UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Eventuali specifiche per l'annotazione di alcune relazioni temporali in *It-TimeML*

Candidato: *Alessio Carli*

Relatore: Prof. *Alessandro Lenci*

Correlatore: Prof.ssa *Maria Simi*

Anno Accademico 2010-2011

Indice generale

1 Introduzione

L'obiettivo principale delle ricerche sul Trattamento Automatico del Linguaggio è quello di far comprendere al computer il modo di parlare e comunicare delle persone così che anch'esso comunichi con noi nello stesso modo, o almeno ci si avvicini. Recentemente è stato oggetto di ricerca, in questo campo e in quello della Linguistica Computazionale, l'identificazione e l'annotazione di entità temporali nel linguaggio naturale al fine di dotare le macchine anche di tale conoscenza. In particolare *TimeML* per la lingua inglese e *It-TimeML* per quella italiana sono i *linguaggi di marcatura*[1] (o di *markup*) che ci permettono di catturare questi elementi temporali in un testo segnandoli con particolari *tag*[2] (etichette). In tale ambito non c'è ancora molta chiarezza su quali siano effettivamente i procedimenti più affidabili per l'annotazione delle varie entità coinvolte in questo tipo di relazioni. A tal proposito, durante il mio tirocinio presso l'ILC al Consiglio Nazionale delle Ricerche a Pisa, ho collaborato insieme al Dott. Tommaso Caselli all'elaborazione di eventuali specifiche per l'annotazione di alcune relazioni temporali. Successivamente sulla base di queste specifiche è stato condotto un esperimento di annotazione di 55 testi in italiano con il relativo calcolo del coefficiente di accordo *k* tra gli annotatori: un buon accordo sarà indizio di buone specifiche, un basso accordo indicherà invece che bisogna rivedere le specifiche.

La tesi è stata strutturata in 4 capitoli principali dopo questo introduttivo:
in particolare il secondo capitolo presenterà una descrizione dei linguaggi a marcatore *TimeML* e *It-TimeML* e analizzerà nello specifico i *tag* di quest'ultimo con alcuni attributi[3] principali e i rispettivi utilizzi. Il terzo capitolo invece sarà una panoramica sull'elaborazione delle specifiche per l'annotazione di alcune relazioni temporali in *It-TimeML* effettuata dal Dott.Tommaso Caselli e me presso l'ILC di

---

1 Un *linguaggio di marcatura* descrive i meccanismi di strutturazione, di semantica o di presentazione del testo attraverso linguaggi che, utilizzando convenzioni standardizzate, possono essere utilizzabili su più sistemi.

2 Con il termine *tag,* in ambito di Trattamento Automatico del Linguaggio, si intende l'etichetta attribuita ad una porzione specifica e limitata di testo nella fase detta di annotazione o marcatura.

3 Con gli attributi, nei linguaggi a marcatore come *It-TimeML*, è possibile specificare alcune caratteristiche o comportamenti dell'elemento marcato con il *tag*. In questa tesi verranno presentati solo quelli più essenziali per la comprensione dell'esperimento descritto nei capitoli successivi.

Pisa. Tale capitolo proseguirà mostrando prima il *tool on-line* utilizzato nelle fasi di annotazioni durante questo periodo di tirocinio (ovvero il *Brandeis Annotation Tool*[4] *o BAT*), poi le specifiche elaborate per l'annotazione di relazioni temporali tra eventi ed espressioni di tempo (la cui valutazione o accordo è stato calcolato successivamente da me), e infine le specifiche elaborate per l'annotazione di relazioni temporali tra espressioni di tempo (dove l'accordo invece è stato calcolato automaticamente dal *tool*).

Il concetto di affidabilità e la sua importanza per la riproducibilità delle specifiche sarà invece il punto di partenza del quarto capitolo che proseguirà con la descrizione del coefficiente *k* di Cohen. Tale capitolo poi si concluderà mostrando da vicino l'esperimento di valutazione delle specifiche per l'annotazione di relazioni temporali tra eventi ed espressioni di tempo attraverso proprio il calcolo di *k*.

Infine nel quinto capitolo ci saranno le dovute conclusioni sui risultati ottenuti dall'esperimento del capitolo precedente.

[4] Il *tool* è disponibile gratuitamente all'indirizzo *web* www.batcaves.org/bat/tool/index.php

2 *TimeML* e *It-TimeML*

Se dovessimo rappresentare le entità coinvolte nella dimensione temporale, sarebbe necessario individuare gli eventi che accadono nel mondo esterno, connessi da qualche tipo di relazione direttamente a un tempo.

A tal proposito *TimeML*[5] è un linguaggio di marcatura che consente di annotare, con particolari *tag*, eventi ed espressioni temporali in un testo. I *tag* principali utilizzati per rappresentare queste entità sono quattro: EVENT per gli eventi temporali, TIMEX3 per le espressioni che indicano il tempo, SIGNAL per i segnali temporali espliciti e LINK[6] per rappresentare le diverse relazioni tra gli elementi temporali. *TimeML* in particolare è stato progettato per affrontare quattro problemi specifici nell'annotazione di eventi ed espressioni di tempo (www.timeml.org):

- Il *Timestamping* di eventi, ovvero l'identificazione di un evento fissandolo nel tempo
- L'ordinamento degli eventi, l'uno rispetto all'altro
- Il ragionamento sui casi di espressioni di tempo che non contengono intrinsecamente tutte le informazioni necessarie per il calcolo del loro valore temporale
- Il ragionamento sulla persistenza degli eventi nel tempo

Per l'annotazione della lingua italiana è possibile inoltre utilizzare *It-TimeML*, un primo adattamento di *TimeML* e delle specifiche *ISO-TimeML*[7] ad un linguaggio differente dall'Inglese (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*)[8]. Esso è stato progettato all'ILC-CNR a Pisa e negli incontri

5 *TimeML* è stato sviluppato nel contesto di tre laboratori e progetti AQUAINT: i laboratori TERQAS e TANGO ed il progetto TARSQI (www.timeml.org).

6 A sua volta il tag LINK si suddivide in TLINK per le relazioni temporali, ALINK per le relazioni aspettuali e SLINK per le relazioni subordinative.

7 Per le specifiche di *TimeML* è possibile consultare www.timeml.org/site/publications/specs.

8 Le linee guida utilizzate per l'elaborazione di questa tesi sono la versione 1.2, ma attualmente (ad Aprile 2011) la versione aggiornata è la 1.3 (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.3*).

ISO TC 37 SC4[9]. Sia *TimeML* quindi, che *It-TimeML* hanno lo stesso scopo: annotare tutte le espressioni che rappresentano intrinsecamente informazioni temporali e le varie relazioni tra di esse.

Di seguito verranno descritti nel dettaglio gli utilizzi per l'annotazione dei *tag* della versione italiana *It-TimeML.*

2.1 Il *tag* EVENT

"The EVENT tag is used to annotate those elements in a text that describe what is conventionally referred to as *eventuality*" (Verhagen e altri, *SemEval-2010 Task 13: TempEval-2*, p.57)[10]. Con questo intendiamo tutto ciò che accade, che si verifica; quindi anche ogni tipo di azione[11] e alcuni stati o situazioni temporanee.
A livello sintattico, gli elementi linguistici che possono essere etichettati come EVENT in *It-TimeML* sono (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2,* p.5):

- *Verbi,* sia nella forma finita che infinita, es.:

(1) a. La polizia *ha sgombrato* la piazza.
b. La nazione mostra i segni della guerra: città *incendiate* o *rivoltate*.

- *Aggettivi,* es.:

(2) Il signor Lugretti, *residente* a Roma, stava trascorrendo un periodo di vacanza in Puglia.

---

9 L'*ISO TC 37 SC4* è un comitato tecnico all'interno dell'organizzazione internazionale per la standardizzazione di risorse linguistiche; esso comprende aree di ricerca della Linguistica Computazionale, della Lessicografia Computerizzata e dell'Ingegneria Linguistica (www.tc37sc4.org). Gli incontri citati si sono svolti alla Brandeis University (Stati Uniti), alla Tiburg University (Olanda) e all'ANFOR (Francia) tra il 2006 e 2007.

10 Traduzione: "Il *tag* EVENT viene utilizzato per annotare quegli elementi in un testo che descrivono ciò che è convenzionalmente indicato come eventualità"

11 Le azioni annotate come EVENT possono essere nel tempo *durative* quando durano per un certo tempo, *momentanee* quando durano o avvengono in un momento, *ripetitive* quando avvengono tante volte, *puntuali* quando avvengano una volta sola e infine *abituali* e *non-abituali*.

❖ *Sezioni predicative,* es.:

(3) Arnold Wainer è il nuovo *presidente* della Digidesign.

❖ *Sintagmi preposizionali,* es*.:*

(4) Un giovane turista tedesco *in villeggiatura* in Sardegna è morto.

❖ *Nomi* che possono manifestare eventualità in modi diversi:

- attraverso una nominalizzazione di un verbo, es.: *fuga*, *crescita*, *arrivo*, *bevuta*, *accordo*...
- quando indicano di per sé un evento, es.: *guerra, uragano*, *assemblea*, *cerimonia*...
- quando possono essere interpretati nel contesto come se si riferissero ad un evento. Per capire ciò prendiamo come esempio la frase citata di seguito (l'esempio 6), dove l'elemento "*il libro*" non verrebbe normalmente considerato un evento, ma contestualizzando la frase esso viene inteso come un processo di lettura che ha temporalmente un' inizio ed una fine, es.:

(5) Ho sospeso *il libro.*

Infine, va ricordato il fatto che anche gli stati possono essere annotati come eventualità a patto che essi siano transitori nel tempo o che partecipino in modo esplicito a una relazione temporale, es.:

(6) Luca è *un tirocinante.*

2.1.1 Gli attributi principali del *tag* EVENT

Gli attributi principali del *tag* EVENT sono[12]:

❖ `id:` è un numero univoco assegnato automaticamente dal *tool* di annotazione ad ogni elemento marcato come evento.

[12] In appendice è riportata la descrizione completa in meta-sintassi BNF del tag EVENT con tutti gli attributi (anche quelli non citati nella tesi).

- `pred:` è l'espressione che esprime il tipo di evento, es.:

(7) La procura di Palermo *ha chiuso* l'inchiesta.
....ha `<EVENT id="e1" pred="chiudere">`chiuso`<EVENT/>`...

- `class:` ogni evento fa parte di una particolare classe di categorie lessicali, identificata ognuna da una combinazione di criteri semantici e sintattici. Le varie classi sono:

  - `REPORTING`, quando l'evento descrive l'azione di una persona oppure organizzazione che dichiara, narra, informa su un evento (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), es.:

(8) Il Telegiornale di ieri sera ha
`<EVENT...class="REPORTING">`detto`</EVENT>` che la guerra è iniziata.

  - `PERCEPTION`, quando l'evento coinvolge la percezione fisica di un altro evento (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), es.:

(9) Dei testimoni hanno `<EVENT...class="PERCEPTION">`visto`</EVENT>` la fuga dei criminali.

  - `ASPECTUAL`, quando l'evento codifica una particolare fase, come l'inizio, la fine, la continuazione oppure il culmine di un altro evento, es.:

(10) La trattativa è già `<EVENT...class="ASPECTUAL">`iniziata`</EVENT>`.

  - `I_ACTION`, quando l'evento è un azione o situazione che introduce un'ulteriore evento come suo argomento; da notare che quest'ultimo deve essere esplicito nel testo, es.:

(11) La Libia ha chiesto di `<EVENT...class="I_ACTION">`ritardare`</EVENT>` l'attacco.

  - `I_STATE`, quando l'evento è una situazione statica che introduce un altro evento come suo argomento; anche qua l'evento-argomento deve essere esplicito nel testo, es.:

(12) <EVENT...class="I_STATE">Sperano</EVENT> che gli abitanti rientreranno nelle loro case dopo l'allarme terremoto.

- OCCURRENCE, la classe che racchiude tutti quegli eventi che non appartengono alle classi precedenti e che descrivono qualcosa che si verifica, che accade nel mondo esterno, es.:

(13) I vari segretari se ne <EVENT...class="OCCURRENCE">tornano</EVENT> nella loro patria.

