

Università di Pisa
Facoltà di Lettere e Filosofia
Corso di Laurea in Informatica Umanistica

Anno Accademico 2005/2006

Elaborato finale

**LA CLASSIFICAZIONE AZIONALE
DEL VERBO ITALIANO:
PRIMI ESPERIMENTI COMPUTAZIONALI**

Candidato
Alessandra Zarcone

Relatore
Prof. Alessandro Lenci

Indice

1	La classificazione azionale	7
1.1	L'Azionalità	7
1.2	Tratti distintivi	8
1.2.1	Duratività	8
1.2.2	Dinamicità	8
1.2.3	Omogeneità	9
1.3	Classi azionali	10
1.3.1	States (stativi)	10
1.3.2	Activities (processi)	10
1.3.3	Accomplishments (realizzazioni)	10
1.3.4	Achievements (conseguimenti)	11
1.3.5	Gli incrementativi	11
1.4	Le 'opposizioni infralessicali'	12
1.5	Un approccio computazionale	13
2	L'annotazione e il modello computazionale	15
2.1	L'annotazione del corpus e il Gold Standard	15
2.2	Apprendimento automatico: la Maximum Entropy	16
2.2.1	La k-fold cross-validation	17
2.2.2	Precisione, accuratezza, f-measure	17
2.3	Classi azionali fondamentali?	18
2.4	Il problema dell'Inter-Tagger Agreement	20
2.5	Una baseline per i tratti	21
3	Marche contestuali	23
3.1	Tratti morfologici tempo-aspettuali e diatesi della testa verbale	24
3.2	Argomenti e complementi	24
3.3	Modificatori della testa verbale	25
3.4	Lessicalizzare alcune marche contestuali	26
4	Primi esperimenti computazionali	27
4.1	La decisione su classi	27
4.1.1	Confronto con altri annotatori	28

4.1.2	Separazione delle marche	28
4.1.3	Errori comuni del sistema	28
4.2	La decisione su tratti	30
4.3	Correlazioni tra marche e classi azionali	31
5	Conclusioni	35
6	Appendice	37

Introduzione

La presente relazione illustra i primi risultati di un'analisi computazionale dell'azionalità del verbo italiano, seguendo l'ipotesi (presentata per l'inglese americano da Siegel e McKeown (2000)[14]) che specifici indicatori contestuali presenti in una frase possano fornire delle informazioni rilevanti sulla categoria azionale del verbo della frase stessa.

Un modello computazionale della classificazione azionale può rivelarsi particolarmente interessante per il trattamento di fenomeni come le “opposizioni infralessicali” e la “commutazione azionale”. Le opposizioni infralessicali (Lucchesi 1971 [8]) sono molto comuni nella lingua italiana: caratterizzano verbi molto frequenti, che mostrano un'alternanza statisticamente significativa tra due interpretazioni azionali (come ad esempio il verbo *conoscere*: “Conoscono bene Carmen” \iff “Quattro mesi fa ho conosciuto Carmen”). La commutazione azionale è invece un fenomeno che può coinvolgere anche verbi caratterizzati da una interpretazione azionale fortemente prototipica, se associati a particolari marche contestuali: tali marche, che interagiscono con il valore semantico del verbo e lo modificano, contribuiscono dunque in gran parte a determinare il valore azionale di quel particolare evento in quella data situazione contestuale (“ibridismo azionale” - Bertinetto 1986 [2]).

L'analisi computazionale di un problema linguistico ha bisogno di alcuni elementi fondamentali: un modello teorico linguistico, un modello sperimentale, un corpus di riferimento annotato adeguatamente.

Il modello di classificazione azionale adottato è quello introdotto da Zeno Vendler [16], con particolare attenzione anche all'elaborazione teorica alle problematiche specifiche del verbo italiano presentate da Pier Marco Bertinetto [2].

