



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

TESI DI LAUREA SPECIALISTICA

**L'annotazione dell'aspetto verbale per il
trattamento automatico della lingua italiana:
esperimenti e valutazioni**

Candidato: *Valentina Bartalesi Lenzi*

Relatore: *Prof. Alessandro Lenzi*

Anno Accademico 2010-2011

Indice

Introduzione	4
Capitolo I: Aspetto e azione verbale.....	8
1.1 Azione verbale	9
<i>1.1.1 Tratti azionali</i>	<i>10</i>
1.1.1.1 Duratività.....	10
1.1.1.2 Dinamicità	11
1.1.1.3 Telicità.....	12
<i>1.1.2 Classi azionali</i>	<i>13</i>
1.1.2.1 State.....	14
1.1.2.2 Activity.....	15
1.1.2.3 Accomplishment.....	15
1.1.2.4 Achievement.....	16
1.1.2.5 Altre categorie azionali	16
<i>1.1.3 Interazione tra azione verbale e contesto</i>	<i>18</i>
1.1.3.1 Ibridismo azionale	18
1.1.3.2 Commutazione azionale	19
1.2 Aspetto verbale.....	20
<i>1.2.1 Perfettivo e imperfettivo</i>	<i>21</i>
1.2.1.1 Aspetto perfettivo	23
1.2.1.1.1 Aspetti aoristico e compiuto.....	24
1.2.1.1.2 Aspetto ingressivo	25
1.2.1.2 Aspetto imperfettivo.....	27
1.2.1.2.1 Aspetto progressivo.....	27
1.2.1.2.2 Aspetto continuo	28
1.2.1.2.3 Aspetto abituale.....	29
1.2.1.2.3.1 Imperfettività gnomica	34
1.3. Interazione tra azione e aspetto verbale.....	37
Capitolo II: I corpora e l'elaborazione dell'informazione temporale	40
2.2 Annotazione di un corpus.....	41
2.3 Verso uno standard internazionale per l'annotazione di risorse linguistiche	43
<i>2.3.1 L'iniziativa ISO e gli altri progetti di standardizzazione</i>	<i>45</i>
<i>2.3.2 Il Linguistic Annotation Framework</i>	<i>46</i>

2.4 Strumenti di annotazione e di estrazione di informazione temporale da testi secondo le specifiche TimeML	47
2.4.1 <i>Strumenti per l'elaborazione temporale in TempEval.....</i>	48
2.4.2 <i>Altri strumenti per l'annotazione e l'estrazione di informazione temporale..</i>	51
2.5 Corpora annotati con informazione inerente azione e aspetto verbale.....	52
Capitolo III: Il progetto TimeML, l'aspetto e l'azione verbale	57
3.1 Lo schema di annotazione	58
3.1.1 <i>Il tag <EVENT></i>	59
3.1.2 <i>Il tag <TIMEX3>.....</i>	60
3.1.3 <i>I tag <TLINK>, <ALINK> e <SLINK>.....</i>	61
3.1.4 <i>Il tag <SIGNAL>.....</i>	61
3.2 IT-TimeML.....	62
3.2.1 <i>Il tag <EVENT></i>	63
3.2.1.1 <i>L'attributo MOOD</i>	63
3.2.1.2 <i>L'attributo VFORM</i>	64
3.2.1.3 <i>I valori degli attributi MODALITY e ASPECT.....</i>	65
3.2.2 <i>Il tag <TIMEX3>.....</i>	65
3.3 L'attributo ASPECT.....	66
3.3.1 <i>Proposta di integrazione dei valori dell'attributo ASPECT.....</i>	72
3.3.2 <i>Specifiche per l'annotazione dei valori aspettuali</i>	75
3.4 L'azione verbale nello schema di annotazione TimeML	80
3.4.1 <i>Proposta di integrazione per la marcatura dell'azione verbale.....</i>	84
Capitolo IV: Crowdsourcing e annotazione dell'aspettualità: un esperimento... 86	
4.1 Ambiti di applicazione e limiti del Crowdsourcing.....	87
4.2 Utilizzo dei sistemi di crowdsourcing per l'annotazione	89
4.2.1 <i>Alcune esperienze di utilizzo di sistemi di crowdsourcing</i>	89
4.2.2 <i>Crowdfower.....</i>	90
4.3 Individuazione dei verbi afferenti alla classe aspettuale della imperfettività gnomica tramite CrowdFlower	91
4.3.1 <i>Esperimenti correlati.....</i>	92
4.3.2 <i>Il data set per l'esperimento.....</i>	92

4.3.2.1 Il gold standard.....	94
4.3.2.2 Descrizione del corpus I-CAB	94
4.3.3 Istruzioni per l'esperimento.....	96
4.3.4 Risultati.....	98
4.3.4.1 Calcolo dell'affidabilità dei lavoratori	99
4.3.4.2 Agreement tra lavoratori	100
4.3.4.3 Analisi quantitativa dei risultati	102
4.3.4.4 Analisi qualitativa dei risultati	105
Capitolo V: Conclusioni.....	117
Appendice A: BNF dei tag <EVENT>, <SIGNAL>, <TIMEX3>.....	123
Appendice B: lista delle frasi che compongono il data set.....	125
Appendice C: lista delle frasi che compongono il gold standard	143
Appendice D: frasi che presentano discordanze nei giudizi.....	145
Bibliografia	150

Introduzione

Si definiscono proprietà aspettuative di un verbo le caratteristiche inerenti la sua costruzione interna e le sue peculiari modalità di svolgimento.

Mentre il tempo verbale esprime la relazione tra la localizzazione temporale di un evento rispetto ad un altro punto temporale di riferimento, l'aspetto verbale riguarda le proprietà strutturali dell'evento stesso.

I linguisti fanno una distinzione formale tra aspetto grammaticale e aspetto lessicale, ovvero tra aspetto e azionalità. L'aspetto lessicale o azionalità ha a che fare con il significato del predicato considerato; l'aspetto grammaticale, invece, è di natura morfologica, è dipendente, in genere, dalla flessione verbale ed esprime un punto di vista prospettico sull'evento.

Questi criteri di distinzione non valgono, però, in tutte le circostanze. L'aspetto, infatti, non è sempre definibile sulla base della morfologia verbale ma in molti casi viene individuato per inferenza piuttosto che per diretta commutazione tra diversi tempi verbali. In italiano, infatti, i tempi verbali sono ambigui rispetto alle valenze aspettuative che esprimono; per esempio, l'imperfetto può assumere una valenza imperfettiva abituale oppure imperfettiva progressiva a seconda del contesto in cui occorre.

Il riconoscimento e l'annotazione dell'aspetto e dell'azione verbale si configurano come task complessi, in cui interagiscono allo stesso tempo vari livelli di elaborazione linguistica. Proprio a causa della difficoltà di risoluzione dei problemi che riguardano il dominio tempo-aspettuale, l'indagine svolta in questa tesi si è sviluppata secondo un processo gerarchicamente ordinato in livelli di difficoltà crescente. Tenendo presente che l'orizzonte di riferimento di questo lavoro è costituito dalla creazione di risorse annotate per sistemi di Natural Language Processing (NLP), si è cercato, innanzitutto, di rispondere a tre domande riguardanti lo stato dell'arte:

1. Esistono schemi di annotazione che contemplano l'etichettatura di azione e aspetto verbale?

2. Se e quali sono le risorse annotate attualmente disponibili che riportano questi tipi di informazione?
3. Quali e quanti sono i sistemi automatici disponibili utilizzati per la creazione di risorse linguistiche di questo tipo o che sfruttano questi dati per il proprio l'addestramento?

Un'attenzione particolare è stata rivolta all'analisi dello schema di annotazione TimeML (Pustejovsky et al., 2003), per la marcatura di eventi, espressioni temporali e delle loro relazioni, nell'ottica di indagare se e in quale misura vengono proposte metodologie per l'annotazione di aspettualità e azionalità e se queste sono sufficientemente esaurienti ed esaustive.

Il lavoro svolto in questa tesi si giustifica proprio alla luce dei risultati ottenuti da quest'analisi. È risultato evidente, infatti, come l'annotazione aspettuale non sia molto diffusa nei corpora, con particolare riferimento alla marcatura dell'aspetto abituale. Tale mancanza è spesso giustificata con l'elevata difficoltà di distinguere le varie tipologie di aspetto verbale: proprio dalla verifica di questa asserzione prende avvio l'esperimento svolto in questa tesi e l'idea di utilizzare la piattaforma di crowdsourcing CrowdFlower per testare la capacità di utenti non linguisticamente addestrati di individuare una particolare tipologia aspettuale, ovvero l'aspetto afferente alla classe dell'imperfettività gnomica (Bertinetto e Lenci, 2011).

La decisione di marcare l'imperfettività gnomica (aspetti abituale, attitudinale, potenziale, generico e Individual Level predicate) è stata dettata dall'interesse per l'individuazione, all'interno dei testi, di frasi che esprimono una generalizzazione di qualche tipo oppure una proprietà che caratterizza un soggetto per un intero periodo della sua vita o per tutta la sua esistenza. Si è cercato, quindi, di capire se possa essere utile/possibile marcare questa particolare classe aspettuale in prospettiva di un'estrazione automatica di informazione di senso comune (Singh, 2002) da testi scritti.

Come afferma Bertinetto (Bertinetto, 1994): "L'aspetto abituale costituisce uno dei problemi più intricati tra quelli che si affollano nel dominio tempo-aspettuale",

un traguardo ambizioso, quindi, quello di indagare le problematiche riguardanti le difficoltà incontrate da persone comuni nell'individuare i verbi afferenti alla classe aspettuale dell'imperfettività gnomica ma proprio per questo un'impresa decisamente interessante.

Struttura della tesi

In questo lavoro si vanno, innanzitutto, ad introdurre gli aspetti linguistici che sono stati al centro dello studio condotto: dai concetti teorici di aspetto e azione verbale allo studio della loro interazione (capitolo 1).

Nel capitolo 2 si affronta il tema del rapporto tra corpora ed elaborazione dell'informazione temporale. Dopo una panoramica sulle procedure di annotazione dei corpora e sui loro usi, viene discusso il problema della standardizzazione delle risorse linguistiche e vengono analizzati in maniera critica gli strumenti di annotazione disponibili che consentono di compiere una marcatura in linea con le specifiche TimeML ed i sistemi utilizzati per l'estrazione di informazione temporale da testi. Viene, infine, presentata una panoramica dei corpora attualmente disponibili annotati con informazione inerente l'azione e l'aspetto verbale.

Il capitolo 3 riguarda il progetto TimeML e l'annotazione dell'aspetto e dell'azione verbale. Lo schema di annotazione viene descritto accuratamente e criticamente analizzato in relazione al trattamento riservato alla marcatura di aspettualità e azionalità. Viene, quindi, proposta un'integrazione dei valori già esistenti per l'attributo indicante la categoria aspettuale del verbo e vengono fornite delle specifiche per gli annotatori in modo da facilitarne il riconoscimento. Viene, inoltre, descritta una proposta di integrazione degli attributi al fine di riconoscere e normalizzare anche l'informazione inerente l'azione verbale.

Il capitolo 4 tratta lo studio sperimentale svolto in questo lavoro. Viene contestualizzato e descritto l'esperimento mirato a testare la capacità di utenti comuni di individuare l'aspetto verbale afferente alla classe dell'imperfettività

gnomica; quindi, è presentata un'approfondita analisi quantitativa e qualitativa sui risultati ottenuti.

Infine, nel capitolo 5 sono raccolte alcune riflessioni conclusive.

Capitolo I

Aspetto e azione verbale

L'aspetto verbale tradizionalmente riguarda ciò che Comrie (1976) chiama "different ways of viewing the internal temporal constituency of a situation". L'intuizione che sta alla base di questa definizione è che mentre il tempo verbale esprime la relazione tra la localizzazione temporale di un evento rispetto ad un altro punto temporale di riferimento, l'aspetto verbale riguarda le proprietà strutturali dell'evento stesso.

I linguisti fanno una distinzione formale tra aspetto grammaticale e aspetto lessicale, ovvero tra aspetto e azionalità (traduzione del tedesco *Aktionsart*, termine originariamente utilizzato in (Agrell, 1908)). In particolare, Smith (1991) riconduce questa distinzione ad un piano semantico in cui si contrappongono quelli che l'autrice definisce come "situation aspect" e "viewpoint aspect" (vedi anche Filip, 1993, 2000). L'aspetto lessicale o azionalità, corrispondente al "situation aspect", ha a che fare con il significato del predicato considerato. L'aspetto grammaticale, invece, è di natura morfologica, è dipendente, in genere, dalla flessione verbale e esprime un punto di vista prospettico sull'evento, il "viewpoint aspect".

Questi criteri di distinzione non valgono, però, in tutte le circostanze. L'aspetto, infatti, come verrà spiegato più approfonditamente nel prosieguo di questo capitolo, non è sempre deducibile dalla morfologia del verbo ma in molti casi viene individuato per inferenza piuttosto che per diretta commutazione tra diversi tempi verbali. L'aspetto non va confuso, dunque, col riferimento temporale così come non vanno confusi i concetti di tempo verbale e di valore tempo-aspettuale: i tempi verbali sono caratterizzati sia da tratti aspettuati che propriamente verbali, inerenti la localizzazione degli eventi, il valore aspettuale, invece, può essere marcato verbalmente come non esserlo e uno stesso tempo verbale può denotare diversi valori aspettuati. Per quanto riguarda l'azione, infine, uno stesso verbo può cambiare di valore a seconda delle differenti condizioni morfosintattiche e contestuali.

1.1 Azione verbale

La prima questione da affrontare nello studio dell'azione verbale è relativa all'interpretazione delle proprietà aspettuali lessicali che possono essere intese o come proprietà di espressioni linguistiche o come proprietà di eventi del mondo reale.

Aristotele nel libro IX della Metafisica e nel X dell'Etica fa una distinzione tra "kinesis" e "energia" ovvero tra attività complete, classificate come ateliche e movimenti incompleti e rivolti a un fine esterno, classificati come telici.

Un forte argomento in favore della teoria secondo la quale gli eventi hanno proprietà intrinseche viene da Kamp (1979), il quale sostiene che il cambiamento è un concetto primitivo e che la distinzione tra eventi statici ed eventi di cambiamento è la suddivisione primitiva che sta alla base di ogni teoria.

Molti studi linguistici, però, ritengono che l'azione verbale riguardi espressioni linguistiche e non proprietà di eventi. In (Krifka, 1998) si afferma che la distinzione tra telicità e atelicità non si evince dalla natura dell'evento descritto ma dalla descrizione dell'evento stesso. Questa teoria viene supportata anche da studi riguardanti la struttura degli eventi come (Parsons, 1990) e (Landman, 2000), in cui si afferma che gli eventi non hanno nessuna struttura atomica interna.

Un altro problema relativo a questa distinzione ma separato, riguarda la natura della classificazione azionale; la classificazione di Vendler (1967) in quattro categorie azionali (*states, activities, achievements, accomplishments*) è stata quella che ha maggiormente influenzato gli ultimi 35 anni di ricerche in questo campo.

Infine, un terzo e ultimo fondamentale problema nello studio dell'azione verbale riguarda l'individuazione dei tratti che determinano le proprietà azionali di un verbo in contesto.

Nell'ambito di questo lavoro si assume che i verbi denotino eventi o stati distinti in classi azionali in relazione alle particolari proprietà del verbo stesso. Nei paragrafi seguenti, verranno trattati in dettaglio i problemi sopra presentati.

1.1.1 Tratti azionali

Secondo la tradizionale classificazione in termini di Aktionsart proposta in Vendler (1967), i predicati vengono raggruppati, come descritto in dettaglio nel prossimo paragrafo, in quattro classi (stativi, trasformativi, risultativi e predicati di attività)¹ sulla base dei loro tratti semantici di dinamicità, duratività e telicità, elicetabili attraverso alcuni test linguistici quali l'applicazione di modificatori temporali. Per esempio, l'avverbiale temporale "a lungo" è compatibile con verbi durativi ma non con verbi puntuali e quindi può essere utilizzato come test sintattico per la verifica della duratività di un verbo.

1.1.1.1 Duratività

Il tratto di duratività distingue fra predicati che denotano processi caratterizzati da uno svolgimento quasi istantaneo, in cui il punto di inizio coincide idealmente con il punto finale, per esempio *incontrare*, *esplodere*, *arrivare*, *cadere* ecc, e verbi che riguardano invece processi che si prolungano nel tempo, per esempio *avere fame*, *crescere*, *dormire*, *scalare*, *amare* ecc.

- (1) Giulia dorme sul divano [durativo]
- (2) Giulia è caduta dal letto [non durativo]

Ci sono casi in cui risulta difficile, però, dire se un verbo sia o meno durativo: il contesto, in queste occasioni, può essere determinante. Per esempio, se *uscire di casa* è sicuramente non durativo, *uscire dalla città* presenta evidentemente un grado maggiore di duratività, come dimostra il seguente esempio:

- (3) A causa del traffico ho impiegato due ore per uscire da Milano

I predicati non durativi sono incompatibili con modificatori avverbiali durativi del tipo "per X tempo" (4a vs 4b) mentre i predicati durativi sono incompatibili con avverbiali puntuali (5a vs 5b).

¹ Seguendo la terminologia proposta in Bertinetto (1986), "trasformativo" e "risultativo" traducono rispettivamente i termini vendleriani "achievement" e "accomplishment".

- (4) a. Mario ha parlato *per due ore*
 b. *Mario è inciampato *per due minuti*²

- (5) a. La bomba è esplosa *alle due in punto*
 b. Giorgio ha dormito *alle due in punto*

1.1.1.2 Dinamicità

Il tratto di dinamicità distingue fra predicati eventivi, che denotano azioni, e predicati stativi, che denotano proprietà, qualità o situazioni in cui tipicamente è assente ogni aspetto di mutamento, per esempio *assomigliare, esistere, stare, credere, preferire, ecc.*

- (6) a. Chiara crede nel destino [non dinamico]
 b. Chiara sta montando la libreria dell'IKEA [dinamico]

In italiano, i predicati stativi sono tipicamente incompatibili con l'imperativo (7a vs 7b) e la perifrasi progressiva (8a vs 8b):

- (7) a. mangia!
 b. *sii alto!

- (8) a. Mario sta mangiando
 b. *Mario sta essendo alto

I verbi non dinamici sono in genere, anche incompatibili con il verbo vicario fare (9), con l'avverbio deliberatamente (10) e con i verbi di persuasione (11):

- (9) ??Angela sapeva la poesia a memoria e Luca faceva lo stesso
 (10) *Gianni sapeva deliberatamente a memoria la poesia

² In questo caso può scattare un'interpretazione iterativa della frase: l'avverbiale "per X TEMPO" indica l'arco temporale in cui si localizzano i sotto-eventi in cui è possibile suddividere l'evento plurinazionale di *inciampare*.

(11) *Maria ha convinto Luca a sapere a memoria la poesia

Esistono, però, eccezioni rispetto alle regole sopra elencate per individuare se un verbo è dinamico o non dinamico: i cosiddetti verbi dinamici agentivi (per esempio *soffrire, ottenere, apprendere una bella notizia*) sono incompatibili con l'imperativo ma non con gli altri test sintattici sopra indicati (12a vs 12b):

(12) a *Laura ottieni un grande successo!

b. Laura sta ottenendo un grande successo

In alcuni usi non standard dell'italiano, per esempio, in italiano regionale meridionale, è ammesso l'uso della perifrasi progressiva con i verbi non dinamici, per esempio "Questo film mi sta proprio piacendo". Esistono, anche, lingue, come l'inglese, che si comportano allo stesso modo:

(13) a.*Luca sta continuamente dimenticando la poesia

(13) b. Luca is continuously forgetting the poem

In italiano, inoltre, i verbi possono assumere una valenza dinamica o non dinamica a seconda del contesto (vedi cap. 3.1.2).

1.1.1.3 Telicità

Il tratto di telicità individua predicati che tendono verso il punto finale di un processo. Se un verbo è telico, è necessario che l'azione sia completata perché possa considerarsi effettivamente avvenuta.

(14) a. Emma prepara la lezione [telico]

(14) b. Emma disegna nello studio [non telico]

I verbi telici quando sono coniugati secondo il paradigma di un tempo imperfettivo, non consentono di derivare alcuna implicazione circa la positiva conclusione del processo, al contrario di ciò che accade quando tali verbi sono

coniugati secondo tempi perfettivi; questo fenomeno è detto paradosso dell'imperfettività.

(15) Giovanni cantava una canzone dei Doors

(16) Giovanni cantò una canzone dei Doors

I predicati atelici non sono compatibili con modificatori avverbiali completivi del tipo “in X tempo” (17a vs 17b).

(17) a. I panni sono asciugati in 10 minuti

(17) b. *Mario ha abitato a Roma in 10 anni

I telici si comportano in modo diverso con l'avverbiale “da X TEMPO” a seconda dell'aspetto verbale: con il valore perfettivo il verbo mantiene la telicità, mentre con il valore imperfettivo l'avverbiale rende il verbo non telico (18° vs 18b):

(18) a. Monica ha steso i panni da mezz'ora (l'evento telico si è concluso mezz'ora fa)

(18) b. Monica sta tostando il pane da mezz'ora (l'evento non telico è iniziato mezz'ora fa)

1.1.2 Classi azionali

Le principali classificazioni verbali in classi azionali sono quelle in (Ryle 1949), (Vendler 1967) e (Kenny 1963). Ryle distingue tra *achievements*, cioè verbi che indicano cambiamenti di stato associati ad uno scopo finale e *accomplishments*, ovvero cambiamenti di stato senza un fine associato. Kenny non tiene in considerazione la classificazione di Ryle e distingue *states*, *activities* e *performances*, quest'ultime indicanti eventi che hanno un naturale punto finale. Vendler, infine, prendendo in considerazione le intuizioni sia di Ryle che di Kenny, distingue i verbi in quattro classi azionali: *states*, *activities*, *achievements*, *accomplishments*.

Di seguito, alcuni esempi di verbi appartenenti alle quattro classi azionali tratti da Dowty (1979):

States: sapere, credere, avere, desiderare, amare, conoscere, essere felici

Activities: correre, camminare, nuotare, guidare la macchina

Achievements: riconoscere, trovare, perdere, raggiungere, morire

Accomplishments: costruire una casa, dipingere un ritratto, disegnare un cerchio

	Duratività	Dinamicità	Telicità
<i>State</i>	+	-	-
<i>Activities</i>	+	+	-
<i>Achievements</i>	-	+	+
<i>Accomplishments</i>	+	+	+

Tabella 1. Rapporto tra classi azionali e tratti azionali

1.1.2.1 State

I predicati stativi, indicano situazioni non dinamiche, non teliche ma durative ed esprimono in genere qualità, abitudini e abilità del soggetto.

(18) Franca è felice

(19) Franca sa la canzone a memoria

Un'ulteriore distinzione interna alla classe degli stativi è quella fra stativi permanenti, verbi cioè non compatibili con alcuni avverbiali temporali come “da X TEMPO”, “dal momento X”, e stativi non permanenti compatibili, invece, con gli avverbiali sopra citati:

(20) *Mauro è alto da due anni

(21) Mauro conosce Chiara da due anni

Vi sono alcuni test utili, di cui, in parte, si è già discusso precedentemente, per individuare i verbi stativi (Rothstein 2004):

1. Gli stativi non occorrono in forma progressiva

(22) *Franca sta sapendo la canzone

2. Un verbo stativo coniugato al presente non può avere una lettura abituale

(23) Livio conosce la risposta

3. I verbi stativi non occorrono insieme a verbi di persuasione o costrizione:

(24) *Gianni costringe Livio a conoscere la risposta

4. Gli stativi non occorrono in forma imperativa

(25) *Conosci la risposta!

5. Gli stativi non occorrono con avverbi come *deliberatamente* o altri avverbi indicanti azioni volontarie:

(26) *Livio deliberatamente conosce la risposta

1.1.2.2 Activity

I predicati di activity (*camminare, nuotare, lavorare, mangiare pesce, ecc...*) rappresentano situazioni dinamiche e durative e, come gli stativi, non prevedono il raggiungimento di un punto terminale del processo [-telici].

(27) Roberto canta mentre cammina

1.1.2.3 Accomplishment

I predicati di accomplishment (*asciugare, imparare, uccidere, mangiare un filetto di pesce, ecc...*) condividono con i predicati di activity i tratti [+dinamico] e

[+durativo], ma a differenza di questi ultimi configurano un punto finale del processo che induce un mutamento di stato ([+telici]). Rientrano in questa classe anche i predicati di activity, come *correre* in (28a), che vengono telicizzati dalla presenza di un oggetto diretto costituito da un lessema (28b)³:

(28) a. *Mario corre in 10 secondi

(28) b. Mario corre i 100 metri in 10 secondi

1.1.2.4 Achievement

Infine, a differenza dei predicati di accomplishment con cui condividono i tratti [+dinamico] e [+telico], i verbi di achievement (*scoppiare, scomparire, trovare, morire, nascere, ecc...*) designano un mutamento di stato puntuale [-durativo].

(29) Federico è nato pochi giorni fa

1.1.2.5 Altre categorie azionali

Alle categorie azionali descritte da Vendler se ne possono aggiungere almeno due di sostanziale rilevanza:

1. La categoria dei semelfattivi o puntuali indicanti un'azione istantanea, non implicante un cambiamento di stato. Questi predicati sono dinamici e non durativi come gli achievement ma, a differenza di questi ultimi, non sono telici (*tossire, starnutire, ecc*).

I puntuali si comportano in maniera opposta rispetto agli achievement, sono, infatti, compatibili con gli avverbiali “per X TEMPO” mentre non

³ L'individuazione del secondo argomento è, infatti, uno dei tratti che determina il grado di telicità (e quindi la classe azionale) di un predicato (Verkuyl 1972, Romagno 2005): “mangiare pesce” è un predicato di *activities* [-telico], compatibile con modificatori del tipo “per X tempo” (“Cesare ha mangiato pesce per un mese”) ma non con modificatori del tipo “in X tempo” (“Cesare ha mangiato pesce in un minuto”), mentre “mangiare un filetto di pesce” è un predicato risultativo [+telico], compatibile con modificatori del tipo “in X tempo” (“Cesare ha mangiato un filetto di pesce in un minuto”) ma non con modificatori del tipo “per X tempo” (“Cesare ha mangiato una filetto di pesce per un mese”). È quindi importante sottolineare che la classe azionale di un predicato è strettamente collegata al particolare contesto in cui esso ricorre: non c'è necessariamente una singola struttura logica soggiacente a tutti gli usi di un certo predicato.

sono accettabili con avverbiali del tipo “in X TEMPO” oppure “ci ha messo X TEMPO”, che fanno riferimento ad un processo di preparazione che culminerà con l’accadimento dell’evento telico:

(30) Viola tossi per tre ore

(31) ?Viola è partita per tre ore

Si noti che l’avverbiale “per X TEMPO” unito a verbi indica il protrarsi di una catena di eventi che iterati hanno, nel complesso, valenza durativa (come *squillare, lampeggiare, rimbombare, tossire*). Nell’esempio seguente, infatti, non è il singolo squillo che dura tre ore ma l’intera serie.

(32) Il telefono ha squillato per tre ore

Con i verbi di achievement, invece, l’avverbiale “per X TEMPO” indica il protrarsi del risultato del verbo telico. L’esempio (31) non indica una serie di partenze durante l’arco delle tre ore ma indica il fatto che Chiara è assente per tre ore.

2. La categoria degli incrementativi, individuata da Bertinetto e Squartini (1995), è costituita da predicati dinamici e durativi che esprimono una graduale progressione verso un nuovo stato non necessariamente definitivo (*ingrassare, aumentare, ecc...*).

Nei tempi perfettivi, come per esempio nella frase “L’orchidea è cresciuta”, questi verbi risultano ambigui, infatti “crescere” può avere due interpretazioni:

1. “diventare grande”, ovvero avere raggiunto l’obiettivo della crescita;
2. “diventare più grande”, ovvero il raggiungimento di un nuovo stato, che non è necessariamente definitivo

Questo tipo di predicati è compatibile con il test di telicità “in X TEMPO” e quindi sono verbi telici. Tuttavia, presentano delle caratteristiche che li distinguono sia dai verbi di activity che da quelli di accomplishment, ovvero: (i) sono compatibili con avverbi come “poco/molto/gradualmente”, mentre i verbi di activity solo con “poco” e “molto” e quelli di accomplishment solo con “gradualmente”; (ii) sono compatibili con avverbi di comparazione come “di molto/di parecchio/ulteriormente/appena/sensibilmente”; (iii) sono caratterizzati dalla possibilità di ricorrere in perifrasi progressiva con il verbo *andare*.

1.1.3 Interazione tra azione verbale e contesto

1.1.3.1 Ibridismo azionale

In italiano i verbi caratterizzati da un'unica categoria azionale sono rari, sono, invece, più frequenti i cosiddetti verbi ibridi (Bertinetto, 1986), come per esempio, il verbo *percepire* che può convogliare un'accezione stativa nella frase *percepire un senso di malessere* oppure un'accezione telica in *percepire una possibilità*, nel senso di “prendere improvvisamente coscienza di”⁴.

L'ibridismo azionale può essere interpretato come una conseguenza della polisemia caratterizzante molti verbi. Per gli esempi (33) e (34) si può parlare infatti di polisemia azionale, per evidenziare come le classi azionali dipendano dalla semantica del verbo e dalle sue interazioni con il contesto:

(33) Conosco Gianpaolo da 5 anni [state]

(34) Ho conosciuto Gianpaolo ieri [achievement]

⁴ Per un'abbondante esemplificazione sul genere di ibridismo azionale qui in discussione, si veda Lucchesi [1971]. Ecco un esempio che ben illustra la situazione:

[i] Leo *impugnò* la pistola; tutt'attorno si fece silenzio [perfettivo, telico]

[ii] Quando Lia entrò, Leo *impugnava* la pistola [imperfettivo, atelico]

[iii] Leo *impugnò* saldamente la pistola per tutta la sparatoria [perfettivo, atelico]

[iv] Quando Leo *impugnava* la pistola, Lia aveva paura [imperfettivo, telico]

Queste oscillazioni verbali, chiamate anche opposizioni infra-lessicali e descritte per la prima volta in (Lucchesi 1971), sono generalmente determinate da elementi di contesto come l'aspetto verbale o l'animatezza del soggetto:

1. aspetto imperfettivo: Bruce Wayne *indossava* la maschera per evitare di essere riconosciuto [state]
2. aspetto perfettivo: Bruce Wayne *indossò* la maschera e si mise al volante della Batmobile [achievement]
3. aspetto imperfettivo – soggetto inanimato: l'edificio *rompeva* l'armonia della piazza [state]
4. aspetto perfettivo – soggetto animato: Marco *ha rotto* l'orologio di suo padre [achievement]

È da sottolineare, però, che per nessuna classe azionale sembra possibile selezionare un insieme di tratti la cui presenza in contesto sia contemporaneamente necessaria e sufficiente a determinare l'appartenenza a quella specifica classe (Lenci e Zarcone, 2009).

Un'altra oscillazione frequente in italiano è quella dei verbi di activity che possono essere resi telici dalla presenza di un complemento oggetto (fatta eccezione per quando questo è un nome massa o un plurale senza articolo). Infatti, verbi come *disegnare*, *cantare*, *fumare* di per sé non sono telici ma ciò che invece li rende tali, in particolari contesti, è proprio la presenza di un complemento oggetto che li finalizza al raggiungimento di un determinato scopo, come avviene in *disegnare un ritratto*, *cantare una canzone*, *fumare una sigaretta*.

1.1.3.2 Commutazione azionale

Il fenomeno della commutazione azionale a differenza dell'ibridismo, coinvolge verbi poco polisemici. Come sostenuto da Pustejovsky, che parla di "coercizione azionale" (Pustejovsky e Bouillon, 1995) e Rothstein, che parla di "Aspectual shift" (Rothstein, 2004), il contesto può, infatti, intervenire sull'interpretazione azionale prototipica di un evento, modificandola.

Per esempio, in italiano o in inglese, un verbo prototipicamente non durativo come *arrivare* può essere utilizzato nella forma di perifrasi progressiva ma questa operazione dà luogo ad una commutazione azionale e il predicato assume una lettura durativa:

(35) Il gatto sta arrivando alla torta [The cat is reaching the cake]

La commutazione quindi, a differenza dell'ibridismo, non è indice di una vera ambiguità azionale del verbo ed è soggetta ad alcune limitazioni:

1. la conclusione dell'evento deve avvenire entro breve tempo;
2. la conclusione dell'evento non può essere esplicitamente falsificata;
3. non sono ammesse pause durante il percorso.

Da notare, però, che secondo Dowty (1977), Lendman (1992) e altri, queste restrizioni indicherebbero che non si tratta di una vera e propria commutazione azionale ma piuttosto di proprietà intrinseche dei predicati di achievement. In tal caso, quindi, non si dovrebbe parlare di "accomplishment derivato" per le forme progressive di achievement che invece dovrebbero avere un'interpretazione a metà strada tra un'accezione prospettica e una "a rallentatore".

1.2 Aspetto verbale

Si definiscono proprietà aspettuali di un verbo le caratteristiche inerenti la sua costruzione interna e le sue peculiari modalità di svolgimento; nell'identificazione di un particolare valore aspettuale, infatti, non vengono considerate la localizzazione nel tempo e la rete di relazioni temporali in cui il verbo è inserito (Bertinetto, 1991).

L'aspetto e il tempo linguistico sono entrambi collegati con il tempo fisico ma in modi differenti. Il tempo linguistico è una categoria deittica e gli eventi risultano localizzati in un determinato tempo fisico: molto spesso, si prendono come riferimenti il tempo di enunciazione o il tempo di svolgimento di un altro evento.

L'aspetto, invece, non riguarda la relazione tra il tempo di un evento e un altro tempo di riferimento ma piuttosto l'organizzazione temporale interna dell'evento stesso.

I valori aspettuali, per l'italiano, possono essere riassunti nel diagramma seguente, le cui etichette saranno oggetto dei prossimi paragrafi.