- STATE, quando l'evento è uno stato o situazione. Come già accennato in precedenza solo alcuni stati vengono annotati (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*): quelli che sono transitori nel tempo, quelli che sono direttamente connessi a un espressione temporale o modificatore come *"ex, recente..."*, e quelli introdotti da un evento di classe REPORTING, I_ACTION oppure I_STATE. Es.:

(14) a. Diversi boss mafiosi sono tutt'ora <EVENT...class="STATE">agli arresti</EVENT>.

b. Berlusconi è l'attuale <EVENT...class="STATE">presidente</EVENT> del consiglio.

c. Ha <EVENT...class="REPORTING">detto</EVENT> che è <EVENT...class="STATE">un bugiardo</EVENT>.

- pos: questo attributo segnala a quale categoria grammaticale appartiene l'evento. I valori che possono essere inseriti sono VERB per gli eventi realizzati da verbi, NOUN per i nomi, ADJECTIVE per gli aggettivi, PREPOSITION per le preposizioni e OTHER se l'evento è realizzato da una categoria grammaticale differente da quelle precedenti. Es.:

(15) La polizia ha <EVENT...pos="VERB">sgombrato</EVENT> la piazza.

- tense: questo attributo viene inserito se l'evento è un verbo e segnala a quale tempo verbale appartiene. In generale i suoi valori[13] possono essere PRESENT

[13] Per le regole dettagliate sull'annotazione dell'attributo aspect vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2,* p.31).

per il Presente Semplice e il Passato Composto, PAST per l'Imperfetto, il Passato Semplice, il Trapassato e Trapassato Prossimo, FUTURE per il Futuro Semplice e Composto, NONE se non serve l'attributo o se non è specificato.

- aspect: questo attributo distingue *l'aspetto verbale*[14] dell'evento. I suoi valori possono essere PERFECTIVE quando l'evento è un azione delimitata nel tempo, IMPERFECTIVE quando l'evento descrive un azione dove non è specificata la durata, e IMPERFECTIVE_PROGRESSIVE quando l'evento specifica l'aspetto imperfettivo dei casi di perifrasi aspettuali[15] (es. *"sto mangiando"*).

- mood: quando l'evento è realizzato da un verbo, questo attributo segnala il modo grammaticale del verbo. I suoi valori possono essere NONE di *default* se il modo è Indicativo, COND se è condizionale, SUBJUNTIVE se è congiuntivo, IMPERATIVE se è imperativo.

- vForm: questo attributo distingue la forma del verbo. I valori che può assumere sono NONE di *default* per i verbi finiti, INFINITIVE per gli infiniti, PARTICIPLE per i participi e GERUND per i gerundi.

2.1.2 L'annotazione del tag EVENT

Per convenienza e semplicità, sia in *It-TimeML* che in *TimeML*, solo la *testa* dell'evento viene marcato con il *tag* EVENT, tutto il resto non viene considerato (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), es.:

(16) a. La polizia ha <EVENT>sgombrato</EVENT> la piazza.
b. La coppia stava <EVENT>trascorrendo</EVENT> una vacanza.
c. <EVENT>Accusando</EVENT>li di <EVENT>omicidio</EVENT>.
d. Berlusconi è l'attuale <EVENT>presidente</EVENT> del consiglio.

14 *L'aspetto verbale* è la proprietà che definisce la durata nel tempo di un azione.
15 Vedi il valore ASPECTUAL nell'attributo class per il *tag* EVENT.

(17) La <EVENT>riunione</EVENT> sta per <EVENT>chiudersi</EVENT>.

Ci sono casi[16] particolari però in cui è necessario annotare in modo differente il *tag*. Se ad esempio siamo in presenza di una preposizione che esprime un evento bisogna includere anch'essa nell'annotazione, es.:

(18) a. La banda è attualmente <EVENT>agli arresti</EVENT>.

b. Un giovane turista tedesco <EVENT>in villeggiatura</EVENT> in Sardegna è morto.

Anche nei casi di *usi metaforici* o simili, *costruzioni a verbo di supporto*[17] o *perifrasi modali*[18] viene marcata nel *tag* più di una parola, es.:

(19) a. Tocca a Del Neri <EVENT>tirare le somme</EVENT>.

b. La nonna è caduta mentre <EVENT>faceva una torta</EVENT>.

c. <EVENT>Siamo in grado</EVENT> di <EVENT>dire</EVENT> che...

Nei casi invece di *costruzioni a verbo di supporto* dove il sostantivo è un verbo sostantivato, oppure quando il verbo *"fare"* sostituisce un verbo che designa un azione, allora vengono creati due *tag* distinti, es.:

(20) a. Il governo ha <EVENT>preso</EVENT> una <EVENT>decisione</EVENT>.

b. Il ladro ha <EVENT>fatto</EVENT> <EVENT>scattare</EVENT> l'allarme.

Infine i *verbi modali,* come *"dovere", "potere", "volere, "sapere"*, devono essere annotati singolarmente, es:

---

16 Ci sono casi in cui è necessario annotare nello stesso *tag* EVENT più parole adiacenti, questo perché tutte insieme, e non singolarmente, esprimono un evento: per una completa casistica riguardante l'annotazione del *tag* EVENT vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*).

17 Grammaticalmente in Italiano la *costruzione a verbo di supporto* avviene quando siamo in presenza di un costrutto composto da "verbo + sostantivo" con un nesso tra loro. Il sostantivo può essere un verbo sostantivato (nome deverbale), ma anche un aggettivo. I verbi di supporto di base sono "*fare, dare, avere, essere, prendere*". Volendo è possibile anche annotare singolarmente il verbo ed il sostantivo specificando successivamente con un TLINK di tipo IDENTITY (vedi il punto 2.4) che siamo di fronte ad una costruzione a verbo di supporto.

18 Le perifrasi modali sono dei costrutti grammaticali che indicano la possibilità o necessità che accada qualcosa. Ad esempio "*essere in grado di + Infinito*", "*c'è da + Infinito*" ecc...

(21) Il governo <EVENT>deve</EVENT> <EVENT>prendere</EVENT> una <EVENT>decisione</EVENT>.

2.2 Il *tag* TIMEX3

In *It-TimeML*, così come in *TimeML*, il *tag* TIMEX3[19] viene utilizzato per annotare tutte quelle parole che denotano esplicitamente espressioni temporali o che contengono direttamente un tempo (www.timeml.org).
Di conseguenza molte parole non esplicitamente riferite al tempo ma che potrebbero avere un significato temporale (es. "*scuola"*, *"incubazione*" etc...) sono automaticamente escluse. In particolare le parole che possono essere marcate con il *tag* TIMEX3 si riferiscono esclusivamente a:

- *Periodi della giornata* (es. "*mezzogiorno*", "*alle 4*", "*la sera*"...)
- *Date* espresse da giorni (es. "*ieri*" , "*3 Marzo 1986*", "*sabato scorso*"...), da settimane (es. "*la prima settimana del mese*", "*la scorsa settimana*" ...), da mesi (es. "*tra sei mesi*", "*Settembre 2010*"...), da stagioni (es. "*il trimestre*", "*la primavera scorsa*"...), da anni (es. "*1984*", "*tre anni fa*"...), ecc...
- *Durate* (es. "*due anni*", "*nei prossimi mesi*", "*nel periodo*", "*20 ore*"...)
- *Avverbiali iterativi* (es. "*due volte alla settimana*", "*ogni sabato*"...)

A livello sintattico le espressioni temporali possono nello specifico essere realizzate da:

- *Nomi* (es. "*alba*", "*agosto*", "*Natale*", "*domenica*", "*estate*", "*Pasqua*"...)
- *Date* (es. "*18/03/2010*", "*2001*", "*12.06*"...)
- *Aggettivi (*es. "*annuale*","*estivo*", "*settimanale*"...)
- *Avverbi* (es. "*oggi*", "*adesso*", "*attualmente*", "*ieri*", "*finora*"...)

[19] Il *tag* TIMEX3 è al suo terzo modello: il primo modello fu TIMEX (Setzer 2001) e il secondo fu TIMEX2 (Ferro e altri 2002).

2.2.1 Gli attributi principali del *tag* TIMEX3

Gli attributi principali del *tag* TIMEX3 sono[20]:

- `id`: è un numero univoco assegnato automaticamente dal *tool* di annotazione ad ogni elemento marcato come espressione temporale.

- `type`: questo attributo specifica di che tipo è l'espressione temporale. I suoi valori possono essere:

  - `DATE`, quando l'espressione temporale è una data; la sua rappresentazione può essere sia nel formato ISO per le date (con gli anni, i mesi, le settimane, i giorni, le ore, i minuti e i secondi), sia espressa con altre strutture che rimandano sempre a una data del calendario, es.:

(22) a. Io sono nato nel <TIMEX3...type="DATE">1984</TIMEX3>.
b. Sono cascato <TIMEX3...type="DATE">giovedì scorso</TIMEX3> .

  - `TIME`, quando l'espressione temporale è un momento oppure un ora della giornata (sia nel formato ISO che non), es.:

(23) a. La rissa è scoppiata <TIMEX3...type="TIME">domenica sera</TIMEX3>.
b. La riunione è alle <TIMEX3...type="TIME">14.00</TIMEX3>.

  - `DURATION`, quando l'espressione temporale è una durata o un periodo di tempo, es.:

(24) a. La transazione dura da <TIMEX3...type="DURATION">più di un anno</TIMEX3>.
b. Il PIL è cresciuto del 10% in <TIMEX3...type="DURATION">tre anni</TIMEX3>.

  - `SET`, quando l'espressione temporale è un'avverbiale iterativo o frequentativo, es.:

[20] In appendice è riportata la descrizione completa in meta-sintassi BNF del *tag* TIMEX3 con tutti gli attributi (anche quelli non citati nella tesi).

(25) a. Più di otto tonnellate <TIMEX3...type="SET">al mese</TIMEX3>.

b. Le riunioni qua si tengono <TIMEX3...type="SET">quasi ogni giorno</TIMEX3>.

- `value`[21]: questo attributo specifica e normalizza il valore temporale corrispondente al valore descritto in `type` (vedi attributo precedente) nei formati standard ISO. Il valore, normalizzato a seconda del tipo di espressione temporale, può essere una data (nel formato ISO "*YYYY-MM-WW-DD*"[22]), un ora dell'orologio (nel formato ISO "*hh:mm:ss*"[23]), delle serie[24]nel tempo oppure dei formati speciali di durate[25] basate sugli standard ISO 8601 e sulle sue estensioni (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*). Es.:

(26) a. <TIMEX3 type="DATE" value="2008-12-02">venerdì due dicembre 2008</TIMEX3>

b. <TIMEX3 type="TIME" value="T16:00">le 16.00</TIMEX3>

c. <TIMEX3 type="SET" value="P1W">una volta a settimana</TIMEX3>

d. <TIMEX3 type="DURATION" value="P4M">4 mesi</TIMEX3>

- `mod`: questo attributo segnala la presenza di modificatori (es. "*circa*", "*primo*", "*tardo*", "*oltre*"...) che influenzano l'interpretazione di `value` (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2,* pp.48-49).

---

21 Per una completa casistica di annotazione di questo attributo vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2,* p.42).

22 Questo è il formato standard ISO per le date dove *Y* sta per anno, *M* sta per mese, *W* sta per settimana e *D* sta per giorno. Nel linguaggio naturale ci sono espressioni che *in It-TimeML* sono considerate date del calendario anche se non si possono ricondurre a questi standard: per analizzare tali casi vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*, p.45).

23 Nello standard ISO *h* sta per ore, *m* sta per minuti ed *s* sta per secondi. Prima di inserire l'ora in questo formato bisogna mettere una *T*. In alcuni casi, quando nel testo l'ora annotata è ancorata ad una data, è necessario inserire anch'essa nello stesso *tag*, sempre nel formato ISO. Inoltre, come per le date, ci possono essere dei casi nel linguaggio naturale che vengono considerate ore della giornata anche non sono rappresentate direttamente in questo formato (Caselli 2010).

24 Per capire come calcolare le serie nel tempo vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*).

25 Per le durate o intervalli di tempo si pone nel campo `value` una *P*. Per capire come calcolarle vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*)

- `temporalFunction`: questo attributo facoltativo specifica se il valore dell'espressione temporale deve essere individuato attraverso una funzione temporale oppure no (ISO 2008). Può capitare infatti che l'espressione di tempo non abbia intrinsecamente tutte le informazioni necessarie per il calcolo del suo valore temporale: ad esempio, nella frase "*Ieri ho dormito sempre*", per sapere a quando si riferisce la parola "*Ieri*" sarebbe necessario calcolare una funzione che individui la sua àncora temporale (vedi il punto seguente) la quale è un'altra espressione di tempo necessaria ad identificarla. Se quindi la data di creazione del testo, su cui si trova la frase, fosse "5/3/2011" ed essa fosse anche l'àncora di tale parola, allora sapremmo attraverso il calcolo della funzione che "Ieri" si riferisce alla data "4/3/2011". In questo caso quindi, visto che è necessaria una funzione temporale, il valore dell'attributo `temporalFunction` dovrebbe essere `true`; se non fosse stato necessario fornire questo attributo il suo valore sarebbe stato `false`.