Il modello sperimentale adottato è l'apprendimento automatico supervisionato detto Maximum Entropy (principio della massimizzazione dell'entropia - Berger *et al.* 1996 [1]). L'uso dell'apprendimento automatico può essere un modo innovativo di modellizzare il problema della classificazione azionale, ma può anche fornire consistenza “sperimentale” alle correlazioni teoriche tra marche contestuali e categorie azionali: la rilevanza delle diverse marche contestuali è valutata in base ai risultati ottenuti dall'agente in grado di apprendere, che, dopo una prima fase di addestramento, deve decidere circa la categoria azionale di una determinata forma verbale. Un analizzatore di questo tipo sfrutta soltanto elementi presenti nel contesto, senza l'ausilio di informazioni semantiche interne al verbo stesso, nè conoscenze afferenti al dominio della pragmatica, che spesso nell'uomo hanno un ruolo rilevante.

Il corpus di riferimento per l'addestramento del sistema è costituito da frasi dal corpus Treebank¹. Quando possibile, sono state reperite nel Treebank anche le frasi di esempio presenti nella relazione. Il lavoro sperimentale è stato preceduto da una fase di annotazione delle frasi utilizzate per quanto riguarda l'azionalità, per permettere lo svolgimento del progetto.

La prima parte della relazione sarà dedicata all'illustrazione del modello vendleriano e a questa prima fase di annotazione. Quindi sarà illustrato il modello sperimentale della Maximum Entropy, e infine saranno presentati i risultati ottenuti.

N.B. convenzioni utilizzate nel testo e nelle tabelle:

STA="states";

ACT="activities";

ACH="achievements";

ACC="accomplishments";

INC="incrementativi";

freq.="frequenza"; prot.="percentuale dell'interpretazione più frequente" (moda); le tabelle con i risultati riportano in orizzontale le "risposte corrette", in verticale le risposte fornite dal sistema, sulla diagonale le risposte che il sistema riesce a "indovinare".

¹Treebank Sintattico Semantica dell'Italiano del progetto SI-TAL (progetto finanziato dal MURST nell'ambito della legge 46/82 art.10 e realizzato sotto la direzione scientifica del Prof. Zampolli)

Capitolo 1

La classificazione azionale

1.1 L’Azionalità

L’Azionalità¹ è un concetto strettamente semantico, intrinseco in ogni forma verbale. Spesso l’Azionalità viene confusa con l’Aspetto², che esprime il modo in cui si guarda a un evento. Non è nell’interesse di questa trattazione addentrarci nel problema delle relazioni tra Aspetto e Azionalità: ci limiteremo a ricordare che, a differenza dell’Aspetto, l’Azionalità non dipende dalla coniugazione e non è genericamente collegata all’informazione morfologica.

Ai fini della presentazione del lavoro svolto è invece interessante sottolineare che il contesto è quindi particolarmente rilevante per l’individuazione della classe azionale di una forma verbale. Per quanto un verbo possa essere considerato un ‘rappresentante prototipico’ di una classe azionale, saranno comunque possibili casi in cui particolari elementi contestuali (come la struttura argomentale o la stessa morfologia del verbo) potranno essere decisivi per la determinazione della classe azionale di una forma specifica di quel verbo, operando una vera e propria ‘commutazione azionale’.

Zeno Vendler [16] introduce un modello di classificazione azionale articolato in quattro classi: *states*, *activities*, *achievements*, *accomplishments*. Ognuna di queste classi è descrivibile anche secondo la presenza o assenza di tre tratti fondamentali ([+/-omogeneo], [+/-durativo], [+/-dinamico]), come in tabella 1.1. Nel basarci sul modello vendleriano, vi abbiamo aggiunto anche una quinta classe, quella dei verbi cosiddetti “incrementativi” (Bertinetto e Squartini 1995 [3]), non compresa nella tassonomia vendleriana, identificato dal tratto [+/-incrementativo].

¹Al concetto di Azionalità si fa riferimento anche con i termini ‘Azione’, ‘carattere del processo’, ‘modo dell’azione’, ‘significato intrinseco’ o anche ‘Aspetto oggettivo’ (in contrapposizione con l’Aspetto propriamente detto, denominato ‘Aspetto soggettivo’) - ‘Azione’ e ‘Azionalità’ sono traduzioni del tedesco *Aktionsart* (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).

²Talvolta invece si fa riferimento all’Azionalità come ‘Aspetto oggettivo’ e all’Aspetto come ‘Aspetto soggettivo’.