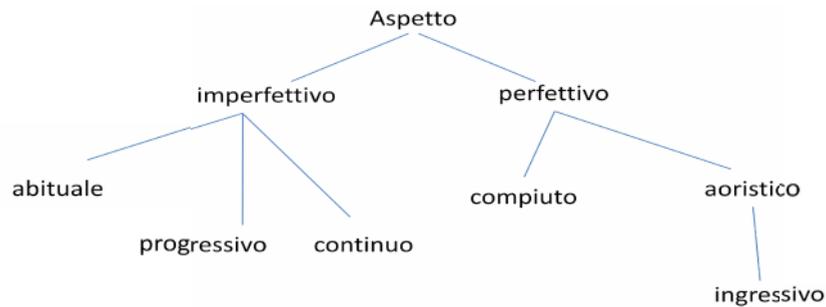


Figura 1. Diagramma dei valori aspettuali per l'italiano

1.2.1 Perfettivo e imperfettivo

Il valore perfettivo di un verbo indica che l'evento da esso denotato viene visto come un tutto, senza la distinzione in fasi interne all'evento stesso. L'imperfettività, al contrario, pone l'attenzione sulla struttura temporale interna dell'evento (Comrie, 1976).

Bertinetto (1986) afferma che l'aspetto perfettivo fa sempre riferimento ad eventi singoli, conclusi e isolati dei quali viene visualizzato l'istante finale, a differenza dell'imperfettivo che invece fa riferimento ad eventi visti nel loro svolgersi, nel loro ripetersi o durare.

Le due definizioni sopra riportate convergono verso un punto comune: nelle definizioni di Bertinetto la chiave che avvicina le due proposizioni sta nel definire l'aspetto perfettivo come facente riferimento ad un evento di cui "viene visualizzato l'istante finale". Questa specificazione, infatti, collega strettamente la visione di Bertinetto con quella di Comrie, in cui l'evento è visto con un tutto, finito e concluso. A mio avviso, proprio l'uso del termine "visualizzazione" da parte di Bertinetto può portare ad avvicinare, in una sorta di metafora, l'aspetto perfettivo all'immagine di un quadro ormai completato in cui viene rappresentata

una scena finita nella quale non può essere aggiunto nient'altro: l'immagine descrive l'azione vista dall'esterno in tutto e per tutto.

Una differenza di valore aspettuale (perfettivo/imperfettivo) è alla base anche del diverso uso tra i tempi passati, di passato prossimo e passato remoto da un lato e dell'imperfetto dall'altro.

L'imperfetto esprime un evento di aspetto imperfettivo. L'aspetto imperfettivo si può manifestare come un evento abituale (imperfetto abituale) o come un evento colto nel suo svolgersi (imperfetto progressivo e imperfetto continuo). Passato prossimo e remoto esprimono, invece, un valore perfettivo. La distinzione d'uso tra i due passati si può ricondurre alla distinzione tra un aspetto perfettivo compiuto e un aspetto perfettivo aoristico.

Di questi particolari aspetti verbali, si tratterà in dettaglio nei paragrafi seguenti.

Una riflessione particolare, però, deve essere riservata all'aspetto dei modi indefiniti, in italiano: participio, gerundio e infinito. Questi, come i modi finiti, possono esprimere opposizioni di tempo, diatesi e appunto di aspetto. L'infinito e il gerundio possono esprimere ambedue l'opposizione anteriorità/non anteriorità; attivo/passivo. Si differenziano, però, per l'aspetto: il gerundio presente è imperfettivo mentre il passato è perfettivo. L'infinito presente secondo la situazione, può esprimere aspetto perfettivo, imperfettivo o rimanere neutro; l'infinito passato, invece, ha aspetto perfettivo. Il participio passato (il participio presente spesso non viene preso in considerazione poiché, sebbene non ancora esclusivamente aggettivo deverbale, raramente assume funzione verbale) non esprime né opposizioni di tempo né di diatesi; il suo tratto distintivo è l'aspetto perfettivo.

In conclusione come sostenuto in (Bertinetto, 2003) le forme non finite composte (infinito e gerundio composti), ma anche il participio passato, sono univocamente orientate ad esprimere l'aspetto compiuto.

Le forme non finite semplici (infinito e gerundio semplici) sono invece aperte ad una molteplicità di letture: aspettualmente, possono indicare valore imperfettivo (progressivo o continuo) e perfettivo-aoristico.

Queste opposizioni si possono illustrare con i seguenti esempi:

(36) ??A mangiare il cioccolato, si è felici

(37) Mangiando il cioccolato, si è felici

(38) Mangiato il cioccolato, si è felici

Nel comparto delle forme finite, la distinzione aspettuale fondamentale riguarda il fatto che l'evento sia visto come completo (perfettività) o incompleto (imperfettività). Nel comparto delle forme non finite la distinzione aspettuale fondamentale sembra essere quella tra stato e dinamismo-eventività. La prima categoria è necessariamente evocata dalle forme composte e dal participio passato, che, esprimendo l'aspetto compiuto, implicano l'esistenza di uno stato risultante, conseguente al compiersi dell'evento. La seconda categoria è invece implicata dalle forme semplici, indipendentemente dall'effettivo valore aspettuale.

1.2.1.1 Aspetto perfettivo

Per quanto riguarda la definizione di perfettività, bisogna, innanzitutto, evidenziare come questa nozione sia stata spesso definita in maniera inadeguata nella letteratura sull'aspetto e nelle grammatiche delle singole lingue (Dostál, 1954).

Alcune volte la forma perfettiva viene definita come indicante eventi di breve durata, mentre la forma imperfettiva, sempre secondo questa prospettiva, è utilizzata per indicare eventi di lunga durata. È semplice, però, trovare esempi, in diverse lingue, che contraddicono questa asserzione: possono essere, infatti, utilizzate entrambe le forme verbali per riferirsi alla stessa durata temporale, senza nessuna precisazione sulla lunghezza o sulla brevità dell'evento considerato (Comrie, 1976).

Per esempio, in francese la differenza tra *il régna trente ans* e *il régnait trente ans* sta nel fatto che nella prima frase il periodo di trent'anni viene visto come un tutto, unico e completo, nella seconda frase invece si fa riferimento alla struttura

interna del regno, che potrebbe spingere a fare considerazioni e ad elencare singoli eventi avvenuti durante questi trent'anni.

Uguualmente, la perfettività non può essere definita come la forma utilizzata per descrivere durate limitate. Un'ora, un giorno, due mesi o trent'anni sono tutti periodi di tempo limitati ma, come dimostra l'esempio sopra riportato, possono essere utilizzate sia forme perfettive che imperfettive per descrivere la durata di uno stesso evento.

Collegata alla precedente definizione se ne trova un'altra anch'essa, però, non abbastanza adeguata, cioè quella per cui la perfettività viene utilizzata per indicare un evento puntuale o momentaneo ma l'esempio riportato prima falsifica anche questa definizione. Si deve però riconoscere che in quest'ultima asserzione è possibile rintracciare un elemento di verità inerente la natura perfettiva: la forma perfettiva non descrive la struttura interna di un evento ma riduce lo stesso ad un tutto, rappresentabile quasi come un singolo punto.

Un'ulteriore, frequente, definizione di perfettività è quella che la caratterizza come indicante un'azione completa. La forma perfettiva, infatti, identifica un evento completo, con un inizio, uno sviluppo e una fine. Con il termine "completo", però, così come spesso utilizzato nelle grammatiche tradizionali, viene data un'enfasi maggiore alla fase terminale di un evento mentre la perfettività tende piuttosto a rappresentare l'evento come un tutto, senza enfatizzare una fase o un'altra della sua struttura interna.

Come risulta dalla Figura 1, l'aspetto perfettivo in italiano può essere suddiviso in tre ulteriori categorie aspettuuali: aoristico, compiuto e ingressivo.

1.2.1.1.1 Aspetti aoristico e compiuto

La differenza d'uso tra passato prossimo e passato remoto può essere spiegata con la distinzione tra un aspetto perfettivo compiuto e un aspetto perfettivo aoristico.

L'aspetto compiuto è tipico dei verbi composti ed esprime il perdurare, nel momento di riferimento, del risultato di un evento svoltosi in precedenza.

L'aspetto aoristico, invece, fissa l'attenzione sul momento finale dell'azione, mostrando l'evento in modo puntuale.

Mentre il passato prossimo esprime una forma di passato inclusivo, in cui l'evento descritto è considerato come concluso ma ripetibile o attuale, il passato remoto esprime un passato definitivo, in cui l'evento è collocato in una dimensione temporale diversa e separata da quella attuale:

(39) Non seppi mai la verità (e ormai è troppo tardi)

(40) Non ho mai saputo la verità (finora)

La contrapposizione dell'aspetto compiuto con l'aoristico risulta evidente attraverso la differente compatibilità tra tempi perfettivi (semplici e composti) con l'avverbiale decorrenziale "da X TEMPO":

(41) Francesco è arrivato da 2 ore

(42) Laura era partita da 3 ore, quando Livio arrivò

(43) *Luca arrivò/partì da due ore

I tempi composti del perfetto presuppongono contemporaneamente un momento nel tempo in cui si localizza l'evento e un successivo momento, rilevante psicologicamente, che viene scelto come punto di valutazione dei risultati dell'evento precedente. Di questi due distinti momenti i tempi semplici contemplanano nel loro sistema di riferimento temporale solo il primo.

1.2.1.1.2 Aspetto ingressivo

In molte lingue che presentano una distinzione tra forme perfettive e imperfettive, le forme imperfettive di alcuni verbi, in particolare stativi, possono essere utilizzate per indicare l'inizio di un evento. Questo particolare aspetto verbale è definito ingressivo e coinvolge i tempi verbali in accezione aoristica e compare quindi anche con gli usi aoristici del presente storico, come nell'esempio (44) in

cui il presente può essere reso con una perifrasi ingressiva come “comincia a muoversi”:

(44) Nessuno arrivava, era veramente nei guai, ma all'improvviso ecco che l'ascensore *si muove*!

L'aspetto ingressivo riguarda sempre tempi in accezione aoristica, con la sola eccezione della valenza abituale dell'imperfetto (vedi par. 1.2.1.2.3).

Va notato, comunque, che l'accezione ingressiva viene maggiormente applicata ai perfetti semplici piuttosto che ai perfetti composti. Questo comportamento appare chiaro soprattutto nei contesti in cui viene evidenziata la compiutezza dell'evento, come in presenza dell'avverbio *finalmente*:

(45) a. Finalmente Giulia parlò

(46) b. Finalmente Giulia ha parlato

Questo comportamento ambiguo del perfetto può essere spiegato nei termini in cui l'uso del perfetto composto conserva il valore di compiutezza di un evento che viene valutato nel perdurare dei suoi risultati, escludendo così l'interpretazione ingressiva.

L'aspetto ingressivo si presenta solo con verbi durativi non-telici, sempre che nel contesto non sia presente un elemento che impedisca di spostare l'attenzione verso l'inizio dell'evento. Bisogna precisare, però, che non tutti i verbi durativi non-telici ammettono l'interpretazione ingressiva, per esempio verbi come *durare* e *abitare* ammettono un'interpretazione di questo tipo solo in particolari contesti, come il seguente:

(47) Finalmente Piero abitò in centro

1.2.1.2 Aspetto imperfettivo

Da ciò che è stato detto sulla perfettività nel paragrafo precedente è possibile definire come elemento caratterizzante dell'aspetto imperfettivo la referenza esplicita alla struttura temporale interna di un evento.

Come risulta dalla Figura 1, l'aspetto imperfettivo in italiano può essere suddiviso in tre ulteriori categorie aspettuali: progressivo, continuo e abituale.

1.2.1.2.1 Aspetto progressivo

L'aspetto progressivo si attiva quando il verbo fa riferimento ad un singolo istante di un processo in svolgimento.

In alcune lingue è obbligatoria la distinzione tra il valore semantico non progressivo e quello progressivo, espresso tramite apposite forme verbali. Questo è, per esempio, il caso dell'inglese in cui le due forme verbali non sono interscambiabili. In italiano, invece, è possibile sostituire la forma progressiva con un'altra forma senza alterare il valore semantico di progressività del verbo. Gli esempi seguenti possono, infatti, essere considerati sinonimi:

(48) In quel momento Marco stava leggendo

(49) In quel momento, Marco leggeva

Il valore semantico progressivo della frase (49) viene reso esplicito in quanto è possibile sostituire il verbo con la perifrasi progressiva “*stare + gerundio*”.

L'aspetto progressivo determina un punto di osservazione all'interno del processo, ovvero un istante di focalizzazione in cui il processo viene osservato dall'interno. Per esempio, nella frase seguente la proposizione dipendente individua un singolo istante di focalizzazione:

(50) Marco leggeva tranquillo quando improvvisamente ci fu un forte boato

Un'altra caratteristica dell'aspetto progressivo è l'indeterminatezza della prosecuzione dell'evento oltre l'istante di focalizzazione: sia che l'evento sia

passato, presente o futuro, la conoscenza sul suo reale svolgimento non viene mai considerata.

Per quanto riguarda il rapporto tra l'aspetto progressivo e gli avverbiali di tempo, è da sottolineare come verbi in questa accezione non siano compatibili con avverbiali del tipo "in/per X TEMPO":

(51) *Quando entrai, Marco leggeva per due ore

(52) *Quando entrai, Marco leggeva in due ore

La causa di questa incompatibilità sta nel fatto che questi avverbiali alludono alla conclusione dell'evento e ciò risulta in conflitto con la concezione del momento dell'accadimento come un intervallo aperto. Per questo stesso motivo, l'aspetto progressivo risulta incompatibile con gli avverbiali culminativi (53), delimitativi (54) e circoscriventi (55):

(53) *Quando entrai, Marco leggeva fino alle 2

(54) *Quando entrai, Marco leggeva dalle 3 alle 4

(55) *Quando entrai, Marco leggeva fra le 3 e le 4

L'aspetto progressivo è invece compatibile con gli avverbiali decorrenziali come, per esempio, "da X TEMPO", soltanto, però, relativamente ai verbi durativi. Nell'esempio seguente, l'avverbiale definisce l'intervallo dall'inizio dell'azione fino all'istante di focalizzazione:

(56) Quando entrai, Marco leggeva da due ore

1.2.1.2.2 Aspetto continuo

Dell'aspetto continuo esistono due varianti: durativa e iterativa. Un esempio della prima variante è riportato nella frase seguente:

(57) Per tutta la durata della conferenza, Marco annuiva all'oratore

In questo caso l'aspetto verbale non può essere abituale data l'unicità della situazione, né progressivo dato che l'imperfetto non può essere sostituito con una perifrasi progressiva.

Per quanto riguarda, invece, la variante iterativa:

(58) Marco entrava e usciva dall'aula in continuazione, tanto che fu richiamato dal professore

In questo esempio, può essere esclusa l'interpretazione progressiva in quanto si fa riferimento ad una serie di avvenimenti e non è possibile fissare un istante di focalizzazione. Non si può neppure avere un'interpretazione abituale in quanto bisogna che l'evento non sia unico ma si sia ripetuto in più occasioni diverse.

Riassumendo, si può quindi affermare che l'aspetto continuo appartiene all'ambito dell'imperfettività perché implica indeterminatezza sia riguardo la prosecuzione del processo sia riguardo il numero di interazioni, nel caso in cui l'accezione sia di tipo iterativo. Inoltre, questo aspetto non contempla un momento di focalizzazione proprio dell'aspetto progressivo né esprime una generalizzazione "gnomica" caratteristica, invece, dell'aspetto abituale (par. 1.2.1.2.3).

1.2.1.2.3 Aspetto abituale

Ogni situazione che può essere protratta sufficientemente nel tempo o che può essere iterata un numero sufficiente di volte in un periodo di tempo abbastanza lungo può essere intesa come abituale, ovvero come una proprietà caratteristica di un intero periodo e non come una proprietà accidentale di un momento. L'abitualità, dunque, è determinata da una combinazione di valori semantico-aspettuali, riguardanti eventi prolungati o iterati nel tempo.

(59) In quel periodo, Manuela prendeva l'autobus alle 7

Per definire l'aspetto abituale la conoscenza del contesto risulta spesso essenziale. Nel caso, infatti, togliessimo la locuzione avverbiale dalla frase precedente

potremmo avere un'interpretazione abituale quanto semelfattiva, riguardante cioè un singolo accadimento. Un test per identificare l'interpretazione abituale di un evento è quello di sostituire il verbo con la perifrasi “*essere solito* + infinito”.

L'abitudine, quindi, presupponendo regolari iterazioni di un evento è strettamente collegata con la nozione di iteratività ma non va confusa con essa, in primo luogo perché una mera ripetizione di una situazione non è sufficiente affinché questa possa essere concepita come abituale. Infatti, se una situazione viene ripetuta un numero limitato di volte può essere vista come un'unica situazione a cui ci si può riferire con una forma verbale perfettiva, sebbene abbia una struttura interna. In italiano, un evento di questo tipo può essere descritto come segue:

(60) Il professore si alzò in piedi, tossì 3 volte e disse...

In secondo luogo, una situazione può essere abituale senza necessariamente essere iterativa. Nell'esempio (61) la naturale interpretazione è che Marco abita a Roma per un intero periodo, senza interruzioni.

(61) Marco è solito vivere a Roma

Questi due differenti concetti di abitudine e iteratività possono essere definiti attraverso la relazione da una parte con la nozione di plurazionalità verbale e dall'altra con quella di imperfettività gnomica (Bertinetto e Lenci, 2011).

La nozione di plurazionalità è stata introdotta originariamente da Newman (1980) ed è stata utilizzata, tra gli altri, da Dressler (1968), Cusic (1981) e Xrakovskij (1997). Questa etichetta può essere utilizzata:

1. per indicare eventi suddivisibili in sotto-eventi che occorrono in una stessa situazione, come nell'esempio (62), in cui l'iteratività è una proprietà intrinseca dell'evento stesso:

(62) Ieri alle due del pomeriggio, i militari spararono per difendersi

2. per indicare uno stesso evento che si ripete in situazioni differenti, come nell'esempio (63):

(63) Emma ogni giorno giocava a tennis

Da notare che le due nozioni di plurazionalità non si escludono vicendevolmente ma possono essere combinate come nella frase seguente:

(64) I militari spararono ogni giorno per difendersi

La plurazionalità intesa come evento che si ripete in diverse situazioni può essere espressa utilizzando differenti strumenti linguistici: morfemi, affissi, avverbi o perifrasi verbali, costruzioni correlative. Un caso particolare di plurazionalità è la *reduplicazionalità*, che si manifesta quando un evento si ripete esattamente due volte: per esprimere ciò, in italiano e nelle lingue romanze in generale, vengono utilizzati morfemi particolari come i prefissi *re/ri*, (es. *riandare*).

Anche il contesto può, alcune volte, attraverso inferenze pragmatiche, suggerire plurazionalità, come nel caso seguente in cui la combinazione di un oggetto diretto plurale e la reciprocità verbale determinano questo tipo di interpretazione:

(65) Livio e Laura si scambiarono molte figurine

Un modo per distinguere l'abitudine dall'iteratività in uno stesso evento che si ripete in situazioni differenti è l'analisi del suo aspetto verbale (66a vs 66b).

(66) a. Negli ultimi mesi, Manuela ha preso spesso l'aereo

(66) b. Quando viveva a Trento, Manuela prendeva abitualmente l'aereo

Entrambi gli esempi precedenti riportano i medesimi eventi plurinazionali, espressi, però, in maniera differente. Questa differenza è messa in luce dall'aspetto verbale: nella prima frase viene utilizzato un aspetto perfettivo per evidenziare la relazione tra un individuo, un oggetto e un intervallo di tempo. Tale relazione mira essenzialmente a esprimere le abitudini del soggetto. Nella seconda

frase invece, l'uso di un aspetto imperfettivo permette di interpretare l'evento di viaggiare, come una proprietà dell'individuo durante un intervallo di tempo.

Per distinguere l'aspetto abituale da quello iterativo si possono analizzare quattro caratteristiche dell'aspetto verbale (Bertinetto e Lenci, 2011): (i) la specificazione delle interazioni, (ii) il dominio temporale, (iii) la localizzazione temporale, (iv) l'enumerabilità.

(i) *La specificazione delle iterazioni*

Caratteristica dell'abitudine è l'incompatibilità con le specificazioni numeriche; è quindi, essenziale che il numero di iterazioni di un evento resti indeterminato e l'indeterminatezza è un tratto distintivo dell'aspetto imperfettivo, infatti:

(67) a. Negli ultimi sei mesi, Manuela ha preso l'aereo 5 volte

(67) *b. Quando viveva a Trento, Manuela prendeva l'aereo 5 volte⁵

La presenza di avverbiali come *raramente* o *spesso* in correlazione con un aspetto verbale imperfettivo, come nell'esempio seguente, consente una lettura abituale della frase, in quanto questi avverbi non si riferiscono al numero di sotto-eventi ma piuttosto alla loro frequenza.

(68) Negli ultimi mesi, Manuela viaggiava raramente in aereo

(ii) *Il dominio temporale*

L'abitudine può occorrere con tutti i domini temporali, compreso il futuro nel passato, mentre l'iteratività non può essere ottenuta nel dominio del presente.

⁵ Da notare che è corretto dire "Quando viveva a Trento, Manuela prendeva l'aereo 5 volte al mese" perché in questo caso l'avverbiale specifica la frequenza dell'evento.

Abitualità

(69) L'anno scorso, l'assessore alla cultura sedeva sempre vicino al sindaco durante le occasioni ufficiali

(70) L'assessore alla cultura siede sempre vicino al sindaco durante le occasioni ufficiali

(71) L'assessore alla cultura siederà sempre vicino al sindaco durante le occasioni ufficiali

Iteratività

(72) Lo scorso anno, Veronica ha perso l'aereo tre volte

(73) *Veronica perde l'aereo tre volte

(74) Prevedo che Veronica, il prossimo anno, perderà l'aereo 3 volte

(iii) *La localizzazione temporale*

Gli avverbiali di tempo apparentemente sembrano avere la stessa funzione sia in contesti abituali che iterativi, indicando l'arco temporale in cui si localizza l'evento plurinazionale:

(75) L'anno scorso, Giovanni ha telefonato a sua madre 20 volte

(76) Durante lo scorso anno, Giovanni telefonava spesso a sua madre

Se gli avverbiali degli esempi precedenti fossero eliminati, la plurazionalità dell'evento verrebbe interpretata in relazione all'intera vita del soggetto. Detto questo però, appare chiaro dagli esempi seguenti che gli avverbiali di tempo in frasi iterative e abituali suggeriscono differenti oggetti del discorso. La stretta delimitazione temporale in (77) è accettabile e delimita esattamente l'evento in esso contenuto, mentre la (78) è semanticamente anomala, in quanto la frase non è di per sé autosufficiente ed infatti lo stesso avverbiale temporale non può delimitare l'argomento del discorso.

(77) Tra il 3 marzo e il 3 aprile, Veronica scrisse tre articoli

(78) ??Tra il 3 marzo e il 3 aprile, Veronica scriveva tre articoli

(79) Durante il semestre scorso, mi sono alzato ogni mattina alle 6

(80) Durante il semestre scorso, mi alzavo ogni mattina alle 6

Gi esempi precedenti mettono in evidenza come gli avverbiali di tempo vengano interpretati in maniera diversa secondo l'aspetto del verbo a cui sono associati: una delimitazione temporale stretta in frasi perfettive-iterative, una delimitazione temporale più vaga nelle frasi imperfettive-abituali.

La funzione di questi avverbiali è quindi quella di restringere la validità temporale di una situazione, a meno che questa sia sempre valida o sia valida per l'intera durata della vita del referente.

(iv) *Enumerabilità*

Ad uno stesso avverbiale di tempo si possono dare due interpretazioni differenti: secondo una lettura perfettiva-iterativa l'avverbiale *durante lo scorso anno*, in (81), delimita un ristretto periodo temporale nel quale il numero di telefonate di Giovanni alla madre può essere esattamente contato. Secondo una lettura imperfettiva-abituale, invece, l'avverbiale non delimita un ristretto periodo di tempo ma piuttosto definisce un periodo di riferimento nel quale viene rilevata la caratteristica di Giovanni di telefonare frequentemente o sporadicamente alla madre.

(81) Durante lo scorso anno Giovanni ha telefonato a sua madre (10 volte)
[iteratività]

(82) Durante lo scorso anno Giovanni telefonava spesso a sua madre [abitualità]

1.2.1.2.3.1 Imperfettività gnomica

Come detto nel paragrafo precedente, le frasi abituali a differenza delle iterative attribuiscono una proprietà ad un determinato referente. Questo le rende simili ad altri tipi di frasi che hanno la funzione di esprimere una generalizzazione di qualche tipo: si può fare riferimento a questa tipologia di frasi come alla classe dell' "imperfettività gnomica" (Bertinetto e Lenci, 2011). A questa classe afferiscono le seguenti tipologie di frasi: abituali, attitudinali, potenziali,

Individual-Level predicate, ovvero predicati che rimangono validi per l'intera esistenza di un individuo (Kratzer, 1995), generiche. Nei linguaggi che presentano esplicite marche aspettuali, questi tipi di frase vengono espressi attraverso l'uso di forme verbali imperfettive.

(83) In quel periodo, Laura si arrabbiava facilmente con le sue colleghe [abituale]

(84) Enrico fuma la pipa [attitudinale]

(85) Stefano parla francese [potenziale]

(86) Gianpaolo è napoletano [Individual-Level predicate]

(87) I gatti hanno la coda [generico]

Non tutte le tipologie di frasi precedenti comportano plurazionalità. Su questo punto le abituali e le generiche sono molto diverse tra loro. Le attitudinali e le potenziali, invece, sono più simili, in quanto entrambe, per poter essere classificate come esprimenti le qualità di un dato referente, devono prevedere un numero, seppur minimo, di ripetizioni della situazione.

Le generiche e i predicati individuali non denotano solo semplici stati ma anche stati permanenti, che persistono cioè oltre l'esistenza del proprio referente, in (88), per esempio, la qualità permanente è delimitata al periodo di esistenza dei *Tyrannosaurus rex*. Al contrario, le abituali denotano spesso eventi piuttosto che stati, come dimostra la loro compatibilità con avverbi agentivi come *deliberatamente* (89):

(88) Il *Tyrannosaurus rex* era un carnivoro [generico]

(89) Gianni lasciava sempre deliberatamente le chiavi nella porta [abituale]

Una caratteristica che oppone le abituali e le generiche ma anche le attitudinali e le potenziali è l'effetto che provoca su di esse la passivizzazione. Tale conversione è possibile per le frasi abituali ma non per generiche, le potenziali e le attitudinali, in quanto questi tre ultimi tipi in forma passiva danno luogo a incongruenze semantiche. Per esempio, in (92) *fumare la pipa* è un proprietà di Enrico e non una proprietà della pipa *farsi fumare* da Enrico.

(90) Le chiavi sono sempre lasciate deliberatamente nella porta da Gianni

(91) *La coda è posseduta dai gatti

(92) *La pipa è fumata da Enrico

(93) *Il francese è parlato da Stefano

Una caratteristica interessante delle attitudinali e potenziali consiste nella loro natura azionale. Queste tipologie di frasi presentano, infatti, in prevalenza predicati eventivi che però possono essere trasformati in stati attraverso il fenomeno della coercizione azionale (vedi cap. 1.3.2). Per esempio, *fumare* è un predicato eventivo in molti contesti ma in (92) lo stesso predicato può avere lettura stativa, così come avviene per il predicato *parlare* in (93).

Il significato lessicale di questi predicati rimane quindi eventivo ma quando rappresentano una proprietà di un dato referente possono assumere un'interpretazione stativa.

Non è possibile generalizzare, però, dicendo che la statività è una caratteristica dell'abitudine: è una condizione necessaria ma non sufficiente. Se fosse condizione sufficiente, infatti, tutti i predicati stativi implementerebbero situazioni abituali ma questo non avviene come dimostra l'esempio (94), che rappresenta una pluralità di situazioni contingenti. Inoltre, predicati stativi possono comparire in contesti perfettivi, chiaramente incompatibili con l'abitudine (95). Infine, perché un predicato stativo, in situazioni contingenti non permanenti, compaia in un contesto abituale è necessario che sia accompagnato esplicitamente da avverbi, come, per esempio *spesso* in (96).

(94) Il libro è sul tavolo

(95) Il libro è stato sul tavolo tutto il giorno

(96) L'anno scorso, Gianni è stato spesso in vacanza

Per concludere, gli stessi predicati possono avere un significato sia stativo contingente che stativo permanente secondo il contesto in cui sono utilizzati:

(97) Il teatro al momento contiene trecento persone [stativo contingente]

(98) Il teatro contiene trecento persone [stativo permanente]

1.3. Interazione tra azione e aspetto verbale

Come già evidenziato nei capitoli precedenti, fra azione e aspetto nel verbo vi sono alcune relazioni. Un esempio si riscontra nel caso dei verbi telici: la telicità viene, infatti, evidenziata dalla commutazione tra tempi perfettivi ed imperfettivi. I verbi telici quando sono coniugati secondo il paradigma di un tempo imperfettivo, non consentono di derivare alcuna implicazione sulla conclusione di un processo, come si evidenzia nell'esempio seguente, in cui il verbo risultativo è coniugato all'imperfetto:

(99) Gianpaolo cantava una canzone di De Andrè

In questo caso non si può dedurre se Gianpaolo abbia o meno portato a termine il processo. Mentre invece, la conclusione del processo risulta chiaramente in (100) in cui compare un tempo perfettivo, il perfetto semplice, che definisce senza alcun dubbio la conclusione del processo del cantare:

(100) Gianpaolo cantò una canzone di De Andrè

In altri casi la commutazione tra tempi perfettivi e imperfettivi evidenzia una struttura polisemica, dei tempi verbali (1.1.3.1). Come già detto in precedenza, si parla in questi casi di opposizioni infra-lessicali ovvero di opposizioni azionali all'interno di uno stesso verbo. Questo, infatti, nel medesimo contesto può possedere un significato statico, ovvero durativo, come in (a) oppure un significato dinamico e non durativo (b), a seconda del tempo utilizzato:

(101) a. Manuela indossava un vestito rosso

b. Manuela indossò un vestito rosso

(102) a. Luca sedeva vicino a Flora

b. Luca sedette vicino a Flora

(103) a. Laura calzava degli stivali

b. Laura calzò degli stivali

Bisogna, però, precisare che è comunque sempre possibile recuperare il senso dinamico di tempi imperfettivi e il senso statico di tempi perfettivi. Viene recuperato il senso dinamico se il verbo si trova in una proposizione dipendente introdotta da congiunzioni come *mentre* o *ogniqualevolta* che introducono rispettivamente le accezioni progressiva e abituale:

(104) Mentre/Ogniqualevolta Luca sedeva vicino a Flora, gli sembrava che il mondo fosse migliore.

Il senso statico viene recuperato, invece, quando un tempo perfettivo viene modificato da un avverbiale di tempo:

(105) Per tutta la sera, Luca sedette vicino a Flora

Esistono, comunque, casi in cui l'opposizione infra-lessicale è più rigida e la staticità o la dinamicità di un verbo dipendono dal suo soggetto:

(106) Il confine correva lungo la cresta della montagna

(107) La finestra dava sul giardino

In frasi come le precedenti, i tempi perfettivi non sono ammessi:

(108) *Il confine corse lungo la cresta della montagna

(109) *La finestra dette sul giardino

Vi sono, infine, verbi che con certi soggetti inanimati assumono un'accezione statica privilegiando i tempi imperfettivi anche se non rifiutano totalmente i tempi perfettivi (110a vs 110b):

(110) a. Il castello cadeva a picco sul mare (statico)

(110) b. Il castello cadde a picco sul mare (dinamico)

Riepilogando, quindi, si può affermare che esistono tre tipi di verbi:

1. verbi che prediligono un'accezione statica o dinamica a secondo del tempo ma possono cambiare la propria valenza a seconda del contesto;
2. verbi che con soggetti inanimati assumono un'accezione statica e ammettono soltanto tempi imperfettivi;
3. verbi che con soggetti inanimati assumono un'accezione statica e privilegiano tempi imperfettivi anche se non rifiutano del tutto i tempi perfettivi.

Si può dire, in conclusione, che c'è una certa relazione tra “evento” durativo e aspetto imperfettivo e “evento” non durativo e aspetto perfettivo. Questa relazione, però, indica solo una tendenza: normalmente, infatti, un verbo può trovarsi coniugato sia con tempi perfettivi che imperfettivi senza cambiare in maniera rilevante il proprio significato.

Capitolo II

I corpora e l'elaborazione dell'informazione temporale

Un corpus è una collezione finita di testi, selezionati secondo specifici criteri allo scopo di mostrare la reale distribuzione di un certo dato linguistico all'interno di una determinata varietà linguistica e di facilitare le analisi linguistiche.

Il ruolo del computer nell'elaborazione dei dati è diventato così cruciale che ormai il termine stesso di “corpus” è sinonimo di “corpus elettronico” come dimostrano anche le due definizioni riportate di seguito.

Crystal (1992) definisce un corpus come “a collection of linguistic data, either compiled as written texts or as a transcription of recorded speech. The main purpose of a corpus is to verify a hypothesis about language – for example, to determine how the usage of a particular sound, word, or syntactic construction varies [...]. Corpus linguistics deals with the principles and practice of using corpora in language study [...]. A computer corpus is a large body of machine-readable texts.”

In (Lenci et al., 2005) il corpus viene definito come “una collezione di testi selezionati e organizzati in maniera tale da soddisfare specifici criteri che li rendono funzionali per le analisi linguistiche. I corpora testuali rappresentano la principale (anche se non esclusiva) fonte di dati per la Linguistica Computazionale. [. . .] Il ruolo del computer nell'uso dei corpora è così cruciale che ormai il termine stesso di corpus è diventato di fatto sinonimo di corpus elettronico, ovvero corpus di testi in formato digitale.”

Sulla base delle precedenti definizioni, le caratteristiche di un corpus in linguistica computazionale possono essere riassunte, quindi, nei seguenti punti:

1. grande collezione di testi in formato elettronico;
2. compilata ad hoc per l'analisi linguistica;

3. progettata per essere (il più possibile) rappresentativa di una particolare varietà linguistica.

In questo contesto l'annotazione si configura come l'attività consistente nell'aggiungere informazione di tipo metalinguistico alle diverse porzioni di un testo: una forma di codifica che nella pratica si traduce nell'attribuzione di una etichetta o tag ad una unità testuale.

Qualunque aspetto dell'analisi linguistica, per esempio, caratteristiche fonologiche e fonetiche, morfologiche, sintattiche, semantiche, pragmatiche ecc... può essere etichettato.

2.2 Annotazione di un corpus

L'annotazione dei corpora è un campo di ricerca della linguistica computazionale diventato importante negli ultimi 20 anni.

Esistono quattro criteri per giudicare il valore di un corpus: (i) la sua grandezza; (ii) la diversità dei tipi di testi contenuti; (iii) la cura con la quale è stato prodotto; (iv) il valore aggiunto determinato dall'annotazione con la quale il corpus è etichettato e che costituisce la conoscenza più esplicita sul testo (Garside et al., 1997).