- `anchorTimeID`: questo attributo è utilizzato per fornire l'ID dell'espressione temporale a cui il TIMEX3 è collegato attraverso una funzione temporale.

- `beginPoint` e `endPoint`: nel caso in cui `type` nell'espressione temporale ha il valore uguale a `DURATION`, tali attributi se presenti hanno come valore l'`id` dell'espressioni temporali che delimitano il punto di inizio o fine della durata stessa.

2.2.2 L'annotazione del *tag* TIMEX3

Come già detto in precedenza, la parola che esprime l'espressione temporale può essere un nome, un aggettivo, un avverbio oppure una data. Normalmente queste parole vengono annotate singolarmente come TIMEX3 anche se ci sono alcuni casi[26] particolari dell'Italiano in cui questo non può essere fatto (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*).

[26] Per una completa casistica riguardante l'annotazione del *tag* TIMEX3 vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*).

Ad esempio, quando l'espressione temporale è accompagnata dalle preposizioni temporali[27] "*circa*", "*intorno*" e "*verso*", anch'esse devono essere inserite nel *tag* TIMEX3, es.:

(27) ...<TIMEX3>verso le 8 di mattina</TIMEX3>...

La stessa cosa avviene quando abbiamo parole multiple come "*per ora*", "*dopo domani*" e "*fin'ora*", che sono considerate singole unità per il loro significato.

Inoltre anche i modificatori[28]che non denotano di per sé un evento e che sono posti prima o dopo l'espressione temporale da loro modificata, devono essere inseriti nello stesso *tag* TIMEX3, es.:

(28) a. ...nel <TIMEX3>primo trimestre</TIMEX3>...
b. ...<TIMEX3>il mese scorso</TIMEX3>...
c. ...<TIMEX3>due anni fa</TIMEX3>...

Anche la parola "*dopo*", nel solo caso in cui ha la funzione di aggettivo, deve essere inclusa nel *tag*, es.:

(29) ...<TIMEX3>quattro settimane dopo</TIMEX3>...

Altri casi particolari possiamo averli quando siamo in presenza di una espressione temporale che esprime un ora dell'orologio, quando ci sono espressioni temporali che differiscono dallo standard Gregoriano e quando ci sono due espressioni temporali consecutive che appartengono alla stessa unità temporale, es.:

(30) a. ...alle <TIMEX3>20 e 30</TIMEX3>...
b. ...<TIMEX3>l'anno fiscale '87 – '88</TIMEX3>...
c. ...<TIMEX3>domenica mattina ore 11</TIMEX3>...

---

[27] Solo le preposizioni temporali citate devono essere inserite nello stesso *tag* TIMEX3, tutte le altre devono essere annotate con il *tag* SIGNAL, descritto successivamente; es. "<SIGNAL>nel</SIGNAL> <TIMEX3>pomeriggio</TIMEX3>".

[28] Anche le proposizioni relative appositive, quelle che aggiungono al sintagma nominale delle informazioni, sono considerate dei modificatori, es. "<TIMEX3>gli anni dei Beatles</TIMEX3>".

2.3 Il *tag* SIGNAL

Il *tag* SIGNAL[29] viene utilizzato per annotare tutte le parti di testo che segnalano chiaramente una relazione tra due elementi (www.timeml.org); questi elementi possono essere due TIMEX3, un TIMEX3 e un EVENT (o viceversa) oppure due EVENT. A livello linguistico le sezioni di testo che possono essere annotate come SIGNAL sono (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*):

- *Preposizioni temporali* (es. "*durante*", "*nel*"...)
- *Congiunzioni temporali* (es. "*quando*", "*mentre*"...)
- *Avverbi temporali* (es. "*intanto*", "*nel frattempo*"...)
- *Caratteri speciali*[30] (es. "-", "/"...)
- *Proposizioni e congiunzioni che segnalano una relazione subordinativa* (es. "se", "*affinché*"...)

Un esempio di annotazione del tag SIGNAL potrebbe essere:

(31) ...<EVENT>riprenderà</EVENT> <SIGNAL>dal</SIGNAL> <TIMEX3>primo Maggio</SIGNAL>...

2.4 Il *link tag* TLINK

TLINK è uno dei tre *link tag* presenti sia in *TimeML* che in *It-TimeML*. Essi non sono *tag* inseribili[31]direttamente nel documento, ma codificano esternamente ad esso le varie relazioni tra le entità temporali annotate nel testo (www.timeml.org). Per

---

29 L'unico attributo del *tag* degno di nota per la tesi è `id`, ovvero un numero univoco assegnato ad ogni elemento marcato come SIGNAL. In appendice comunque è riportata la descrizione completa in meta-sintassi BNF del *tag*.

30 Questi tipi di SIGNAL vengono utilizzati quando ad esempio siamo in presenza di espressioni temporali che denotano *range* temporali (Caselli 2010), es. "18 – 25 Ottobre".

31 I *link tag* vengono creati esternamente al testo.

riuscire a comprendere pienamente quando due entità temporali sono connesse da un certo tipo di relazione sarà quindi necessario leggere attentamente il testo e solo dopo creare al di fuori del documento tutti i vari *link tag*.

Nello specifico TLINK (o *temporal link*) indica la relazione temporale che può sussistere tra due eventi, due espressioni temporali o tra un evento ed una espressione temporale (o viceversa).
Per fare un semplice esempio pratico prendiamo la frase:

(32) Giovanni ha <EVENT id="e1">*giocato*</EVENT> e <SIGNAL id="s1">*dopo*</SIGNAL> ha <EVENT id="e2">*mangiato*</EVENT>.

Se analizziamo il senso della frase, il primo evento (*e1*) "*giocato*" avviene temporalmente prima rispetto al secondo evento (*e2*) "*mangiato*". Ad aiutarci nel capire questo c'è per fortuna anche un segnale (*s1*) ovvero la parola "*dopo*": non sempre infatti il testo è così esplicito nel farci capire la successione temporale degli eventi. A questo punto, al di fuori del testo, verrà creato un TLINK il quale specificherà che l'evento *e1* avviene temporalmente prima dell'evento *e2* ed anche che questa relazione viene espressa tramite il segnale *s1*, es.:

(33) <TLINK eventID="e1" relatedToEvent="e2" signal="s1" reltype="BEFORE">[32]

2.4.1 Gli attributi principali del *link tag* TLINK

Gli attributi principali del *link tag* TLINK sono[33]:

- `id`: è un numero univoco assegnato ad ogni relazione temporale.

[32] Per comprendere appieno l'esempio vedi successivamente gli attributi di TLINK al punto 2.4.1.
[33] In appendice è riportata la descrizione completa in meta-sintassi BNF del *link tag* TILINK con tutti gli attributi (anche quelli non citati nella tesi).

- `eventID` o `timeID`: a seconda che la relazione temporale sia creata da un evento oppure da un espressione temporale, questi due attributi segnalano rispettivamente i loro `id`. In pratica quindi indicano l'elemento di partenza nella relazione temporale.

- `signalID`: se ad esplicitare la relazione temporale c'è un segnale (SIGNAL), allora questo attributo specifica la sua presenza con il suo `id`.

- `relatedToEvent` o `relatedToTime`: a seconda che la relazione temporale sia riferita a un evento oppure a una espressione temporale, questi due attributi segnalano rispettivamente i loro `id`. In pratica quindi indicano l'elemento di arrivo a cui è riferito il primo di partenza nella relazione temporale.

- `relType`: questo attributo indica il tipo di relazione temporale che sussiste tra i due elementi coinvolti in essa. I suoi valori possono essere:

  - `SIMULTANEOUS`: quando le entità coinvolte nella relazione accadono nello stesso momento, oppure quando una di esse è percepita che accada nello stesso momento, es.:

(34) `<EVENT id="e1">`Mangio`</EVENT> <SIGNAL id="s1">`quando`</SIGNAL> <EVENT id="e2">`guardo`</EVENT>` un film.
`<TLINK eventID="e1" relatedToEvent="e2"...relType="SIMULTANEOUS">`

  - `BEFORE`: quando un entità è temporalmente prima di un'altra, es.:

(35) Il ragazzo ha `<EVENT id="e1">`ucciso`</EVENT>` la compagna. `<SIGNAL id="s1">`Dopo`</SIGNAL>` si è `<EVENT id="e2">`tolto`</EVENT>` la vita.
`<TLINK eventID="e1" relatedToEvent="e2"...relType="BEFORE">`

  - `AFTER`: quando un entità è temporalmente dopo un'altra: è l'inverso di `BEFORE`.

- IBEFORE o IAFTER: quando un entità nel tempo avviene immediatamente prima oppure dopo un'altra, es.:

(36) Nell'<EVENT id="e1">impatto</EVENT> tutti i passeggeri sono <EVENT id="e2">morti</EVENT>.
<TLINK eventID="e1" relatedToEvent="e2"...relType="IBEFORE">

- DURING: quando un entità ha come durata un'altra entità (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), es.:

(37) Giovanni <EVENT id="e1">gioca</EVENT> <SIGNAL id="s1">per</SIGNAL> <TIMEX3 id="t1">due ore</TIMEX3> a basket.
<TLINK eventID="e1" relatedToTime="t1"...relType="DURING">

- INCLUDES o IS_INCLUDED: quando un entità include oppure è inclusa temporalmente in un'altra, es.:

(38) Giovanni <TIMEX3 id="t1">lunedì</TIMEX3> ha <EVENT id="e1">giocato</EVENT> a basket.
<TLINK timeID="t1" relatedToEvent="e1"...relType="INCLUDES">

- BEGINS o BEGUN_BY: quando un entità fa iniziare oppure viene iniziata temporalmente da un'altra entità, es.:

(39) <SIGNAL id="s1">Da</SIGNAL> <TIMEX3 id="t1">lunedì</TIMEX3> si <EVENT id="e1">cerca</EVENT> un accordo.
<TLINK timeID="t1" relatedToEvent="e1"...relType="BEGINS">

- ENDS o ENDED_BY: quando un entità fa concludere o viene conclusa temporalmente da un'altra entità, es.:

(40) Giovanni <SIGNAL id="s1">fino</SIGNAL> alle <TIMEX3 id="t1">7</TIMEX3> ha <EVENT id="e1">giocato</EVENT> a basket.
<TLINK timeID="t1" relatedToEvent="e1"...relType="ENDS">

- IDENTITY: questo è un attributo che si utilizza solo in casi particolari: ad esempio, quando abbiamo delle frasi causative con due eventi e nel mezzo ad essi

il verbo “causare”, IDENTITY specifica la posizione dell’evento soggetto che causa il secondo, es.:

(41) Lo <EVENT id=”e1”>scoppio</EVENT> ha <EVENT id=”e2”>causato</EVENT> <EVENT id=”e3”>l’incendio</EVENT>.
<TLINK eventID=”e1” relatedToEvent=”e2”...relType=”IDENTITY”>
<TLINK timeID=”e1” relatedToEvent=”e3”...relType=”BEFORE”>

Oppure quando abbiamo casi di costruzioni a verbo di supporto, IDENTITY collega il verbo con il suo sostantivo, es.:

(42) Gianni ha<EVENT id=”e1”>fatto</EVENT> una <EVENT id=”e2”>torta</EVENT>.
<TLINK eventID=”e1” relatedToEvent=”e2”...relType=”IDENTITY”>

Infine quando si duplica un evento perché si riferisce contemporaneamente ad altri elementi nel testo, IDENTITY viene utilizzato per collegarli, es.:

(43) Gianni ha<EVENT...id=”e1”>insegnato</EVENT> <TIMEX3 id=”t1”>lunedì</TIMEX3> e <TIMEX3 id=”t1”>martedì</TIMEX3>.
<EVENT...id=”e2”>
<TLINK eventID=”e1” relatedToEvent=”e2”...relType=”IDENTITY”>
<TLINK eventID=”e1” relatedToTime=”t1”...relType=”IS_INCLUDED”>
<TLINK eventID=”e2” relatedToTime=”t2”...relType=”IS_INCLUDED”>

2.5 Il *link tag* ALINK

ALINK (o *aspectual link*) è il secondo *link tag* presente in *TimeML* e *It-TimeML*; esso indica esclusivamente la relazione che sussiste tra un EVENT di classe

aspettuale (ASPECTUAL)[34] ed un altro EVENT che funge da suo attributo. Come ogni *link tag* anche ALINK non è marcabile direttamente nel testo ma viene creato al di fuori del documento, es.:

(44) La commissione <EVENT id="e1"...class="ASPECTUAL">inizierà</EVENT> alle <TIMEX3 id="t1"...>4</TIMEX3> la <EVENT id="e2"...>riunione</EVENT>.
<ALINK eventID="e2" relatedToEvent="e1" relType="INITIATES">

2.5.1 Gli attributi principali del *link tag* ALINK

Gli attributi principali del *link tag* ALINK sono[35]:

- id: è un numero univoco assegnato ad ogni relazione aspettuale.
- eventID: questo attributo indica l'id dell'evento da cui parte la relazione aspettuale.
- relatedToEvent: questo attributo indica l'id dell'evento-argomento relativo al primo di partenza della relazione aspettuale.
- signalID: se ad esplicitare la relazione aspettuale c'è un segnale (SIGNAL), allora questo attributo specifica la sua presenza con il suo id.
- relType: questo attributo indica il tipo di relazione aspettuale che sussiste tra i due eventi coinvolti in essa. I suoi valori possono essere:
  - INITIATES: quando la relazione aspettuale esprime la fase iniziale di un evento, es.:

[34] Vedi nel punto 2.1.1 l'attributo class="ASPECTUAL".

