrated the results of a M. E. tagger for determining the semantics of temporal signals in constructions of the kind "eventuality + signal + temporal expression". The various models we have experimented have shown that there is a mismatch in features' relevance between linguistic theory and actual applications. With the exception of the All-feature model, the best feature configuration is represented by the 10-feature model. As for the AKTIONSAART, we have shown that its presence is almost irrelevant to the good performance of the tagger. Nevertheless, we have also noticed that it can have a certain influence, in particular for the identification of the *after* relation with the *in* [in] signal. We claim that the AKTIONSAART feature could be substituted with a reduced set of very simple features, namely features which deal with expressed information on the subject (definiteness, countable/uncountable, human/non-human) and on the direct object (presence, cardinality, definiteness). This information represents a strategy to by-pass the computation of the lexical aspect and exploit more surface based information on the surrounding elements of the eventuality which are well known to influence the final assignment of the lexical aspectual value.

An interesting element which has emerged is that the information expressed by the temporal expression is extremely important. In particular, we have identified in the feature QUANTIFIER one of the distinctive features for the good performance of the tagger. From a certain point of view, and also on the basis on the results of the 4-feature model, we can state that the information provided by the temporal expressions offers a sort of necessary and sufficient condition for the identification of the right semantics of the signals, while the information obtained from the eventuality are only necessary.

A further observation is related to the difference between the quantitative analysis of the features and their quality. We can identify a restricted set of core features which cannot be excluded for a good, though simple, performance of the tagger. These core features have been identified with the 5-feature model, and they are PREP, INSTANT, INTERVAL, TIMEX, QUANTIFIER. The relevance of the remaining features is linked to the correct tagging of specific temporal relations. For instance, for the identification of the *after* relation of the *in* signal, features from the eventuality, such as TENSE, ASPECT and AKTIONSAART are required. Similarly, for the *end* relation of the *a* signal the PRECEDED SIGNAL+TIMEX feature is highly important.

Provided the good results obtained, we aim at developing a complete automatic system for the identification of temporal relations in presence of signals. This requires the integrations of different tools, from a dependency parser up to a temporal expression tagger.

Future research will be oriented to the identification of the relevant features for the other two types of constructions. In particular, an extensive corpus study should be performed in order to identify constructions of the kind "eventuality + signal + eventuality", and also of the required features.

## References


[1] R. Passonneau, "A computational model of the semantics of tense and aspect," *Computational Linguistics*, vol. 14, no. 2, pp. 44–60, 1988.

[2] B. Webber, "Tense as discourse anaphor," *Computational Linguistics*, vol. 14, no. 2, pp. 61–73, 1988.

[3] H. Kamp and U. Reyle, *From Discourse to Logic. Introduction to the Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse representation Theory*. Kluwer, 1993.

[4] J. Hitzeman, M. Moens, and C. Glover, "Algorithms for analysing the temporal structure of discourse," in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL-95)*, 1995, pp. 253–60.

[5] A. Kehler, "Resolving temporal relations using tense meaning and discourse interpretation," in *Formalizing the dynamics of information*, M. Faller, S. Kaufmann, and M. Pauly, Eds. CSLI Publications, 2000, pp. 189–207.

[6] I. Mani, M. Verhagen, B. Wellner, C. M. Lee, and J. Pustejovsky, "Machine learning of temporal relations," in *Proceedings of the ACL-EACL 2001 Workshop for Temporal and Spatial Information Processing*, 2006, pp. 753–60.

[7] F. Schilder and C. Habel, "From temporal expressions to temporal information: Semantic tagging of news messages," in *Proceedings of the ACL-EACL 2001 Workshop for Temporal and Spatial Information Processing*, 2001.

[8] R. Marinelli, L. Biagini, R. Bindi, S. Goggi, M. Monachini, P. Orsolini, E. Picchi, S. Rossi, N. Calzolari, and A. Zampolli, "The italian `PAROLE` corpus: an overview," *Linguistica Computazionale, Computational Linguistics in Pisa, special Issue*, vol. XVIII-XIX, pp. 401–23, 2003.

[9] N. Ruimy, M. Monachini, E. Gola, N. Calzolari, M. D. Fiorentino, M. Ulivieri, and S. Rossi, "A computational semantic lexicon of italian: *SIMPLE*," *Linguistica Computazionale, Computational Linguistics in Pisa, special Issue*, vol. XVIII-XIX, pp. 821–64, 2003.

[10] L. Ferro, I. Mani, B. Sundheim, and G. Wilson, "Tides temporal annotation guidelines, v. 1.0.2," MITRE, MITRE Technical report MTR01W0000041, 2001.

[11] J. Pustejovsky, J. Castao, R. Ingria, R. Saurì, R. Gaizauskas, A. Setzer, and G. Katz, "Timeml: Robust specification of event and temporal expressions in text," in *Fifth International Workshop on Computational Semantics (IWCS-5)*, 2003a.

[12] I. Mani and B. Schiffman, "Temporally anchoring and ordering events in news," in *Time and Event Recognition in Natural Language*, J. Pustejovsky and R. Gaizauskas, Eds. John Benjamins Publishing, 2007.

[13] A. Berger, S. D. Pietra, and V. D. Pietra, "A maximum entropy approach to natural language processing." *Computational Linguistics*, 1996.