Il Part of Speech (PoS) tagging, per esempio, è una delle forme più comuni di annotazione testuale consistente nell'associare attraverso un apposito tag una parte del discorso a ciascuna parola del testo, come nell'esempio seguente:

Il/A

gatto/N

miagola/V

Generalmente è possibile definire l'annotazione come interpretativa e metalinguistica.

Interpretativa in quanto l'annotazione dipende dal livello di conoscenza del testo da parte dell'annotatore. Per esempio, l'annotatore potrebbe chiedersi se la parola *futuro* in "il suo futuro sposo" è un nome o un aggettivo e un volta deciso il livello

di dettaglio che deve essere raggiunto inserire o meno nel testo quest'informazione.

L'annotazione è metalinguistica nel senso che fornisce informazioni aggiuntive sul linguaggio rispetto a quelle espresse attraverso il semplice testo puro non annotato. Questa distinzione tra annotazione interpretativa e metalinguistica è facile da rilevare in testi scritti, più difficile per il linguaggio parlato trascritto per il quale il trascrittore stesso deve necessariamente attivare un sistema di interpretazione del discorso.

Esistono tre metodi di annotazione:

- 1) annotazione manuale, di tipo tradizionale, che è svolta da persone che appongono specifiche etichette a porzioni di testo sulla base di valutazioni metalinguistiche più o meno condivise e standardizzate;
- 2) annotazione automatica, che procede senza l'intervento umano attraverso applicazioni di elaborazione del linguaggio naturale;
- 3) annotazione semi-automatica, costituita da una prima fase automatica e una successiva fase di revisione manuale.

L'annotazione dei corpora è importante per diverse ragioni la prima delle quali è sicuramente la sua utilità in prospettiva di un processo automatico di estrazione di informazione da testi. Sfruttando, infatti, l'etichettatura testuale è possibile compiere ricerche specifiche sui corpora.

Un altro elemento che rende importante la fase di annotazione è costituito dalla possibilità di riutilizzo del testo etichettato. L'annotazione di un corpus, infatti, è un'attività dispendiosa in termini di costi e tempo ma, una volta costruito, il corpus diventa una risorsa preziosa e soprattutto riutilizzabile per diversi scopi di ricerca.

Un corpus annotato, inoltre, è multifunzionale. L'annotazione, infatti, ha spesso differenti scopi e applicazioni. Il PoS tagging, per esempio, può essere utilizzato sia in ambito lessicografico (indica la classe grammaticale di appartenenza di una determinata parola), sia della sintesi del parlato (il nome "pesca", per esempio, è

pronunciato diversamente a seconda che si tratti della voce del verbo “pescare” oppure del nome del frutto).

Gli usi che si fanno dei corpora annotati sono essenzialmente due:

- 1) vengono utilizzati per l'addestramento di modelli di machine learning supervisionati;
- 2) nella valutazione di sistemi di NLP possono essere utilizzati come gold standard.

Vi sono dei requisiti formali che garantiscono la possibilità di riutilizzo e certificano la qualità di un corpus: innanzitutto, la separazione e l'autonomia del materiale grezzo del corpus dalle codifiche e annotazioni linguistiche; quindi, la disponibilità di una documentazione riguardante: (i) lo schema di annotazione; (ii) come e da chi il processo di annotazione è stato svolto; (iii) la qualità dell'annotazione. La codifica e l'annotazione, inoltre, devono essere conformi ad uno standard e vi deve essere un'indicazione esplicita dei principi di annotazione (regole di attribuzione di un'etichetta ad un token). Anche i criteri linguistici alla base dell'etichettatura devono essere generalmente condivisi e deve essere garantita la possibilità di eseguire specifiche ricerche sulle etichette. Infine, è necessario che la fruibilità del corpus e dell'annotazione sia indipendente da specifici sistemi operativi e dalle caratteristiche dell'hardware.

2.3 Verso uno standard internazionale per l'annotazione di risorse linguistiche

Le risorse linguistiche sono costituite da insiemi di dati in formato elettronico il cui scopo è quello di supportare la ricerca e le applicazioni di elaborazione del linguaggio naturale. Tipicamente i dati vengono annotati con varie informazioni linguistiche come le categorie morfo-sintattiche, le strutture sintattiche o del discorso, l'informazione circa la coreferenza ecc.

Soprattutto negli ultimi 15-20 anni, ricercatori e sviluppatori hanno iniziato ad avvertire il bisogno di definire pratiche e formati comuni per le risorse

linguistiche che costituiscono i dati di riferimento per l'addestramento dei modelli statistici del linguaggio.

Alcuni fondamentali principi di rappresentazione sono stati adottati in maniera generalizzata, per esempio l'uso di un'annotazione stand-off (Ide e Priest-Dorman, 1996), l'utilizzo di XML, l'impiego di particolari grafici di etichettatura (Bird e Liberman, 2001). Soprattutto, però, i formati di annotazione tendono ancora a differenziarsi da risorsa a risorsa, spesso per soddisfare i vincoli imposti dai software.

Sebbene, dunque, nella pratica vi sia un consenso diffuso circa alcuni principi fondamentali di rappresentazione per le risorse linguistiche, non sono stati ancora stabiliti, però, standard definitivi. Inoltre, i progressi nelle tecnologie insieme con l'emergere di standard *web-based*⁶ hanno ridefinito molte delle idee circa le metodologie con le quali le risorse devono essere raccolte ed i modi con i quali vi si deve accedere.

La continua evoluzione delle tecnologie e dei domini applicativi induce, quindi, a ipotizzare che sia ancora prematuro tentare una standardizzazione delle risorse linguistiche; al contrario, però, la crescita del web, l'esplosione del numero di documenti elettronici da elaborare e mantenere all'interno soprattutto dei settori industriali, ha creato un immediato e urgente bisogno di componenti generici, ovvero non specifici per una particolare tipologia di linguaggio né per una particolare lingua, per l'elaborazione del linguaggio in grado, per esempio, di indicizzare e classificare documenti, di sviluppare sistemi per l'estrazione di informazione da testi o applicazioni di *summarization* e *topic detection*, in ambienti mono e multi-lingua.

Per costruire un ambiente informativo globale, quindi, è necessario per le nuove generazioni di applicazioni di elaborazione del linguaggio che i dati linguistici siano riutilizzabili e le risorse standardizzate possano essere usate in maniera flessibile per diversi scopi di ricerca.

⁶ Per standard *web-based* si intendono le specifiche e buone pratiche utilizzate per la costruzione di applicazioni accessibili via web per mezzo di un network, come ad esempio una intranet o Internet. Questo è il caso di strumenti di annotazione accessibili tramite web come per esempio il tool per l'annotazione di testi BAT (per un approfondimento su BAT si veda: http://www.lrecconf.org/proceedings/lrec2010/pdf/740_Paper.pdf).

Al fine di creare una nuova infrastruttura per lo sviluppo e l'utilizzo di risorse linguistiche, l'International Organization for Standardization (ISO) ha deciso di formare un sotto-comitato (SC4)⁷ del più ampio Technical Committee 37 (TC37, *Terminology and Other Language Resources*) che si occupa della gestione delle risorse linguistiche. Lo scopo di ISO/TC 37/SC 4 è quello di definire standard internazionali e linee guida per la gestione da parte dei sistemi applicativi delle risorse linguistiche già esistenti.

In particolare, la commissione si è data i seguenti obiettivi:

- fornire mezzi per l'uso e il riuso dei dati linguistici, a tutti i livelli di descrizione, dal markup del livello superficiale delle risorse all'elaborazione multi-livello dei risultati;
- facilitare il mantenimento di un coerente ciclo di vita di un documento attraverso vari livelli di elaborazione, così da poter arricchire i dati esistenti con nuove informazioni e incrementare la costruzione di sistemi di elaborazione del linguaggio.

In conclusione, il comitato ha il compito di sviluppare principi e metodi per la creazione, la codifica, l'elaborazione e la gestione di varie risorse linguistiche come corpora di testi scritti, di parlato, dizionari e schemi di classificazione. Il focus del lavoro è posto sulla definizione del markup, delle modalità di interscambio dei dati e della valutazione. Gli standard definiti dalla ISO/TC 37/SC 4 dovrebbero essere utilizzati per migliorare la gestione dell'informazione all'interno di organizzazioni industriali e centri di ricerca al fine di accrescere l'efficienza della comunicazione linguistica supportata dalle macchine.

2.3.1 L'iniziativa ISO e gli altri progetti di standardizzazione

Come già in parte anticipato nel precedente paragrafo, negli ultimi 20 anni sono stati portati avanti numerosi progetti ed iniziative volti ad individuare una base di standardizzazione della rappresentazione e dell'annotazione delle risorse, come per esempio la Text Encoding Initiative (TEI)⁸, il Corpus Encoding Standard

⁷ <http://www.tc37sc4.org>

⁸ <http://www.tei-c.org>

(CES e XCES)⁹, l' Expert Advisory Group on Language Engineering Standards (EAGLES)¹⁰ e l'International Standard for Language Engineering (ISLE). Allo stesso modo sono state create piattaforme software per la realizzazione, l'annotazione e l'utilizzo delle risorse, come MULTEXT¹¹, LT XML¹², GATE¹³, NITE¹⁴, ATLAS¹⁵.

Molti di questi progetti attualmente si occupano di ciò che riguarda la dimensione multipla della rappresentazione delle risorse linguistiche, inclusi i seguenti aspetti:

- la definizione dei formati di rappresentazione come XML, dei grafici di annotazione o di particolari formati di database;
- l'annotazione dei contenuti, incluse categorie di annotazione per i fenomeni linguistici, come l'aspetto verbale o la modalità, e i valori che possono essere associati a queste categorie;
- i principi delle architetture generali per le risorse linguistiche.

L'iniziativa ISO, in questo contesto, cerca di recuperare i risultati ottenuti dai precedenti progetti riprendendo elementi che possono essere estesi o modificati e traendo vantaggio dalla convergenza incrementale di opinioni su particolari aspetti del processo di standardizzazione. Per questa ragione, il gruppo ISO ha stabilito collaborazioni con i maggiori gruppi di standardizzazione a livello internazionale al fine di assicurarsi che lo sviluppo di standard per la gestione delle risorse linguistiche incorpori e rifletta le pratiche già esistenti.

2.3.2 Il Linguistic Annotation Framework

Il Linguistic Annotation Framework (LAF) è stato sviluppato dal comitato ISO/TC 37/SC 4, come strumento di base per armonizzare le risorse linguistiche esistenti e per crearne di nuove. Lo scopo di LAF è quello di sviluppare infrastrutture standard per la rappresentazione delle risorse e per loro marcatura tenendo conto che l'annotazione dei dati linguistici comprende diversi livelli di

⁹ <http://www.xml-ces.org>

¹⁰ <http://www.ilc.cnr.it/EAGLES96/home.html>

¹¹ <http://www.lpl.univ-aix.fr/projects/multext>

¹² <http://www.ltg.ed.ac.uk/software/xml>

¹³ <http://gate.ac.uk/>

¹⁴ <http://www.dfki.de/nite/main.html>

¹⁵ <http://www.nist.gov/speech/atlas/>

etichettatura e che l'etichettatura dei livelli più bassi tipicamente costituisce l'input per l'annotazione dei livelli più alti, in un processo di complessità incrementale.

Attualmente molte risorse annotate sono costituite da entità atomiche, non ulteriormente decomposte in sotto-entità, che si riferiscono ad un unico livello di annotazione e sono usate principalmente per l'addestramento di software come per esempio accade nel campo della linguistica dei corpora o in lessicografia. Nel contesto del Web semantico, però, le annotazioni per i vari livelli linguistici dovranno necessariamente essere accessibili in formati per il web in modo tale da poter essere utilizzati da agenti software in grado di compiere inferenze e recuperare informazioni da testi.

Il LAF, dunque, non solo si basa sulle tecnologie come RDF o OWL per la rappresentazione delle annotazioni ma anche sull'utilizzo di livelli di annotazione che sono allo stesso tempo separabili, in modo tale che i software di analisi possano accedere ai soli livelli richiesti per i loro scopi, integrabili, così che due o più tipi di annotazione possano essere combinati, e dinamici, nel senso che sia possibile modificare o aggiungere nuova informazione.

LAF consente di utilizzare ogni formato di annotazione coerente con le caratteristiche dei modelli di dati che possono essere usati per definire un formato di interscambio. Ciò suggerisce uno scenario futuro nel quale gli annotatori potranno creare ed editare annotazioni in un formato proprietario, convertirle attraverso appositi strumenti nel formato di interscambio o di elaborazione e, se necessario, riportarle al formato originale.

2.4 Strumenti di annotazione e di estrazione di informazione temporale da testi secondo le specifiche TimeML

L'utilità di sistemi di elaborazione temporale risulta evidente soprattutto se vengono presi in considerazione gli importanti e recenti sviluppi delle applicazioni di Natural Language Processing (NLP) come il Question Answering (Q/A), l'Information Extraction e la Summarization. Per quanto riguarda il Q/A, per esempio, una domanda plausibile da porre ad un sistema potrebbe essere

quella di individuare il nome del Presidente degli Stati Uniti nel gennaio 1983 avendo come dati di riferimento articoli di giornale che parlano dell'attività del Presidente dal 1980 al 1985. In questo caso, come in altri simili, può essere utile l'integrazione del sistema di Q/A con altri strumenti di elaborazione capaci di estrarre inferenze temporali da testi in modo da fornire la risposta corretta anche se questa non è esplicitamente presente.

Anche in ambito medico, i sistemi per l'elaborazione temporale possono essere sfruttati proficuamente. Per quanto riguarda la registrazione della storia clinica dei pazienti, infatti, i medici generalmente non riportano le informazioni in ordine cronologico. L'estrazione e la ricostruzione di una struttura temporale delle informazioni mediche relative ai pazienti può essere, invece, estremamente utile per delinearne facilmente la storia clinica ed effettuare diagnosi.

Infine, per persone che hanno problemi nella lettura e nella comprensione dei testi, come per esempio i pazienti affetti da dislessia o i parlanti non nativi, la ricostruzione della struttura temporale di un documento può essere utile per seguire con più facilità l'evolversi degli eventi in esso descritti.

2.4.1 Strumenti per l'elaborazione temporale in TempEval

Recentemente, gli strumenti di elaborazione temporale sono stati in gran parte utilizzati per identificare automaticamente, seguendo le specifiche dello standard di annotazione ISO TimeML (per un approfondimento su questo tipo di annotazione si rimanda al cap. 3 di questa tesi), le espressioni temporali, gli eventi e le loro relazioni presenti nei testi.

A questo proposito un'iniziativa interessante è costituita da TempEval (Verhagen et al., 2007; Verhagen et al., 2009). Si tratta di un task di elaborazione temporale inserito all'interno della più vasta campagna di valutazione SemEval¹⁶.

In particolare, la prima edizione dei TempEval, nel 2007, consisteva in un esercizio di valutazione per il riconoscimento e l'ordinamento di espressioni temporali. La seconda edizione si è svolta nel 2009 ed è risultata più elaborata rispetto alla precedente in quanto: (i) si è trattato di un task multilingue; (ii) c'è

¹⁶ http://aclweb.org/aclwiki/index.php?title=SemEval_Portal

stato un incremento del numero degli elementi temporali e dei rispettivi attributi da riconoscere e annotare.

In particolare, in TempEval-2 sono state prese in considerazione 5 lingue per ognuna delle quali un gruppo di organizzatori ha preparato indipendentemente i dati per l'esperimento. Le specifiche di annotazione e le linee guida, invece, sono state sviluppate in collaborazione tra i vari gruppi, utilizzando come punto di riferimento le linee guida TimeML sviluppate per l'inglese (per un approfondimento su queste linee guida si veda il cap. 3 di questa tesi).

L'annotazione delle risorse linguistiche per questo task è stata svolta in due fasi: una prima fase in cui due annotatori hanno etichettato separatamente tutti i documenti, una seconda in cui i casi di discordanza sono stati risolti da un terzo annotatore definito giudice.

Per l'annotazione è stato utilizzato BAT - the *Brandeis Annotation Tool* (Verhagen, 2010), un generico strumento di annotazione *web based* che consente di destrutturare task complessi di annotazione temporale in sottotask più semplici.

Otto gruppi hanno partecipato al task di TempEval-2, sottomettendo i risultati di 18 sistemi.

Per l'inglese, i due sistemi che hanno ottenuto migliori prestazioni sono stati TRIP-TRIOS (Uzzaman e Allen, 2010) e TIPSem (Llorens et al., 2010).

TRIP-TRIOS

Il sistema TRIP-TRIOS sviluppato dall'Università di Rochester utilizza un approccio ibrido a metà tra l'uso di regole linguistiche predefinite e sistemi di apprendimento automatico.

Il parser semantico TRIPS (Allen et al., 2008) è stato utilizzato per produrre le strutture logiche del testo che, in larga parte, sono state usate in processi di elaborazione statistica come il PoS tagging, l'etichettatura dei costituenti, l'interpretazione di parole sconosciute attraverso l'utilizzo di WordNet e il riconoscimento di entità nominate. Quindi, è stato utilizzato un insieme di regole manualmente definite per estrarre eventi ed espressioni temporali dalle forme

logiche prodotte dal parser. In parallelo, attraverso strumenti di classificazione e di apprendimento automatico, sono stati classificati gli eventi, identificati i valori dei loro attributi e classificate le relazioni temporali. Il sistema risultante dall'utilizzo del parser TRIPS, dei sistemi di estrazione degli attributi e di classificazione prende il nome di TRIOS.

Questo sistema ibrido combina, quindi, tecniche di elaborazione del testo indipendenti dal dominio di applicazione con strumenti di apprendimento automatico capaci di estrarre una quantità significativa di informazione temporale dai documenti. Da notare, però, che il sistema risulta dipendente dai classificatori automatici che a loro volta dipendono dalla disponibilità di corpora per l'addestramento.

Gli scopi principali del progetto sono due e possono essere riassunti come segue: (i) creare un sistema per l'estrazione di informazione temporale da testi indipendente dal dominio applicativo; (ii) utilizzare TRIOS come strumento semi-automatico per raccogliere ed annotare risorse linguistiche secondo lo schema TimeML.

TIPSem

TIPSem (the Temporal Information Processing based on Semantic information) è un sistema per l'estrazione di informazione temporale da testi in inglese e spagnolo, sviluppato dall'Università di Alicante.

Il sistema si focalizza sulla etichettatura dei ruoli semantici: questi identificano tutti i costituenti di ciascun predicato, definendone argomenti e aggiunti.

Il sistema è stato costruito secondo un approccio *data-driven*, sfruttando i Conditional Random Fields (CRF), l'efficiente tecnica di Machine Learning per l'annotazione di sequenze supervisionate di dati (Lafferty et al., 2001) al fine di costruire modelli di apprendimento a partire da risorse di addestramento. I CRF sono stati usati in considerazione del fatto che gli approcci *data-driven* hanno ottenuto buoni risultati in molti task di NLP e sono stati utilizzati con successo nel risolvere problemi di etichettatura come quelli proposti nella campagna di valutazione TempEval-2.

2.4.2 Altri strumenti per l'annotazione e l'estrazione di informazione temporale

TARSQI Toolkit

La Brandeis University ha sviluppato il TARSQI Toolkit (TTK)¹⁷, un sistema modulare per l'annotazione automatica di espressioni temporali, eventi e loro relazioni in testi in linguaggio naturale. Le componenti di TTK utilizzano TimeML (Pustejovsky et al., 2003; Pustejovsky et al., 2005) come linguaggio di annotazione. In particolare, TTK è costituito da un modulo che compie una pre-elaborazione sul testo puro utilizzando approcci standard per la tokenizzazione, il PoS tagging e il chunking. Per quanto riguarda le espressioni temporali, viene utilizzato il tagger GUTime per riconoscerle e normalizzarle. Quindi il tool Evita (Saurì et al., 2005) è utilizzato per il riconoscimento degli eventi e per l'analisi delle loro caratteristiche grammaticali come l'aspetto e il tempo verbale.

TTK include anche un modulo per l'identificazione di potenziali conflitti tra le relazioni temporali attraverso la costruzione automatica del grafo delle relazioni presenti in un documento.

Un sistema per l'annotazione di testi secondo lo schema TimeML per il francese

Un altro sistema per il riconoscimento e l'annotazione di eventi ed espressioni temporali secondo le specifiche TimeML in testi di lingua francese è stato sviluppato dall'Università Paris Diderot (Bittar, 2009).

Il sistema è costituito da due moduli: un primo tagger per l'annotazione delle espressioni temporali; un secondo per l'annotazione degli eventi.

Il tagger per le espressioni temporali è basato su una successione di automi a stati finiti a larga copertura che applicati al testo puro riconoscono date, ore, durate e intervalli temporali e annotano questi elementi con l'apposito tag definito nello schema di annotazione TimeML.

Il sistema di riconoscimento degli eventi prende in input un insieme di testi su cui precedentemente sono stati applicati un sistema di PoS tagging e un parser per l'analisi morfologica e sintattica. In particolare, questo tagger consiste in due

¹⁷ <http://timeml.org/site/tarsqi/toolkit/>

principali livelli di elaborazione: un primo livello di analisi lessicale per l'individuazione di nomi e verbi; un secondo livello di elaborazione contestuale consistente nell'applicazione di euristiche per il riconoscimento degli eventi e per la loro classificazione.

Da testi in italiano a documenti TimeML utilizzando un parser a dipendenze

Per quanto riguarda l'italiano, l'Università di Torino in collaborazione con il CNR-ILC (Robaldo et al., 2011), ha sviluppato un prototipo di sistema che a partire da testo puro crea documenti xml annotati secondo lo schema TimeML.

Innanzitutto, il sistema prevede che sia effettuato un parsing del testo attraverso il parser a dipendenze TULE, sviluppato dall'Università di Torino e scaricabile gratuitamente dal web¹⁸.

Il testo così elaborato viene dato in input ad un programma scritto in Java che esplora gli alberi di dipendenza costruiti dal parser TULE e che, per ogni nodo, utilizzando un insieme di regole predefinite, decide se si tratta di un evento o di un'espressione temporale e nel caso di giudizio positivo li annota.

Le regole sono molto semplici e consentono di valutare se i nodi dell'albero sintattico appartengono ad una lista precostituita di parole e locuzioni che soddisfano particolari vincoli. Questa lista è stata creata a partire dal corpus La Repubblica (Baroni et al., 2004), sulla base di documenti annotati usati per un'annotazione sperimentale di testi secondo lo standard TimeML adattato alla lingua italiana. In conclusione, le regole analizzano semplicemente il contenuto di un nodo e dei suoi più prossimi modificatori controllando se contengono o meno parole chiave.

2.5 Corpora annotati con informazione inerente azione e aspetto verbale

Da una ricerca svolta prevalentemente attraverso internet, sembra non siano molti i corpora in cui viene annotata l'informazione circa l'azione e l'aspetto verbale.

¹⁸ <http://www.tule.di.unito.it/>

Per ciò che concerne i corpora riguardanti l'azione verbale, un'analisi dei predicati nella prospettiva di una categorizzazione vendleriana è stata svolta all'interno del progetto SenSem¹⁹ (Vázquez et al., 2005) il cui scopo era quello di creare una base di dati che riflettesse le caratteristiche sintattiche e semantiche delle spagnolo. Il corpus di riferimento era costituito da circa un milione di parole organizzate in frasi ciascuna delle quali contenente una delle 250 forme verbali più frequenti in spagnolo. Alla fine del processo di creazione della risorsa ad ogni forma verbale corrispondevano circa 100 frasi. Queste frasi sono state selezionate casualmente da un corpus di circa 13 milioni di parole estratte dalle versioni on-line di due quotidiani spagnoli: *El Periodico* e *La Vanguardia*.

Seguendo l'approccio tradizionale nel campo della ricerca sull'azione verbale e sulla struttura eventiva (Comrie, 1976; Vendler, 1957; Pustejovsky, 1995), sono state prese in considerazione tre classi azionali: eventi, processi e stati. Per ogni entrata lessicale presente nella base di dati è stata indicata, quindi, la classe azionale di appartenenza. In particolare, dopo aver scelto per ogni verbo il significato corretto gli sono state automaticamente associate le informazioni relative alle categorie azionali. Gli annotatori successivamente hanno modificato l'etichettatura nei casi in cui gli elementi contestuali modificavano la classe azionale assegnata ai verbi del corpus.

Una classificazione in base alle categorie azionali è stata effettuata anche da uno studio condotto presso il Dipartimento di Inglese della Tamkang University a supporto della cosiddetta *Aspect Hypothesis*, cioè della influenza della classi azionali su chi apprende una nuova lingua quando questa fa uso di marche tempo-aspettuali. Lo scopo di questo studio è quello di usare dati recuperati da un corpus elettronico contenente un milione circa di parole e prodotto da nativi di Taiwan che apprendono l'inglese come seconda lingua. I dati raccolti nel corpus sono stati recuperati attraverso la piattaforma di scrittura on-line IWiLL (Intelligent Web-based Interactive Language Learning) (Wible et al., 2001). Attraverso questa piattaforma gli insegnanti hanno assegnato compiti scritti agli allievi. Tutte le parole prodotte dagli studenti nello svolgimento dei compiti sono state poi

¹⁹ Databank Sentential Semantics: "Creación de una Base de Datos de Semántica Oracional". MCyT (BFF2003- 06456).

raccolte nell'English Taiwan Learner Corpus (TLC). Lo studio si è focalizzato sulle marche tempo-aspettuali usate in 24 verbi del TLC (tali verbi sono stati scelti in quanto rappresentati i predicati generalmente utilizzati come esempio per introdurre le categorie vendleriane) classificati secondo la suddivisione azionale di Vendler: *states* (4271), *activities* (248), *accomplishments* (266) e *achievements* (1416).

L'analisi dei verbi appartenenti a ciascuna categoria ha dimostrato la validità dell'*Aspect Hypothesis* in quanto è risultata evidente la relazione tra la categoria azionale del verbo e l'uso di marche temporali da parte di chi apprende una seconda lingua. In particolare, i verbi telici (*accomplishments* e *achievements*) sono prevalentemente flessi al passato (55% e 59%) mentre *states* e *activities* presentano una percentuale molto più bassa di marche passate (31% e 25%).

Per quanto riguarda l'italiano una classificazione dei verbi secondo le classi azionali vendleriane è stata effettuata nell'ambito di una ricerca condotta dall'Università di Pisa e dalla Scuola Normale Superiore (Zarcone e Lenci, 2008). L'ipotesi di ricerca di questo studio è che l'interpretazione del valore azionale di un verbo possa essere definita tramite un processo di integrazione di vincoli morfologici, sintattici, semantici.

Lo studio riguarda due modelli computazionali per l'identificazione automatica dell'azione verbale in italiano: (i) il modello supervisionato della Maximum Entropy (MaxEnt; Berger et al., 1996); (ii) le Self-Organizing Maps (SOMs; Kohonen, 1997) per definire ed identificare le classi azionali in maniera non supervisionata.

Per quanto riguarda la MaxEnt, il peso dei diversi vincoli linguistici che determinano la classe azionale di un verbo viene calcolato sulla base di un algoritmo di apprendimento automatico basato appunto sul principio della "massimizzazione dell'entropia" che consente di collegare le diverse classi azionali con i tratti contestuali dei verbi estratti da un corpus annotato. Il metodo della massimizzazione dell'entropia richiede dunque la selezione preventiva di un insieme di tratti linguistici rilevanti al fine della classificazione. Per gli esperimenti che sono stati effettuati nell'ambito di questo studio è stato estratto un insieme di verbi da TreSSI (Montemagni et al., 2003), un corpus dell'italiano contemporaneo annotato a livello morfosintattico e sintattico. Dato che il metodo di apprendimento automatico della

MaxEnt è supervisionato, cioè ogni verbo del corpus usato per l'addestramento deve essere marcato con l'effettivo valore azionale, le occorrenze dei verbi in TreSSI sono state annotate manualmente secondo il valore azionale che queste assumevano in contesto. Dal lemmario del corpus sono stati selezionati 28 verbi e sono quindi state estratte tutte le frasi in cui questi ricorrevano nel corpus, per un totale di 3129 occorrenze.

Il metodo della MaxEnt è stato testato essenzialmente in due differenti esperimenti:

- 1) l'addestramento di un modello stocastico per identificare le quattro classi azionali vendleriane;
- 2) l'addestramento di modelli stocastici per l'individuazione dei singoli tratti azionali: [+/- telico], [+/- durativo], [+/- dinamico].

Il secondo modello computazionale investigato nello studio, riguarda l'utilizzo delle Self-Organizing Maps per l'individuazione della classe azionale dei verbi. Si tratta di un particolare tipo di rete neurale ad approccio non supervisionato. Un campione di 40 verbi è stato selezionato per l'alto livello di prototipicità rispetto alla quattro classi azionali vendleriane. Seguendo l'approccio di Lagus e Airola (2005), ogni verbo è stato rappresentato come un vettore, registrando le frequenze di co-occorrenza rispetto ad un certo numero di tratti linguistici contestuali. In linea con l'ipotesi distribuzionale di Harris (1968), è stato ipotizzato che due verbi avessero classe azionale simile se mostravano una simile distribuzione dei tratti linguistici contestuali.

Successivamente ai due precedenti esperimenti è stata condotta un'annotazione manuale della classe azionale di verbi in contesto che ha messo in evidenza la rilevante difficoltà di svolgimento del task anche da parte di annotatori esperti.

I risultati di questo esperimento di annotazione manuale sono stati poi comparati con quelli ottenuti dai due modelli computazionali e il loro confronto ha evidenziato la singolare vicinanza tra le prestazioni degli annotatori umani e dei modelli computazionali, rendendo ancor più significativi i risultati conseguiti da questi ultimi.

Per quanto riguarda, invece, i corpora annotati con informazioni inerenti l'aspetto verbale, per adesso, risultano disponibili solo dati, in varie lingue, annotati secondo lo schema TimeML.

Un esempio di questo tipo di corpora è la TimeBank per l'inglese, nella sua versione più recente, la 1.2.

La TimeBank 1.2 contiene 183 articoli di quotidiani annotati seguendo le specifiche TimeML 1.2.1²⁰. Questo corpus è scaricabile gratuitamente dal sito del Linguistic Data Consortium²¹.

Un altro esempio è costituito dalla TimeBank per il francese che contiene approssimativamente 61000 token per una grandezza complessiva di 1.2 TB. L'annotazione è svolta seguendo le linee guida TimeML adattate alla lingua francese. Anche questo corpus è scaricabile gratuitamente sul sito del progetto INRIA²².

Infine, un ultimo esempio è rappresentato dalla Ita-TimeBank (Caselli et al., 2011), il corpus di riferimento per l'italiano, costituito, a sua volta, da due corpora distinti: il CELCT corpus e l'ILC corpus. I due corpora sono stati sviluppati in parallelo rispettivamente dal Centro per la valutazione del linguaggio e le tecnologie della comunicazione (CELCT)²³ e dall'Istituto di Linguistica Computazionale (ILC)²⁴ del CNR.

Il CELCT corpus contiene circa 90000 token estratti da articoli del quotidiano L'Adige mentre l'ILC corpus contiene 68000 token recuperati dalla Italian Syntactic-Semantic Treebank, dal corpus PAROLE e dal web.

²⁰ http://www.timeml.org/site/publications/timeMLdocs/timeml_1.2.1.html

²¹ <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T08>

²² <https://gforge.inria.fr/projects/fr-timebank/>

²³ <http://www.celct.it>

²⁴ <http://www.ilc.cnr.it/indexnoflash.html>

Capitolo III

Il progetto TimeML, l'aspetto e l'azione verbale

Negli ultimi anni si è assistito ad un rinnovo di interesse nella comunità di Natural Language Processing per i processi di elaborazione temporale; questo è avvenuto, grazie, soprattutto, al successo ottenuto dallo schema di annotazione TimeML (Pustejovsky et al., 2003) e alla disponibilità di relative risorse annotate come le TimeBank per il francese (Bittar, 2008) e per l'inglese (Pustejovsky et al., 2003b) e il corpus di TempEval (Verhagen et al., 2010).

TimeML è uno schema di annotazione per gli eventi che ha lo scopo di facilitare il ragionamento e l'inferenza sulla loro organizzazione temporale nei testi, in modo da darne un'immagine più completa possibile. L'attenzione di TimeML, dunque, è rivolta principalmente agli eventi, alle espressioni temporali e alle loro relazioni reciproche.

Questo schema è stato elaborato nel 2002 durante il workshop TERQAS (Time and Event Recognition for Question Answering Systems)²⁵ organizzato da James Pustejovsky. L'organizzazione di TERQAS aveva lo scopo di proporre idee per migliorare le performance di sistemi di Question Answering, basati sul linguaggio naturale, nel rispondere a domande riguardanti le relazioni temporali tra eventi presenti in articoli di giornale. A partire da questo workshop è stata definita la versione 1.0 delle linee guida TimeML e successivamente è stato creato il TimeBank corpus (Pustejovsky et al., 2006), una risorsa di riferimento per l'annotazione di informazioni temporali, basato su una versione successiva delle linee guida per l'inglese (1.2.1).

Nel corso degli anni sono state prodotte versioni aggiornate delle linee guida e nel maggio del 2009 la versione del 2007 è diventata uno standard ISO (ISO/DIS 24617-1).

²⁵ <http://www.timeml.org/site/terqas/index.html>

Per chiarezza nella lettura, nel seguito di questo capitolo, quando verranno citate le linee guida TimeML per l'inglese si farà sempre riferimento alla versione più recente, diversa dallo standard ISO, ossia alle TimeML Annotation Guidelines Version 1.3, del 22 febbraio 2010.

Negli ultimi anni iniziative come ISO TC 37 / SC 4 (Ide e Romary, 2001) e TempEval hanno contribuito allo sviluppo di schemi di annotazione conformi a TimeML in varie lingue oltre all'inglese, tra cui lo spagnolo, il cinese, il coreano, il francese e l'italiano. Una volta che tutti i corpora relativi alle lingue sopra elencate saranno completati e resi disponibili alla comunità scientifica, si concretizzerà la possibilità di accedere a differenti risorse linguistiche caratterizzate da un comune livello di annotazione. La disponibilità di questo tipo di risorse mira ad incoraggiare gli studi in *multilingual temporal processing* ed a migliorare la performance di complessi sistemi di NLP multilingue come, per esempio, i già citati sistemi di Question Answering²⁶ o i sistemi di Textual Entailment²⁷.