[35] In appendice è riportata la descrizione completa in meta-sintassi BNF del *link tag* ALINK con tutti gli attributi (anche quelli non citati nella tesi).

(45) Il <EVENT id="e1"...>dibattito</EVENT> <EVENT id="e2"...class="ASPECTUAL">inizia</EVENT> <TIMEX3 id="t1"...>adesso</TIMEX3>.
<ALINK eventID="e1" relatedToEvent="e2" relType="INITIATES">

- CULMINATES: quando la relazione aspettuale esprime la fase culminante di un evento: più precisamente abbiamo questo tipo di relazione aspettuale quando siamo in presenza di verbi telici (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), ovvero verbi che descrivono eventi tendenti a un fine, al raggiungimento di un risultato, es.:

(46) Chiara ha <EVENT id="e1"...class="ASPECTUAL">finito</EVENT> di <EVENT id="e2"...>preparare</EVENT> la torta.
<ALINK eventID="e1" relatedToEvent="e2" relType="CULMINATES">

- TERMINATES: quando la relazione aspettuale esprime la fase terminale di un evento: a differenza di CULMINATES abbiamo questo tipo di relazione aspettuale quando nonostante ci siano eventi come "*finire*", "*terminare*"...ecc., il verbo principale è atelico (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), ovvero non descrive un evento tendente a un fine, es.:

(47) Chiara ha <EVENT id="e1"...class="ASPECTUAL">finito</EVENT> di <EVENT id="e2"...>scrivere</EVENT>.
<ALINK eventID="e1" relatedToEvent="e2" relType="TERMINATES">

- CONTINUES: quando la relazione aspettuale esprime la fase di continuazione di un evento, es.:

(48) La <EVENT id="e1"...>trattativa</EVENT> <EVENT id="e2"...class="ASPECTUAL">dura</EVENT> <SIGNAL>da</SIGNAL> <TIMEX3>un anno</TIMEX3>.
<ALINK eventID="e2" relatedToEvent="e1" relType="CONTINUES">

- `REINITIATES`: quando la relazione aspettuale esprime una nuova fase iniziale di un evento.

2.6 Il *link tag* SLINK

SLINK (o *subordination link*) è l'ultimo *link tag* presente in *TimeML* e *It-TimeML*; esso rende esplicita una relazione di subordinazione che può sussistere tra due eventi. Più precisamente SLINK si utilizza sia in casi dove eventi di classe `I_ACTION`, `I_STATE`, `PERCEPTION` o `REPORTING` innescano altri eventi subordinati a essi, i quali di solito sono i loro complementi (www.timeml.org), sia in casi di costruzioni grammaticali di subordinata finale dove un evento posto nella reggente indica un fine oppure un obiettivo, e sia nei casi di costruzioni condizionali tra l'evento posto nella principale e quello posto nella consequenziale (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*). Un esempio di annotazione[36] di un *link tag* SLINK potrebbe essere, es.:

(49) `<EVENT id="e1"...>`Richiamerò`</EVENT>` Luca
`<SIGNAL>`perché`</SIGNAL id="s1">` vi `<EVENT`
`id="e2"...>`raggiunga`</EVENT>` al campo da calcio.
`<SLINK eventID="e1" subordinatedEvent="e2" signalID="s1"`
`relType="INTENSIONAL">`

2.6.1 Gli attributi principali del *link tag* SLINK

Gli attributi principali del *link tag* SLINK sono[37]:

---

[36] Per usi specifici di annotazione di SLINK vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*, p.63).

[37] In appendice è riportata la descrizione completa in meta-sintassi BNF del *link tag* SLINK con tutti gli attributi (anche quelli non citati nella tesi).

- `id`: è un numero univoco assegnato ad ogni relazione di subordinazione.

- `eventID`: questo attributo indica l'`id` dell'evento da cui parte la relazione di subordinazione.

- `subordinatedEvent`: esso indica l'`id` dell'evento subordinato.

- `signalID`: questo attributo indica l'`id` del SIGNAL che, se presente, rende esplicita la relazione di subordinazione.

- `relType`[38]: questo attributo indica il tipo di subordinazione che sussiste tra i due eventi coinvolti nella relazione. I suoi valori possono essere `INTENSIONAL` quando un evento apre lo scenario ad una possibilità, `EVIDENTIAL` quando si da la prova di una dichiarazione affermativa, `NEG_EVIDENTIAL` quando si da la prova di una dichiarazione negativa, `FACTIVE` quando un evento nella reggente comporta la veridicità dell'evento-argomento posto nella subordinata, `COUNTER_FACTIVE` quando un evento nella reggente comporta la non veridicità dell'evento-argomento posto nella subordinata e `CONDITIONAL` quando siamo in presenza di costruzioni condizionali.

3 L'elaborazione di specifiche per annotazioni di alcuni TLINK

Nel Maggio del 2010, durante un periodo di tirocinio all'Istituto di Linguistica Computazionale del CNR a Pisa, ho collaborato insieme al Dott. Tommaso Caselli

[38] Per un approfondimento dei valori di `relType` in SLINK vedi (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*, p.25)

all'elaborazione di specifiche per l'annotazione di due particolari casi[39] di TLINK: il primo quando gli elementi coinvolti sono un EVENT e una TIMEX3 e il secondo quando nella relazione ci sono due TIMEX3. Se infatti sono state già formulate delle linee guida per l'annotazione dei singoli elementi temporali in un testo in italiano (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*), non ci sono attualmente dei riferimenti esaustivi per l'annotazione delle relazioni temporali in *It-TimeML*. Tutto ciò va quindi considerato come un punto di partenza nella speranza di rendere più semplice e completa l'annotazione dei diversi *link tag*.

Per arrivare alla formulazione di queste specifiche si sono succedute diverse fasi di lavoro:

- La prima fase consisteva nell'annotazione di relazioni temporali[40], da parte di due annotatori in modo indipendente senza seguire alcuna specifica ben precisa, su dei corpora[41] nei quali erano già stati annotati precedentemente tutti i possibili elementi temporali (EVENT, TIMEX3, SIGNAL). Tutto ciò è stato utile infatti per vedere quali punti in comune potevano esistere a priori nelle conoscenze personali degli annotatori.
- Sulla base degli elementi in comune, trovati nelle annotazioni degli annotatori, è stato controllato se questi risultati identici potevano essere dovuti al fatto di aver di fronte forme grammaticali simili: in tal caso queste forme venivano segnate come interessanti.
- Partendo da tali forme grammaticali sono state elaborate delle possibili specifiche, utilizzate poi come punto di riferimento per una seconda fase di annotazione dei corpora.

---

[39] Il lavoro più grosso, dove sono state concentrate maggiormente le energie, è il primo caso, quello tra EVENT e TIMEX3.

[40] Una prima annotazione riguardava tutte le relazioni temporali che potevano sussistere tra un EVENT e una TIMEX3 (o viceversa) e poi tra due TIMEX3.

[41] Questi corpora sono attualmente reperibili nel sito www.batcaves.org/bat/tool, dove tra l'altro si trova anche il *tool on-line* gratuito necessario per l'annotazione dei vari testi (*BAT*). In particolare i corpora utilizzati per le prove di annotazione sono due: *TimeBank_Italian_part1* e *TimeBank_Italian_part2*; essi sono rispettivamente 55 (28463 *token*) e 56 (20293 *token*) testi provenienti da quotidiani italiani. Al momento della prima fase di lavoro su di essi era già stata effettuata un annotazione da altri annotatori ed erano già visibili tutti gli EVENT, TIMEX3 e SIGNAL possibili.

- Dopo diverse discussioni riguardo i risultati ottenuti dalla seconda fase di annotazione, su tali specifiche, è stato condotto un esperimento[42] per calcolare il grado di accordo tra gli annotatori: ciò è essenziale per verificare l'affidabilità dei risultati per una futura riproducibilità delle specifiche.

3.1 Il *Brandeis Annotation Tool*

Il *Brandeis Annotation Tool* (o *BAT)* è lo strumento *on-line* gratuito utilizzato per l'annotazione dei corpus durante il mio periodo di tirocinio presso l'ILC del CNR a Pisa.
Esso ha essenzialmente tre compiti principali (www.batcaves.org/bat/tool):

- La selezione di elementi (o sequenze di elementi) testuali attribuibili a una certa categoria (EVENT, TIMEX3, SIGNAL) di *TimeML*.
- La possibilità di aggiungere attributi agli elementi annotati.
- L'etichettatura delle relazioni che possono sussistere tra due elementi annotati.

L'interfaccia del *tool* è molto semplice ed intuitiva (vedi la figura 1 di seguito): ogni frase nel testo è numerata sequenzialmente a sinistra con la lettera "s" (che sta per *sentence*) seguita da un numero, che è il numero della frase nel testo. Ogni elemento, annotato manualmente dall'annotatore, viene poi colorato automaticamente di verde dal *too*l ed inserito tra due parentesi quadre, così da essere subito reperibile per altre eventuali annotazioni: il numero posto in alto a destra ad ognuno di essi (anche questo in modo automatico da *BAT*) indica l'*id*[43] numerico univoco assegnato ad ogni elemento della stessa categoria (EVENT, TIMEX3, SIGNAL) nel medesimo testo. A destra infine vengono visualizzate, se sono presenti, le relazioni[44] tra gli

42 L'esperimento sarà analizzato in dettaglio nel capitolo 4 di questa tesi.

43 L'*id* numerico posto sull'elemento annotato nel testo non necessariamente corrisponde con l'attributo `ID` di *TimeML*. L'*id* numerico nel testo è utile infatti se, ad esempio, nella stessa frase vogliamo distinguere due elementi che nonostante siano rappresentati graficamente nello stesso modo non rappresentano lo stesso evento, la stessa espressione temporale o lo stesso segnale.

44 Le relazioni visualizzate rappresentano di volta in volta un *link tag* tra due elementi della frase, i quali sono selezionati manualmente dall'annotatore.

elementi annotati, dove è anche possibile inserire manualmente in una cella il loro valore. La scelta degli elementi da inserire di volta in volta in una relazione spetta all'annotatore, che può decidere in qualsiasi momento di eliminarla tramite un pulsante rosso (con una "x") posto alla sinistra della relazione stessa.

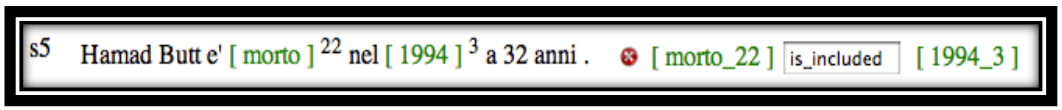

Figura 1. L'annotazione tramite *BAT* di una relazione temporale.

3.2 Specifiche per annotazioni di TLINK fra EVENT e TIMEX3

Osservando attentamente le annotazioni degli annotatori, nel caso delle relazioni temporali in cui sono coinvolti un evento ed una espressione di tempo, sono state elaborate delle specifiche, le quali si possono sostanzialmente suddividere in due sottogruppi:

- quando l'evento coinvolto nella relazione temporale è un nome, un aggettivo oppure un sintagma preposizionale
- quando l'evento coinvolto nella relazione temporale è un verbo

Di seguito verranno mostrate le eventuali specifiche per questi casi di TLINK.