3.1 Lo schema di annotazione

Nello standard ISO viene definito evento “something that can be said to obtain or hold true, to happen or to occur” (ISO (2008): 1). Questa definizione include tutti i tipi di azioni (puntuali o durative) e di stati (Bach, 1986).

Nel rispetto dei precedenti schemi di annotazione (tra gli altri, Katz-Arosio, 2001; Filatova e Hovy, 2001; Setzer e Gaizauskas, 2001), in TimeML gli elementi linguistici che sintatticamente possono ricorrere come eventi sono: aggettivi, nomi, verbi e frasi preposizionali.

Come spiegato in dettaglio nei seguenti sottoparagrafi, lo schema di annotazione prevede le seguenti funzionalità: l'annotazione di eventi, espressioni temporali, “segnali”, la loro normalizzazione e la codifica delle relazioni esistenti tra eventi ed espressioni temporali.

²⁶ I sistemi di Question Answering (Q/A) sono strumenti di recupero automatico delle informazioni, destinati a rispondere a domande poste in linguaggio naturale. Diversamente dagli attuali motori di ricerca, i sistemi di Q/A non ricercano interi documenti ma forniscono risposte specifiche situate in piccoli frammenti di testo.

²⁷ I sistemi di Recognising Textual Entailment (RTE) permettono di riconoscere, dati due frammenti di testo, denominate Text e Hypothesis, se il significato dell'uno è inferito dell'altro.

3.1.1 Il tag <EVENT>

Gli eventi sono marcati con il tag <EVENT> e la loro annotazione si basa sulla nozione di unità minima²⁸, ovvero si prevede che debba essere annotata solo la loro testa sintattica.

Informazioni sulla polarità (attributo POLARITY), sulla parte del discorso (Part-Of-Speech, POS) sul tempo verbale (attributo TENSE), sull'aspetto (attributo ASPECT, discusso più in dettaglio in 3.3), sulla modalità (attributo MODALITY), sui modi verbali (VFORM e MOOD) e sulla classe²⁹ di appartenenza dell'evento (attributo CLASS) sono rappresentati come attributi del tag <EVENT>.

Di seguito due esempi per esplicitare in maniera più chiara il significato degli attributi sopra elencati:

(1) Non si sono mai [tirati indietro]

```

CLASS=OCCURRENCE
POS =VERB
TENSE=PRESENT
ASPECT=PERFECTIVE
VFORM=NONE
POLARITY=NEG
MOOD=NONE
MODALITY=NONE

```

(2) [Vorrei] [andare] a casa

vorrei	andare
CLASS=I_STATE	CLASS=OCCURRENCE
POS=VERB	POS=VERB
TENSE=PRESENT	TENSE=NONE
ASPECT=IMPERFECTIVE	ASPECT=NONE
VFORM=NONE	VFORM=INFINITIVE

²⁸ In linguaggio tecnico, *minimal chunk*

²⁹ TimeML prevede sette differenti classi di eventi; per una descrizione completa di queste classi si rimanda alla lettura di (Saurí et al., 2005), per una descrizione sintetica si veda il paragrafo 3.5 di questo capitolo.

POLARITY=POS	POLARITY=POS
MOOD=CONDITIONAL	MOOD=NONE
MODALITY=VOLERE	MODALITY=NONE

In Appendice A si trova la Backus-Naur Form (BNF) del tag, così come riportata nelle linee guida per l'inglese, dove per ogni attributo è presente la lista dei suoi possibili valori.

3.1.2 Il tag <TIMEX3>

Le espressioni temporali sono annotate con il tag <TIMEX3> e sono divise in 4 classi: date (DATE, *il 15 gennaio, 15.10.2011*), ore (TIME, *15:20, il pomeriggio*), durate (DURATION, *5 giorni, due anni*) e intervalli temporali (SET, *tutti i giorni, ogni anno*).

TimeML permette, inoltre, la rielaborazione delle espressioni temporali anaforiche, come *lunedì prossimo* o *l'anno precedente*, il cui valore deve essere determinato in rapporto ad un punto temporale di riferimento.

Di seguito alcuni esempi in cui sono riportati i valori per l'attributo *type*, indicante il tipo di classe a cui appartiene l'espressione temporale:

- (3) Colombo è arrivato in America nel
<TIMEX3...type="DATE">1492</TIMEX3>
- (4) La riunione è terminata alle <TIMEX3...type="TIME">20:00</TIMEX3>
- (5) La guerra va avanti ormai da <TIMEX3...type="DURATION">oltre un mese</TIMEX3>
- (6) La manifestazione si tiene oramai <TIMEX3 ...type="SET">ogni anno</TIMEX3>

In Appendice A è riportata la BNF del tag, così come da linee guida per l'inglese.

3.1.3 I tag <TLINK>, <ALINK> e <SLINK>

Gli eventi e le espressioni temporali sono messi in relazione attraverso tre tipi di legami (LINK): legami temporali (TLINK), aspettuali (ALINK) e di subordinazione (SLINK).

I TLINK catturano le relazioni temporali tra due eventi (7), due espressioni temporali (8) o tra un evento e un'espressione temporale (9):

(7) <EVENT>Ascolto</EVENT> sempre musica quando
<EVENT>scrivo</EVENT>

(8) <TIMEX3>Sabato</TIMEX3> alle <TIMEX3>20:00</TIMEX3>

(9) Marco è <EVENT>arrivato</EVENT> a Pisa <TIMEX3>lunedì</TIMEX3>

Gli ALINK vengono utilizzati per marcare la relazione aspettuale tra due eventi (10); questa relazione si attua in presenza, per esempio, di verbi aspettuali (per la loro annotazione vedere il cap. 3.4) come “finire”, “terminare”, “iniziare”:

(10) Marco ha <EVENT>finito</EVENT> di <EVENT>leggere</EVENT> il libro

Gli SLINK servono per marcare la presenza di qualsiasi relazione di subordinazione tra due eventi (11):

(11) Marco non <EVENT>vuole</EVENT> <EVENT>venire</EVENT>

3.1.4 Il tag <SIGNAL>

Gli elementi testuali che hanno la funzione di segnalare esplicitamente la presenza dei vari tipi di link sono annotati con il tag <SIGNAL>. Le espressioni linguistiche che possono essere considerate segnali sono ristrette alle seguenti:

- Preposizioni temporali: semplici o articolate seguite da espressioni temporali;

- Congiunzioni temporali: *prima, quando, mentre, ecc;*
- Avverbi temporali: *intanto, nel frattempo, ecc;*
- Caratteri speciali: “-” o “√”, in espressioni temporali che denotano periodi (per esempio, *26 - 28 Ottobre 2011*);
- Preposizioni e congiunzioni che segnalano relazioni subordinate: *per, affinché, ecc;*

Qui sotto è riportato un esempio semplificato di annotazione con i tre principali tag e un TLINK tra l’evento e l’espressione temporale.

(12) Luca è arrivato a Pisa nel febbraio 2011

```
Luca     è     <EVENT   ...id="e1">arrivato</EVENT>     a     Pisa
<SIGNAL>nel</SIGNAL> <TIMEX3 ...id="t1">febbraio 2011</TIMEX3>
<TLINK ...event="e1" time="t1"/>
```

3.2 IT-TimeML

Applicare uno schema di annotazione ad un linguaggio diverso da quello per il quale è stato sviluppato richiede uno studio approfondito delle specifiche relative ai fenomeni linguistici caratterizzanti la nuova lingua presa in considerazione (Im et al., 2009; Bittar, 2008).

Le seguenti sottosezioni illustreranno i due più importanti tag definiti in TimeML, <EVENT> e <TIMEX3>, concentrandosi sui loro adattamenti alla lingua italiana. A questo proposito il riferimento teorico è costituito dall’ultima versione delle linee guida per l’italiano “It-TimeML: TimeML Annotation Scheme for Italian Version 1.3.1”³⁰ (da qui in avanti definite semplicemente come linee guida per l’italiano) pubblicata nel settembre 2010 dal CNR-ILC³¹ di Pisa, in collaborazione con il Centro per la valutazione del linguaggio e le tecnologie della comunicazione CELCT³² di Trento.

³⁰ <http://puma.isti.cnr.it/dfdownload.php?ident=/cnr.ilc/2010-TR-002&langver=it&scelta=Metadata>

³¹ <http://www.ilc.cnr.it/indexnoflash.html>

³² <http://www.celct.it>

3.2.1 Il tag <EVENT>

Nell'adattamento dello schema di annotazione TimeML all'italiano, sia la procedura di mark-up che la nozione di *minimal chunk* per l'individuazione dell'estensione dei tag sono state preservate così come presentate nello standard ISO.

Le più consistenti differenze rispetto all'inglese riguardano la lista e i valori degli attributi del tag <EVENT>.

Per l'italiano, infatti, sono stati previsti dodici attributi³³ rispetto ai 10 riportati nelle linee guida per l'inglese. I due attributi ulteriori sono MOOD e VFORM che catturano la distinzione chiave del sistema tempo-modo-aspettuale dell'italiano. Questi due attributi sono previsti anche negli adattamenti ad altre lingue come lo spagnolo, il catalano, il francese ed il coreano.

3.2.1.1 L'attributo MOOD

L'attributo MOOD marca le espressioni grammaticali caratterizzanti le differenti modalità di presentazione di un evento realizzato da un verbo.

La presenza di questo attributo è importante poiché il livello grammaticale ha un forte impatto nell'identificazione delle relazioni temporali e dei loro valori. In italiano, l'attributo esprime la morfologia verbale che, a differenza dell'inglese, è una caratteristica indipendente dall'utilizzo di verbi ausiliari (in inglese, per esempio, il modo condizionale viene espresso tramite l'utilizzo dell'ausiliare *would*).

Di seguito, i valori che l'attributo MOOD può assumere in italiano (in inglese vengono considerati solo i valori "subjunctive" e "none"):

- NONE: è usato come valore di default e corrisponde all'indicativo.

(13) L'esercito ha <EVENT ...mood="NONE">schierato</EVENT> 500 carro armati.

- CONDITIONAL: utilizzato per segnalare il modo condizionale.

³³ Per una visione più dettagliata degli attributi di TimeML e maggiori informazioni sui loro si rimanda alla lettura delle linee guida.

(14) <EVENT...mood="CONDITIONAL">Mangerei</EVENT> del dolce.

- SUBJUNCTIVE: utilizzato per il modo congiuntivo.

(15) <EVENT ...mood="NONE">Voglio</EVENT> che tu <EVENT ...mood="SUBJUNCTIVE">resti</EVENT>.

- IMPERATIVE: utilizzato per segnalare l'imperativo.

(16) <EVENT mood="IMPERATIVE">Taci</EVENT>!

3.2.1.2 L'attributo VFORM

L'attributo VFORM è stato introdotto per distinguere la forme verbali finite da quelle non finite. I valori che può assumere sono i seguenti:

- NONE: è il valore di default che segnala forme verbali finite.

(17) L'esercito ha <EVENT ...vForm="NONE">schierato</EVENT> 500 carri armati.

- INFINITIVE: per segnale l'utilizzo del modo infinito (presente e passato).

(18) <EVENT ...vForm="NONE">Vorrei</EVENT> <EVENT...vForm="INFINITIVE">mangiare</EVENT> del dolce.

- GERUND: per segnale la presenza del gerundio (presente e passato).

(19) <EVENT ...vForm="GERUND">Andando</EVENT> in treno, ha <EVENT...vForm="NONE">evitato</EVENT> la coda.

- PARTICIPLE: per segnale l'utilizzo del modo participio (presente e passato).

(20) <EVENT ...vForm="PARTICIPLE">Mangiato</EVENT> il dolce, se ne <EVENT ...vForm="NONE">andò</EVENT>.

3.2.1.3 I valori degli attributi MODALITY e ASPECT

Per quanto riguarda i valori degli attributi, i cambiamenti più rilevanti riguardano quelli di MODALITY e ASPECT.

In inglese i verbi modali non sono annotati come eventi e l'attributo MODALITY è associato al verbo principale (il valore è il token corrispondente al verbo modale). In italiano, invece, i verbi modali sono considerati in tutto e per tutti simili agli altri verbi ed è quindi possibile assegnare loro valori di tempo e aspetto. Ogni istanza di verbi modali in italiano (“dovere”, “potere”, “volere”), dunque, deve essere annotata con il tag <EVENT>, il cui attributo CLASS deve assumere categoricamente il valore I_STATE e il valore dell'attributo MODALITY deve corrispondere al lemma del verbo modale stesso.

L'attributo ASPECT cattura la distinzione standard in categorie grammaticali dell'aspetto verbale o “event viewpoint” (Smith, 1991). Questo attributo può assumere uno dei seguenti valori: i) PROGRESSIVE; ii) PERFECTIVE; iii) IMPERFECTIVE, iv) NONE (per gli eventi non espressi da verbi). In linea con la filosofia di TIMEML di annotare solo il livello superficiale del testo, e in prospettiva di una annotazione automatica, l'uso di marche più precise per l'aspetto verbale è stato scartato in favore di una distinzione più generica tra aspetto perfettivo, imperfettivo e una sotto classe dell'imperfettività, ovvero l'aspetto progressivo³⁴.

3.2.2 Il tag <TIMEX3>

Il tag <TIMEX3> si basa sullo schema di annotazione TIDES TIMEX2 (Ferro et al., 2005) al quale cerca di essere il più conforme possibile. L'adattamento all'italiano per l'annotazione delle espressioni temporali presenti nei testi è riportato in (Magnini et al., 2006). L'unica differenza rispetto all'annotazione delle TIMEX2 riguarda il trattamento delle preposizioni articolate che in TimeML

³⁴ Per una trattazione più completa di questa problematica si veda il cap. 3.3

vengono annotate come SIGNAL (21) mentre in (Magnini et al., 2006) sono considerate parte della realizzazione testuale delle espressioni temporali (22).

(21) <SIGNAL ...>nel</SIGNAL> <TIMEX3 ...>pomeriggio</TIMEX3>

(22) <TIMEX2 ...>nel pomeriggio</TIMEX2>

È stato, inoltre, deciso di seguire le specifiche dello standard TIMEX2 per l'annotazione di aggettivi come espressioni temporali (per esempio, *recente*, *ex*, ecc...) e di includere modificatori come *che rimane* in "l'anno che rimane" nell'estensione del tag TIMEX3 dato che l'informazione portata dal modificatore è essenziale per la normalizzazione dell'espressione temporale.

3.3 L'attributo ASPECT

L'attributo ASPECT, così come l'attributo TENSE, influenza, in tutte le lingue, solo gli eventi espressi tramite verbi, e dunque come regola generale è stato stabilito che per altre parti del discorso il valore assegnato a questo attributo è NONE.

Nelle linee guida per l'inglese si specifica che l'attributo ASPECT può assumere uno dei seguenti valori:

- PROGRESSIVE: per eventi che sono generalmente descritti come in corso di svolgimento, realizzati con l'uso dell'ausiliare *be* seguito dalla forma – *ing* del verbo.
- PERFECTIVE: per eventi che possono generalmente essere descritti come completi, realizzati con l'uso dell'ausiliare *have* seguito da un participio passato.
- PERFECTIVE PROGRESSIVE: per eventi che sono espressi in forma sia perfettiva che progressiva.
- NONE: per eventi che sono nelle forme del *simple present*, *past*, o *future*, senza la presenza di marche progressive o perfettive.

Di seguito sono riportate le regole per l'assegnazione del valore all'attributo ASPECT:

- **Diatesi attiva**

tense= "PRESENT"

verbo	aspect=
teaches	NONE
has taught	PROGRESSIVE
has been taught	PERFECTIVE
has been teaching	PERFECTIVE PROGRESSIVE

tense= "PAST"

verbo	aspect=
taught	NONE
was teaching	PROGRESSIVE
had taught	PERFECTIVE
had been teaching	PERFECTIVE PROGRESSIVE

tense= "FUTURE"

verbo	aspect=
will teach	NONE
is going to teach	NONE
will be teaching	PROGRESSIVE
is going to be teaching	PROGRESSIVE
will have taught	PERFECTIVE PROGRESSIVE
will have been teaching	PERFECTIVE PROGRESSIVE

- **Diatesi passiva**

tense= "PRESENT"

verbo	aspect=
is taught	NONE
is being taught	PROGRESSIVE
has been taught	PERFECTIVE

tense= "PAST"

verbo	aspect=
was taught	NONE
was being taught	PROGRESSIVE
had been taught	PERFECTIVE

tense= "FUTURE"

verbo	aspect=
will be taught	NONE
is going to be taught	NONE
will have been taught	PERFECTIVE

- **Verbi preceduti da *have to* e *ought to***

tense= "PRESENT"

verbo	aspect=
has to teach	NONE
has to be teaching	PROGRESSIVE
has to have taught	PERFECTIVE
has to have been teaching	PERFECTIVE PROGRESSIVE

tense= "PAST"

verbo	aspect=
had to teach	NONE
had to be teaching	PROGRESSIVE

tense= "FUTURE"

verbo	aspect=
will have to teach	NONE
will have to be teaching	PROGRESSIVE

- **Verbi preceduti da altri ausiliari come *must, may, might, can, could, shall, should, e would.***

tense= "NONE"

verbo	aspect=
could teach	NONE
could be teaching	PROGRESSIVE
could have taught	PERFECTIVE
could have been teaching	PERFECTIVE PROGRESSIVE

- **Modi non finiti:**

- Present Participle

tense= "NONE"

verbo	aspect=
seeking	NONE

- Past Participle

tense= "PASTPART"

verbo	aspect=
found	NONE

- Infinitive

tense= "INFINITIVE"

verbo	aspect=
(to) release	NONE
(to) be releasing	PROGRESSIVE
(to) have released	PERFECTIVE
(to) have been releasing	PERFECTIVE PROGRESSIVE

Nelle linee guida per l'italiano, come già anticipato nel cap. 1.2.1.3, si specifica che i valori ammissibili per l'attributo ASPECT sono i seguenti: PROGRESSIVE, PERFECTIVE, IMPERFECTIVE e NONE. Non vengono considerati tutti i valori aspettuali possibili per la lingua italiana (vedi diagramma al cap. 1.2) ma questi

vengono ridotti ai quattro sopra citati, specificando che il valore PROGRESSIVE deve essere assegnato solo in presenza esplicita di perifrasi aspettuali (per esempio, *sto giocando*). Questa semplificazione viene giustificata con l'affermazione che l'assegnazione del tipo di aspetto deve essere eseguita in prospettiva di un'annotazione automatica, che per adesso, però rimane irrealizzata, e secondo una filosofia di analisi superficiale delle forme verbali.

Di seguito sono riportate le regole di annotazione presenti nelle linee guida per l'italiano:

- **Verbi di modo finito:**

tense= "PRESENT"	aspect=
gioca	IMPERFECTIVE
sta giocando	PROGRESSIVE
ha mangiato	PERFECTIVE
è mangiato	IMPERFECTIVE
(che) mangi	IMPERFECTIVE

tense= "PAST"	aspect=
giocò	PERFECTIVE
ebbe l'abitudine di giocare	PERFECTIVE
fu mangiato	PERFECTIVE
è stato mangiato	PERFECTIVE
(che) abbia mangiato	PERFECTIVE
aveva giocato	PERFECTIVE
ebbe giocato	PERFECTIVE
era stata mangiata	PERFECTIVE
(che) avesse mangiato	PERFECTIVE
giocava	IMPERFECTIVE
stava giocando	PROGRESSIVE
aveva l'abitudine di giocare	IMPERFECTIVE
era mangiata	IMPERFECTIVE

(che) mangiasse	IMPERFECTIVE
-----------------	--------------

tense= "PAST"	aspect=
giocherà	PERFECTIVE
avrà giocato	PERFECTIVE
sarà mangiata	PERFECTIVE
starà mangiando	PROGRESSIVE

tense= "NONE"	
mood="CONDITIONAL/SUBJUNCTIVE/IMPERATIVE"	

	aspect=
mangerebbe	IMPERFECTIVE
sarebbe mangiato	IMPERFECTIVE
avrebbe mangiato	PERFECTIVE
sarebbe stato mangiato	PERFECTIVE
starebbe mangiando	PROGRESSIVE
mangia!	NONE

- **Modi non finiti:**

tense= "NONE"	
vform="INFINITIVE/PARTICIPLE/GERUND"	

	aspect=
giocare	NONE
aver giocato	PERFECTIVE
stare giocando	PROGRESSIVE
giocando	NONE
avendo giocato	PERFECTIVE
giocante	NONE
giocato	PERFECTIVE

Come spiegato in dettaglio nel cap. I, la lingua italiana necessita di informazioni semantiche e pragmatiche per determinare il corretto valore aspettuale; l'aspetto,

infatti, viene definito dalla costruzione della frase, dal tempo verbale utilizzato e da avverbiali che vengono aggiunti per chiarire il punto di vista sull'evento. Le linee guida per l'italiano, non tengono conto di queste variabili che caratterizzano la lingua: non risulta, infatti, chiara la motivazione della scelta di utilizzare come valori aspettuali PERFECTIVE, IMPERFECTIVE e PROGRESSIVE. Il perfetto e l'imperfetto rappresentano la distinzione di base dei valori aspettuali ma il progressivo è un sotto-valore della classe degli imperfettivi, al pari dell'aspetto abituale e continuo.

Nelle linee guida è precisato, inoltre, che il valore progressivo viene assegnato, ricalcando l'approccio adottato per l'inglese, solo in presenza di perifrasi, seguendo la logica TimeML di un'analisi superficiale delle forme verbali. In italiano, però, il valore progressivo può essere espresso con altre forme verbali, per esempio con verbi coniugati all'imperfetto in particolari contesti (vedi cap. 1.2.1.2.1)

Infine per quanto riguarda i modi non finiti, è stato deciso il valore perfetto per le forme verbali composte e quello progressivo in presenza di una perifrasi verbale. Per le forme verbali semplici è stato deciso, invece, il valore NONE, in quanto, non prendendo in considerazione il contesto non è possibile compiere analisi più approfondite.

3.3.1 Proposta di integrazione dei valori dell'attributo ASPECT

In questo lavoro si tenta di proporre una classificazione più dettagliata dei valori aspettuali in modo da integrare quella già esistente. Nella prospettiva di facilitare il lavoro degli annotatori, si è cercato di stilare una lista di regole chiare e facili da applicare per individuare il corretto valore aspettuale e, in quest'ottica, sono state sviluppate delle specifiche di annotazione piuttosto che delle linee guida. La distinzione tra specifiche e linee guida risulta, in questo contesto, di particolare importanza in quanto permette di distinguere il livello teorico, astratto e formale, dal livello pratico di effettiva annotazione attraverso l'ideazione di un nuovo livello di *best practice* per la creazione di risorse linguistiche annotate semanticamente (Calzolari-Caselli, 2009; Calzolari et al. 2010).

Lo scopo è quello di riuscire a classificare alcuni tipi di contrasti aspettuali rilevanti per l'italiano, la cui individuazione può essere utile anche in un'ottica di elaborazione computazionale dei testi.

Per quanto riguarda l'aspetto imperfettivo la proposta è quella di distinguere tra le seguenti sottoclassi: *aspetto progressivo*, *aspetto gnomico* e *aspetto continuo*. Per quello che concerne, invece, l'aspetto perfettivo le sottoclassi sono costituite dall'*aspetto ingressivo*, e da una classe perfettiva generica che riunisce il perfettivo aoristico e il perfettivo compiuto. La lingua italiana così come è ambigua nella distinzione delle varie forme di imperfettività, ad eccezione della forma progressiva che può essere resa esplicita, lo è anche rispetto ai due ultimi tipi di aspetto sopra citati; basti pensare all'uso aoristico del passato prossimo tipico dei parlanti nativi settentrionali oppure all'utilizzo del passato remoto in accezione compiuta caratterizzante i parlanti nativi meridionali, in particolare, siciliani. Questa ambiguità si ritrova anche in frasi come (23) dove sebbene i due verbi siano coniugati al passato prossimo tendono ad assumere un'accezione aoristica.

(23) L'ha raggiunta, ha tirato fuori l'arma da sotto il giubbotto e ha sparato

Gli aspetti compiuto e aoristico hanno caratteristiche diverse: il primo ha un ancoraggio temporale dinamico e complesso con il tempo verbale di riferimento, rispetto al quale esprime una relazione temporale come in (24), il secondo tende, invece, ad avere un rapporto più lineare e statico con i verbi con cui si relaziona: successioni di verbi in accezione aoristica esprimono generalmente semplici sequenze temporali come in (25) e in (26).

(24) L'uomo era stato rinchiuso nel manicomio criminale di Castiglione delle Stiviere da cui però fuggì

(25) Arrivai a fine giugno 2003 e qualche giorno dopo giunse anche Simona Pari

(26) Riflessioni di grande suggestione che si dilatarono, affinarono e moltiplicarono con l'affermarsi della psicanalisi

Con la decisione di non distinguere tra i due tipi di aspetto non si vuole negare, comunque, che ci possano essere casi in cui sarebbe interessante marcare questa distinzione, per esempio in task volti alla ricostruzione delle relazioni temporali fra gli eventi presenti nei testi.

La decisione di marcare l'imperfettività gnomica è dettata dall'interesse nell'individuare frasi che esprimono o una generalizzazione di qualche tipo (cap. 1.2.1.2.3.1) oppure proprietà non accidentali di un momento ma che caratterizzano un intero periodo (cap. 1.2.1.2.3). Tale specifico interesse trova origine nel tentativo di dimostrare l'utilità di marcare la classe aspettuale della "imperfettività gnomica" in prospettiva di un'estrazione automatica di informazione di senso comune da testi scritti (Singh, 2002).

Infine, in questo contesto, potrebbe risultare interessante adottare uno specifico attributo sovra-verbale per marcare le strutture abituali caratterizzanti intere frasi. Il verbo incluso nella testa di una struttura di questo tipo, infatti, può avere aspetto abituale ma la dipendente può comunque contenere un verbo di aspetto perfettivo utilizzato per esprimere una relazione di causalità come in (27) oppure di temporalità come in (28):

(27) Tiene abitualmente conferenze in tutto il mondo poiché i suoi libri hanno avuto ovunque molto successo

(28) Ogni giorno sparano sui civili e hanno già occupato l'ospedale

Nella definizione delle specifiche di annotazione sono stati riscontrati due problemi rilevanti: la determinazione del valore aspettuale per il futuro semplice e per i trapassati.

Per quanto riguarda il futuro semplice la difficoltà è determinata dalla caratteristica di questo tempo verbale di avere molteplici letture a seconda del contesto in cui appare; oltre che il più comune valore perfettivo, può assumere, infatti, anche aspetto imperfettivo progressivo come in (29), imperfettivo abituale (30) e imperfettivo continuo (31):

(29) Arriverai al castello dove molte guardie dormiranno

(30) Luca siederà sempre vicino al sindaco durante le occasioni ufficiali

(31) Mentre Giulia parlerà, Luca proietterà il video

Anche il trapassato prossimo, può assumere diversi valori aspettuali: oltre alla più comune lettura perfettiva, può avere anche lettura imperfettiva abituale come in (32):

(32) Non appena aveva terminato il suo giro di ispezione, la guardia si rimetteva sulla sua poltrona a dormire

La soluzione che è stata adottata nelle specifiche, secondo una logica di semplificazione, è stata quella di stabilire un trattamento particolare per questi tempi generalizzando il loro valore aspettuale più comune: ai futuri, senza distinzione tra semplici e composti, e ai piuccheperfetti, senza distinguere prossimi e remoti, si è deciso di assegnare sempre, indipendentemente dal contesto, il valore PERFECTIVE, senza ulteriore specificazione.

3.3.2 Specifiche per l'annotazione dei valori aspettuali

Nella prospettiva di realizzare un'annotazione semantica dei testi e quindi di interpretare semanticamente i diversi valori aspettuali che i verbi possono assumere in contesto, è presentata di seguito una breve guida per interpretare, e dunque disambiguare, le categorie aspettuali che non hanno una realizzazione testuale univoca.

In particolare, ci si propone di individuare il corretto valore aspettuale attraverso una serie di passi successivi, ordinati per complessità. Inizialmente l'annotatore può ricorrere a quelli che di seguito sono indicati come test di primo livello, ovvero test di parafrasi, sfruttando il fatto che ad un determinato valore aspettuale può corrispondere un'esplicita parafrasi. Per esempio, l'aspetto abituale può essere individuato sostituendo il verbo con la perifrasi verbale “*essere solito + infinito*”.

I test di secondo livello, invece, si basano su due diverse considerazioni sul testo: (i) la verifica di particolari segnalatori che permettono di dare una determinata lettura aspettuale al verbo; (ii) la possibilità di integrare in maniera appropriata segnalatori aspettuati nel testo, anche quando non esplicitamente presenti.

Di seguito sono riportate le specifiche di annotazione distinte secondo i tempi verbali dei modi finiti e non finiti.

Valori corrispondenti ai tempi verbali dei modi finiti:

tempo verbale	aspetto=
futuro semplice e anteriore	PERFECTIVE
piuccheperfetti (trapassato prossimo e remoto)	PERFECTIVE
perfetti (passato prossimo e remoto) se non ingressivi	PERFECTIVE

Imperfettivo_progressivo

- **Tempo verbale:** imperfetto, presente, perifrasi progressive
- **Test di primo livello:** è possibile sostituire il verbo con la perifrasi progressiva “*stare + gerundio*”
- **Test di secondo livello:**
 - ✓ Compatibilità con gli avverbiali decorrenziali come “da X TEMPO” in presenza di verbi durativi:
 - Quando entrai, Giulia dormiva da due ore
 - Laura legge da due ore
 - ✓ Incompatibilità con avverbiali del tipo “in/per X TEMPO”:
 - *Quando entrai, Giulia dormiva per due ore
 - *Quando entrai, Giulia dormiva in due ore
 - *Laura legge per due ore
 - *Laura legge in due ore
 - ✓ Incompatibilità con gli avverbiali culminativi, delimitativi e circoscriventi:
 - *Quando entrai, Giulia dormiva fino alle 2
 - *Quando entrai, Giulia dormiva dalle 3 alle 4

- *Quando entrai, Giulia dormiva fra le 3 e le 4
- *Laura legge (sta leggendo) fino alle 2
- *Laura legge (sta leggendo) dalle 3 alle 4
- *Laura legge (sta leggendo) fra le 3 e le 4
- ✓ Testare la compatibilità del verbo con una frase incidentale del tipo “mentre X dormiva/mangiava/lavorava ecc...”. Tale frase può non essere presente nel testo ma se l’aspetto del verbo principale è progressivo può essere aggiunta.

Imperfettiva_gnomica

- **Tempo verbale:** imperfetto, presente
- **Test di primo livello:** per l’individuazione dell’aspetto abituale, che costituisce un sotto tipo di imperfettività gnomica, è possibile sostituire il verbo con la perifrasi “*essere solito* + infinito”
- **Test di secondo livello**
 - ✓ Per quanto riguarda l’aspetto abituale, compatibilità con avverbiali come *raramente, spesso, abitualmente, sempre, ogniqualvolta* che in correlazione con un aspetto verbale imperfettivo consentono una lettura abituale della frase:
 - Negli ultimi mesi, Manuela viaggiava raramente in aereo
 - Laura si arrabbia spesso con le colleghe
 - ✓ Incompatibilità con le specificazioni numeriche, il numero di iterazioni di un evento deve rimanere indeterminato:
 - Negli ultimi sei mesi, Manuela ha preso l’aereo 5 volte [non-abituale]
 - *Quando viveva a Trento, Manuela prendeva l’aereo 5 volte
 - Durante lo scorso anno Giovanni ha telefonato a sua madre (10 volte) [non-abituale]
 - Durante lo scorso anno Giovanni telefonava spesso a sua madre [abituale]
 - Enrico fuma la pipa [attitudinale]
 - Stefano parla francese [potenziale]

- Valentina è toscana [IL-predicate]
- I cani hanno quattro zampe [generico]

Imperfettivo_continuo

N.B. si differenziano dall'aspetto abituale perché si riferiscono ad un'unica situazione che non si ripete. Non possono avere lettura progressiva.

- **Tempo verbale:** imperfetto, presente
- **Test di primo livello:** è possibile sostituire il verbo con le perifrasi “*non fare altro che + infinito*”, “*andare + gerundio*”, “*continuare a + infinito*”.
- **Test di secondo livello:**
 - ✓ Accadimento singolo in presenza degli avverbiali “per X TEMPO”, “durante X TEMPO”
 - Per tutta la durata della conferenza, Marco annuiva all'oratore
 - Durante il ricevimento, Veronica accoglie gli ospiti
 - ✓ Accadimento singolo in possibile presenza, esplicita o sottintesa, di avverbiali come “in continuazione”, “continuamente”, “di continuo”
 - Durante la riunione, Marco entrava e usciva dall'aula in continuazione, tanto che fu richiamato dal professore

Perfettivo_ingressivo

- **Tempo verbale:** perfetti (passato prossimo e remoto), presente storico
- **Test di primo livello:** è possibile sostituire il verbo con la perifrasi ingressiva “*cominciare a + infinito*”
- **Test di secondo livello:**
 - ✓ Questo valore può essere assegnato solo in presenza di verbi durativi non telici
 - Finalmente Giulia parlò
 - In quel momento, il treno si mosse
 - Improvvisamente l'acqua è zampillata dal rubinetto

Come spiegato nel cap. 1.2.1, le forme non finite composte sono univocamente orientate ad esprimere l'aspetto perfettivo.

Le forme non finite semplici (infinito, gerundio e participio semplici) sono aperte ad una molteplicità di letture: aspettuale, possono indicare valore imperfettivo (progressivo o continuo) e perfettivo aoristico. A causa delle difficoltà del determinare il valore aspettuale di queste forme verbali e per semplificare il lavoro dell'annotatore si è stabilito di assegnare alle forme semplici di gerundio e participio il valore NONE.

Per quanto riguarda l'infinito semplice, pur senza alcuna pretesa di esaustività, si è deciso di compiere un'analisi un poco più approfondita circa il suo valore aspettuale tenendo in considerazione le due regole di seguito elencate: quando il verbo è retto da verbi dichiarativi o incluso in contesti abituali assume valore PERFECTIVE. Nei restanti casi il valore assegnato all'infinito semplice deve essere NONE: individuare, infatti, ulteriori livelli di granularità nei valori risulterebbe eccessivo sia per il carico cognitivo che dovrebbe essere sostenuto dall'annotatore, sia, soprattutto, nell'ottica di un'annotazione automatica.

Valori corrispondenti ai tempi verbali dei modi non finiti:

tempo verbale	aspetto=
Gerundio composto	PERFECTIVE
Gerundio semplice	NONE
Participio passato	PERFECTIVE
Participio presente	NONE
Infinito composto	PERFECTIVE

Infinito semplice:

- Infinito retto da verbi dichiarativi (dire, affermare, dichiarare, certificare, giurare, narrare, rimproverare) → valore perfettivo
- in contesti abituali con tempi reggenti imperfettivi e perfettivi:
Ogni giorno, Marta gli chiedeva l'auto per essere in tempo all'appuntamento/ ??Ogni giorno, Marta gli chiese l'auto per essere in tempo all'appuntamento → valore perfettivo

- Negli altri casi (infinito retto da altri verbi, da aggettivi, da nomi)
inserire NONE

3.4 L'azione verbale nello schema di annotazione TimeML

Nello standard ISO-TimeML, così come nelle linee guida per l'italiano, non è previsto un attributo specifico per marcare l'azione verbale. Da un'analisi più dettagliata degli attributi e dei loro valori è possibile, però, rintracciare alcune informazioni relative all'azione tra i valori dell'attributo CLASS. Questo attributo è obbligatorio e contempla sette valori diversi, gli stessi per l'inglese e per l'italiano. Di seguito sono descritti sinteticamente (negli esempi relativi alle diverse classi sono annotati solo gli eventi di cui viene fornita la spiegazione):

1. REPORTING: "Reporting events describe the action of a person or an organization declaring something, narrating an event, informing about an event, etc" (ISO (2008): 48); questo valore riguarda verbi come: *dire, spiegare, raccontare, affermare, parlare, ecc*;

(33) Il presidente ha <EVENT
...CLASS="REPORTING">spiegato</EVENT> che la fusione si farà

2. PERCEPTION: "Events involving the physical perception of another event" (ISO (2008): ibid.); per esempio: *vedere, guardare, osservare, ascoltare, sentire, ecc*;

(34) Luca ha <EVENT ...CLASS="PERCEPTION">visto</EVENT>
Angela uscire di casa

3. ASPECTUAL: verbi che codificano informazione circa una particolare fase di svolgimento di un evento, per esempio: *iniziare, ricominciare, terminare, finire, continuare ecc*;

(35) Gianni ha <EVENT ...CLASS="ASPECTUAL">iniziato</EVENT>
a leggere il libro

4. I_ACTION: "I_ACTION stands for intensional action. I_ACTIONS describe an action or situation which introduces another event as its argument, which must be in the text explicitly. Explicit performative predicates [...] are also included in this class" (ISO (2008): 49). A questa classe appartengono verbi come: *cercare, provare, tentare, ricercare, progettare, ostacolare, impedire, prevenire, cancellare, chiedere, ordinare, persuadere, comandare, promettere, assicurare, proporre, nominare, eleggere, ecc.* Nei seguenti esempi l'evento introdotto è marcato con la sottolineatura.

(36) Manuela ha <EVENT ...CLASS="I_ACTION">chiesto</EVENT> a
Silvia di rimandare l'appuntamento

(37) Il sindaco ha <EVENT ...CLASS="I_ACTION">assicurato</EVENT> che il nuovo asilo nido
sarà aperto a breve

(38) L'assemblea ha <EVENT ...CLASS="I_ACTION">proposto</EVENT> di nominare un nuovo
presidente

5. I_STATE: "They are similar to the events in the previous class. I_STATES also select for another event as their argument, but contrary to I_ACTIONS, they denote stative situations" (ISO (2008): ibid.). A questa classe appartengono verbi come: *credere, pensare, immaginare, sperare, desiderare, temere, odiare, avere paura, avere bisogno, dovere, potere, volere, riuscire, ecc.* Come per la classe precedente, anche nei seguenti esempi l'evento introdotto è marcato tramite sottolineatura.

(39) Marco <EVENT ...CLASS="I_STATE">deve</EVENT> portare
fuori il cane

(40) <EVENT ...class="I_STATE">Temevano</EVENT> di morire

(41) <EVENT ...class="I_STATE">Sperano</EVENT> di arrivare in tempo

6. OCCURRENCE: sono inclusi in questa classe tutti gli eventi che descrivono situazioni che accadono o occorrono nel mondo.

(42) L' uragano ha <EVENT

...CLASS="OCCURRENCE">provocato</EVENT> danni per miliardi di dollari

(43) Tre giornalisti della BBC sono stati <EVENT

...CLASS="OCCURRENCE">arrestati</EVENT> in Libia

(44) Il comando è <EVENT

...CLASS="OCCURRENCE">arrivato</EVENT> all'improvviso

7. STATE: questa classe descrive le circostanze in cui qualcosa viene ottenuto o rimane valido.

(45) Intere città <EVENT ...CLASS="STATE">distrette</EVENT> dall'uragano

(46) I profughi a Pantelleria <EVENT

...CLASS="STATE">sono</EVENT> migliaia ormai

(47) Luca <EVENT ...CLASS="OCCURRENCE">ha</EVENT> due gatti

L'individuazione del corretto valore dell'attributo CLASS si basa sulla valutazione sia di informazione sintattica che semantica. Proprio per questo motivo si assiste ad una certa variabilità tra i diversi tipi di classi: gli stessi verbi possono appartenere a classi diverse a seconda del contesto in cui compaiono. Per esempio, il verbo *cercare* negli esempi seguenti assume due valori diversi:

- (48) Marco ha <EVENT ...CLASS="OCCURRENCE">cercato</EVENT> il gatto ovunque
- (49) Marco ha <EVENT ...CLASS="I_ACTION">cercato</EVENT> di chiedere informazioni

Prendendo come punto di riferimento la classificazione di Vendler (1967) nelle quattro categorie azionali di *states*, *activities*, *achivements* e *accomplishments* si nota subito che la prima categoria coincide con il valore STATE dell'attributo CLASS. In particolare: i) alla classe vendleriana appartengono i predicati stativi, cioè quelli indicanti situazioni non dinamiche, non teliche ma durative ed esprimono, in genere, qualità, abitudini e abilità del soggetto; ii) il valore STATE viene utilizzato per descrivere circostanze in cui alcuni elementi rimangono validi. Si nota come le due definizioni sopra riportate tendano a sovrapporsi nei loro punti essenziali: entrambe, infatti, fanno riferimento a situazioni che durano nel tempo e che caratterizzano il soggetto a cui si riferiscono. Tralasciando il trattamento dei nomi e degli aggettivi eventivi, nell'annotazione della TimeBank per l'italiano (Caselli e Prodanof, 2010), così come nell'annotazione del corpus I-CAB (Magnini et al., 2006) secondo le specifiche TimeML, i punti in comune tendono a sfumare e, mentre per l'attribuzione della classe STATE ad alcuni verbi si è fatto coincidere il concetto di predicato stativo vendleriano con quello del valore STATE dell'attributo CLASS, come avviene per il verbo *conoscere* nell'esempio (50), per altri verbi questa logica non è stata seguita, come accade per il predicato *amare* in (51). Inoltre, mentre, vengono trattati in maniera corretta casi di ibridismo azionale caratterizzanti alcuni predicati, per cui a seconda del contesto in cui sono inseriti possono assumere valore STATE oppure OCCURRENCE, come accade a *trattare* negli esempi (52) e (53), per altri verbi non viene fatto lo stesso e il valore che possono assumere è univoco, indipendentemente dal contesto. Questo è il caso del *portare* negli esempi (54) e (55) o del verbo *spaccare* dell'esempio (56), in cui i verbi sono annotati come OCCURRENCE sebbene, in questi i casi, assumano una lettura stativa.

- (50) Da tempo gli agenti del commissariato <EVENT ...CLASS="STATE">conoscevano</EVENT> Lidia Valentini
- (51) Un filosofo che <EVENT ...CLASS="OCCURRENCE">ama</EVENT> il teatro
- (52) Gli Stati Uniti hanno <EVENT...CLASS="OCCURRENCE">trattato</EVENT> con Mosca
- (53) Si <EVENT ...CLASS="STATE">tratta</EVENT> di un dato positivo
- (54) La giovane <EVENT ...CLASS="OCCURRENCE">porta</EVENT> dentro di sé tutto il peso del proprio passato
- (55) La strada che <EVENT ...CLASS="OCCURRENCE">porta</EVENT> alla discarica
- (56) Il nuovo confine croato-sloveno che <EVENT ...CLASS="OCCURRENCE">spacca</EVENT> in due le popolazioni italiane dell'Istria

I verbi che nei due corpora di riferimento per l'italiano, compaiono sempre con valore STATE sono: *essere*, *avere*, il procomplementare *esserci*, *stare*, i participi passati di qualsiasi verbo quando non sono presenti complementi d'agente o causa efficiente.

Vi sono, infine, verbi che denotano stati secondo la definizione vendleriana per i quali, però, lo schema di annotazione TimeML prevede classi diverse da STATE. Questo è il caso del verbo *sapere* che non può mai avere lettura stativa ma, a seconda del contesto, può appartenere alle due classi di I_STATE od OCCURRENCE.

3.4.1 Proposta di integrazione per la marcatura dell'azione verbale

La mancanza di regole certe e precise nell'attribuzione del valore STATE all'attributo CLASS, come spiegato nel paragrafo precedente, costituisce un limite consistente nell'analisi del testo: se, infatti, nell'annotazione fossero marcati come STATE solo i predicati considerati stativi secondo la definizione vendleriana sarebbe possibile estrarre automaticamente almeno tutti i verbi appartenenti a questa classe azionale.

Per quanto riguarda le rimanenti tre classi vendleriane, ovvero activity, achievement e accomplishment, non esistono valori dell'attributo CLASS dedicati. Vengono tutte raggruppate nel generico valore OCCURRENCE. Esiste, quindi, una disparità di trattamento tra i verbi stativi per i quali è stato pensato, con i limiti evidenziati, un valore ad hoc, e le restanti classi azionali per cui non sono contemplati valori specifici. Questa scelta è stata dettata, forse, da una logica di semplificazione secondo la quale è più facile individuare un predicato stativo rispetto ad un predicato appartenente ad una delle altre tre classi azionali in quanto solo gli stati sono caratterizzati positivamente da un unico tratto ovvero la duratività, non presentano invece tratti telici o dinamici.

Alla luce di quello che è stato detto finora, un esperimento interessante sarebbe quello di arricchire l'annotazione prevista dallo schema TimeML in modo tale da integrare l'informazione relativa al valore azionale dei verbi eventivi. Secondo una prospettiva di annotazione automatica, e per facilitare la scelta tra i valori da parte degli annotatori, si potrebbe pensare di ampliare l'annotazione inserendo quattro ulteriori attributi: tre attributi differenti corrispondenti ai tre tratti azionali necessari ad identificare una classe vendleriana, ovvero dinamicità, duratività e telicità, ed un ultimo attributo per marcare la classe azionale vera e propria.

L'annotazione dell'azione verbale potrebbe essere suddivisa, quindi, in due fasi: la prima che preveda l'assegnazione da parte dell'annotatore di un valore booleano per ogni attributo corrispondente ad un tratto azionale. La seconda in cui tre valori booleani vengano combinati in modo tale che, dato che ogni classe azionale presenta una differente composizione dei tratti, sia un annotatore, sia un sistema automatico, possano derivare il corretto valore azionale del verbo preso in considerazione e possano inserirlo, manualmente o automaticamente, come valore dell'attributo introdotto per marcare le classi azionali.

È evidente come questa proposta di integrazione abbia limiti consistenti in termini sia di complessità del task che di impiego di tempo e risorse umane. Potrebbe comunque, essere ritenuta una modalità applicativa valida in task finalizzati ad individuare un unico tipo di classe azionale oppure per l'identificazione di casi di ibridismo o di commutazione azionale.

Capitolo IV

Crowdsourcing e annotazione dell'aspettualità: un esperimento

Il termine crowdsourcing è stato coniato nel 2006, quando Jeff Howe lo ha utilizzato in un articolo su Wired³⁵. Fare crowdsourcing, secondo questa definizione, significa appaltare un compito ad un vasto ed indefinito gruppo di persone (*crowd*, la folla), tramite una chiamata aperta a cui chiunque può rispondere.

Il termine crowdsourcing definisce, quindi, un modello di business basato sul lavoro distribuito attraverso il web. Un'azienda o un'istituzione richiede lo sviluppo di un progetto, di un servizio o di un prodotto ad un insieme distribuito di persone non già organizzate in una comunità virtuale.

Il crowdsourcing ha avuto la sua genesi nel movimento dei Software Open Source (Linux), ma ormai è utilizzato in diversi ambiti, dal marketing commerciale (Zooppa³⁶) al *marketing research*, dal settore amministrativo (Co-Create London³⁷) al settore creativo/culturale, tanto che Wikipedia³⁸ viene considerata da molti un esempio di crowdsourcing volontario.

Nell'ambito di questo lavoro, la piattaforma di crowdsourcing CrowdFlower³⁹ è stata utilizzata per portare a termine un task linguistico consistente nel valutare la capacità di individuare un particolare tipo di aspetto verbale, ovvero quello afferente alla classe dell'imperfettività gnomica, da parte di soggetti non esperti le cui uniche conoscenze in materia derivano dalle istruzioni che sono state loro fornite per svolgere il compito.

³⁵ <http://www.wired.com/wired/archive/14.06/crowds.html>

³⁶ <http://zooppa.com/>

³⁷ <http://www.cocreatelondon.com/>

³⁸ http://it.wikipedia.org/wiki/Pagina_principale

³⁹ <http://crowdfower.com/>

4.1 Ambiti di applicazione e limiti del Crowdsourcing

Howe (2009) individua quattro tipologie di crowdsourcing:

1. *Crowd-wisdom*, ovvero l'intelligenza collettiva⁴⁰: consiste nel mettere a frutto la conoscenza dei gruppi, in quanto superiore alla conoscenza dei singoli;
2. *Crowd-creation*: utilizza non solo la conoscenza ma anche l'energia creativa di persone comuni per lo svolgimento di attività;
3. *Crowd-voting*: adopera le scelte e i giudizi delle persone comuni per organizzare le informazioni (l'esempio più noto è Google);
4. *Crowd-funding*: permette ai gruppi di raccogliere auto-finanziamenti.

L'autore sottolinea come spesso i progetti più fortunati derivino dalla combinazione di questi quattro approcci.

Anche Jenkins (2006) individua quattro diverse modalità di cultura partecipativa:

1. *Affiliation*: creazione di comunità, formali ed informali, accentrate intorno a diverse forme di media (per esempio Facebook);
2. *Expression*: produzione collettiva di contenuti e nuove forme creative;
3. *Collaborative problem-solving*: lavoro di gruppo allo scopo di portare a termine obiettivi e sviluppare la conoscenza (per esempio Wikipedia);
4. *Circulation*: dare valore al flusso dei media, come nel caso dei blog o dei podcasting.

In realtà dalle categorie sopra elencate non è chiaro quali siano i limiti di ciò che è definibile come crowdsourcing. Alcuni ritengono che il crowdsourcing si possa configurare semplicemente con le azioni volte a produrre una diffusione di progetti altrui. Per esempio, un *Mi piace* su Facebook crea un *output* di visibilità. La visibilità consente la condivisione della creatività e favorisce la creazione di

⁴⁰ L'*intelligenza collettiva*, termine coniato da Pierre Lèvy, è la capacità delle comunità virtuali di utilizzare la competenza combinata dei loro membri; l'organizzazione dei pubblici in quelle che Lèvy chiama *comunità di sapere* consente loro di esercitare un potere aggregato maggiore (Jenkins, 2006).

un'organizzazione sociale del sapere. La *folla*, diffondendo prodotti che giudica meritevoli, crea un ulteriore valore aggiunto per quei prodotti. Google stesso è considerato da Howe un prodotto creato dalla *Crowd-wisdom*, in quanto i risultati di ricerca sono selezionati e ordinati anche in base al numero di visite ricevute da un sito. Benkler (2007), riguardo a Google, evidenzia come la “somma degli effetti delle azioni individuali, anche quando non consapevolmente cooperative, [ha permesso] di produrre un nuovo e più ricco ambiente informazionale”.

Vi sono prodotti *crowdsourced* che sono emersi in maniera spontanea e auto-organizzata. Tuttavia, come sostiene Shirky (2008), specie per i compiti più complessi, è necessario un certo livello di controllo e coordinamento. A tal fine è stato sviluppato un nuovo approccio al lavoro attraverso la creazione dell'ambiente e delle condizioni che permettono a persone comuni di esprimere le proprie competenze.

L'utilizzo del crowdsourcing pone, però, una serie di problemi di carattere etico e legale, come l'incerto inquadramento contrattuale dei lavoratori (sono da considerarsi lavoratori indipendenti o impiegati?), il potenziale sfruttamento (anche se secondo Biewald, CEO di Crowdfunder, le opportunità che si aprono andranno a compensare il possibile *dumping*⁴¹), il salario minimo (come stimare equamente il valore dei micro compiti? Oggi lavorando a tempo pieno si guadagna circa un migliaio di dollari).

Secondo Howe, comunque, le aziende che guardano alle persone comuni solo come ad risorsa di lavoro a basso costo sono destinate a fallire: una comunità cosciente di essere sfruttata, infatti, priva l'azienda di credibilità e fiducia, le due caratteristiche indispensabili per portare avanti qualsiasi azione di collaborazione tra pari.

⁴¹ Con il termine *dumping* si indica, nell'ambito del diritto internazionale una procedura di vendita di un bene o di un servizio su di un mercato estero (mercato di importazione) ad un prezzo inferiore rispetto quello di vendita (o, addirittura, a quello di produzione) del medesimo prodotto sul mercato di origine (mercato di esportazione).

4.2 Utilizzo dei sistemi di crowdsourcing per l'annotazione

Molti progetti riguardanti l'annotazione su larga scala come TreeBank (Marcus et al., 1993), PropBank (Palmer et al., 2005), TimeBank (Pustejovsky et al., 2003), FrameNet (Baker et al., 1998), SemCor (Miller et al., 1993), hanno giocato un ruolo importante per lo sviluppo di nuove idee, compiti ed algoritmi. La costruzione dei relativi *dataset* è risultata, però, estremamente costosa in termini di ore di lavoro e di costi; la prestazione di molti task di NLP è, infatti, limitata dalla quantità e dalla qualità dei dati di training a disposizione (Banko e Brill, 2001). Per sopperire a questo limite, una promettente alternativa, almeno per alcuni task, è costituita da collezioni di dati raccolti ed elaborati da annotatori non esperti tramite strumenti di crowdsourcing.

4.2.1 Alcune esperienze di utilizzo di sistemi di crowdsourcing

L'idea di collezionare annotazioni tramite contributi volontari è stata utilizzata per una vasta varietà di task (Snow et al., 2008). Luis von Ahn è stato il primo a collezionare dati attraverso giochi in rete come ESPGame per l'annotazione di immagini (von Ahn e Dabbish, 2004) e Verbosity per l'annotazione di relazioni tra parole (von Ahn et al., 2006).

Anche l'iniziativa Open Mind (Stork, 1999) ha utilizzato un approccio simile, cercando di rendere i task di annotazione del senso delle parole (Chklovski e Mihalcea, 2002) e di relazioni tra parole di senso comune (Singh, 2002) sufficientemente semplici e "divertenti" da indurre gli utenti ad annotare volontariamente i dati.

Negli ultimi anni si è assistito ad un incremento consistente di esperimenti nei quali per l'annotazione sono stati utilizzati sistemi di crowdsourcing come Amazon Mechanical Turk⁴²(AMT).

AMT è un *online labor market* attraverso il quale i lavoratori portano a termine semplici task per i quali vengono pagati con esigue somme di denaro. Per quanto riguarda le prestazioni del sistema, in (Su et al., 2007) si evidenzia come utilizzando il Mechanical Turk per task di riconoscimento di entità nominate

⁴² <http://mturk.com>

riguardanti nomi di hotel e di estrazione di attributi quali l'età o la marca e il modello di un prodotto, i risultati ottenuti siano caratterizzati da un altro grado di accuratezza se comparati con il gold standard.

In (Kittur et al., 2008) vengono messe a confronto le valutazioni ottenute tramite AMT sulla qualità di articoli estratti da Wikipedia e articoli prodotti da professionisti, individuando test di validazione importanti per assicurare buoni risultati nella valutazione.

Infine, in diversi studi recenti AMT è stato anche utilizzato senza effettuare successivamente una comparazione con un gold standard. Kaisser e Lowe (2008), per esempio, hanno utilizzato l'AMT per la costruzione di un *dataset* per un task di *Question Answering*, individuando, tra un vasto insieme di frasi, le risposte a 8107 domande.

In generale, i dati ottenuti tramite il lavoro di volontari non esperti si presentano in quantità più consistenti ma sono caratterizzati da maggiori inconsistenze rispetto a quelli prodotti da esperti. Queste caratteristiche possono comunque essere sfruttate per aggregare annotazioni svolte da soggetti differenti al fine di ottenere elevati livelli di affidabilità.

4.2.2 Crowdfower

Crowdfower è un'azienda specializzata in servizi di crowdsourcing. I suoi collaboratori, tra cui circa diecimila italiani, sono sparsi in tutto il mondo e si prendono carico ciascuno di un piccolo task, semplice e ripetitivo. In genere, si ricorre a Crowdfower per svolgere compiti che è difficile o impossibile automatizzare efficacemente ma che, una volta esternalizzati, possono essere svolti con rapidità e accuratezza.

Il punto di forza di CrowdFlower è la qualità del risultato, ovvero i meccanismi studiati per accertare l'affidabilità del lavoro svolto e la validazione delle singole risposte. Infatti, se due persone autonomamente forniscono la medesima risposta, la probabilità che questa sia corretta è molto alta; se un collaboratore che comincia a lavorare produce risultati progressivamente più complessi, identici a quelli di un suo collega risultato nel tempo affidabile, si avrà un primo riscontro anche

dell'affidabilità del nuovo arrivato. In particolare, CrowdFlower per valutare l'affidabilità dei lavoratori prevede la costruzione di un gold standard: un lavoratore è considerato affidabile quando, giudicato un numero minimo di frasi appartenenti al gold (di default è 4, ma il numero può essere cambiato), fornisce risposte corrette ad almeno il 70% delle domande proposte.

Il riferimento è costituito, dunque, dal Mechanical Turk di Amazon, ma il valore aggiunto da CrowdFlower risiede nel modo in cui sono stati studiati e articolati i compiti.

L'uso di CrowdFlower per l'annotazione di risorse linguistiche richiede un'accurata definizione delle specifiche dei task. Gli individui che svolgeranno questi compiti non hanno conoscenze linguistiche specialistiche né una particolare esperienza in campo linguistico, per questo è necessario che le istruzioni fornite ai lavoratori per portare a termine i task siano semplici e chiare.

4.3 Individuazione dei verbi afferenti alla classe aspettuale della imperfettività gnomica tramite CrowdFlower

Nell'ambito di questo lavoro, si è deciso di utilizzare CrowdFlower per valutare la capacità di utenti generici, senza, quindi, conoscenze linguistiche specialistiche, di individuare frasi che presentano verbi caratterizzati da aspetto afferente alla classe dell'imperfettività gnomica, all'interno di un corpus contenente frasi i cui verbi presentano anche le seguenti tipologie di aspetto verbale: imperfettivo progressivo, imperfettivo continuo, perfettivo iterativo ed episodico.

Come già introdotto nel cap. 3.3.1, la scelta di focalizzare l'esperimento sul riconoscimento di questo tipo di aspetto verbale è dettata dal fatto che a questa classe aspettuale appartengono molte informazioni cosiddette di senso comune (Singh, 2002). L'aspetto verbale potrebbe configurarsi, quindi, come una caratteristica discriminante per individuare questo tipo di informazione e, conseguentemente, potrebbe essere sfruttata utilmente in strumenti che consentono l'estrazione automatica di informazione di senso comune da testi.

4.3.1 Esperimenti correlati

Riguardo all'identificazione delle caratteristiche verbali, sono stati svolti esperimenti mirati alla distinzione tra forme generiche e specifiche di predicati nominali (Suh, 2006), alla determinazione della classe lessicale dei verbi (Brent 1990; Siegel, 1999) ed anche alla classificazione supervisionata di frasi abituali ed episodiche (Mathew e Katz, 2009). Come nei task proposti da Suh e da Mathew e Katz, anche l'esperimento svolto in questa tesi riguarda una classificazione dei verbi in contesto tanto che potrebbe essere definito come un particolare task di word sense disambiguation, anche se nel caso specifico non si tratta di un task di classificazione automatica, bensì manuale.

Il presente esperimento si differenzia, comunque, da quello di Brent e Siegel che classificano solo due particolari classi verbali: gli eventi e gli stati.

Da notare, però, che esiste una correlazione abbastanza stretta tra l'approccio di Siegel e quello di Mathew e Katz: la distinzione tra predicati eventivi e stativi, infatti, ha molto in comune con l'individuazione di frasi abituali ed episodiche su cui si concentra il compito di Mathew e Katz che infatti, per rendere più efficiente il loro sistema, riutilizzano molte delle caratteristiche individuate da Siegel.

4.3.2 Il data set per l'esperimento

Le frasi che compongono il data set costruito per l'esperimento sono in totale 375 e si suddividono in quattro categorie:

1. 150 frasi i cui verbi afferiscono alla classe dell'imperfettività gnomica;
2. 75 frasi perfettive iterative;
3. 75 frasi perfettive episodiche;
4. 75 tra frasi che presentano verbi di aspetto imperfettivo continuo (36) e progressivo (39).

Tutte le frasi del data set presentano tempi al passato: questa scelta è motivata dal fatto che solo nel passato è possibile marcare tutte le tipologie di aspetto e visualizzarne lo spettro completo: dal perfettivo all'imperfettivo.

La lista completa delle frasi del data set è riportata in Appendice B.

Per quanto riguarda le ultime tre categorie aspettuali elencate precedentemente, tutte le frasi presenti nel corpus sono estratte da I-CAB (Italian Content Annotation Bank), una collezione di testi costituita da 525 articoli estratti dal quotidiano trentino L'Adige⁴³ (per una trattazione più dettagliata si rimanda al cap. 4.3.2.2). Per quanto concerne, invece, la classe dell'imperfettività gnomica circa metà delle frasi è stata estratta da I-CAB, l'altra è stata ripresa da due manuali di storia: "Medioevo. I caratteri originali di un'età di transizione" di Giovanni Vitolo (2000) e "Storia di Roma" di Giovanni Brizzi (1997). Questa integrazione si è resa necessaria a causa della particolare natura del corpus di riferimento: negli articoli di giornale, infatti, difficilmente compare un numero cospicuo di frasi afferenti alla classe dell'imperfettività gnomica: espressioni abituali, leggi e regolarità generali non si addicono ad articoli in cui, per la maggior parte, si riportano e descrivono eventi puntali avvenuti nel passato. In questo contesto, la scelta di manuali di storia come testi integrativi è stata dettata dalla relativa semplicità di individuare nell'ambito di questa disciplina affermazioni universalmente valide o frasi che esprimono abitudini che si sono protratte per un arco di tempo significativo.

Riguardo la classe dell'imperfettività gnomica, bisogna evidenziare altre due caratteristiche del corpus per l'esperimento: (i) tutte le frasi contengono un unico tempo verbale: l'imperfetto; (ii) le frasi sono bilanciate tra quelle che presentano una costruzione copulativa (69) e quelle in cui tale costruzione non è presente (81).

Un bilanciamento simile a quello descritto nel paragrafo precedente è stato effettuato anche per le frasi con aspetto imperfettivo progressivo: 16 su 39 presentano una perifrasi progressiva esplicita mentre le altre esprimono l'aspetto progressivo attraverso il tempo verbale dell'imperfetto.

⁴³ <http://www.ladige.it/>

Le frasi del data set sono state individuate per la maggior parte attraverso un lavoro manuale di ricerca e lettura attenta di ciascun file o pagina di manuale. Solo in alcuni casi inerenti le frasi estratte da I-CAB, in particolare per l'individuazione dell'aspetto progressivo e abituale, è stato possibile utilizzare strumenti di ricerca testuale, come il comando *grep*, per trovare in maniera semi-automatica particolari segnalatori (per esempio, gli avverbi “mentre” o “spesso”) che possono comparire in concomitanza con questi due tipi di aspetto verbale. A questa fase è seguita poi un'attenta revisione dei risultati ottenuti e una scelta accurata delle frasi corrette e più idonee in termini semantici ad essere collezionate.

4.3.2.1 Il gold standard

Parallelamente alla costruzione del corpus è stato creato un gold standard composto da 35 frasi così bilanciate:

- 11 frasi afferenti alla classe aspettuale dell'imperfettività gnomica;
- 5 con aspetto imperfettivo progressivo;
- 5 con aspetto imperfettivo continuo;
- 7 perfettive iterative;
- 7 perfettive episodiche.

La lista completa delle frasi del data set è riportata in Appendice C.

4.3.2.2 Descrizione del corpus I-CAB

Negli ultimi anni sono state avviate numerose iniziative finalizzate alla creazione di corpora per la valutazione di task di NLP.

Recentemente, all'interno del Programma ACE (Automatic Content Extraction)⁴⁴, è stato sviluppato un insieme di schemi di annotazione riguardanti task inerenti il campo dell'Estrazione di Informazione da testi. Tali schemi prendono in considerazione espressioni temporali, entità e relazioni tra entità. Sulla base delle risorse ottenute sono state organizzate con successo diverse campagne di valutazione (TERN 2004 e 2005, ACE 2000-2006).

⁴⁴ <http://www.nist.gov/speech/tests/ace/>

In questa prospettiva si colloca I-CAB, un corpus italiano di articoli di giornale annotato semanticamente. In particolare, I-CAB contiene annotazioni relative a Espressioni Temporalì (<tre anni>), entità Geo-Politiche (<Italia>, <Trento>), di tipo Persona (<George W. Bush>, <la tua [famiglia]>), Organizzazione (<La [Microsoft]>, <Università di Pisa>) e Luogo (<Il [Po]>, <Via Nazionale, 12>).

Per l'annotazione del corpus è stato utilizzato il software Callisto⁴⁵, sviluppato e distribuito gratuitamente dalla MITRE Corporation.

La creazione di I-CAB è parte del progetto triennale Ontotext⁴⁶ promosso dalla Fondazione Bruno Kessler⁴⁷ e finanziato dalla Provincia Autonoma di Trento.

Questo progetto aveva come scopo lo studio e lo sviluppo di tecnologie innovative per l'estrazione dell'informazione e della conoscenza nell'ambito del Web Semantico.

Il corpus I-CAB è stato annotato manualmente e vuole configurarsi come un corpus di riferimento per task di riconoscimento e normalizzazione di espressioni temporali, entità e relazioni tra entità. Seguendo una politica di riutilizzo dei linguaggi di annotazione già disponibili, sono stati adottati i formalismi sviluppati all'interno del programma ACE ma a causa delle notevoli differenze morfo-sintattiche tra l'inglese e l'italiano è stata necessaria una revisione e un adattamento delle linee guida inglesi.

I-CAB è composto di 525 articoli del quotidiano locale "L'Adige" distribuito nella Provincia di Trento. Gli articoli sono tratti da 4 differenti giornate (7-8 settembre 2004 e 7-8 ottobre 2004) e sono raggruppati in 5 categorie: Attualità (87 articoli), Cultura (72 articoli), Economia (54 articoli), Sport (123 articoli) e Trento (189 articoli).

I-CAB si divide in una sezione di training e in una sezione di test, contenenti rispettivamente 335 e 190 file. In totale, il corpus è composto da circa 182500 parole: 113000 nella sezione di training e 69000 nella sezione di test.

⁴⁵ <http://callisto.mitre.org/>

⁴⁶ <http://tcc.itc.it/projects/ontotext/>

⁴⁷ <http://www.fbk.eu>

4.3.3 Istruzioni per l'esperimento

Ai lavoratori di CrowdFlower è stato chiesto di individuare, tra quelle presenti nel corpus, le frasi esprimenti leggi, regolarità generali o abitudini.

In particolare, è stato chiesto loro di indicare scegliendo tra due sole opzioni, vero o falso, se la frase letta esprimeva una proprietà generale, una legge, una regolarità o un'abitudine che nel passato è rimasta valida per un certo periodo di tempo.

In Figura 1 un'immagine della pagina di CrowdFlower contenente le istruzioni così come presentate ai lavoratori.



Figura 1. Pagina delle istruzioni dal sito di CrowdFlower

Le istruzioni fornite ai lavoratori per svolgere questo task sono riportate di seguito:

ISTRUZIONI: individua leggi, regolarità generali e abitudini

In questo compito devi decidere se è vero o falso che le frasi proposte esprimono una legge, una regolarità generale o un'abitudine che nel passato è rimasta valida per un certo arco di tempo.

ATTENZIONE: per decidere concentrati sull'evento espresso dal verbo in grassetto nella frase.

In generale, le frasi che esprimono **leggi, regolarità generali**, oppure **abitudini** che hanno caratterizzato un certo arco di tempo nel passato appartengono alle seguenti categorie:

1) **Nozioni storiche, geografiche, mediche:**

- Es: *Giulio Cesare **discendeva** da un'antica famiglia patrizia*
- Es: *La mummia del Similaun **era** probabilmente un uomo di alto rango*
- Es: *La Gallia **corrispondeva** ad un'ampia regione dell'Europa centro-occidentale*
- Es: *Cartagine si **trovava** nell'odierna Tunisia*
- Es: *Prima della scoperta del vaccino, il vaiolo **era** una malattia mortale*

2) **Qualità generali e caratteristiche che hanno caratterizzato individui o organizzazioni:**

- Es: *Secondo la mitologia, Elena **era** la più bella donna del mondo*
- Es: *Giovanni Falcone **era** un magistrato italiano*
- Es: *“Un ponte per...” **era** un'organizzazione molto attiva in Iraq*

3) **Abitudini che non hanno riguardato un singolo accadimento ma sono perdurate nel tempo:**

- Es: *Luca **giocava** a tennis ogni giorno*
- Es: *La ragazza **sedeva** spesso davanti a casa*
- Es: *Simona **viveva** a Roma*
- Es: *Gianna **lavorava** in centro*
- Es: *L'organizzazione umanitaria italiana **inviava** aiuti alimentari alla popolazione civile*
- Es: *La compagnia teatrale **proponeva** spesso rappresentazioni tratte dai romanzi di Daniel Pennac*
- Es: *Laura **teneva** abitualmente corsi di fotografia*

NON esprimono leggi, regolarità o abitudini frasi che riportano:

1) Avvenimenti passati descritti però come in corso di svolgimento:

- Es: *La ragazza **stava uscendo** di casa*
- Es: *La bomba è scoppiata mentre tutti **dormivano***

2) Singoli accadimenti e situazioni specifiche che hanno avuto luogo in un tempo passato:

- Es: *Durante l'interrogatorio, la ragazza **rispondeva** tranquilla*
- Es: *Quella sera l'uomo **entrava ed usciva** di casa in preda all'agitazione*
- Es: *Durante la riunione, la dirigente **interveniva** spesso*
- Es: *Per l'intera durata dello spettacolo, Marco **diceva** solo di voler tornare a casa*
- Es: *Pamuk **ha vinto** il premio Nobel per la letteratura*
- Es: *Luisa **ha sostenuto** ieri l'esame di linguistica computazionale*
- Es: *Luca **ha incontrato** Marco ieri pomeriggio*

3) Avvenimenti che si sono ripetuti nel tempo un certo numero di volte:

- Es: *Luca è **caduto** due volte durante la gara*
- Es: *Il terremoto **ha interessato** spesso l'Italia*
- Es: *Gianni **ha chiesto** raramente aiuto al padre*
- Es: *Tra luglio e agosto l'autrice **ha presentato** il suo libro in varie manifestazioni culturali*

4.3.4 Risultati

Per questo esperimento è stato richiesto un minimo di 3 giudizi per ogni frase. Il canale scelto per lo svolgimento del task è stato Mechanical Turk di Amazon (AMT).