3.2.1 Se l'EVENT è un nome, un aggettivo o un sintagma preposizionale

Le specifiche mostrate di seguito possono essere tutte applicabili, sia se l'evento nella relazione temporale è un *nome*, un *aggettivo* oppure un *sintagma preposizionale*. Per semplicità verrà mostrato solo il caso in cui l'evento è un *nome*.

Se quindi l'EVENT è un *nome*[45] può essere in relazione temporale con una TIMEX3 nei casi in cui troviamo delle forme grammaticali del tipo:

- [*nome* (EVENT) + *aggettivo temporale* (TIMEX3)] o viceversa, ovvero quando un nome è modificato da un aggettivo temporale, es.:

(50) La <EVENT>*mostra*</EVENT> <TIMEX3>*biennale*</TIMEX3>...

- [*nome* (EVENT) + di + (TIMEX3)][46]; per capire che siamo di fronte a questa forma grammaticale basta vedere se essa risponde alla domanda "quando ?" oppure "per quanto tempo ?" e la risposta è la TIMEX3 coinvolta nella relazione, es.:

(51) a. L'<EVENT>*assemblea*</EVENT> di <TIMEX3>*ieri*</TIMEX3>...
b. La <EVENT>*trattativa*</EVENT> di <TIMEX3>*3 anni*</TIMEX3>...

- [(TIMEX3) + di + *nome* (EVENT)][47]; questa forma risponde alla domanda "per quanto tempo ?" e la risposta è la TIMEX3 coinvolta nella relazione, es.:

(52) <TIMEX3>*3 giorni*</TIMEX3> di <EVENT>*scontri*</EVENT>...

- [a + (TIMEX3) + da + *nome* (EVENT)][48]; questa forma risponde alla domanda "dopo quanto tempo ?" e la risposta è la TIMEX3 coinvolta nella relazione, es.:

(53) A <TIMEX3>*vent'anni*</TIMEX3> dalla <EVENT>*chiusura*</EVENT>...

- [*nome incrementativo*[49] (EVENT) + (TIMEX3 situato nella stessa frase)[50]], es.:

(54) Il <EVENT>*crollo*</EVENT> della borsa nel <TIMEX3>*1924*</TIMEX3>...

---

45 L'EVENT può quindi anche essere un *aggettivo* oppure un *sintagma preposizionale*.

46 Di solito il valore di questo TLINK è `includes` oppure `is_included`; se però il TIMEX3 è una durata esso è `measure`.

47 Il valore di questo TLINK è `measure`.

48 Di solito il valore di questo TLINK è `measure`.

49 Alcuni esempi di *nomi incrementativi* posso essere "crescita", "crollo", "aumento", "perdita", "abbassamento" o simili...

50 Il TIMEX3 in questione può anche non essere adiacente al *nome incrementativo*, ma deve comunque essere situato nella medesima frase. In casi di dubbio è consigliato non creare la relazione di tempo: non è detto che un *nome incrementativo* nella medesima frase di un TIMEX3 sia inevitabilmente in relazione con esso; è sempre consigliabile motivare le proprie scelte.

- [*nome* (EVENT) + (*frase relativa* o *participio passato* con TIMEX3)], ovvero un nome modificato da una frase relativa o participio passato che contiene un'espressione di tempo, es.:

(55) L'<EVENT>*assemblea*</EVENT>, che si è tenuta <TIMEX3>*ieri*</TIMEX3>, ha approvato il piano.

- [*nome* che sia *soggetto* o *complemento oggetto* di *verbi che misurano tempo* (EVENT) + (TIMEX3)][51], es.:

(56) a. L'<EVENT>*assemblea*</EVENT> è durata <TIMEX3>*3 ore*</TIMEX3>.

b. L'<EVENT>*assemblea*</EVENT> è stata rinviata al <TIMEX3>*3 maggio*</TIMEX3>.

- [*nome funzionale*[52] o *nome che denota uno stato temporaneo sortale*[53] (EVENT) + (TIMEX3)] o viceversa, es.:

(57) a. Il <EVENT>*bilancio*</EVENT> <TIMEX3>*2011*</TIMEX3>.

b. L'<TIMEX3>*ex*</TIMEX3> <EVENT>*presidente*</EVENT>...

3.2.2 Se l'EVENT è un verbo

Se l'EVENT è un *verbo* può essere coinvolto in una relazione temporale con una TIMEX3 a patto che entrambi gli elementi siano o nella *stessa frase principale* oppure nella *stessa subordinata*: non è accettabile ad esempio che un verbo in una *subordinata non temporalmente esplicita*[54] abbia una relazione temporale con una TIMEX3 nella frase principale. Quindi in definitiva:

---

51 Se nella frase c'è il verbo "durare" il valore del TLINK fra i due elementi coinvolti è sempre `measure`. Per verbi che misurano tempo si intendono anche verbi come "prolungare", "ritardare", "convocare", "rinviare", "anticipare", "seguire", "slittare", "annullare", "prevedere", "posporre" o simili...

52 Con *nomi funzionali* si intendono nomi che rappresentato "qualcosa di misurabile" (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*) come "temperatura", "l'utile", "bilancio", "popolazione", "taglia" o simili...

53 Con *nomi che denotano uno stato temporaneo sortale* (Caselli, *It-TimeML: TimeML Annotation Guidelines for Italian Version 1.2*) si intendono nomi come "presidente", "consigliere", "amministratore", "ambasciatore" o simili...

54 Solo nel caso in cui siamo di fronte ad una subordinata temporale esplicita tale affermazione può essere smentita (vedi a p.31).

- ❖ se la TIMEX3 è nella *frase principale* allora può essere in relazione temporale *solo* con il *verbo* (o i *verbi*) della principale

(58) L'assemblea, che si è tenuta ieri, ha <EVENT>*approvato*</EVENT> il piano in <TIMEX3>*3 ore*</TIMEX3>.

- ❖ se la TIMEX3 è nella *subordinata* allora può essere in relazione temporale *solo* con il *verbo* (o i *verbi*) della subordinata

(59) L'assemblea, che si è <EVENT>*tenuta*</EVENT> <TIMEX3>*ieri*</TIMEX3>, ha approvato il piano in 3 ore.

In realtà c'è un solo caso dove un TIMEX3 nella *frase principale* può essere in relazione temporale con un *verbo-evento* nella *subordinata*:

- ❖ quando la *subordinata* è *temporalmente esplicita* essendo introdotta da parole come "*quando*", "*mentre*", "*prima di*", "*dopo*", "*non appena*", "*come*" o simili, es.:

(60) Ha mangiato alle <TIMEX3>*16.00*</TIMEX3>, *dopo* è <EVENT>*caduto*</EVENT>.

Esistono poi altre due regole da tener presente per un caso particolare: quando nella medesima frase abbiamo una TIMEX3 ed una sequenza di *verbi-eventi in coordinazione* o *giustapposizione*[55]:

- ❖ se la TIMEX3 indica un' *intervallo generico,* come un giorno, un momento della giornata o una durata, allora *tutti* i verbi in coordinazione o giustapposizione hanno una relazione temporale con il TIMEX3, es.:

(61) <TIMEX3>*Stamani*</TIMEX3> ho <EVENT>*mangiato*</EVENT> e <EVENT>*bevuto*</EVENT>.

- ❖ se la TIMEX3 indica un *istante,* come l'ora di un giorno, allora *solo il primo* evento della coordinazione o della giustapposizione ha una relazione temporale con la TIMEX3, es.:

[55] Gli *eventi in giustapposizione* sono eventi uno dietro l'altro separati solo dalla virgola.

(62) Alle <TIMEX3>*7.00*</TIMEX3> ho <EVENT>*mangiato*</EVENT>, bevuto e guardato la televisione.

Infine è importante notare alcuni casi:

- se siamo in presenza di [*verbo modale* (EVENT) + *nome* (EVENT) + (TIMEX3)], allora la relazione temporale è *solo* tra il *verbo modale* e la TIMEX3, es.:

(63) Il governo <EVENT>*deve*</EVENT> *prendere* una *decisione* <TIMEX3>*oggi*</TIMEX3>.

- se siamo in presenza di [*verbo aspettuale* (EVENT) + *nome* (EVENT) + (TIMEX3)], allora la relazione temporale è *solo* tra il *verbo aspettuale* e la TIMEX3, es.:

(64) E'<EVENT>*iniziata*</EVENT> l'*assemblea* <TIMEX3>*adesso*</TIMEX3>.

- se siamo in presenza di [*verbo di supporto* (EVENT) + *nome*/*sintagma preposizionale* + (TIMEX3)], allora la relazione temporale è *solo* tra il *verbo di supporto* e la TIMEX3, es.:

(65) Marta ha <EVENT>*fatto*</EVENT> la torta <TIMEX3>*oggi*</TIMEX3>.

- se siamo in presenza di [*copula* (EVENT) + *nome*/*aggettivo*/*sintagma preposizionale* + (TIMEX3)], allora la relazione temporale è *solo* tra la *copula* e la TIMEX3, es.:

(66) Il cielo <EVENT>*è*</EVENT> rosso <TIMEX3>*oggi*</TIMEX3>.

In generale diciamo che, se siamo di fronte ad una scelta, è sempre consigliato dare priorità a degli *eventi-verbo* rispetto agli *eventi-nome/aggettivo/sintagma preposizionale*. Naturalmente possono esserci delle eccezioni rispetto ai punti citati: il consiglio è quello di leggere attentamente più volte il testo e di volta in volta cercare la soluzione più appropriata, motivando sempre la scelta.

3.3 Specifiche per annotazioni di TLINK fra TIMEX3

Nel corso del tirocinio all'ILC di Pisa, ho collaborato anche all'elaborazione di alcune specifiche per l'annotazione di relazioni temporali nel caso in cui gli elementi coinvolti siano due espressioni di tempo nello stesso documento. Rispetto agli altri casi, citati nel punto 3.2, il tempo per lavorare a queste specifiche è stato purtroppo ridotto: nonostante sia opportuno in futuro analizzare con più calma il caso, alcuni piccoli risultati sono stati comunque trovati e successivamente anche su tali specifiche, come con le altre, è stato comunque calcolato il grado di accordo per una futura riproducibilità.
In generale le regole sviluppate sono essenzialmente quattro:

- Se siamo in presenza di una TIMEX3 il cui valore temporale è esplicito (come ad esempio "*2/10/2011* ", "*10 maggio*"...) allora essa è sempre in relazione temporale con la *data di creazione del documento* (che è a sua volta è un'altra TIMEX3, di solito non nella stessa frase), es.:

(67) <TIMEX3 id='t1'>*10/3/2011*</TIMEX3>, Roma.
Il <TIMEX3 id='t2'>*5 marzo*</TIMEX3> è successo l'incidente...
<TLINK timeID="t2" relatedToTime="t1"...>

- Se la TIMEX3 è una durata o periodo di tempo (come ad esempio "*2 ore*", "*un anno*"...) e se essa ha un punto di inizio e/o un punto di fine, che sono altre due TIMEX3 nel documento[56], allora tale espressione di tempo avrà rispettivamente un TLINK con ognuna di quest'ultime, es.:

(68) <TIMEX3 id='t1'>*10/3/2011*</TIMEX3>, Libia.
Il <TIMEX3 id='t2'>*3 marzo*</TIMEX3> è scoppiata la guerriglia e solo <TIMEX3 id='t3'>*oggi* </TIMEX3> è finita.
<TIMEX3 id='t4' beginPoint='t2' endPoint='t3'...>*7 giorni*</TIMEX3> di scontri non sono bastati per...
<TLINK timeID="t4" relatedToTime="t2"...>
<TLINK timeID="t4" relatedToTime="t3"...>

[56] Le TIMEX3 coinvolte possono anche non essere nella medesima frase: è essenziale in questo caso leggere attentamente il testo e comprendere se la durata o periodo di tempo ha un punto di inizio e/o fine nel documento annotato.