L'unità base di lavoro su AMT è chiamata Human Intelligence Task (HIT) e nel caso di questo esperimento è costituita da 5 frasi. Per ogni HIT i lavoratori sono

stati pagati \$0,05. Per quanto riguarda, infine, la nazionalità dei lavoratori si è scelto di includere solo i residenti in Italia.

L'esperimento è stato portato a termine da 4 lavoratori che, casualmente, sono esponenti delle principali macrozone in cui si suddivide il territorio italiano: Trento a nord, Pisa e Jesi al centro, Francavilla Di Sicilia al sud. Il task è stato completato in 4 giorni.

4.3.4.1 Calcolo dell'affidabilità dei lavoratori

In Tabella 1 è riportato per ciascun lavoratore il numero di frasi valutate e la sua affidabilità (secondo i criteri di CrowdFlower) espressa in percentuale. Come si nota dai dati riportati in tabella, l'affidabilità dei quattro lavoratori è alta, oscilla, infatti, tra 96% e il 100%. Dalla lettura dei dati si evidenzia un calo, che può definirsi fisiologico, di affidabilità in relazione all'aumentare del numero di frasi giudicate.

Worker ID	# frasi giudicate	Affidabilità
2139638	432	96%
1835736	425	96%
2798057	308	97%
2812049	115	100%

Tabella 1. Numero di frasi giudicate e percentuale di affidabilità per ciascun lavoratore

In Figura 2 è riportato il grafico delle frasi giudicate da ogni lavoratore. Due lavoratori hanno portato a termine il compito per intero mentre per quanto riguarda i restanti due: uno ha svolto circa 3/4 del task e l'altro circa 1/4.

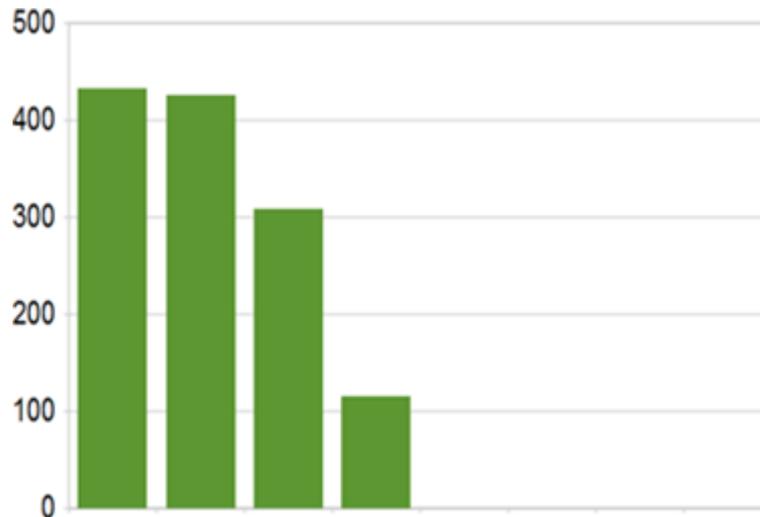


Figura 2. Numero di frasi giudicate da ciascun lavoratore

La media di affidabilità per i quattro lavoratori corrisponde al 97%. Come si vede in Tabella 2 non ci sono stati lavoratori inaffidabili, tutti hanno superato la soglia del 70% di affidabilità stabilita da CrowdFlower.

Affidabilità media per il Gold Standard	
Affidabilità dei lavoratori	97%
Tutti i lavoratori	97%
Lavoratori inaffidabili	0%

Tabella 2. Affidabilità media per il gold standard

4.3.4.2 Agreement tra lavoratori

In Tabella 3 sono riportati i calcoli relativi all'agreement delle sole frasi con confidence=1, ottenuti applicando la seguente formula:

$$agreement = \frac{\text{veri positivi} + \text{veri negativi}}{\text{numero totale di frasi del corpus senza ripetizioni}}$$

agreement totale	88%
agreement sulle frasi vere	92,7%
agreement sulle frasi false	84,9%
agreement sulle frasi gnomiche	97,1%
agreement sulle frasi con aspetto progressivo	98,4%
agreement sulle frasi con aspetto continuo	98,4%
agreement sulle frasi iterative	95,2%
agreement sulle frasi episodiche	99%

Tabella 3. Agreement tra lavoratori distinto per tipologia di frase

Dai risultati mostrati in tabella, si nota come ci sia un alto tasso di accordo tra i giudizi espressi dai lavoratori. Il valore dell'agreement su tutte le frasi giudicate, infatti, raggiunge l'88%. I valori massimi sono registrati per le frasi episodiche (99%) e imperfettive progressive e continue (98,4% per entrambe). Il valore di agreement più basso riguarda invece le frasi iterative (95,2%). L'accordo sulle frasi afferenti alla classe dell'imperfettività gnomiche è del 97,1% e si pone, quindi, come intermedio tra le percentuali di accordo per le precedenti tipologie di frase. Da notare però che, sebbene da una valutazione superficiale le percentuali di agreement relative alle frasi iterative ed episodiche possano sembrare decisamente diverse, un calcolo del Chi Quadrato fornisce un valore pari a 1,86. Ciò significa che la discrepanza tra i due risultati non è significativamente alta, in quanto il numero risultante è minore di 3,841, ossia del valore del Chi Quadrato con un grado di libertà al livello del 5%. In sintesi, il fatto che la percentuale di agreement sulle frasi iterative sia inferiore a quella sulle episodiche non indica, in realtà, che il task è più semplice in relazione alla seconda tipologia di frasi piuttosto che alla prima. Ciò vuol dire che non si possono fare generalizzazioni sulla base della differenza tra le due percentuali.

4.3.4.3 Analisi quantitativa dei risultati

Per quanto riguarda l'analisi dei risultati ottenuti dall'esperimento, il parametro di riferimento utilizzato per la valutazione è stato quello dell'accuracy, ossia la percentuale di risultati veri (sia veri positivi che veri negativi) nella popolazione, definita come:

$$accuracy = \frac{\text{veri positivi} + \text{veri negativi}}{\text{veri positivi} + \text{falsi positivi} + \text{veri negativi} + \text{falsi negativi}}$$

L'accuracy totale ottenuta nell'ambito di questo esperimento è del 95%. Questo dato dimostra la capacità di soggetti non linguisticamente addestrati nel distinguere diverse tipologie di aspettualità. Conferma, inoltre, la validità delle istruzioni presentate ai lavoratori.

Per un'analisi più fine sono stati calcolati anche due valori di accuracy distinti: uno per le frasi che esprimono leggi, regolarità generali o abitudini, e che quindi hanno giudizio vero, e uno per le altre frasi, cosiddette false. I risultati sono riportati in Tabella 4.

Accuracy per le frasi vere	97%
Accuracy per le frasi false	93%

Tabella 4. Accuracy per frasi vere e false

Sebbene a prima vista le due percentuali appaiano simili, un calcolo del Chi Quadrato fornisce un valore di 7,08. Questo vuol dire che la differenza tra i due risultati è significativamente diversa, in quanto il risultato è maggiore di 6,635, ossia del valore del Chi Quadrato con un grado di libertà dell'1%. Il fatto, dunque, che la percentuale di accuracy sulle frasi vere sia inferiore a quella sulle false, denota come davvero in questo task sia più facile riconoscere le frasi vere rispetto alle false. Ciò potrebbe dipendere dalla struttura delle istruzioni oppure da caratteristiche intrinseche del task.

La Figura 3 mostra la visualizzazione grafica delle percentuali di accuracy totale, di accuracy per le frasi vere e per le frasi false.

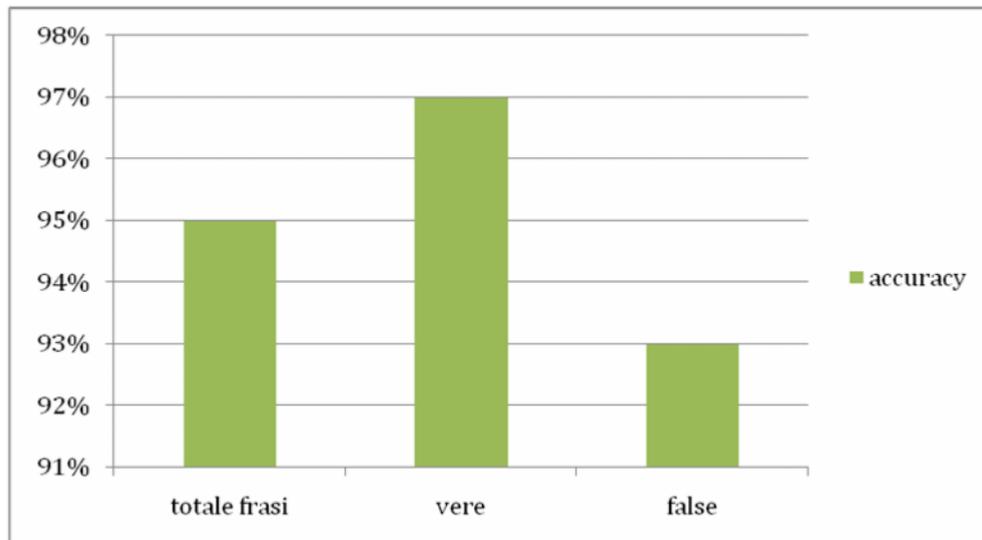


Figura 3. Percentuali dell'accuracy totale, per le frasi vere e false

In Tabella 5 sono riportate le diverse percentuali di accuracy per le 5 tipologie di frasi presenti nel corpus, mentre in Figura 4 le stesse percentuali sono rappresentate graficamente.

Tipologia di frasi	Accuracy
frasi afferenti alla classe dell'imperfettività gnomica	97%
frasi con verbi di aspetto progressivo	94%
frasi con verbi di aspetto continuo	95%
frasi iterative	88%
frasi episodiche	98%

Tabella 5. Accuracy per tipologia di frasi

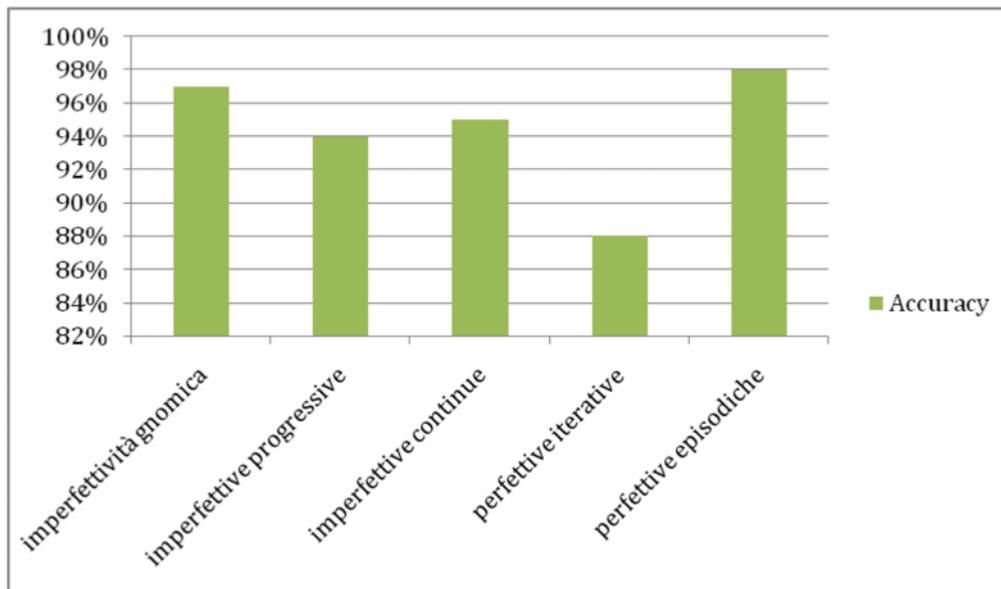


Figura 4. Percentuale di accuracy per tipologia di frase

L'accuracy più elevata (98%) riguarda le frasi episodiche, quelle cioè il cui verbo indica un evento specifico, che ha carattere occasionale e si presenta senza regolarità.

Queste frasi sono caratterizzate da verbi al passato prossimo e molto spesso anche dalla presenza di espressioni temporali che tendono a collocare l'evento espresso dal verbo in un unico ed irripetibile spazio temporale, per esempio "ieri", "l'anno scorso", "nel 1994", "alcuni giorni fa". Probabilmente, almeno a livello intuitivo, i lavoratori hanno percepito queste caratteristiche come distanti da quelle che denotano leggi, regolarità ed abitudini.

Dall'analisi dei dati si nota, poi, che il secondo valore di accuracy più elevato (97%) riguarda proprio la tipologia di frasi per cui era stato chiesto il riconoscimento. Questo dato conferma, come già quello sull'accuracy complessiva, la capacità di lavoratori non esperti di individuare un particolare tipo di aspetto verbale quando vengono loro fornite istruzioni chiare e semplici da seguire corredate da un buon numero di esempi che chiariscono in maniera più concreta le teorie sulla base delle quali queste istruzioni sono state redatte. Tali esempi, nell'ambito di questo esperimento, sono stati opportunamente classificati in categorie distinte in modo da illustrare e spiegare meglio la semantica delle

frasi in cui compare con maggiore probabilità un verbo il cui aspetto è riferibile alla classe dell'imperfettività gnomica.

Le frasi caratterizzate da verbi con aspetto imperfettivo, rispettivamente continuo e progressivo, presentano percentuali di accuracy molto simili: 94% per i verbi con aspetto progressivo, 95% per i verbi con aspetto continuo. È interessante notare come questa classe di frasi sia stata più efficacemente riconosciuta come diversa e distinta dalla classe dei verbi con aspetto imperfettivo gnomico rispetto alla classe delle frasi iterative sebbene queste ultime siano caratterizzate dal tempo verbale del passato prossimo mentre nelle frasi precedenti, così come in quelle afferenti alla classe dell'imperfettività gnomico, sia presente, nella maggior parte dei casi, l'imperfetto.

Le frasi iterative, infine, sono quelle che presentano il valore di accuracy più basso (88%). Questo dato è spiegabile con la difficoltà oggettiva nel distinguere queste frasi dalle abituali. Entrambe, infatti, denotano avvenimenti che si sono ripetuti nel tempo ma mentre per le iterative è possibile indicare il numero di volte in cui l'evento espresso dal verbo è accaduto, per le abituali non è possibile specificare il numero delle iterazioni. I dati dimostrano che questa sottile differenza è risultata poco intuitiva per i lavoratori che non hanno prestato molta attenzione neppure alla differenza di tempo verbale (passato prossimo vs imperfetto) che caratterizza queste due diverse tipologie di frasi.

Delle 375 frasi proposte ai lavoratori solo per 45 sono stati espressi giudizi contrastanti. Ognuna di queste 45 frasi è stata giudicata da 3 lavoratori.

In appendice D sono raccolti i casi di discordanza.

4.3.4.4 Analisi qualitativa dei risultati

Di seguito viene presentata una lista degli elementi testuali che hanno contribuito a fare in modo che gli annotatori potessero identificare con maggiore o minore facilità un aspetto verbale come appartenente alla classe dell'imperfettività gnomico, così come richiesto dall'esperimento.

Tali elementi possono essere classificati secondo una sorta di gerarchia di complessità: ve ne sono alcuni la cui presenza è un indizio forte dell'appartenenza del predicato ad una determinata classe aspettuale, altri che per poter essere interpretati in maniera corretta devono essere combinati tra di loro, altri ancora che si presentano in costruzioni sintattiche particolari.

Di seguito viene proposta un'analisi dettagliata dei casi in cui gli annotatori hanno espresso giudizi discordanti ordinati secondo la gerarchia di complessità degli elementi linguistici presenti nelle frasi.

Gli avverbi quantificazionali e iterativi

Da un'analisi qualitativa delle frasi che presentano discordanze si nota come in 16 frasi su 45 sia presente l'avverbio "spesso".

Le frasi che presentano discordanze in relazione a questo elemento appartengono a due differenti tipologie:

1. frasi perfettive iterative scambiate per imperfettive abituali;
2. frasi imperfettive abituali non riconosciute come tali.

Relativamente alla prima tipologia di frasi ne sono state riscontrate 12 in cui è presente l'avverbio "spesso" e per cui almeno un annotatore su 3 ha dato un giudizio sbagliato:

- (1) *Durante il rapimento il Presidente Ciampi ha telefonato **spesso** alle famiglie delle ragazze sequestrate*
- (2) *Simona Pari ha raccontato **spesso** a Emiliano della sua esperienza in Iraq*
- (3) *L'organizzazione non governativa "Un ponte per..." ha operato **spesso** in Iraq*
- (4) *La Polizia postale di Reggio Calabria ha combattuto **spesso** il fenomeno delle truffe all'Inps*
- (5) *L'organizzazione non governativa "Un ponte per..." ha inviato **spesso** i propri volontari in Iraq*
- (6) *Il medico del pronto soccorso è intervenuto **spesso** in caso di incidenti stradali*
- (7) *Il pub Simposio ha proposto **spesso** spettacoli teatrali*
- (8) *I talenti i "cervelli" italiani sono migrati **spesso** verso ambienti di ricerca più liberi e meritocratici*

- (9) *L'associazione Apeiron ha gestito spesso progetti di solidarietà in Nepal*
(10) *La qualità dei servizi è apparsa spesso in ribasso*
(11) *Negli ultimi anni l'acquisto del corredo scolastico ha dissanguato spesso il portafoglio delle famiglie*
(12) *Spesso le donne sono state escluse dalla storia della scienza*

Nella maggior parte degli esempi precedenti i lavoratori hanno associato l'avverbiale “spesso” al concetto di abitudine non prendendo in considerazione il fatto che tale avverbio può presentarsi anche in correlazione con tempi verbali perfettivi come accade nelle frasi iterative.

Mentre, infatti, in frasi come “L'organizzazione non governativa Un ponte per... ha operato spesso in Iraq” oppure come “Il pub Simposio ha proposto spesso spettacoli teatrali” è possibile determinare il numero di iterazioni dell'evento espresso dal verbo, la presenza dell'avverbiale “spesso” in correlazione con un aspetto verbale imperfettivo consente solo una lettura abituale, in quanto questo avverbio non si riferisce al numero di sotto-eventi espressi dal verbo ma piuttosto alla loro frequenza.

Un'attenzione particolare meritano anche gli esempi (13) e (14) in cui compaiono due diversi avverbiali quantificazionali: “quotidianamente” e “raramente”. È possibile ipotizzare, dopo quanto detto sopra per l'avverbiale “spesso”, che la presenza di questa tipologia di avverbi, oltre all'esplicita indicazione dell'arco temporale in cui si localizza l'evento plurazionale, abbia indotto i lavoratori a dare giudizi sbagliati, in quanto, probabilmente, questi avverbiali vengono percepiti come indicanti frasi di tipo abituale.

- (13) *Rapimenti e rivolte sono avvenuti quotidianamente durante la guerra in Iraq*
(14) *Il terzo mondo raramente ha trovato gli spazi adeguati per parlare*

Anche per la frase seguente si assiste ad una confusione tra iterazione e abitudine. L'avverbio “sempre”, infatti, nell'esempio (15) è stato probabilmente percepito dai lavoratori come indicante l'immutabilità o la ripetizione dell'evento espresso

dal verbo e secondo questa logica la frase è stata interpretata come abituale senza tenere conto della presenza di un tempo verbale perfettivo.

(15) *Purtroppo gli Stati Uniti hanno risposto sempre alle varie "chiamate alle armi"*

Relativamente, invece alla seconda tipologia di frasi, ovvero le abituali non riconosciute come tali, ne sono state riscontrate 4 con discordanze:

(16) *I giovani rubavano spesso cosmetici*

(17) *La ragazza andava spesso a trovare la madre*

(18) *Simonetta si recava spesso presso gli uffici di via Poma*

(19) *La ragazza era spesso a casa della madre negli ultimi tempi*

Oltre alla presenza dell'avverbio "spesso" è da notare in (20) la discordanza dei giudizi anche nel caso di utilizzo di un altro avverbio quantificazionale "frequentemente":

(20) *La donna litigava frequentemente col marito a causa del fumo*

Un lavoratore su tre, infatti, ha giudicato l'esempio (20), così come le frasi da (16) a (19), come NON appartenenti alla classe dell'imperfettività gnomica non tenendo conto del fatto che in tutti gli esempi precedenti viene utilizzato un tempo imperfettivo.

In conclusione, l'abitudine presupponendo regolari iterazioni di un evento è strettamente collegata con la nozione di iteratività ma non va confusa con essa. La ripetizione di una situazione, infatti, non è sufficiente affinché questa possa essere concepita come abituale: se una situazione viene ripetuta un numero limitato di volte può essere vista come un'unica situazione a cui ci si può riferire con una forma verbale perfettiva, sebbene abbia una propria struttura interna.

Un dato interessante su cui riflettere è rappresentato dal fatto che su un totale di 18 frasi iterative che presentano discordanze, in ben 15 è presente un avverbio

quantificazionale. Questo dato evidenzia come, in questo caso, le informazioni tempo-aspettuali portate dalle marche morfologiche abbiano una minore rilevanza rispetto a quelle portate dagli avverbiali. Dall'esperimento, infatti, si rileva come il concetto di abitudine sia strettamente collegato alla presenza di avverbiali quantificazionali che compaiono in correlazione con il verbo piuttosto che all'analisi delle marche morfologico-aspettuali.

Ancora dall'analisi delle 18 frasi iterative che presentano discordanze è possibile rilevare come non siano presenti esempi in cui compaiono avverbiali iterativi, come "2 volte", "più volte" ecc. Questo dato dimostra come tali avverbiali costituiscano un elemento linguistico efficace affinché anche persone senza conoscenze linguistiche specialistiche riescano a distinguere frasi iterative da frasi abituali. Questo è possibile grazie alla caratteristica propria di questa tipologia di avverbiali di specificare in maniera evidente ed oggettiva il numero di iterazioni dell'evento espresso dal verbo.

Le espressioni temporali

Come dimostrano le seguenti frasi iterative, per le quali sono stati riscontrati giudizi discordanti, la presenza di espressioni esplicite che individuano l'arco temporale in cui si svolgono i sotto-eventi espressi dal verbo ha contribuito ad indurre in errore i lavoratori che hanno interpretato tali espressioni come indicanti un periodo di riferimento nel quale viene rilevata la persistenza di una determinata caratteristica del soggetto. In questi casi gli avverbiali di tempo non sono stati interpretati tenendo in considerazione il tempo verbale a cui sono associati: secondo questa logica, infatti, i lavoratori avrebbero dovuto attribuire una delimitazione temporale stretta a frasi perfettive-iterative, una delimitazione temporale più vaga a frasi imperfettive-abituali.

(21) *Durante l'ultimo anno circa 3 milioni di italiani hanno scaricato musica senza pagare*

(22) *Le guide hanno organizzato escursioni sul monte Bondone da marzo a settembre*

(23) *Tra il 2003 e il 2004 sono stati effettuati controlli sui conti della Federazione*

(24) *Rapimenti e rivolte sono avvenuti quotidianamente durante la guerra in Iraq*

Una riflessione simile alla precedente può essere riservata alle frasi con verbi di aspetto imperfettivo continuo.

In frasi come le seguenti, infatti, si può supporre che la presenza esplicita di espressioni temporali che delimitano il periodo in cui si svolge l'azione espressa dal verbo, oltre alla presenza di tempi all'imperfetto, abbia indotto almeno un lavoratore su tre a considerare le frasi come afferenti alla classe dell'imperfettività gnomica e, molto probabilmente, ad attribuire loro un'accezione abituale.

Da notare comunque come le frasi in questione siano genuinamente ambigue rispetto ad un'accezione imperfettiva continua e un'interpretazione gnomica comunque possibile. Il fatto che i lavoratori non siano stati univoci nell'esprimere un giudizio è, dunque, molto coerente.

(25) *Durante la performace gli attori si rivolgevano al pubblico*

(26) *Durante le lezioni l'insegnante interrogava gli alunni anche per un'ora consecutiva*

(27) *Durante l'incontro il ragazzo spiegava le sue ragioni*

(28) *In attesa all'ambulatorio il ragazzo tossiva in continuazione*

(29) *Durante la proiezione del film il pubblico applaudiva in continuazione*

(30) *Per tutta la durata dell'interrogatorio il terrorista guardava davanti a sé con lo sguardo perso nel vuoto*

La forma perfettiva

Di seguito sono riportati tre esempi interessanti in cui almeno un annotatore ha confuso frasi perfettive episodiche con frasi imperfettive gnomiche.

(31) *Violante Placido ha esordito al cinema giovanissima*

(32) *Fabio Baldato ha conquistato la prima tappa del tour della Polonia*

(33) *Simona Torretta si è innamorata di Baghdad nel 1994 al suo primo viaggio*

La confusione è giustificabile con il fatto che le frasi precedenti sono episodiche ma presentano un verbo di cambiamento di stato che però ha un risultato di fatto permanente. Dunque, in questi casi i lavoratori si sono focalizzati sullo stato risultante dell'evento interpretandolo come una caratteristica immutabile della persona.

Abitualità senza iteratività

Una situazione può essere abituale senza necessariamente essere iterativa. Negli esempi (34) e (35) si riscontrano valutazioni errate: la corretta interpretazione delle frasi è quella secondo cui Simona Pari e Simona Torretta “hanno lavorato” e “hanno girato con il velo in testa” per un intero periodo, senza interruzioni.

I lavoratori, probabilmente, non hanno riconosciuto l'aspetto verbale abituale che caratterizza questi due esempi in quanto il verbo non appare in contesto in correlazione con avverbiali quantificazionali che caratterizzano invece la maggior parte delle frasi abituali presenti nel corpus.

(34) *Simona pari e Simona Torretta lavoravano tra mille difficoltà*

(35) *Simona Pari e Simona Torretta a Baghdad giravano con il velo in testa*

Costruzioni copulative nelle frasi afferenti alla classe dell'imperfettività gnomica

Tra le frasi imperfettive gnomiche che presentano discordanze non compaiono costruzioni copulative. Evidentemente questo tipo di struttura evidenzia con maggiore forza rispetto a costruzioni non copulative il valore aspettuale gnomico del verbo coinvolto.

Da notare che le istruzioni per l'esperimento sono state scritte con lo scopo di evidenziare proprio questo aspetto: si indicano, infatti, come afferenti alla classe dell'imperfettività gnomiche frasi esprimenti qualità generali che hanno caratterizzato individui od organizzazioni, la cui realizzazione sintattica avviene, come dimostrano gli esempi riportati, mediante l'uso di costruzioni copulative.

Forma progressiva

I lavoratori hanno espresso giudizi discordanti anche sulle frasi i cui verbi presentano aspetto imperfettivo progressivo, espresso sia attraverso l'uso della perifrasi sia tramite l'utilizzo dell'imperfetto.

(36) *Umberto Berardi si stava recando come ogni mattina a Monza dove avrebbe preso l'autobus per andare a lavoro*

(37) *La squadra di Albarello stava lavorando molto bene*

(38) *Gli sforzi della squadra davano i primi significativi frutti*

(39) *Sul furto indagavano i Carabinieri di Imer*⁴⁸

(40) *Le ragazze furono sequestrate a Baghdad mentre lavoravano nel loro ufficio*

(41) *Il fotografo li ha sorpresi mentre facevano la quotidiana passeggiata nel parco*

Nelle istruzioni si specifica che avvenimenti passati descritti come in corso di svolgimento non esprimono mai leggi, regolarità o abitudini. Da notare, però, l'esiguità del numero di esempi riportati: solo due, di cui uno caratterizzato dal tempo verbale dell'imperfetto e l'altro dalla perifrasi progressiva. Probabilmente perché il concetto fosse più chiaro ai lavoratori sarebbe stato opportuno riportare un numero più cospicuo di esempi.

Sottospecificazione delle categorie aspettuali afferenti alla classe dell'imperfettività gnomica

Per quanto riguarda le frasi che presentano discordanze il cui aspetto verbale afferisce alla classe dell'imperfettività gnomica, come riportato in Tabella 6, su un totale di 11 frasi si riscontra che: 10 sono classificabili come abituali e solo una può essere classificata, con la dovuta cautela (per approfondimenti vedere la sezione seguente "Casi particolari"), come Individual-Level predicate.

⁴⁸L'esempio (39) all'atto della costruzione del corpus è stato inserito nella categoria delle imperfettive progressive sebbene un'interpretazione gnomica sia comunque ammissibile.

Frase gnomiche con discordanze	Tipologie di frasi
<i>Ormai da anni mio figlio andava a scuola con dei supplenti come insegnanti</i>	Abituale
<i>I giovani rubavano spesso cosmetici</i>	Abituale
<i>Simona Pari e Simona Torretta a Baghdad giravano con il velo in testa</i>	Abituale
<i>Estroteatro proponeva abitualmente spettacoli di Stefano Benni</i>	Abituale
<i>Simona pari e Simona Torretta lavoravano tra mille difficoltà</i>	Abituale
<i>Amos Gitai girava due film ogni anno</i>	Abituale
<i>La ragazza andava spesso a trovare la madre</i>	Abituale
<i>La donna litigava frequentemente col marito a causa del fumo</i>	Abituale
<i>Simonetta si recava spesso presso gli uffici di via Poma</i>	Abituale
<i>La ragazza era spesso a casa della madre negli ultimi tempi</i>	Abituale
<i>Javier Bardem interpretava Ramon Sampedro nel film "Mare dentro"</i>	Individual Level-predicate

Tabella 6. Frasi afferenti alla classe dell'imperfettività gnomiche che presentano discordanze

Le altre frasi di tipo Individual-Level predicate, così come le attitudinali, potenziali e generiche presenti nel corpus, sono state giudicate tutte correttamente come appartenenti alla classe dell'imperfettività gnomiche.

Le frasi generiche rientrano nella categoria delle "Nozioni storiche, geografiche, mediche" che, come da istruzioni, esprimono sempre leggi o regolarità generali. Le attitudinali, potenziali e Individual-Level predicate sono state fatte rientrare nella categoria delle "Qualità generali e caratteristiche che hanno caratterizzato individui o organizzazioni". Entrambe le categorie sono corredate di un ricco numero di esempi che ha evidentemente aiutato i lavoratori nella comprensione di

concetti la cui spiegazione teorica è ben più complessa di quella riportata nelle istruzioni.

Casi particolari

Vi sono tre esempi interessanti, che si configurano come *borderline*, tra le frasi che presentano discordanze. Il primo è il (42), unico esempio tra quelli presenti nel corpus per cui tutti i lavoratori sono stati in accordo nell'esprimere un giudizio errato:

(42) *Javier Bardem interpretava Ramon Sampedro nel film "Mare dentro"*

In questo caso la frase non è stata interpretata come afferente alla classe dell'imperfettività gnomica nella cui categoria invece era stata pensata durante la costruzione del corpus. Piuttosto, si può ipotizzare che i lavoratori, senza tenere conto del tempo verbale all'imperfetto, abbiano giudicato la frase come episodica.

Tale esempio può essere interpretato, infatti, in due modi diversi: (i) come un Individual Level-predicate se il focus viene posto sul film: in questo caso il protagonista è, è stato e rimarrà Javier Bardem; (ii) nel caso in cui, invece, ci si focalizzi sulla vita di Bardem come arco temporale per interpretare l'evento, la partecipazione come attore al film "Mare dentro" non è che un punto preciso del tutto e non caratterizza sicuramente la vita nel suo complesso.

Un altro esempio interessante di discordanza è rappresentato dalla frase (43). In questo caso la presenza di una espressione temporale e di una specificazione numerica, che però non è riferita all'evento espresso dal verbo ma al suo complemento oggetto, hanno probabilmente causato confusione facendo in modo che alla frase fosse attribuita un'accezione iterativa invece che abituale.

(43) *Amos Gitai girava due film ogni anno*

L'ultimo esempio che vale la pena citare in quanto i tre giudizi espressi sono tutti errati è il seguente:

(44) *Spesso le donne sono state escluse dalla storia della scienza*

Si tratta di una frase perfettiva iterativa scambiata per una afferente alla classe dell'imperfettività gnomica. La motivazione dell'errore può essere ricercata nella presenza dell'avverbio quantificazionale "spesso". Come scritto in precedenza, infatti, questa tipologia di avverbiali viene collegata dai lavoratori, nella maggior parte dei casi, a frasi con aspetto abituale.

Una visione d'insieme: discordanze e tipologie di frase

In Tabella 7 viene riportato per ciascuna tipologia di aspetto verbale presente nel corpus il numero di frasi che presentano discordanze.

Tipologia di frasi	# di frasi con discordanze
frasi gnomiche	11
frasi con verbi di aspetto progressivo	6
frasi con verbi di aspetto continuo	6
frasi iterative	18
frasi episodiche	4

Tabella 7. Numero di frasi con discordanze distinte per tipologia aspettuale

Il maggior numero di errori riguarda la valutazione delle frasi iterative che spesso, come già spiegato nel corso di questo capitolo, sono state scambiate per abituali. Sebbene, infatti, le frasi iterative in totale siano 75 sono state espresse ben 18 valutazioni discordanti a fronte delle 11, su un totale di 150 frasi, espresse per gli esempi afferenti alla classe dell'imperfettività gnomica.