- ❖ Tutte le TIMEX3 il cui valore temporale non è esplicito, se hanno un' àncora temporale nel documento, la quale non è altro che un'altra TIMEX3 a cui essa è temporalmente legata, allora tra di loro ci sarà una relazione temporale, es.:

(69) Nel <TIMEX3 id='t2'>*giugno del 1984*</TIMEX3> c'è stato a Clermont-Ferrand un incontro tra i sindacati italiani e francesi e già <TIMEX3 id='t3' anchorTimeID='t2'...>*allora*</TIMEX3> si prevedeva...
<TLINK timeID="t3" relatedToTime="t2"...>

- ❖ Se due TIMEX3 nella stessa frase sono separate da un segnalatore (SIGNAL), ovvero [(TIMEX3) + (SIGNAL) + (TIMEX3)][57], allora c'è un TLINK tra le due espressioni di tempo, es.:

(70) <TIMEX3 id='t2'>*Oggi* </TIMEX3> <SIGNAL>*dalle*</SIGNAL> <TIMEX3 id='t3'>*8*</TIMEX3>...
<TLINK timeID="t3" relatedToTime="t2"...>

Anche in questo caso, come in quello del punto 3.2, è opportuno rileggere più volte il documento per riuscire a "scoprire" le varie relazioni temporali tra le TIMEX3, soprattutto per quelle che, nonostante siano legate da un TLINK, non sono nella stessa frase e quindi sono più difficili da trovare.

4 L'accordo e l'affidabilità

Ormai da anni, nel campo della Linguistica Computazionale (ma anche in altri campi), si ritiene opportuno verificare l'affidabilità delle specifiche o dei modelli computazionali elaborati per assicurarsi che siano attendibili nella loro futura riproducibilità: questo è pertanto un requisito preliminare per dimostrare la validità del codice elaborato (Artstein & Poesio 2008). Il presupposto quindi che sta alla base

[57] Se non sussiste questa forma non si annota la relazione temporale.

di tutto ciò, per quanto riguarda il nostro caso[58] (ovvero l'etichettatura di elementi testuali con determinati *tag*)[59] è quello che se gli annotatori possono dimostrare di concordare (o essere in accordo), in una classificazione statistica, su delle categorie assegnate agli elementi annotati, allora le specifiche si possono ritenere affidabili. Se invece gli annotatori non concordano, e quindi sono incoerenti tra loro, allora o qualcuno di essi sta sbagliando qualcosa, oppure le specifiche sono inappropriate[60]. Per misurare questo grado di accordo tra gli annotatori sono stati proposti nel corso degli anni diversi metodi di misurazione, come ad esempio la *k* (*kappa*) di Cohen (Cohen 1960), la π di Scott (Scott 1955) oppure l'α (*alfa*) di Krippendorff (Krippendorff 1980): ognuno di essi ha dei pro e dei contro a seconda del fenomeno che andiamo ad analizzare, ma generalmente la *k* di Cohen è il metodo più utilizzato, soprattutto se gli annotatori sono due (Artstein & Poesio 2008) come nel nostro caso. Una cosa però, sulla quale la letteratura è decisamente concorde, è che sia fondamentale l'uso di metodi di misura (come quelli precedentemente citati) che riescano a calcolare l'accordo tra gli annotatori, escludendo l'accordo dovuto al caso.

L'accordo osservato (A_o) ad esempio, che è la percentuale delle volte in cui due annotatori[61] (*t*) concordano nell'assegnazione degli elementi (*i*) nelle categorie (*c*), è uno dei metodi di misura più semplici, ma anche uno di quelli che escludono il computo dell'accordo per caso. Esso può essere espresso come il numero di elementi sui quali concordano gli annotatori diviso il numero totale degli elementi. Precisamente la sua formula è:

$$A_o = \frac{1}{i}\sum_{i \in I} agr_i$$

---

58 Con *i* intenderemo il numero degli *elementi* annotabili, con *c* il numero delle *categorie* di classificazione assegnabili a questi elementi e con *t* il numero degli *annotatori*.

59 In realtà nel nostro caso calcoleremo l'accordo su quante volte, secondo gli annotatori, esiste una relazione temporale nei testi annotati e quante volte (in queste relazioni) essi hanno assegnato lo stesso valore. Nei punti successivi vedremo come sarà adattato questo procedimento ed in particolare quali sono gli elementi e le categorie coinvolte.

60 E' importante notare che non sempre tutto ciò può essere ritenuto vero: un buon accordo può non garantire la validità delle specifiche se ad esempio gli annotatori sono due e condividono gli stessi pregiudizi: in linea generale sarebbe opportuno in fase di annotazione avere più annotatori possibili. Se comunque gli studi sono su piccola scala allora può essere sufficiente anche l'annotazione di solo due annotatori (Artstein & Poesio, *Inter-Coder Agreement for Computational Linguistics*).

61 Nella tesi verrà utilizzato il termine "annotatore" anche per le descrizioni delle formule statistiche, ma in generale si può parlare di "classificatori" visto che tali formule possono essere applicabili in più campi scientifici.

Per capire la formula assumiamo che, per ogni elemento, se i due annotatori assegnano *i* alla stessa categoria, il *valore di accordo* agr_i sarà uguale ad 1, se invece i due annotatori assegnano *i* a categorie diverse, allora agr_i sarà uguale a 0.
L'accordo osservato (A_o) quindi è la media aritmetica dei *valori di accordo* (agr_i) per tutti gli *elementi* $i \in I$.
Così facendo però, come già detto, non si calcola il fatto che a volte l'accordo tra i due annotatori possa avvenire casualmente; per rappresentare quindi con una formula *l'accordo dovuto al caso* (A_e) possiamo scrivere:

$$A_e = \sum_{c \in C} P(c|t_1) \cdot P(c|t_2)$$

Per capire la formula assumiamo che i due annotatori (t_1 e t_2) siano indipendenti nell'annotazione l'uno rispetto all'altro, ovvero, il caso che gli annotatori concordino sull'assegnazione di una determinata categoria (c) è il prodotto della probabilità (P), per ciascuno di essi, nell'assegnare un determinato elemento a tale categoria: $P(c|t_1) \cdot P(c|t_2)$[62]. *L'accordo dovuto al caso* è quindi la probabilità (P) di t_1 e t_2 di concordare su ogni categoria, che è la somma del prodotto su tutte le categorie (c) (Artstein e Poesio, *Inter-Coder Agreement for Computational Linguistics*).

Per calcolare il grado di accordo tra gli annotatori e quindi l'affidabilità delle specifiche descritte nel capitolo 3 è stato utilizzato il metodo di Cohen, il quale è in grado tener conto anche del caso: nel punto successivo verrà quindi descritto nello specifico l'uso generale di k[63] ed infine verrà mostrato come quest'ultimo è stato applicato per il calcolo dell'accordo sulle specifiche elaborate.

---

62 Il calcolo di $P(c|t_x)$ cambia a seconda del metodo di misurazione utilizzato. Nel punto 4.1 verrà descritto come calcolarlo se si utilizza il metodo *k* di Cohen.

63 Per semplicità *l'accordo osservato* sarà indicato con A_o mentre *l'accordo dovuto al caso* con A_e.

4.1 Il coefficiente *k* di Cohen

Il coefficiente statistico *k* è la misura di accordo creata dallo scienziato Jacob Cohen[64] (Cohen 1960). Esso ci indica che proporzione di possibile accordo al di là del caso è stato effettivamente osservato in un esperimento di classificazione statistica. La sua formula è[65]:

$$k = \frac{A_o - A_e}{1 - A_e}$$

dove il valore 1-Ae indica quanto accordo sia raggiungibile oltre al caso, mentre il valore Ao-Ae indica quanto accordo al di là del caso è stato effettivamente trovato (Artstein & Poesio, 2008). Il rapporto tra Ao-Ae/1-Ae ci indicherà quindi l'accordo che esiste tra gli annotatori in una classificazione scientifica, al di là del caso.
Va specificato che se si utilizza il coefficiente *k* di Cohen, per calcolare la distribuzione individuale $P(c|t_x)$ in *Ae*, ovvero la probabilità che un annotatore (*tx*) classifichi un elemento arbitrario in una tale categoria, bisogna dividere il numero totale di assegnazioni di tale categoria da parte dell'annotatore, ovvero $n_{t_x c}$, con il numero totale di elementi (*i*). Quindi:

$$P(c|t_i) = \frac{n_{t_x c}}{i}$$

Detto questo, visto che anche nel coefficiente *k* la probabilità che i due annotatori assegnino un determinato elemento ad una certa categoria è la probabilità congiunta di ogni annotatore di fare questa assegnazione in modo indipendente, la formula di A_e per tale coefficiente diventa:

$$A_e = \sum_{c \in C} P(c|t_1) \cdot P(c|t_2) = \sum_{c \in C} \frac{n_{t_1 c}}{i} \cdot \frac{n_{t_2 c}}{i}$$

64 Jacob Cohen (1923-1998) si occupò di statistica e psicologia nella sua vita. E' conosciuto soprattutto per aver dato il nome ai coefficienti statistici *k* e *d*.

65 Per comprendere come calcolare *l'accordo osservato Ao* e *l'accordo dovuto al caso Ae* vedi l'introduzione del capitolo 4.

Per riuscire a calcolare *k* in una classificazione statistica è possibile utilizzare una *matrice di confusione o di accordo* (vedi la figura 2 di seguito) che rappresenta l'accuratezza di tale classificazione. In pratica essa si costruisce in base a quante categorie sono coinvolte nella classificazione: se ad esempio le categorie sono due (*c1* e *c2*), come nel caso dell'esperimento di questa tesi, avremo allora due righe e due colonne. Le righe rappresentano le assegnazioni delle categorie del primo annotatore mentre le colonne quelle del secondo. Nelle celle che vanno a formare una prima diagonale nella matrice (quelle colorate in verde nella figura 2) si inserisce il numero di casi, per ogni categoria, in cui i due annotatori sono in accordo: la somma dei numeri nelle celle rappresenta la proporzione delle assegnazioni concordanti tra gli annotatori. Viceversa nelle celle che vanno a formare l'altra diagonale (quelle colorate di grigio nella figura 2) si inserirà il numero di casi per ogni categoria in cui i due annotatori non sono d'accordo: la somma dei numeri in tali celle rappresenta la proporzione delle assegnazioni non concordanti tra gli annotatori.

		ANNOTATORE 2	
	Categoria 1 (*c1*)	Categoria 2 (*c2*)	*TOTALE*
ANNOTATORE 1 Categoria 1 (*c1*)	Numero di casi in accordo sulla categoria 1 tra i due annotatori	Numero di casi non in accordo dell'annotatore 1 con l'annotatore 2	*Totale delle assegnazioni della categoria 1 per l'annotatore 1*
Categoria 2 (*c2*)	Numero di casi non in accordo dell'annotatore 2 con l'annotatore 1	Numero di casi in accordo sulla categoria 2 tra i due annotatori	*Totale delle assegnazioni della categoria 2 per l'annotatore 1*
TOTALE	*Totale delle assegnazioni della categoria 1 per l'annotatore 2*	*Totale delle assegnazioni della categoria 2 per l'annotatore 2*	*Numero Totale delle possibili assegnazioni (i)*

Figura 2. La *matrice di confusione o accordo* per l'esperimento della tesi.

Una cosa importante da notare, per semplificare il calcolo, è quella che alla fine di ogni colonna e di ogni riga abbiamo il numero totale di assegnazioni su una certa categoria da parte di un annotatore (ovvero $n_{t_x c}$). Nell'ultima cella in basso a destra della matrice invece abbiamo il numero totale delle possibili assegnazioni da parte degli annotatori (ovvero *i*).

Alla fine di tutti i calcoli i valori di *k* potranno essere compresi nell'intervallo che va da -1 a 1: un valore di *k* uguale ad 1 significherà che *l'accordo è perfetto*, un valore uguale a 0 significherà che *l'accordo è uguale al caso* ed un valore invece uguale a -1 significherà che c'è un *disaccordo perfetto*. In linea generale comunque un valore di *k* che va da 0,67 fino a 0,8 in letteratura è considerato accettabile ed è già possibile con esso esprimere conclusioni preliminari sull'esperimento; con un valore finale uguale o superiore a 0,8 invece è possibile dare conclusioni definitive, essendo considerato tale valore una soglia che rappresenta decisamente un buon accordo tra gli annotatori.

4.2 L'esperimento sulle specifiche elaborate per l'annotazione dei TLINK

Sulla base delle specifiche elaborate insieme al Dott. Tommaso Caselli all'ILC di Pisa, descritte nel terzo capitolo di questa tesi, è stato condotto un esperimento per verificare la loro affidabilità calcolando il grado di accordo in una annotazione di testi[66] in italiano tra due annotatori. Per quanto riguarda l'annotazione di TLINK fra TIMEX3 l'accordo è stato calcolato automaticamente dal *tool BAT*[67], mentre per l'annotazione di TLINK fra EVENT e TIMEX3 il calcolo è stato effettuato da me attraverso calcoli matematici: in ogni caso il coefficiente utilizzato è la *k* di Cohen.

Sia per i TLINK fra TIMEX3 sia per quelli tra EVENT e TIMEX3 è stato comunque calcolato *k* su due differenti prove, ovvero *Related to Time* e *Relation to Time*:

66 Per i TLINK fra TIMEX3 i testi annotati sono stati 20 su 55 mentre per i TLINK fra EVENT e TIMEX3 sono 55 su 55.

67 I dati sull'accordo calcolati automaticamente dal *tool BAT* mi sono stati concessi gentilmente dal Dott. Tommaso Caselli. Tali risultati verranno rappresentati successivamente nella tesi escludendo però i vari passaggi su come il *tool* ci sia arrivato: gli unici passaggi descritti saranno solo quelli per i TLINK fra EVENT e TIMEX3; se comunque l'accordo sui TLINK fra TIMEX3 fosse stato anch'esso calcolato manualmente, i procedimenti per il suo calcolo sarebbero stati identici a quelli per l'accordo tra EVENT e TIMEX3. Per maggiori informazioni su come poter far calcolare l'accordo a *BAT* e sulle varie limitazioni al riguardo vedi www.batcaves.org/bat/tool.

- *Related to Time* è la prova per capire quante volte i due annotatori sono d'accordo sull'esistenza di una relazione temporale (risponde alla domanda "quanti sono, secondo gli annotatori, i TLINK[68] in comune?").
- *Relation to Time* è la prova per capire quanti TLINK, fra quelli su cui gli annotatori hanno concordato l'esistenza in *Related to Time*, hanno lo stesso valore temporale (risponde alla domanda "quanti, dei TLINK in accordo sull'esistenza, hanno lo stesso valore temporale?").

Se su ogni prova, sia per l'annotazione di TLINK fra TIMEX3 che tra EVENT e TIMEX3, avremo un risultato accettabile compreso tra 0,67 e 0,8, potremo esprime delle conclusioni preliminari sulle specifiche; se invece il valore sarà al di sotto dello 0,67 le specifiche non saranno completamente affidabili; al contrario, un valore uguale o superiore a 0,8 ci confermerà l'affidabilità di quest'ultime.

4.2.1 I dati di partenza

Di seguito verranno proposti i dati di partenza con cui è stato possibile calcolare *k* per le prove *Related to Time* e *Relation to Time*, sia riguardo l'annotazione di TLINK fra EVENT e TIMEX3, sia per l'annotazione di TLINK fra TIMEX3[69]. Questi dati, alcuni gentilmente concessi dal Dott. Tommaso Caselli, alcuni calcolati manualmente da me in base a dei file statistici creati dal *tool* dopo l'annotazione dei testi, sono essenziali per comprendere i calcoli descritti nei punti successivi della tesi.

---

68 Per calcolare *k* in questa prova si è tenuto conto del numero di *token* (elementi) coinvolti nei TLINK (sia i TLINK in comune tra gli annotatori sia quelli che esistono per un solo annotatore): per questo è necessario tener presente che un TLINK è composto da due elementi/*token,* in questo caso un EVENT ed un TIMEX3. Pertanto nei dati esposti di seguito verranno anche inseriti non solo il numero di TLINK in comune, ma anche il numero di *token*/elementi che sono coinvolti in questi TLINK.

69 I dati di partenza per i TLINK fra TIMEX3 sono nettamente inferiori a quelli per i TLINK fra EVENT e TIMEX3 perché sono gli unici riportati dal *tool BAT*. Esso infatti ha calcolato automaticamente il risultato di *k* per tali specifiche direttamente dai file annotati.

Dati di partenza per il calcolo di k sull'annotazione di TLINK fra EVENT e TIMEX3:

Nome del corpora	Timebank Italian Part 1
Numero totale di testi del corpora	55
Numero di testi annotati dai due annotatori	55
Numero totale di token[70] *del corpora*	28463
Numero totale di token dei file annotati	28463
Numero totale di EVENT	4800
Numero totale di TIMEX3	625
Numero di elementi/token nel testo che possono essere coinvolti in un TLINK[71]	5425
Numero totale di TLINK secondo l'annotatore 1	523 (di cui 523 *token* sono EVENT e 468 TIMEX3=991)[72]
Numero totale di TLINK secondo l'annotatore 2	511 (di cui 511 *token* sono EVENT e 451 TIMEX3=962)[73]
Numero di volte (o di TLINK) in cui solo l'annotatore 1 ritiene esserci un TLINK mentre l'annotatore 2 no	53 (di cui 53 *token* sono EVENT e 18 TIMEX3=71)[74]

70 Un *token* è l'unità di analisi minima in cui può essere scomposta una sequenza di caratteri in un testo.

71 Per calcolare questo dato è necessario sommare il numero totale di EVENT (4800) con il numero totale di TIMEX3 (625).

72 Il numero totale di *token* coinvolti nei 523 TLINK sono quindi 991.

73 Il numero totale di *token* coinvolti nei 511 TLINK sono quindi 962.

74 Il numero totale di *token* coinvolti nei 53 TLINK sono quindi 71.

Numero di volte (o di TLINK) in cui solo l'annotatore 2 ritiene esserci un TLINK mentre l'annotatore 1 no	65 (di cui 65 *token* sono EVENT e 35 TIMEX3=100)[75]
Numero di volte (o di TLINK) in comune in cui i due annotatori hanno ritenuto esserci l'esistenza di una relazione temporale	458 (di questi TLINK, 458 *token* sono EVENT e 433 TIMEX3)
Numero di token che per entrambi gli annotatori sono coinvolti in un TLINK	891[76]
Numero di casi in cui i due annotatori sono d'accordo sull'esistenza di un TLINK ma non sono d'accordo sul suo valore temporale	32[77]
Numero totale assoluto di TLINK[78]	576 (composti da 1062 elementi/*token*)
Numero di volte in cui il valore temporale, fra i TLINK su cui gli annotatori hanno condiviso l'esistenza, è uguale	426
Numero di casi in cui i due annotatori non sono d'accordo ne sull'esistenza di un TLINK, ne sul suo valore temporale[79]	118
Numero di token che non sono coinvolti in un TLINK per entrambi gli annotatori	4363[80]

75 Il numero totale di *token* coinvolti nei 65 TLINK sono quindi 100.

76 Questi 891 *token*/elementi sono quelli coinvolti nei 458 TLINK in comune tra gli annotatori (458 EVENT + 433 TIMEX3)

77 Per calcolare questo dato è necessario sottrarre al numero di volte in cui i due annotatori hanno ritenuto esserci l'esistenza di un TLINK (458), il numero di volte in cui i due annotatori sono anche d'accordo sul valore temporale dato ad esso (426).

78 Per calcolare il numero assoluto di annotazioni di TLINK è necessario sommare il numero di casi in cui i due annotatori hanno concordato l'esistenza di un TLINK (458), il numero di volte in cui solo l'annotatore 1 ritiene esserci un TLINK (53) ed il numero di volte in cui solo l'annotatore 2 ritiene esserci un TLINK (65). Sommando per ogni dato il relativo numero di *token* il risultato è quindi 1062.

79 Per calcolare questo dato è necessario sommare il numero di volte in cui solo il primo annotatore ritiene esserci l'esistenza di un TLINK (53), con il numero di volte in cui solo il secondo annotatore ritiene esserci l'esistenza di un TLINK (65).

80 Per calcolare questo dato bisogna prima sommare il numero totale di EVENT nel corpora (4800) con il numero totale di TIMEX3 (625), e da questo risultato bisogna poi sottrarre il numero totale di *token* (1062) che vanno a formare il numero totale assoluto di TLINK trovati dagli annotatori (ovvero 576). Questi sono quindi tutti i *token* che per entrambi gli annotatori non sono coinvolti in un TLINK.

Dati di partenza per il calcolo di k sull'annotazione di TLINK fra TIMEX3[81]*:*

Nome del corpora	Timebank Italian Part 1
Numero totale di testi del corpora	55
Numero di testi annotati dai due annotatori	20
Numero totale di token del corpora	28463
Numero totale di token dei file annotati	8064

4.2.2 Il calcolo di *k* sulle specifiche di annotazione per TLINK fra EVENT e TIMEX3

Di seguito verranno descritti i passaggi per il calcolo di *k*[82] sulle prove *Related to Time* e *Relation to Time* riguardanti l'annotazione di TLINK fra EVENT e TIMEX3 con le specifiche descritte nel capitolo 3.

- *Related to Time*

La prima cosa da controllare, sia per *Related to Time* che per *Relation to Time*, è il numero di categorie (*c*) e di elementi (*i*) coinvolti nella prova analizzata, visto che saranno fondamentali per costruire le *matrici di confusione* e quindi per calcolare *k*.

81 I dati di partenza, presentati per i TLINK fra TIMEX3, sono i soli riportati da *BAT;* essi sono stati comunque presentati perché sono indicativi nel confronto con il risultato finale di *k*, calcolato automaticamente dal *tool.* Per maggiori informazioni al riguardo si consiglia di visitare il sito di *BAT* www.batcaves.org/bat/tool.

82 Per comprendere i vari passaggi del calcolo di *k* è necessario guardare i dati di partenza presentati nel punto precedente.

In questo caso, visto che la prova ci chiede quante volte gli annotatori sono d'accordo sull'esistenza di un TLINK, possiamo affermare che le categorie coinvolte sono due (*c1* e *c2*): una per indicare l'esistenza di un TLINK (*c1*), una per indicare il contrario (*c2*). Ad essere precisi, visto che abbiamo tenuto conto per questa prova del numero di *token* coinvolti nei TLINK, possiamo affermare che una categoria (*c1*) ci indicherà il numero di *token* in comune tra gli annotatori (ovvero il numero di *token* che sono coinvolti nei TLINK in comune tra gli annotatori), mentre l'altra categoria (*c2*) ci indicherà tutti quei *token* che per entrambi gli annotatori non sono coinvolti in un TLINK. Il numero di elementi/*token* (*i*) che possono essere annotati da i due annotatori come TLINK sono invece 5425.
Partendo da questi dati, inseriremo nella prima cella di accordo della matrice (quella in verde, in alto a sinistra, per la categoria 1, nella figura 3 di seguito), il numero di *token* in comune fra i due annotatori (891), ovvero quelli che sono coinvolti nei TLINK su cui gli annotatori hanno concordato la loro esistenza, mentre nella cella che va a formare con quella precedente la *diagonale di accordo* (quella in verde, in basso a destra), inseriremo il numero di *token* che, per entrambi gli annotatori, non sono coinvolti in un TLINK (ovvero 1062). Nelle celle rimanenti (in grigio), quelle che vanno a formare la *diagonale di disaccordo* nella matrice, inseriremo rispettivamente 100 e 71, ovvero il numero di *token* che sono coinvolti in un TLINK solo per un annotatore e non per l'altro, e quindi il numero di casi in cui uno solo di essi ha ritenuto esserci l'esistenza di un TLINK.

ANNOTATORE 2

ANNOTATORE 1		Esistenza TLINK (*c1*)	Inesistenza TLINK (*c2*)	*TOTALE*
	Esistenza TLINK (*c1*)	891	100	*991*
	Inesistenza TLINK (*c2*)	71	4363	*4434*
	TOTALE	*962*	*4463*	*5425 (i)*

Figura 3. La *matrice di confusione* per la prova *Related to Time* sull'annotazione di TLINK fra EVENT e TIMEX3.