Per quanto riguarda, invece, gli aspetti progressivo e continuo, 12 frasi, su un totale di 75, presentano valutazioni discordanti. La difficoltà di riconoscimento per questa tipologia di frasi si colloca, quindi, in una posizione intermedia tra quella relativa alle frasi iterative e quella riguardante le imperfettive gnomiche.

Dall'analisi dei dati si nota, inoltre, come l'individuazione dell'aspetto imperfettivo gnomico risulti piuttosto semplice se il focus viene posto sulla valutazione fatta dai lavoratori rispetto alle frasi episodiche: su 75 frasi, infatti, solo 4 presentano giudizi discordanti. Il verbo in queste frasi, forse perché coniugato al passato prossimo e spesso in correlazione con espressioni temporali puntuali, viene riconosciuto con facilità dai lavoratori come indicante un evento specifico, che ha carattere occasionale e che si presenta senza regolarità.

Capitolo V

Conclusioni

In questo lavoro si è tentato di approfondire la natura problematica del dominio tempo-aspettuale e di muovere qualche passo in avanti nello studio delle metodologie da adottare per creare, sulla base di quelle già esistenti, risorse linguistiche annotate con informazioni inerenti l'aspetto e l'azione verbale, in modo da poterle utilizzare come dati di addestramento per sistemi di NLP.

Un approfondimento particolare è stato dedicato all'indagare, attraverso l'uso di strumenti di crowdsourcing, la capacità di utenti non linguisticamente addestrati di individuare l'aspetto verbale afferente alla classe dell'imperfettività gnomica (Bertinetto e Lenci, 2011) nel tentativo di dimostrare l'utilità di marcare questa particolare classe aspettuale in prospettiva di un'estrazione automatica di informazione di senso comune da testi scritti (Singh, 2002).

In particolare, nell'ambito di questo lavoro si è cercato di analizzare il trattamento e l'utilizzo che i più recenti sistemi di NLP fanno dell'aspetto e dell'azione verbale. Innanzitutto, è stata svolta una ricerca mirata a delineare lo stato dell'arte dell'annotazione inerente aspettualità ed azionalità e le relative risorse attualmente disponibili.

Da questa ricerca è emerso chiaramente come siano pochi sia gli strumenti di NLP capaci di annotare automaticamente questi tipi di informazione sia i data set su cui questi sistemi compiono il proprio addestramento. Se il numero dei corpora in cui viene marcata l'azione verbale è esiguo, ancora meno consistente è il numero di quelli che riportano informazioni inerenti l'aspetto verbale, la cui marcatura, infatti, è presente solo in data set, di varie lingue, annotati secondo lo schema TimeML (Pustejovsky et al., 2003).

Nello specifico TimeML è uno schema di annotazione per gli eventi che ha lo scopo di facilitare il ragionamento e l'inferenza sulla loro organizzazione temporale nei testi in modo da darne un'immagine più completa possibile.

In questo lavoro è stata svolta un'analisi approfondita sul trattamento che questo schema di annotazione riserva all'azione e all'aspetto verbale e sono state proposte delle integrazioni.

Per quanto riguarda l'azione verbale TimeML non ne contempla la marcatura, anche se, da un'analisi più dettagliata degli attributi e dei loro valori è possibile rintracciare alcune informazioni relative all'azionalità tra i valori dell'attributo CLASS il cui scopo è quello di indicare la classe sintattica di appartenenza di un evento.

Queste informazioni azionali, dunque, risultano confuse con altri parametri linguistici per cui non è possibile utilizzarle per stabilire la corretta classe azionale dei verbi. Data questa confusione, nell'ottica di un arricchimento dei testi con informazioni inerenti l'azione verbale, seguendo una prospettiva di annotazione automatica e cercando di facilitare la scelta tra i possibili valori azionali da parte degli annotatori, è stato proposto un ampliamento dello schema TimeML con quattro ulteriori attributi: tre corrispondenti ai tratti azionali necessari ad identificare una classe vendleriana, ovvero dinamicità, duratività e telicità, ed un quarto per la marcatura della classe azionale vera e propria, che può essere determinata automaticamente sulla base del valore dei tratti azionali già marcati.

Per quanto riguarda, invece, la marcatura dell'aspetto verbale, prendendo come punto di riferimento l'annotazione definita dallo schema TimeML, è stata proposta una classificazione più dettagliata dei valori aspettuali in modo da integrare i tre già esistenti (perfettivo, imperfettivo e progressivo).

Nella prospettiva di facilitare il lavoro degli annotatori, si è cercato di stilare una lista di regole chiare e di facile applicazione per individuare il corretto aspetto di un verbo e, in quest'ottica, sono state sviluppate delle apposite specifiche di annotazione.

In particolare, con questa integrazione ci si propone di individuare il corretto valore aspettuale attraverso una serie di passi successivi, ordinati per complessità.

Lo scopo è quello di riuscire a classificare alcuni tipi di contrasti aspettuali rilevanti per l'italiano, la cui individuazione può essere utile in un'ottica di elaborazione automatica dei testi.

Per quanto riguarda la parte sperimentale di questa tesi, la decisione di marcare l'imperfettività gnomica, come accennato precedentemente, è stata dettata dall'interesse per l'individuazione e il riconoscimento, all'interno dei testi, di frasi che esprimono una generalizzazione di qualche tipo oppure una proprietà non accidentale di un soggetto ma che lo caratterizza per un periodo della sua vita o per l'intera sua esistenza.

L'aspetto verbale gnomico, infatti, potrebbe configurarsi come una caratteristica discriminante per recuperare informazione di senso comune da testi scritti e, conseguentemente, potrebbe essere sfruttata con profitto in strumenti di NLP.

Proprio nella prospettiva di testare la difficoltà da parte di persone non esperte di individuare verbi afferenti alla classe dell'imperfettività gnomica, la piattaforma CrowdFlower, che sfrutta il paradigma del crowdsourcing, è stata utilizzata per portare a termine un esperimento linguistico. A lavoratori senza conoscenze specialistiche sono state presentate 375 frasi contenenti verbi appartenenti a diverse tipologie di aspetto verbale (imperfettivo continuo, progressivo e gnomico, perfettivo episodico ed iterativo) ed è stato chiesto loro di riconoscere le sole frasi il cui aspetto afferiva alla classe dell'imperfettività gnomica, ovvero, più semplicemente, di individuare le frasi esprimenti leggi, regolarità generali o abitudini.

Un'attenzione particolare è stata rivolta, in questo contesto, al lavoro di definizione del corpus. La sua costruzione, infatti, è andata nella direzione di rendere disponibile ai lavoratori uno spettro completo delle categorie aspettuali in modo tale da poter valutare il livello di accordo tra i giudizi espressi in relazione a tutte le tipologie di aspetto presenti nel corpus. La ricchezza dei parametri di osservazione per l'individuazione delle diverse categorie aspettuali ha imposto il ricorso a tecniche di analisi quantitativa che tenessero conto di tutti i tipi di aspetto in modo da poter valutare, almeno a livello generale, il grado di complessità nel riconoscimento oltre che della classe aspettuale gnomica anche delle altre categorie.

I soddisfacenti risultati quantitativi illustrati in questo lavoro confermano l'importanza metodologica di continuare a porre una costruttiva attenzione nei confronti di ogni ipotesi riguardante la capacità di individuazione delle categorie aspettuali da parte di utenti inesperti, anche e soprattutto attraverso l'utilizzo di strumenti di crowdsourcing.

L'analisi qualitativa dei risultati dell'esperimento ha permesso, inoltre, di mettere in luce degli aspetti interessanti per quanto riguarda la capacità intuitiva di soggetti non esperti di individuare l'aspetto verbale afferente alla classe dell'imperfettività gnomica.

In particolare, l'esperimento mostra la capacità da parte di persone comuni di individuare l'aspettualità gnomica utilizzando le sole istruzioni pratiche che sono state messe loro a disposizione, riuscendo a discernere intuitivamente tra varie tipologie di aspetto verbale attraverso una superficiale analisi del contesto, nel quale alcuni elementi più di altri fanno propendere per la pronuncia di un determinato giudizio.

L'esperimento evidenzia proprio come l'analisi del contesto risulti un elemento essenziale per il corretto svolgimento del compito. In particolare, i risultati ottenuti mostrano come la presenza negli esempi di avverbiali quantificazionali abbia aiutato i lavoratori ad individuare la classe aspettuale richiesta anche nei casi in cui l'interpretazione aspettuale del predicato era più ambigua, per esempio nella distinzione tra la classe dell'imperfettività abituale e quella dell'imperfettività continua entrambe espresse tramite il tempo verbale dell'imperfetto. Questo dato mette in luce come in casi di difficile interpretazione aspettuale gli avverbiali quantificazionali siano stati considerati dagli annotatori un elemento fondamentale per il corretto svolgimento del compito.

Il concetto di aspetto gnomico che appare così teoricamente complesso risulta, quindi, se delineato nei suoi punti essenziali, un'informazione linguistica relativamente facile da indovinare anche da parte di persone prive di conoscenze specialistiche.

In conclusione, da un'analisi approfondita dei risultati ottenuti dell'esperimento condotto in questa tesi, è possibile estrapolare alcune regole basilari per il riconoscimento dell'aspetto gnomico, regole che potrebbero essere utilizzate da un sistema automatico per l'individuazione di questa particolare categoria di aspetto.

Dallo studio dei dati, sembra, infatti, possibile basarsi sulla valutazione positiva dei seguenti parametri per cercare di stabilire, almeno in linea generale, se un aspetto verbale afferisce o meno alla classe dell'imperfettività gnomico:

- presenza nella frase di avverbi quantificazionali in correlazioni con tempi verbali imperfettivi;
- presenza di costruzioni copulative in correlazione con tempi verbali imperfettivi;
- presenza di espressioni temporali che delimitano un preciso arco di tempo in presenza di tempi verbali imperfettivi ed eventuali avverbiali quantificazionali.

Rimangono, comunque, alcune questioni aperte di difficile risoluzione. Tra queste la più rilevante è sicuramente quella inerente l'identificazione dell'aspettualità gnomico in frasi in cui il verbo esprime una situazione abituale in assenza di avverbiali quantificazionali, come per esempio accade nella frase "Roberto viveva a Pisa", la cui naturale interpretazione è che Roberto abbia abitato a Pisa per un intero periodo di tempo, senza interruzioni. In casi come questo l'interpretazione semantica della frase è determinante per disambiguare l'etichettatura aspettuale del verbo: è, quindi, necessaria una conoscenza più vasta rispetto alle informazioni sulla sintassi della frase.

Si può pensare di proseguire nell'indagine delle dinamiche tempo-aspettuali cercando di affrontare le problematiche aperte e sollevate anche da questo lavoro.

Un interessante contributo futuro potrebbe consistere nel riproporre questo esperimento per una lingua diversa dall'italiano, per esempio l'inglese. In questa lingua, infatti, l'opposizione aspettuale tra imperfettivo abituale e perfettivo non è marcata per cui lo stesso *simple past* viene utilizzato sia per esprimere l'aspetto

abituale sia per esprimere l'aspetto perfettivo. Sarebbe, quindi, interessante comparare i risultati dei due esperimenti per individuare gli elementi che avvicinano e fanno differire il riconoscimento dell'aspettualità gnomica in queste due differenti lingue.

Un altro studio particolarmente interessante potrebbe consistere, infine, nell'indagare le sottocategorie aspettuative dell'imperfettività gnomica. In particolare, si potrebbe pensare di definire un task di crowdsourcing volto a valutare la capacità di persone comuni di distinguere tra frasi abituali, attitudinali, potenziali, Individual-Level predicate e generiche, così da validare o smentire i risultati dell'esperimento svolto in questa tesi, secondo cui la difficoltà maggiore per i lavoratori consiste nel riuscire a riconoscere l'aspetto abituale rispetto alle altre sottocategorie gnomiche.

Appendice A: BNF dei tag <EVENT>, <SIGNAL>, <TIMEX3>

Backus-Naur Form (BNF) del tag <EVENT>:

```

attributes ::= eid eiid class tense aspect pos polarity
[modality]
eid ::= e<integer>
eiid ::= ei<integer>
class ::= 'REPORTING' | 'PERCEPTION' | 'ASPECTUAL' |
'I_ACTION' | 'I_STATE' | 'STATE' | 'OCCURRENCE'
tense ::= 'PAST' | 'PRESENT' | 'FUTURE' | 'NONE' |
'INFINITIVE' | 'PRESPART' | 'PASTPART'
aspect ::= 'PROGRESSIVE' | 'PERFECTIVE' |
'PERFECTIVE_PROGRESSIVE' | 'NONE'
pos ::= 'ADJECTIVE' | 'NOUN' | 'VERB' | 'PREP' |
'OTHER'
polarity ::= 'NEG' | 'POS' {default, if absent, is
'POS'}
modality ::= CDATA

```

Backus-Naur Form (BNF) del tag <TIMEX3>:

```

attributes ::= tid type
[functionInDocument][beginPoint][endPoint][quant][freq]
[temporalFunction] (value | valueFromFunction)
[mod][anchorTimeID]
tid ::= ID
{tid ::= TimeID
TimeID ::= t<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' | 'APPROX'
anchorTimeID ::= IDREF
{anchorTimeID ::= TimeID}
```

Appendice B: lista delle frasi che compongono il data set

Di seguito la lista delle 375 frasi che compongono il dataset. La valutazione dei lavoratori ha riguardato l'evento espresso dal verbo in grassetto.

1. Oetzi **era** sicuramente un uomo di rango.
2. Giovanni Bensi **era** trentino.
3. L'epica **cantava** le gesta degli eroi in guerra.
4. Quello delle truffe all'INPS **era** un fenomeno diffuso.
5. I Kriegsmaler **documentavano** attraverso le loro opere la vita al tempo della Grande Guerra.
6. Egon Schiele **faceva parte** dei Kriegsmaler.
7. Oetzi **era** uno sciamano un uomo di religione.
8. Lorenzi **era** corretto e sincero anche se duro a volte.
9. Lorenzi **era** un uomo deciso e intransigente.
10. Ormai da anni mio figlio **andava** a scuola con i dei supplenti come insegnanti.
11. L'uomo **era** malato da tempo.
12. **Teneva** abitualmente conferenze in tutto il mondo.
13. I giovani **rubavano** spesso cosmetici.
14. Gli anziani **rubavano** di solito alimentari.
15. Le signore di solito **rubavano** prodotti piccoli ad alto prezzo come i profumi.
16. Ogni giorno in negozio **circolava** un uomo della security.
17. **Trascorreva** ogni notte in case lasciate momentaneamente vuote.
18. La piccola Denise **giocava** spesso davanti a casa.
19. Valerio Mosele **giocava** a tennis ogni giorno.
20. In quel periodo Mosele **prendeva** sempre la palla.

21. Da trent'anni **insegnava** nello stesso circolo.
22. Simona Torretta **viveva** a Baghdad.
23. Simona Pari **lavorava** con i bambini.
24. Gli imam italiani **si riunivano** insieme in preghiera ogni anno.
25. Simona Pari e Simona Torretta **lavoravano** in Iraq per motivi umanitari.
26. Simona Pari e Simona Torretta a Baghdad **giravano** con il velo in testa.
27. Simona Pari e Simona Torretta **lavoravano** a contatto con la gente.
28. La figlia **aiutava** la madre nelle faccende domestiche.
29. Putin **era** solito usare la politica della forza.
30. Estro teatro **proponeva** abitualmente spettacoli di Stefano Benni.
31. In Cecenia spesso le famiglie sfollate **vivevano** in fattorie semidistrutte.
32. Marco **abitava** a Trento sud.
33. Marco **andava** a scuola in centro.
34. Simona pari e Simona Torretta **lavoravano** tra mille difficoltà.
35. Amos Gitai **girava** due film ogni anno.
36. Il "Tavolo Trentino per l'Iraq" **era** un coordinamento informale di istituzioni e associazioni di volontariato.
37. Javier Bardem **interpretava** Ramon Sampedro nel film "Mare dentro".
38. Valerio Mosele **era** un bambinetto piccolo e magro con un ciuffetto sbarazzino sopra gli occhietti furbi.
39. Nell'azienda molti laureati **lavoravano** come Co.co.co..
40. I lavoratori degli uffici stampa spesso **vivevano** situazione al di fuori delle norme contrattuali.
41. Jal al-D n R m **era** un poeta mistico persiano.
42. Shams-i Tabr z **era** un derviscio vagante.
43. In Trentino i controllori dell'autobus **facevano** molte multe.
44. Publio Virgilio Marone **era** uno fra i principali poeti dell'epoca augustea.
45. Il poeta Orazio **cantava** il senso civico la concordia familiare e l'onestà dei costumi.
46. Lo storico Tito Livio **narrava** le vicende di Roma dalla sua fondazione

fino ai suoi tempi.

47. Oltre ai poeti e agli scrittori Mecenate **protegeva** anche gli artisti.
48. Nel 1348 Avignone **era** la sede papale.
49. Molti medici di fronte alla peste **fuggivano**.
50. Contro la peste i medici frequentemente **ricorrevano** ad infusi di erbe aromatiche.
51. Durante la peste i flagellanti **percorrevano** in processione le strade delle città.
52. Durante la peste il papa **invitava** il clero a porre gli ebrei sotto la sua protezione.
53. Tiberio Sempronio Gracco **era** nipote di Scipione Africano.
54. La Numidia **era** un antico regno berbero.
55. La numidia **occupava** i territori del Nord Africa corrispondenti all'attuale Algeria.
56. Dalla Spagna Cartagine **importava** grandi quantità di argento.
57. Il Governo cartaginese **era** un'oligarchia.
58. I Cartaginesi **veneravano** molti dei.
59. Nel periodo di massimo splendore Cartagine **ospitava** un grande numero di divinità provenienti dalle civiltà greca egizia ed etrusca.
60. Ninive antica capitale dell'Assiria **si trovava** nei pressi dell'odierna città di Mossul (Iraq settentrionale).
61. Il Regno dei Visigoti **si estendeva** dalla Spagna alla Gallia meridionale.
62. La peste **era** una malattia incurabile.
63. L'Etruria **corrispondeva** ad un'area comprendente la Toscana il Lazio fino al Tevere l'Umbria parte delle Liguria e della Lombardia.
64. Gli Etruschi **estraevano** il ferro dall'isola d'Elba.
65. I Greci **chiamavano** gli Etruschi "tirreni".
66. Gli opliti **erano** i soldati etruschi a servizio permanente.
67. Il complesso di Sasso Pisano **era** un importante centro termale in epoca etrusca.

68. La donna nella società etrusca **aveva** una posizione particolarmente rilevante.
69. L'etrusco **era** una lingua non indoeuropea.
70. Nella cultura etrusca la divinazione **occupava** un ruolo fondamentale.
71. Tarquinio Prisco **era** un ricchissimo e noto abitante della città etrusca di Tarquinia.
72. Tarquinio Prisco **era** originario di Tarquinia.
73. La divinazione etrusca **si divideva** in due branche principali.
74. Populonia **era** l'unica città etrusca sul mare.
75. Enea **era** figlio del mortale Anchise e di Venere dea della bellezza..
76. Assur **era** l'antica capitale dell'Assiria.
77. L'Assira **corrispondeva** all'estrema regione settentrionale dell'attuale Iraq.
78. L'Assiria**era** già popolata durante il Paleolitico.
79. Babilonia **sorgeva** sulle rive del fiume Eufrate.
80. All'epoca di Alessandro Magno Babilonia **contava** un milione di abitanti.
81. I giardini pensili di Babilonia **erano** una delle sette meraviglie del mondo.
82. Con il termine Mesopotamia i greci **intendevano** la zona che si estendeva tra il Tigri e l'Eufrate.
83. I Babilonesi **utilizzavano** le ziggurat anche per le osservazioni astronomiche.
84. I Babilonesi **erano** particolarmente esperti in astronomia.
85. Ur **era** un'antica città della bassa Mesopotamia.
86. La città di Ur **si trovava** alla foce del Tigri e dell'Eufrate.
87. Il re Nabucodonosor **attribuiva** il merito delle sue vittorie a Marduk il principale dio di Babilonia.
88. Le corporazioni di arti e mestieri nel Medioevo **controllavano** la qualità dei prodotti i prezzi e la formazione dei nuovi addetti.
89. Maometto **predicava** un dio unico "Allah".
90. Shams-i Tabr z **aveva** spesso delle visioni.
91. Nei comuni medievali il podestà **esercitava** i poteri esecutivo di polizia e

- giudiziario .
92. In epoca comunale il capitano del popolo un magistrato spesso forestiero **restava** in carica per sei mesi o un anno.
93. Il podestà in epoca comunale **durava** in carica di solito sei mesi o un anno.
94. Il podestà **giurava** fedeltà agli statuti comunali.
95. Il Comune **era** una forma di governo locale in età medievale.
96. Gli Etruschi **erano** un popolo dell'Italia antica affermatosi in un'area denominata Etruria.
97. La città greca di Cuma **ospitava** il tempio di Giove.
98. Cuma **godeva** di un'ottima posizione strategica per le azioni militari.
99. Anchise **era** il padre di Enea.
100. Assurnasirpal II **era** un sovrano assiro.
101. Il codice di Hammurabi **ammetteva** anche il risarcimento in denaro fatto molto innovativo per l'epoca.
102. Nel codice di Hammurabi la pena **era** proporzionata alla classe sociale a cui appartenevano le parti coinvolte.
103. L'azione di governo di Hammurabi **si incentrava** sull'unità dello stato.
104. I genitori di Luca **erano** portatori sani di Talassemia Mayor.
105. Una volta l'autonomia dei ricercatori **era** un requisito indispensabile.
106. La ragazza **andava** spesso a trovare la madre.
107. Piero Valentini **viveva** a Jupiter.
108. In quegli anni i Verdi di solito **erano** un po' fondamentalisti.
109. I Carabinieri frequentemente **svolgevano** servizi di viabilità davanti alle scuole.
110. Note Italiane nel mondo **era** una manifestazione organizzate da Radio Italia.
111. Antonio Schizzerotto **conosceva** bene il Trentino.
112. La Mesopotamia **si estendeva** tra il Tigri e l'Eufrate.
113. L'uomo **era** un gran fumatore.

114. La donna **litigava** frequentemente col marito a causa del fumo.
115. Enzo Baldoni **era** un giornalista freelance.
116. Enzo Baldoni **svolgeva** principalmente l'attività di copywriter in una sua società.
117. Enzo Baldoni **era** un volontario della Croce Rossa.
118. Baldoni nei suoi articoli **era solito** far passare concetti crudi in forma lieve e dissacrante.
119. Simonetta Cesaroni **era** una bella ragazza romana che viveva nel quartiere di Don Bosco.
120. Il padre di Simonetta **era** un dipendente dell'azienda tramviaria di Roma e provincia.
121. Simonetta **si recava** spesso presso gli uffici di via Poma.
122. La Reli Sas **era** uno studio commerciale sito in zona Casilina a Roma.
123. La ragazza **era** spesso a casa della madre negli ultimi tempi.
124. In quel periodo i suicidi in carcere **erano** molto frequenti.
125. Il Basso Egitto **aveva** come capitale la città di Buto.
126. Nekhen **era** la capitale dell'Alto Egitto.
127. Eliopoli **era** una delle più rilevanti località dell'antico Egitto legate al culto solare.
128. Eliopoli **era** un'importante città dell'antico Egitto.
129. Cleopatra **era** figlia del faraone Tolomeo XII Aulete.
130. Nel 42 a.C. Marco Antonio **era** uno dei triumviri che governavano Roma.
131. La ragazza **prendeva** abitualmente l'autobus per andare a scuola.
132. Biblioteca reale di Alessandria **era** uno dei principali poli culturali ellenistici.
133. La Biblioteca di Alessandria **era** la più grande e ricca biblioteca del mondo antico.
134. La Biblioteca di Alessandria **era** molto famosa nel mondo antico.
135. La Sibilla Cumana **era** una sacerdotessa di Apollo.

136. La Sibilla Cumana **svolgeva** la sua attività oracolare nei pressi del Lago d'Averno in una caverna conosciuta come l'"Antro della Sibilla".
137. L'importanza della Sibilla Cumana nel mondo italico **era** pari a quella del celebre oracolo di Delfi in Grecia.
138. Secondo la religione greca e poi romana il lago di Averno **era** un accesso all'Oltretomba regno del dio Plutone..
139. La Lega peloponnesiaca **era** l'alleanza militare costituitasi verso la metà del VI secolo a.C. tra le poleis del Peloponneso con a capo Sparta.
140. La città di Sparta **esisteva** già all'epoca della guerra di Troia.
141. Al contrario dei cretesi i Micenei **erano** un popolo molto bellicoso.
142. Agamennone **era** il comandante in capo dei greci durante la guerra di Troia.
143. Achille **era** figlio del mortale Peleo e della ninfa Teti.
144. Teti per rendere immortale il piccolo Achille lo **ungeva** ogni giorno con l'ambrosia.
145. Ettore **era** sposo di Andromaca e padre di Astianatte.
146. Troia **era** una città dell'Asia Minore.
147. Giulio Cesare **era** un esponente dei populares.
148. Cesare **era** un discendente della gens Iulia.
149. La famiglia di origine di Giulio Cesare non **era** ricca per gli standard della nobiltà romana.
150. Enzo Baldoni **era** originario di Città di Castello in Umbria.
151. L'addetto della Provincia **falciava** l'erba sulla statale delle Dolomiti quando è stato investito.
152. L'uomo **stava attraversando** la strada quando è accaduto l'incidente.
153. Il mezzo pesante **stava trasportando** un escavatore quando è stato fermato dai Carabinieri.
154. Quando sono arrivati i Carabinieri il ladro **stava forzando** con un piede di porco il distributore di sigarette in via Rosmini .
155. Il ladro **stava scappando** con un paio di cassette di monetine trafugate da una lavanderia.

156. La maestra **stava spiegando** come si riproducono gli animali quando i bambini hanno cominciato a fare domande.
157. L'auto **stava svoltando** a sinistra quando la moto è sopraggiunta all'improvviso.
158. Si è lussato il piede mentre **stava lavorando**.
159. Umberto Berardi **si stava recando** come ogni mattina a Monza dove avrebbe preso l'autobus per andare a lavoro.
160. La bambina ieri pomeriggio **stava giocando** insieme ad alcuni compagni di asilo.
161. Un pauroso incendio **stava minacciando** il levante genovese.
162. L'ira per la stage di Beslan **stava prendendo** pieghe pericolose in Ossezia.
163. La squadra di Albarello **stava lavorando** molto bene.
164. Alcune specie come orso e lince **stavano ricomparendo** sulle nostre montagne.
165. In quegli anni gli Americani **stavano costruendo** un oleodotto in Georgia.
166. E' stato arrestato dalla Polizia di frontiera mentre **attraversava** illegalmente il confine.
167. I Carabinieri della stazione di Moena **stavano conducendo** le indagini quando appresero la notizia della fuga dell'indagato.
168. Le indagini sul rapimento di Denise Pipitone **si stanno concentrando** su quattro persone.
169. Gli sforzi della squadra davano i primi significativi frutti.
170. Sul furto **indagavano** i Carabinieri di Imer.
171. I bambini **stavano giocando** in palestra quando fece irruzione il commando.
172. Patty Pravo in quel momento **si esibiva** in playback.
173. In quel momento l'inflazione **toccava** punte record.
174. Mentre l'ambulanza con la ragazza ferita **si dirigeva** al pronto soccorso i vigili tentarono di identificare il pirata della strada.

175. Al momento dell'esplosione i fedeli **pregavano**.
176. Il primo sparo mentre la donna **chiedeva** aiuto al telefono.
177. Il terzo e il quarto sparo mentre la donna **scappava** verso l'uscio di casa.
178. Al momento del rapimento le ragazze **lavoravano** nei loro uffici.
179. Al momento dell'esplosione i militari **si dirigevano** verso la loro base a Nassiriya.
180. L'uomo è stato bloccato da due agenti mentre **scappava**.
181. Le ragazze furono sequestrate a Baghdad mentre **lavoravano** nel loro ufficio.
182. La bimba è scomparsa mentre **giocava** davanti a casa.
183. E' stata bloccata dai Carabinieri mentre **usciva** di casa.
184. Il giovane è scomparso mentre **eseguiva** su commissione delle fotografie panoramiche.
185. Il motociclista è caduto mentre **si dirigeva** a Canazei.
186. I malviventi sono entrati nell'abitazione mentre i proprietari **dormivano**.
187. Maurizio è stato travolto da un ciclomotore mentre **attraversava** la strada.
188. Il fotografo li ha sorpresi mentre **facevano** la quotidiana passeggiata nel parco.
189. Proprio mentre la ragazza **apriva** lo sportello dell'auto è passata una Fiat Seicento che l'ha centrata in pieno.
190. Durante la riunione **cambiava** opinione in continuazione.
191. Per tutta la durata del Welcome party gli studenti **distribuivano** volantini.
192. Durante la fuga i Carabinieri **intercettavano** il telefono del malvivente.
193. Durante la performance gli attori **si rivolgevano** al pubblico.
194. Durante il collegamento l'inviato **intervistava** molte persone.
195. Durante le lezioni l'insegnante **interrogava** gli alunni anche per un'ora consecutiva.
196. Durante l'incontro il ragazzo **spiegava** le sue ragioni.

197. La donna **piangeva** senza sosta durante l'interrogatorio dei Carabinieri.
198. Durante la trasmissione Andrea **usciva** frequentemente dallo studio.
199. Durante la lezione la maestra **telefonava** frequentemente.
200. Durante l'allenamento Chiellini **si toccava** il ginocchio in continuazione.
201. Per l'intera durata dell'interrogatorio il sospettato **negava** ogni coinvolgimento nella vicenda.
202. Dopo l'operazione il paziente **riposava** tranquillamente.
203. Durante il concerto dei REM gli spettatori giovani e meno giovani **cantavano** tutte le canzoni.
204. In attesa all'ambulatorio il ragazzo **tossiva** in continuazione.
205. Durante l'intervento di Dellai Mauro Bondi **sorrideva** tranquillo.
206. In un lungo interrogatorio la brigatista Cinzia Banelli **ammetteva** la propria responsabilità.
207. Quella sera dalle 21 alle 23 Estro teatro **metteva in scena** l'atto unico di Stefano Benni.
208. Durante l'addestramento i bambini **sbadigliavano** continuamente.
209. Durante la proiezione del film il pubblico **applaudiva** in continuazione.
210. Dopo l'incidente **si spostava** nella zona ma mai troppo lontano dal ferito.
211. Dopo l'incidente Debiasi **si metteva** in contatto con i soccorritori.
212. Dellai quel pomeriggio **tossiva** di continuo a causa dell'influenza.
213. Quel pomeriggio Margherita Cogo **entrava** ed usciva dalla sala congressi in continuazione.
214. Quel giorno la donna **faceva** avanti e indietro tra casa sua e quella della madre a causa della malattia di quest'ultima.
215. Quella sera la donna **entrava** e **usciva** da casa senza un apparente motivo.
216. Quel giorno il navigatore dello sfortunato automobilista non **funzionava**.
217. In sole tre ore la ditta **concludeva** il trasloco.
218. Durante la riunione il delegato provinciale **interveniva** spesso.
219. Per tutta la durata dell'operazione i parenti **attendevano** ansiosi in sala

- d'aspetto.
220. Durante l'interrogatorio la ragazza **controllava** il telefonino di continuo.
221. Dalle 21 alle 22 la ragazza **passeggiava** ignara nel parco in compagnia del malvivente.
222. Durante il settimo giorno dell'operazione i Palestinesi **lanciavano** razzi di continuo.
223. Durante la prima seduta del processo l'imputato palesemente nervoso **chiedeva** continue spiegazioni al suo avvocato.
224. In poche ore i rappresentanti del Comune e dell'azienda **concludevano** l'accordo per il trasferimento della Ignis a Trento.
225. Per tutta la durata dell'interrogatorio il terrorista **guardava** davanti a sé con lo sguardo perso nel vuoto.
226. Negli ultimi giorni i Verdi **hanno detto** cose diverse.
227. Rapimenti e rivolte **sono avvenuti** quotidianamente durante la guerra in Iraq.
228. Durante la presa della scuola di Beslan **sono state eseguite** molte esecuzioni sommarie.
229. Durante il rapimento il Presidente Ciampi **ha telefonato** spesso alle famiglie delle ragazze sequestrate.
230. Durante gli ultimi anni la mostra è **stata allestita** più volte in luoghi diversi.
231. La ragazza **ha subito** più volte violenza durante la guerra in Kosovo.
232. Il pittore trentino **ha preso** più volte la parola durante l'inaugurazione della mostra.
233. Durante la prigionia è **evaso** ben due volte.
234. Nel mese di maggio **ha aperto** il negozio solo tre volte.
235. L'associazione umanitaria **ha aiutato** più volte il popolo iraqueno durante l'embargo.
236. Durante l'attacco i militari americani **hanno sparato** contro un convoglio di automezzi.
237. L'attaccante **si è fermato** più volte durante l'allenamento.

238. Durante la partita l'arbitro **ha estratto** 3 volte il cartellino rosso.
239. L'uomo **ha sparato** più volte durante la prima giornata di caccia.
240. La Lega Nord **ha organizzato** spesso manifestazioni di piazza.
241. I REM **hanno suonato** due volte a Bolzano.
242. Le colline intorno a Genova **hanno preso fuoco** tre volte quest'estate.
243. Durante il dibattito il sindaco **ha telefonato** spesso.
244. Tra oggi e giovedì **sono andati in scena** tre spettacoli.
245. Negli ultimi mesi Baldini **ha firmato** molti autografi.
246. Prodi **ha partecipato** a vari convegni durante il suo mandato di
Presidente del Consiglio.
247. Le guide **hanno organizzato** escursioni sul monte Bondone da marzo a
settembre.
248. Tra il 2003 e il 2004 **sono stati effettuati** controlli sui conti della
Federazione.
249. Tra luglio e settembre l'autore **ha presentato** il suo libro nell'ambito di
diversi eventi culturali.
250. Tra il 2003 e il 2004 il l'Amministratore Delegato **ha redatto** cinque
rapporti sulla situazione patrimoniale della Federazione.
251. Nella partita contro la Moldava Flores **ha segnato** due volte al 35"e al
93" del secondo tempo.
252. Durante l'ultima stagione Chiellini **ha battuto** sette rigori.
253. Gilardino **è caduto** due volte in area di rigore.
254. Ieri sera Alessandro del Piero **ha fatto** due gol importantissimi per l'Italia
di Lippi.
255. Negli ultimi mesi i Carabinieri **hanno arrestato** il ragazzo diverse volte.
256. Tra maggio e giugno il cervo **era fuggito** dal recinto già due volte.
257. Il ragazzo domenica **ha telefonato** più volte al fratello .
258. Negli ultimi sette anni di attività sportiva Baldini **ha vinto** diverse
medaglie.
259. Negli ultimi 5 anni Luca **è stato sottoposto** a trasfusione due volte al

mese.