A questo punto dobbiamo calcolare *l'accordo osservato Ao* e *l'accordo dovuto al caso Ae*, seguendo le formule descritte all'inizio di questo capitolo, per avere i dati necessari al calcolo di *k*.
Quindi per il calcolo di *Ao* basterà sommare i numeri nelle celle in verde della matrice, quelle che vanno a formare la *diagonale di accordo*, e dividere tale risultato con il numero di elementi nel testo che possono essere annotati come TLINK (ovvero *i*):

$$A_o = \left[(891 + 4363) : 5425\right] = 0{,}9684792$$

Alla fine di ogni colonna ed ogni riga della matrice invece abbiamo l'assegnazioni totali che ogni annotatore ha fatto per ogni categoria, ovvero $n_{t_x c}$; questi quattro valori ci serviranno per calcolare le distribuzioni individuali $P(c|t_x)$ in *Ae*:

$$A_e = \left[(991 : 5425) \cdot (962 : 5425)\right] + \left[(4434 : 5425) \cdot (4463 : 5425)\right] = 0{,}7047855$$

A questo punto non ci rimane che inserire i valori di *Ao* ed *Ae* nella formula per calcolare *k* su *Related to Time*:

$$k_{\mathrm{Re}latedToTime} = \frac{0{,}9684792 - 0{,}7047855}{1 - 0{,}7047855} = 0{,}8932274$$

❖ *Relation to Time*

Anche per la prova *Relation to Time* abbiamo due categorie (*c1* e *c2*): una (*c1*) ci indica l'esistenza, per i due annotatori, di TLINK in comune con gli stessi valori, e l'altra (*c2*) indica il contrario, ovvero l'inesistenza di TLINK in comune con gli stessi valori. Il valore invece che attribuiamo ad *i*, è il numero totale assoluto di annotazioni di TLINK, ovvero 576.
Nella prima cella di accordo, quella per la prima categoria (in verde nella figura 4 di seguito), inseriremo il numero di volte in cui per i due annotatori esiste un TLINK con il medesimo valore temporale, ovvero 426. Nella seconda cella di accordo,

quella per la seconda categoria (che va a creare con la precedente la *diagonale di accordo*), inseriremo invece il numero di casi che ci indica l'accordo, per i due annotatori, sull'inesistenza di TLINK con il medesimo valore, ovvero 118[83]. Successivamente, in una delle due celle in grigio (le quali vanno a formare la *diagonale di disaccordo)*, inseriremo il numero di casi in cui per gli annotatori esiste un TLINK ma non c'è accordo sul suo valore, ovvero 32. Nell'altra cella in grigio invece inseriremo un valore fittizio pari a 0, che sta semplicemente ad indicare che l'annotatore di riferimento è solo uno (ovvero l'altro): i TLINK totali su cui non c'è accordo sul valore infatti sono 32, quindi è indifferente[84] ai fini del calcolo sapere chi ha fatto certe assegnazioni.

		ANNOTATORE 2		
		Esistenza TLINK con lo stesso valore (*c1*)	Inesistenza TLINK con lo stesso valore (*c2*)	*TOTALE*
ANNOTATORE 1	Esistenza TLINK con lo stesso valore (*c1*)	426	0	*426*
	Inesistenza TLINK con lo stesso valore (*c2*)	32	118	*150*
	TOTALE	*458*	*118*	*576 (i)*

Figura 4. La *matrice di confusione* per la prova *Relation to Time* sull'annotazione di TLINK fra EVENT e TIMEX3.

Come per l'altra prova, anche per *Relation to Time* sarà sufficiente calcolare, tramite i valori ottenuti nella matrice, *l'accordo osservato Ao* e *l'accordo dovuto al caso Ae*. Inserendo tali risultati nella formula di *k* avremo l'accordo anche per questa prova; quindi:

83 Questo valore è il numero di volte in cui solo il primo annotatore ritiene esserci l'esistenza di un TLINK (53) sommato al numero di volte in cui solo il secondo annotatore ritiene esserci l'esistenza di un TLINK (65). Tutto ciò è il totale dei casi dove non viene riconosciuta, da entrambi gli annotatori, l'esistenza del medesimo TLINK e di conseguenza nemmeno il suo valore: capovolgendo il punto di vista quindi possiamo anche dire che esso indica l'accordo, tra gli annotatori, sull'inesistenza di TLINK con gli stessi valori.

84 Per bilanciare questa cosa (e quindi per non avere un unico annotatore di riferimento), avremmo potuto anche assegnare, ad esempio, un valore pari a 16 a tutte e due le celle: il risultato sarebbe stato comunque lo stesso.

$$A_o = \left[(426 + 118) : 576\right] = 0{,}9444444$$

$$A_e = \left[(426 : 576) \cdot (458 : 576)\right] + \left[(150 : 576) \cdot (118 : 576)\right] = 0{,}6414205$$

$$k_{\mathrm{Re}\,lationToTime} = \frac{0{,}9444444 - 0{,}6414205}{1 - 0{,}6414205} = 0{,}8450675$$

4.2.3 Il calcolo di *k* sulle specifiche di annotazione per TLINK fra TIMEX3

Come già accennato nei punti precedenti, il calcolo dell'accordo fra i due annotatori, riguardo l'annotazione di TLINK fra TIMEX3 con le specifiche elaborate, è stato automaticamente calcolato dal *tool BAT*[85], lo stesso *tool* utilizzato per l'annotazione dei testi. Il coefficiente è sempre *k* e le prove effettuate sono le stesse: *Related to Time* e *Relation to Time*.
I risultati finali calcolati da *BAT* sono quindi i seguenti:

$$k_{\mathrm{Re}\,latedToTime} = 0{,}95$$
$$k_{\mathrm{Re}\,lationToTime} = 0{,}94$$

85 Per maggiori informazioni riguardo il calcolo di *k* tramite *BAT* vedi www.batcaves.org/bat/tool.

5 Conclusioni

Visto che i valori di *k*, inerenti alle prove *Related to Time* e *Relation to Time* per le specifiche di annotazioni di TLINK fra EVENT e TIMEX3 e per quelle per i TLINK fra soli TIMEX3, sono al di sopra della soglia dello 0,8[86], possiamo affermare che il lavoro effettuato all'ILC del CNR a Pisa è stato più che buono e che quindi le specifiche sono affidabili. Bisogna comunque considerare il fatto che avendo avuto più tempo, più annotatori e più testi su cui poter effettuare l'esperimento, il risultato sarebbe stato sicuramente più preciso. Con ciò non siamo certo arrivati alla fine di un percorso, considerando anche il fatto che l'annotazione di relazioni temporali tra eventi ed espressioni di tempo nell'ambito del *Trattamento Automatico del Linguaggio* è ancora all'inizi: possiamo affermare però che questo risultato, soprattutto in ambito italiano con *It-TimeML,* è un punto di partenza nella speranza di avere nel corso dei prossimi anni delle specifiche che fungano da punto di riferimento per le annotazioni dei testi al riguardo.

[86] Per i TLINK fra TIMEX3 i risultati sono ottimi, attestandosi per tutte e due le prove al di sopra dello 0,9.

6 Bibliografia

Artstein, Ron, e Massimo Poesio. 2008. *Inter-coder agreement for computational linguisics*. "Computational Linguistics".

Caselli, Tommaso. 2010. *It-TimeML: TimeMl Annotation Guidelines for Italian Version 1.2.*

Caselli, Tommaso. 2010. *It-TimeML: TimeMl Annotation Guidelines for Italian Version 1.3.*

Cohen, Jacob. 1960. *A coefficient of agreement for nominal scales*. "Educational and Psychological Measurement".

Ferro, Lisa, Inderjeet Mani, Beth Sundheim, George Wilson, Robyn Kozierok e Laurie Gerber. 2002. *Instruction Manual for the Annotation of Temporal Expressions*. MITRE, Washington C3 Center, McLean, Virginia.

ISO, S. W. G. 2008. *ISO DIS 24617-1: Language resource management – Semantic annotation frame work – Part1: Time and events*. ISO Central Secretariat, Ginevra.

Krippendorff, Klaus. 1980. *Content Analysis: An introduction to Its Methodology*. Beverly Hills, CA.

Scott, William A. 1955. *Reliability of content analysis: The case of nominal scale coding*. "Public opinion quarterly".

Setzer, A. 2001. *Temporal information in newswire article: An annotation scheme and a corpus study*. University of Sheffield.

Verhagen, Marc, e James Pustejovsky. 2010. *SemEval-2010 Task 13: Evaluating events, time expressions, and relations (tempeval-2)*.

Siti web consultati:

www.batcaves.org/bat/tool
www.tc37sc4.org
www.timeml.org

7 Appendici

Descrizione BNF del *tag* EVENT:

```
attributes ::= id anchor pred class type tense aspect pos
               polarity mood [modality]
id ::= e<integer>
anchor ::= IDREF
{IDREF ::= (token<integer>)*}
{default, if absent, is an empty string}
pred ::= CDATA
class ::= REPORTING | PERCEPTION | ASPECTUAL | I_ACTION |
          I_STATE | OCCURRENCE
pos ::= ADJECTIVE | NOUN | VERB | PREPOSITION | OTHER
tense ::= FUTURE | PAST | PRESENT | NONE
aspect ::= IMPERFECTIVE_PROGRESSIVE | PERFECTIVE | IMPERFECTIVE
           | NONE
vform ::= INFINITIVE | GERUND | PARTICIPLE | NONE
          {default, if absent, is NONE}
polarity ::= NEG | POS {default, if absent, is POS}
mood ::= SUBJUNCTIVE | CONDITIONAL | IMPERATIVE | NONE
         {default, if absent, is NONE}
modality ::= CDATA
comment ::= CDATA
```

Descrizione BNF del *tag* TIMEX3:

```
attributes ::= id anchor type
               [functionInDocument][beginPoint][endPoint][quant]
               [freq][temporalFuction](value|valueFromFunction)
               [mod][anchorTimeID]
id ::= ID
{ID ::= TimeID
TimeID ::= t<integer>}
anchor ::= IDREF
{anchor ::= (token<integer>)}
type ::= DATE | TIME | DURATION | SET
```

```
beginPoint ::= IDREF
{beginPoint ::= TimeID}
endPoint ::= IDREF
{endPoint ::= timeID}
quant ::= CDATA
freq ::= CDATA
functionInDocument ::= CREATION_TIME | EXPIRATION_TIME |
                       MODIFICATION_TIME | PUBLICATION_TIME |
                       RELEASE_TIME | RECEPTION_TIME | NONE
                       {default, if absent, is NONE}
temporalFunction ::= true | false {default, if absent, is
                     false}
{temporalFunction ::= boolean}
value ::= CDATA
{value ::= duration | dateTime | time | date | gYearMonth | gYear
          | gMonthDay | gDay | gMonth}
valueFromFunction ::= IDREF
{valueFromFunction ::= TemporalFunctionID
TemporalFunctionID ::= tf<integer>}
mod ::= BEFORE | AFTER | ON_OR_BEFORE | ON_OR_AFTER | LESS_THAN
        | MORE_THAN | EQUAL_OR_LESS | EQUAL_OR_MORE | START | MID
        | END | APROX
anchorTimeID ::= IDREF
{anchorTimeID ::= TimeID}
```

Descrizione BNF del *tag* SIGNAL:

```
attributes ::= id anchor
id ::= ID
{ID ::= s<integer>}
anchor ::= IDREF
{ANCHOR ::= (token<integer>)*}
```

Descrizione BNF del *link tag* TLINK:

```
attributes ::= [id][origin](eventID | timeID) [signalID]
               (relatedToEvent | realatedToTime) relType
id ::= ID
{id ::= LinkID
```

```
LinkID ::= l<integer>}
origin ::= CDATA
eventID ::= IDREF
{eventID ::= EventID}
timeID ::= IDREF
{timeID ::= TimeID}
signalID ::= IDREF
{signalID ::= SignalID}
relatedToEvent ::= IDREF
{relatedToEvent ::= EventID}
relatedToTime ::= IDREF
{relatedToTime ::= TimeID}
relType ::= BEFORE | AFTER | INCLUDES | IS_INCLUDED |
            SIMULTANEOUS | DURING | DURING_INV | IAFTER |
            IBEFORE | IDENTITY | BEGINS | ENDS | BEGUN_BY |
            ENDED_BY
temporalDistance ::= IDREF
{temporalDistance ::= TimeID}
```

Descrizione BNF del *link tag* ALINK:

```
attributes ::= [id][origin] eventID [signalID] relatedToEvent
               relType
id ::= ID
{ID ::= LinkID
LinkID ::= l<integer>}
origin ::= CDATA
eventID ::= ID
{ID ::= EventID}
signalID ::= IDREF
{signalID ::= SignalID}
relatedToEvent ::= IDREF
{relatedToEvent ::= EventID}
relType ::= INITIATES | CULMINATES | TERMINATES | CONTINUES |
            REINITIATES
```

Descrizione BNF del *link tag* SLINK:

```
ATTRIBUTES ::= [id][origin] eventID [signalID]
               subordinatedEvent relType
id ::= ID
{id ::= LinkID
LinkID ::= l<integer>}
origin ::= CDATA
eventID ::= IDREF
{eventID ::= EventID}
subordinatedEvent ::= IDREF
{subordinatedEvent ::= EventID}
```