260. Lo spettacolo è **stato replicato** fino alla metà di dicembre.
261. Maurizio Dini e Isabella Turso **si sono esibiti** più volte su questo palco.
262. Tosi e Bonvicini **si sono accordati** più volte ma sempre con scarso successo.
263. In questi giorni Gianni Letta **ha incontrato** spesso i rappresentanti di "Un ponte per...".
264. Emiliano Bertoldi **ha incontrato** spesso Simona Pari durante la sua missione in Medioriente.
265. La Florida negli ultimi mesi è **stata colpita** da violenti uragani.
266. Nel giro di un paio di anni **sono arrivate** le licenze per cinque centri commerciali.
267. In questi anni i presidenti **hanno presentato** al ministro Calderoli numerosi emendamenti al testo di legge.
268. Gli uffici del Comune **hanno notificato** nell'ultimo anno circa cinquemila contravvenzioni.
269. Purtroppo gli Stati Uniti **hanno risposto** sempre alle varie "chiamate alle armi".
270. Simona Pari **ha raccontato** spesso a Emiliano della sua esperienza in Iraq.
271. La polizia di Beslan **ha compiuto** diversi arresti nei giorni seguenti la strage.
272. Il Presidente **ha incontrato** spesso il Ministro Calderoli negli ultimi mesi.
273. Garbari **ha confermato** in varie occasioni che la forte concorrenza dell'Alto Adige è dannosa per il Trentino.
274. Brugnetti **ha partecipato** già due volte al Memorial Dordoni di Piacenza.
275. Il socio **ha pagato** le rate per l'acquisto della casa per almeno cinque anni.
276. L'organizzazione non governativa "Un ponte per..." **ha operato** spesso in Iraq.

277. La Polizia postale di Reggio Calabria **ha combattuto** spesso il fenomeno delle truffe all'Inps.
278. Le truppe americane **hanno compiuto** molti attacchi ad obiettivi sensibili durante la guerra in Iraq.
279. L'organizzazione non governativa "Un ponte per..." **ha inviato** spesso i propri volontari in Iraq.
280. Il medico del pronto soccorso **è intervenuto** spesso in caso di incidenti stradali.
281. Il pub Simposio **ha proposto** spesso spettacoli teatrali.
282. Durante l'ultimo anno circa 3 milioni di italiani **hanno scaricato** musica senza pagare.
283. I talenti i "cervelli" italiani **sono migrati** spesso verso ambienti di ricerca più liberi e meritocratici.
284. L'associazione Apeiron **ha gestito** spesso progetti di solidarietà in Nepal.
285. La Banda cittadina **ha promosso** spesso corsi di strumenti musicali.
286. Il terzo mondo raramente **ha trovato** gli spazi adeguati per parlare.
287. Il terremoto **ha interessato** raramente Buenos Aires.
288. Spesso le donne **sono state escluse** dalla storia della scienza.
289. La qualità dei servizi **è apparsa** spesso in ribasso.
290. I nuovi politici **hanno denunciato** più volte la grave situazione dei conti pubblici.
291. Galesi **ha sparato** almeno 8 colpi contro Massimo D'Antona.
292. **È caduto** due volte durante la gara.
293. Il prigioniero **ha scritto** poesie durante il periodo di detenzione.
294. La ragazza **è stata sottoposta** due volte ad intervento chirurgico.
295. A Demozzi la Procura **ha contestato** più volte il reato di appropriazione indebita.
296. I talebani **hanno condannato** a morte diversi civili cattolici.
297. Ogni anno il Presidente **ha definito** un piano operativo.
298. Negli ultimi anni l'acquisto del corredo scolastico **ha dissanguato** spesso

- il portafoglio delle famiglie.
299. I genitori di Luca **hanno chiesto** spesso aiuto ai medici .
300. Del Piero **ha segnato** due gol importanti durante la partita contro la Moldavia.
301. A Beslan in una giornata fredda e piovosa **hanno ricevuto** l'ultimo saluto altre 192 vittime della strage della scuola numero 1.
302. Il ministro **ha dato** le dimissioni alla vigilia di un rimpasto di governo.
303. L'esplosione **ha ucciso** tre uomini della Guardia Nazionale irachena.
304. Clinton **ha subito** un intervento al cuore.
305. L'operazione **ha avuto** l'esito sperato.
306. Il padre della piccola rapita **ha lanciato** ieri un nuovo appello ai sequestratori.
307. Una frase infelice di Giovanardi **ha scatenato** la reazione del centrosinistra.
308. Polizia e Guardia di Finanza **hanno inferto** un duro colpo agli affari della cosca.
309. Ieri una giovane donna **si è lanciata** dal balcone a Napoli.
310. L'uomo **ha ucciso** la moglie a fucilate.
311. Ieri pomeriggio **è divampato** all'improvviso un pauroso incendio nella pineta di Sestri Levante.
312. Sabato scorso **si è conclusa** l'ultima edizione del Festival pianistico Busoni.
313. La biografia romanzata di Alcide de Gasperi **ha incontrato** i favori di pubblico e critica.
314. Violante Placido **ha esordito** al cinema giovanissima.
315. Il Numtel **ha segnato** un progresso dello 050%.
316. L'azienda **ha affrontato** un aumento di capitale di 32 milioni di euro.
317. L'immobiliare **ha aumentato** del 40% il proprio fatturato.
318. La Disciplinare di serie C **ha inflitto** al Catanzaro 5 punti di penalizzazione.

319. Alessandro Petacchi **ha vinto** la tappa di domenica.
320. L'attaccante **ha vinto** il pallone d'oro lo scorso anno.
321. Fabio Baldato **ha conquistato** la prima tappa del tour della Polonia.
322. Contro la Norvegia la Under 21 di Gentile **ha vinto** 2-0.
323. L'assemblea dei soci del 21 dicembre **ha stabilito** la messa in liquidazione dell'azienda.
324. L'imprenditore **ha presentato** una denuncia nei confronti del commercialista.
325. Chini **ha deciso** di denunciare Signifredi.
326. Gli inquirenti **hanno acquisito** tutte le cartelle cliniche della donna.
327. La Legionella **ha provocato** la morte di una paziente .
328. Il Rettore **ha lasciato** la carica.
329. Luca **si è iscritto** ieri all'Università.
330. Durante la mattinata di ieri gli studenti **hanno incontrato** il personale dell'Ufficio Accoglienza Stranieri.
331. I sanitari del 118 arrivati sul posto **hanno deciso** di chiamare l'elisoccorso.
332. Sul luogo dell'incidente **è intervenuta** la polizia stradale per i rilievi.
333. A seguito dell'incidente il motociclista **è stato trasportato** al Pronto Soccorso.
334. La moto **ha allargato** una curva finendo contro un autocarro.
335. In pochi minuti **è arrivata** sul luogo dell'incidente l'auto medica della Croce Rossa.
336. Ieri in tarda mattinata un turista **è caduto** nei boschi di Cavalese.
337. Ieri mattina il ladro **ha forzato** un distributore di sigarette.
338. Ieri mattina il giovane **è stato processato** per direttissima.
339. Un autoarticolato ieri pomeriggio **si è ribaltato** sulla strada che porta alla discarica.
340. La conferenza dei capigruppo **si è riunita** ieri a palazzo Thun.
341. Un commando **ha prelevato** le donne dal loro ufficio.

342. Simona Pari **è arrivata** a Baghdad un anno fa.
343. L'operazione non **è durata** più di 5 minuti.
344. L'attacco **è avvenuto** alle 21 a 6 Km. a nord-ovest di Nassirya.
345. Simona Torretta **si è innamorata** di Baghdad nel 1994 al suo primo viaggio.
346. Gli esponenti religiosi **hanno condannato** con fermezza il rapimento delle due volontarie italiane.
347. Un camionista turco **è stato ucciso** ieri durante un attacco ad un convoglio di automezzi.
348. In un lungo interrogatorio Cinzia Banelli **ha raccontato** la sua verità.
349. Romano Prodi **ha preso** i suoi bagagli ed è partito.
350. Il Professore **ha cenato** da solo con Enrico Letta.
351. Un detenuto di 36 anni **si è ucciso** ieri nel carcere di Livorno.
352. Un uomo di 65 anni **ha ucciso** la moglie a coltellate al termine di una violenta lite.
353. L'uomo **ha ucciso** la sua ex-convivente sul molo di Viareggio.
354. I Carabinieri **hanno trovato** la donna in stato confusionale.
355. L'omicida **ha atteso** i Carabinieri sulla soglia di casa.
356. Beslan **ha vissuto** un nuovo giorno di lacrime e funerali.
357. La TV **ha trasmesso** ieri sera un video girato dal commando.
358. Un forte tifone **ha colpito** ieri il Giappone.
359. L'influenza dei polli **ha ucciso** ieri un neonato in Vietnam.
360. La Mostra del Cinema **ha ospitato** ieri il noir del regista francese.
361. Il Mibtel **ha segnato** un calo dello 016%.
362. L'aumento **ha portato** il capitale a 3 milioni di euro.
363. L'Enel **ha revocato** dei lavori importanti.
364. Ieri il consiglio di amministrazione **ha approvato** il bilancio semestrale.
365. Ronaldo **ha chiesto** la mano di Daniella Ciccarelli.
366. La Procura federale **ha confermato** le richieste di sanzione.
367. La nazionale della Guinea **ha battuto** per 4-0 il Botswana.

368. Petacchi **ha vinto** in volata la quarta tappa della Vuelta a Espana.
369. Al torneo di Norton Vijay Singh **ha battuto** Tiger Woods.
370. Il Ministro **ha elaborato** una nuova proposta di legge.
371. L'ispettrice del Provveditorato **ha assistito** ad una delle lezioni della maestra.
372. Alcuni giorni fa il bagnino del Lido di Bolzano **ha vietato** la piscina a due madri musulmane.
373. Il tribunale del riesame **ha annullato** le ordinanze nei confronti di Spagnolo e Trentini.
374. La donna e il marito **hanno deciso** di fare causa all'aggressore.
375. L'ambulanza **ha soccorso** il conducente della vettura coinvolta nell'incidente.

Appendice C: lista delle frasi che compongono il gold standard

Di seguito la lista delle 35 frasi che compongono il gold standard. La valutazione dei lavoratori ha riguardato l'evento espresso dal verbo in grassetto.

1. Valerio Mosele **giocava** a tennis ogni giorno.
2. Simona Torretta **viveva** a Baghdad.
3. Nel 1348 Avignone **era** la sede papale.
4. La Numidia **era** un antico regno berbero.
5. La ragazza **prendeva** abitualmente l'autobus per andare a scuola.
6. La donna **litigava** frequentemente col marito a causa del fumo.
7. La Sibilla Cumana **era** una sacerdotessa di Apollo.
8. L'uomo **stava attraversando** la strada quando è accaduto l'incidente.
9. Enzo Baldoni **era** un volontario della Croce Rossa.
10. Al momento del rapimento le ragazze **lavoravano** nei loro uffici.
11. La bambina ieri pomeriggio **stava giocando** insieme ad alcuni compagni di asilo.
12. Al momento dell'esplosione i fedeli **pregavano**.
13. Gli Etruschi **estraevano** il ferro dall'isola d'Elba.
14. L'uomo è stato bloccato da due agenti mentre **scappava**.
15. Durante l'allenamento Chiellini **si toccava** il ginocchio in continuazione.
16. Durante l'addestramento i bambini **sbadigliavano** continuamente.
17. Quella sera la donna **entrava** e **usciva** da casa senza un apparente motivo.
18. Durante la riunione il delegato provinciale **interveniva** spesso.
19. Babilonia **sorgeva** sulle rive del fiume Eufrate.
20. Durante la lezione, la maestra **telefonava** frequentemente.
21. Estro teatro **proponeva** abitualmente spettacoli di Stefano Benni.
22. Prodi **ha partecipato** a vari convegni durante il suo mandato di Presidente

del Consiglio.

23. Lo spettacolo è **stato replicato** 5 volte.
24. Simona Pari **ha raccontato** spesso a Emiliano della sua esperienza in Iraq.
25. Durante gli ultimi anni la mostra è **stata allestita** più volte in luoghi diversi.
26. **E' caduto** due volte durante la gara.
27. Del Piero **ha segnato** due gol importanti durante la partita contro la Moldavia.
28. Tra luglio e settembre l'autore **ha presentato** il suo libro nell'ambito di diversi eventi culturali.
29. Clinton **ha subito** un intervento al cuore.
30. L'attaccante **ha vinto** il pallone d'oro lo scorso anno.
31. I Carabinieri **hanno trovato** la donna in stato confusionale.
32. Un forte tifone **ha colpito** ieri il Giappone.
33. La donna e il marito **hanno deciso** di fare causa all'aggressore.
34. L'Enel **ha revocato** dei lavori importanti.
35. Il Ministro **ha elaborato** una nuova proposta di legge.

Appendice D: frasi che presentano discordanze nei giudizi

Accanto ad ogni frase sono riportati il giudizio corretto e i giudizi dati dai lavoratori.

Fraasi con disagreement	Giudizio corretto	Giudizi dei lavoratori
Ormai da anni mio figlio andava a scuola con i dei supplenti come insegnanti.	VERO	1. VERO 2. VERO 3. FALSO
I giovani rubavano spesso cosmetici.	VERO	1. VERO 2. VERO 3. FALSO
Simona Pari e Simona Torretta a Baghdad giravano con il velo in testa.	VERO	1. VERO 2. VERO 3. VERO 3. FALSO
Estro teatro proponeva abitualmente spettacoli di Stefano Benni.	VERO	1. VERO 2. VERO 3. FALSO
Simona pari e Simona Torretta lavoravano tra mille difficoltà .	VERO	1. VERO 2. VERO 3. FALSO
Amos Gitai girava due film ogni anno.	VERO	1. VERO 2. VERO 3. FALSO
Javier Bardem interpretava Ramon Sampedro nel film "Mare dentro".	VERO	1. FALSO 2. FALSO 3. FALSO

La ragazza andava spesso a trovare la madre.	VERO	1. VERO 2. VERO 3. FALSO
La donna litigava frequentemente col marito a causa del fumo.	VERO	1. VERO 2. VERO 3. FALSO
Simonetta si recava spesso presso gli uffici di via Poma.	VERO	1. VERO 2. VERO 3. FALSO
La ragazza era spesso a casa della madre negli ultimi tempi.	VERO	1. VERO 2. VERO 3. FALSO
Umberto Berardi si stava recando come ogni mattina a Monza dove avrebbe preso l'autobus per andare a lavoro.	FALSO	1. VERO 2. FALSO 3. FALSO
La squadra di Albarello stava lavorando molto bene.	FALSO	1. VERO 2. VERO 3. FALSO
Gli sforzi della squadra davano i primi significativi frutti.	FALSO	1. VERO 2. VERO 3. FALSO
Sul furto indagavano i Carabinieri di Imer.	FALSO	1. VERO 2. FALSO 3. FALSO
Le ragazze furono sequestrate a Baghdad mentre lavoravano nel loro ufficio.	FALSO	1. VERO 2. FALSO 3. FALSO
Il fotografo li ha sorpresi mentre facevano la quotidiana passeggiata nel parco.	FALSO	1. VERO 2. FALSO 3. FALSO

Durante la performace gli attori si rivolgevano al pubblico.	FALSO	1. VERO 2. FALSO 3. FALSO
Durante le lezioni l'insegnante interrogava gli alunni anche per un'ora consecutiva.	FALSO	1. VERO 2. FALSO 3. FALSO
Durante l'incontro il ragazzo spiegava le sue ragioni.	FALSO	1. VERO 2. FALSO 3. FALSO
In attesa all'ambulatorio il ragazzo tossiva in continuazione.	FALSO	1. VERO 2. FALSO 3. FALSO
Durante la proiezione del film il pubblico applaudiva in continuazione.	FALSO	1. VERO 2. FALSO 3. FALSO
Per tutta la durata dell'interrogatorio il terrorista guardava davanti a sé con lo sguardo perso nel vuoto.	FALSO	1. VERO 2. FALSO 3. FALSO
Rapimenti e rivolte sono avvenuti quotidianamente durante la guerra in Iraq.	FALSO	1. VERO 2. VERO 3. FALSO
Durante il rapimento il Presidente Ciampi ha telefonato spesso alle famiglie delle ragazze sequestrate.	FALSO	1. VERO 2. FALSO 3. FALSO
Le guide hanno organizzato escursioni sul monte Bondone da marzo a settembre.	FALSO	1. VERO 2. FALSO 3. FALSO
Tra il 2003 e il 2004 sono stati effettuati controlli sui conti della Federazione.	FALSO	1. VERO 2. FALSO 3. FALSO

Purtroppo gli Stati Uniti hanno risposto sempre alle varie "chiamate alle armi".	FALSO	1. VERO 2. VERO 3. FALSO
Simona Pari ha raccontato spesso a Emiliano della sua esperienza in Iraq.	FALSO	1. VERO 2. FALSO 3. FALSO
L'organizzazione non governativa "Un ponte per..." ha operato spesso in Iraq.	FALSO	1. VERO 2. VERO 3. FALSO
La Polizia postale di Reggio Calabria ha combattuto spesso il fenomeno delle truffe all'Inps.	FALSO	1. VERO 2. FALSO 3. FALSO
L'organizzazione non governativa "Un ponte per..." ha inviato spesso i propri volontari in Iraq.	FALSO	1. VERO 2. FALSO 3. FALSO
Il medico del pronto soccorso è intervenuto spesso in caso di incidenti stradali.	FALSO	1. VERO 2. FALSO 3. FALSO
Il pub Simposio ha proposto spesso spettacoli teatrali.	FALSO	1. VERO 2. VERO 3. FALSO
Durante l'ultimo anno circa 3 milioni di italiani hanno scaricato musica senza pagare.	FALSO	1. VERO 2. VERO 3. FALSO
I talenti i "cervelli" italiani sono migrati spesso verso ambienti di ricerca più liberi e meritocratici.	FALSO	1. VERO 2. FALSO 3. FALSO
L'associazione Apeiron ha gestito spesso progetti di solidarietà in Nepal.	FALSO	1. VERO 2. VERO 3. FALSO

Il terzo mondo raramente ha trovato gli spazi adeguati per parlare.	FALSO	1. VERO 2. FALSO 3. FALSO
Spesso le donne sono state escluse dalla storia della scienza.	FALSO	1. VERO 2. VERO 3. VERO
La qualità dei servizi è apparsa spesso in ribasso.	FALSO	1. VERO 2. FALSO 3. FALSO
Negli ultimi anni l'acquisto del corredo scolastico ha dissanguato spesso il portafoglio delle famiglie.	FALSO	1. VERO 2. FALSO 3. FALSO
L'operazione ha avuto l'esito sperato.	FALSO	1. VERO 2. FALSO 3. FALSO
Violante Placido ha esordito al cinema giovanissima.	FALSO	1. VERO 2. FALSO 3. FALSO
Fabio Baldato ha conquistato la prima tappa del tour della Polonia.	FALSO	1. VERO 2. FALSO 3. FALSO
Simona Torretta si innamorò di Baghdad nel 1994 al suo primo viaggio.	FALSO	1. VERO 2. FALSO 3. FALSO

Bibliografia

Agrell S. 1908. Aspektänderung und Aktionsartbildung beim polnischen Zeitworte: Ein Beitrag zum Studium der indogermanischen Präverbia und ihrer Bedeutungsfunktionen. Lunds Universitets Arsskrift, new series, I, iv.2.

Ahn L. V., Dabbish L. 2004. Labeling Images with a Computer Game. In ACM Conference on Human Factors in Computing Systems, CHI 2004. New York, NY, USA.

Ahn L.V., Kedia M., Blum M. 2006. Verbosity: A Game for Collecting Common-Sense Knowledge. Negli atti di ACM Conference on Human Factors in Computing Systems, CHI Notes 2006. Stork D.G. 1999. The Open Mind Initiative. IEEE Expert Systems and Their Applications, 16-20.

Allen J., Swift M., de Beaumont W. 2008. Deep semantic analysis of text. Negli atti del Symposium on Semantics in Systems for Text Processing (STEP). Venezia, Italia.

Bach, E. 1986. The algebra of events. Linguistics and Philosophy 9: 5–16.

Baker C. F., Fillmore C. J., Lowe J. B. 1998. The Berkeley FrameNet project. Negli atti di COLING-ACL 1998. Montreal, CA.

Bertinetto P.M. Il verbo. Renzi L. – Salvi G. (a cura di). 1991. Grande grammatica italiana di consultazione, vol. II, I sintagmi verbale, aggettivale, avverbiale. La subordinazione. Il Mulino, Bologna.

Bertinetto P.M. 1994. Le perifrasi abituali in italiano ed in inglese. Quaderni dei Laboratorio di Linguistica 8.32-41.

Banko M, Brill E. 2001. Scaling to Very Very Large Corpora for Natural Language Disambiguation. Negli atti di ACL-2001. Stroudsburg, PA, USA.

Baroni M., Bernardini S., Comastri F., Piccioni L., Volpi A., Aston G., Mazzoleni M. 2004. Introducing the la Repubblica corpus: a large, annotated, TEI(XML)-compliant corpus of newspaper Italian. A cura di Lino M., Xavier M., Ferreira F., Costa R., Silva R., atti di LREC 2004. Lisbona, Portogallo.

Benkler, Y. 2007. La ricchezza della Rete. La produzione sociale trasforma il mercato e aumenta le libertà Università Bocconi editore, Milano.

Berger A., Della Pietra S., Della Pietra V. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

Bertinetto P.M. 2003. Sulle proprietà tempo-aspettuali dell'Infinito in italiano. Atti del XXXV congresso internazionale di studi della Società di Linguistica Italiana (SLI), Parigi, 20-22 settembre 2001. Roma, Bulzoni.

Bertinetto P.M., Lenci A. 2011. Pluriactionality, habituality and gnomic imperfectivity.

Bertinetto P.M., Squartini M. 1995. An attempt at defining the class of 'gradual completino verbs', in P.M. Bertinetto – V. Bianchi – J. Higginbotham – M. Squartini (eds), *Temporal reference. Aspect and actionality*, Rosenberg & Sellier, 11-26, Torino.

Bittar A. 2008. Annotation des informations temporelles dans des textes en français,. Negli atti RECITAL 2008. Avignon, France.

Bittar A. 2009. Annotation of Events and Temporal Expressions in French Texts.. Negli atti di LAW III. Singapore.

Brent, M. 1991. Automatic semantic classification of verbs from their syntactic contexts: an implemented classifier for stativity. Negli atti di the fifth conference on European chapter of the Association for Computational Linguistics: 222-226.

Brizzi G. 1997. *Storia di Roma*. Patron, Bologna.

Caselli T., Prodanof I. 2010. Annotating Event Anaphora: A Case Study. Negli atti di LREC 2010 - Seventh conference on International Language Resources and Evaluation (Valletta, Malta, 17-23 May 2010). Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner, Daniel Tapias (eds.). La Valletta, Malta.

Chklovski T., Mihalcea R. 2002. Building a sense tagged corpus with Open Mind Word Expert. Negli atti di Word Sense Disambiguation: Recent Successes and Future Directions, ACL 2002. Philadelphia, U.S.A

Cohen, J. 1960. A coefficient of agreement for nominal scales. In *Educational and Psychological Measurement*. New York: Sage Publications, 20, pp. 37-46.

Comrie B. 1976. *Aspect: An Introduction to Verbal Aspect and Related Problems*. Cambridge University Press.

Crystal, D. 1992. *An Encyclopedic Dictionary of Language and Languages*. Oxford.

Cusic D.D. 1981. *Verbal Plurality and Aspect*. PhD dissertation, Stanford University.

Dostál A. 1954. *Studie o vidovém systému v staroslověnině*. SNP, Praha.

Dowty, D. R. 1977. Toward a semantic analysis of verb aspect and the English 'imperfective progressive'. *Linguistics and Philosophy*. 1: 45-78.

Dowty D.R. 1979. *Word Meaning and Montague Grammar*. Kluwer, Dordrecht.

Dressler W.U. 1968. Studien zur Verbalen Pluralität. Wien: Österreichische Akademie der Wissenschaften.

Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson G. (2005) TIDES 2005 Standard for the Annotation of Temporal Expressions.

http://fofoca.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf

Filip H. 1993. Aspect, Situation Types and Nominal Reference (Ph.D. Dissertation). Berkeley (CA), University of California.

Filip H. 2000. "The Quantization Puzzle." James Pustejovsky, and Carol L. Tenny (eds.). Events as Grammatical Objects, from the Combined Perspectives of Lexical Semantics and Syntax. Stanford: CSLI Press, 3-60.

Filatova E., Hovy E. 2001. Assigning Time-Stamps To Event-Clauses. Negli atti del workshop ACL on Temporal and Spatial Information Processing. Toulouse, France.

Garside, R., Leech, G. and McEnery, T. (Eds.) (1997). Corpus Annotation: Linguistic Information from Computer Text Corpora. London: Longman. Editore, Roma.

Harris Z. S. 1968. Mathematical Structures of Language. Wiley, New York.

Howe, J. 2009. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. Crown Publishing Group.

Ide, N., Romary, L. 2002. Standards for Language Resources. Negli atti di LREC 2002. Las Palmas, Gran Canaria, 59-65.

ISO: Language Resource Management – Semantic Annotation Framework (SemAF) - Part 1: Time and Events. Secretariat KATS, August 2007. ISO Report ISO/TC37/SC4 N269 version 19 (ISO/WD 24617-1).

Ide N., Romary L. and de la Clergerie E. 2003. International standard for a Linguistic Annotation Framework. Negli atti di HLT-NAACL'03 - Workshop on The Software Engineering and Architecture of Language Technology. Edmonton, Canada.

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

Jenkins H. 2006. White paper *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. http://digitalllearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF

Kaisser M. Lowe J. 2008. A Research Collection of QuestionAnswer Sentence Pairs. Negli atti di LREC 2008. Marrakech, Marocco.

Kamp H. 1979. Events, Instants and Temporal Reference. In *Semantics from Different Points of View*, eds. Rainer Bäuerle, Urs Egli and Arnim von Stechow. Berlin: Springer.

Katz G., Arosio F. 2001. The Annotation Of Temporal Information In Natural Language Sentences. Negli atti del workshop ACL on Temporal and Spatial Information Processing. Toulouse, France, 104-111.

Kenny A. 1963. *Actions, Emotions, and Will*. Humanities Press, London.

Kittur A., Chi E.H., Suh B. 2008. Crowdsourcing user studies with Mechanical Turk. Negli atti di CHI-2008.

Kohonen T. 1997. Self-organizing Maps. Springer, New York, USA.

Kratzer, A. (1995). Stage Level and Individual Level Predicates. In Carlson, G.; Pelletier, F.J. (eds.), *The Generic Book*. Chicago: The University of Chicago Press.

Lagus K., Airola A. 2005. Semantic clustering of verbs-analysis of morphosyntactic contexts using the som algorithm. In A. Lenci, S. Montemagni e V. Pirrelli editori. *Acquisition and Representation of Word Meaning: Theoretical and computational perspectives*. *Linguistica Computazionale XXII-XXIII*. IEPI, Pisa-Roma.

Lafferty J.D., McCallum A., Pereira F.C.N. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Negli atti di ICML pages 282–289. Williams College, Williamstown, MA, USA.

Landman, F. 1992. The progressive. *Natural Language Semantics*. 1: 1-32.

Landman F. 2000. *Events and Plurality: The Jerusalem Lectures*. Kluwer, Dordrecht.

Lenci A., Montemagni S., Pirrelli V. 2005. *Testo e computer*. ISBN 88-430-3425-1. Carrocci Editore.

Linguistic Data Consortium. 2004. ACE (Automatic Content Extraction) English Annotation Guidelines for Entities, version 5.6.1 2005.05.23.

http://projects.ldc.upenn.edu/ace/docs/English-Entities-Guidelines_v5.6.1.pdf

Llorens H., Saquete E., Navarro B. 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval 2. Negli atti di the 5th International Workshop on Semantic Evaluation, pp. 284–291. Uppsala, Svezia.

Lucchesi V. 1971. Fra grammatica e vocabolario. Studio sull'aspetto del verbo italiano. *Studi di grammatica italiana*, 1:179–270.

Magnini B., Pianta E., Girardi C., Negri M., Romano L., Speranza M., Bartalesi Lenzi V., Sprugnoli V. 2006. I-CAB: the Italian Content Annotation Bank. Negli atti del convegno LREC 2006, Genova, Italia.

Marcus M. P., Marcinkiewicz M. A., Santorini B. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics* 19:2.

Mathew T., Katz G. 2009. Supervised Categorization of Habitual and Episodic Sentences. In *Sixth Midwest Computational Linguistics Colloquium*, Indiana University. Bloomington, Indiana.

Miller G. A., Leacock C., Teng R., Bunke R.T. 1993. A semantic concordance. In *Negli atti di HLT-Human Language Technology Workshop*. San Francisco, USA.

Montemagni S., Barsotti F., Battista, M., Calzolari N., Corazzari O., Lenci A., Zampolli A., Fanciulli F., Massetani M., Raffaelli R., Basili R., Pazienza M.T., Saracino D., Zanzotto F., Mana N., Pianesi F., Delmonte R.. 2003. Building the italian syntacticsemantic treebank. In A. Abeill'e, editor, *Treebanks. Building and Using Parsed Corpora*, pages 189–210. Kluwer, Dordrecht.

Newman P. 1980. *The classification of Chadic within Afroasiatic*. Universitaire Press, Leiden.

Parsons T. 1990. *Events in the Semantics of English*. Cambridge, Mass, MIT Press.

Palmer M., Gildea D., Kingsbury P. 2005. The Proposition Bank: A Corpus Annotated with Semantic Roles. *Computational Linguistics*, 31:1.

Pustejovsky J., Castaño J., Ingria R., Saurí R., Gaizauskas R., Setzer A., Katz G. 2003a. Timeml: Robust Specification of Event and Temporal Expressions in Text. Negli Atti di IWCS-5, Fifth International Workshop on Computational Semantics. Tilburg, The Netherlands.

Pustejovsky J., Hanks P., Saurí R., See A., Gaizauskas R., Setzer A., Radev D., Sundheim B., Day D., Ferro L., Lazo M. 2003b. The Timebank Corpus. *Corpus Linguistics*, 647–656.

Pustejovsky J., Knippen R., Littman J., Saurí R. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39:123-164. Springer-Verlag Berlin.

Pustejovsky J., Littman J., Saurí R., Verhagen M. 2006. TimeBank 1.2 Documentation. <http://timeml.org/site/timebank/documentation-1.2.html>

Robaldo L, Caselli T., Russo I., Grella M. 2011. From Italian Text to TimeML Document via Dependency Parsing, Negli atti di CicLing-2011. Tokyo, Giappone.

Romagno D. 2005. La codificazione degli attanti nel mediterraneo romanzo: accordo del participio e marcatura dell'oggetto. *Archivio Glottologico Italiano*, 90,1: 90-113.

Rothstein S. 2004. Structuring Events - A Study in the Semantics of Lexical Aspect. Blackwell Publishing, Oxford.

Rheingold, H. 2002. *Smart Mobs: The Next Social Revolution*. Basic Books.

Ryle G. 1949. *The concept of mind*. Hutchinson & Company, London.

Saurí R., Knippen R., Verhagen M, Pustejovsky J. 2005. Evita : a robust event recognizer for q/a systems. Negli atti di hlt/emnlp 2005, 700–707.

Setzer, A., Gaizauskas. R. 2001. A pilot study on annotating temporal relations in text. Negli atti del workshop ACL on Temporal and Spatial Information Processing. Toulouse, France, 73–80.

Siegel, E. 1999. Corpus-Based Linguistic Indicators for Aspectual Classification. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics: 112-119. San Francisco: Morgan Kaufmann.

Singh P. 2002. The public acquisition of commonsense knowledge. Negli atti di AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA.

Shirky C. 2008. Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Press.

Smith C. 1991. The Parameter of Aspect. Dordrecht: Kluwer.

Snow R., O'Connor B., Jurafsky D., Ng A. 2008. Cheap and fast - but is it good?: evaluating non-expert annotations for natural language tasks. Negli atti di EMNLP 2008. Morristown, NJ, USA.

Su Q., Pavlov D., Chow J.H., Baker W. C. 2007. Internet-Scale Collection of Human-Reviewed Data. In Proc. of WWW-2007.

Suh, S. 2006. Extracting Generic Statements for the Semantic Web. MSc thesis In Artificial Intelligence, School of Informatics, University of Edinburgh.

UzZaman N., Allen J. 2010. TRIOS-TimeBank Corpus: Extended TimeBank Corpus with Help of Deep Understanding of Text. Negli atti di LREC 2010. Malta.

Vendler Z. 1967. *Linguistics in Philosophy*. Ithaca. Cornell, New York.

Verhagen M. 2010. *The Brandeis Annotation Tool*. Negli atti di LREC 2010. Malta.

Verhagen M., Gaizauskas R., Schilder F., Hepple M., Katz G., Pustejovsky J. 2007. *Semeval-2007 task 15: Tempeval temporal relation identification*. Negli atti di Fourth Int. Workshop on Semantic Evaluations (SemEval-2007), pages 75–80. Praga, Repubblica Ceca.

Verhagen M., Gaizauskas R., Schilder F., Hepple M., Moszkowicz J., Pustejovsky J. 2009. *The tempeval challenge: identifying temporal relations in text*. *Language Resources and Evaluation. Journal of Language Resources and Evaluation* 43(2), 161-179.

Verhagen M., Saurí R., Caselli T., Pustejovsky J. 2010. *SemEval-2010 Task 13: TempEval-2*. Negli atti del 5th International Workshop on Semantic Evaluation. Uppsala, Sweden.

Verkuyl H.J. 1972. *On the compositional nature of the aspects*. Reidel, Dordrecht.

Vitolo G. 2000. *Medioevo. I caratteri originali di un'età di transizione*. Sansoni, Firenze.

Xrakovskij V.S. 1997. *Semantic types of the plurality of situations and their natural classification*. In Id. (Ed.), *Typology of iterative constructions*. Lincom, München-Newcastle.

Zarcone, A., Lenci A. 2008. *Computational Models of Event Type Classification in Context*. Negli atti di LREC-08. Marrakech, Marocco.