	[omogeneo]	[durativo]	[dinamico]	[incrementativo]
STA	+	+	-	-
ACT	+	+	+	-
ACC	-	+	+	-
ACH	-	-	+	-
INC	-	+	+	+

Tabella 1.1: tabella riassuntiva dei tratti caratterizzanti di ogni classe

1.2 Tratti distintivi

1.2.1 Duratività

La distinzione più intuitiva è quella tra verbi come *lavorare*, *amare*, *spingere una carriola* e verbi come *partire*, *tornare*, *esplodere*: i primi esprimono eventi che sono percepiti come prolungati nel tempo, gli altri fanno riferimento ad eventi percepiti come puntuali, in cui il punto iniziale e quello finale tendono a coincidere.

(1.1) Sono quindici anni che **lavoro** nel Dubai.

(1.2) **Partirà** a novembre la tv dei Benetton.

Un tipico criterio di verifica sintattico per l'opposizione [+/-durativo] è l'uso di espressioni temporali durative o, di contro, puntuali. Espressioni introdotte da *'finché'* o come *'per x tempo'*, non compatibili con eventi percepiti come 'istantanei', non sono ammissibili come grammaticali per i non-durativi, mentre al contrario espressioni di localizzazione temporale puntuale come *'alle 3 in punto'* non sono compatibili con il tratto [+durativo]:

(1.3) I soccorritori **hanno dovuto lavorare** per una quindicina di minuti.

(1.4) ***Si sfracellò** per un'ora.

D'altra parte in alcuni casi espressioni di questo tipo possono operare una commutazione azionale su un verbo prototipicamente non-durativo, restituendone un'accezione iterativa (e quindi durativa), come nel seguente esempio:

(1.5) **Starnutì** per tutta la giornata.

(1.6) I turisti stranieri sono ancora numerosi e **continuano ad arrivare** in Istria.
[‘continuare’, soggetto plurale]

1.2.2 Dinamicità

I verbi caratterizzati dal tratto [+dinamico] presuppongono genericamente lo svolgersi di un evento dinamico, i verbi [-dinamici] invece descrivono il protrarsi di uno stato. Per il momento ci limitiamo a notare una serie di caratteristiche tipiche dei

verbi non dinamici, come *assomigliare*, *trovarsi*, *conoscere* (nel senso di ‘essere a conoscenza’). In primo luogo, a differenza dei verbi [+dinamici], essi non ammettono soluzioni di continuità (‘condizione di densità’³: non possono essere interrotti dal soggetto durante il loro svolgersi, mentre se una persona sta ‘ridendo’ o ‘lavorando’ questi processi possono subire delle brevi interruzioni). In secondo luogo, se passiamo a esaminare i criteri di verifica sintattici, tale classe è caratterizzata dalla bassa compatibilità con l’imperativo e con la perifrasi progressiva:

(1.7) ***Assomigliami!**

(1.8) *Sanare la giustizia in crisi è quindi urgente e il piano immediato **sta comprendendo** quattro leggi di riforma.

1.2.3 Omogeneità

Un’altro tratto è quello che distingue eventi che si svolgono in maniera omogenea ed eventi che invece tendono verso un fine, una condizione terminale necessaria perché si dica che l’evento è effettivamente avvenuto. Il tratto di omogeneità viene così descritto da Vendler (1967):

If it is true that someone has been running for half an hour, then it must be true that he has been running for every period within that half an hour. But even if it is true that a runner has run a mile in four minutes, it cannot be true that he has run a mile in any period which is a real part of that time, although it remains true that he was running, or that he was engaged in running a mile, during any substretch of those four minutes. Similarly, in case I wrote a letter in an hour, I did not write it, say, in the first quarter of that hour. It appears, then, that running and its kind go on in time in an homogeneous way; any part of the process is of the nature of the whole. Not so with running a mile or writing a letter; they also go on in time, but they proceed towards a terminus which is logically necessary to their being what they are. Somehow this climax casts its shadow backward, giving a new color to all that went before.[16]

Gli eventi caratterizzati dal tratto [–omogeneo] (detti ‘telici’), hanno bisogno quindi di un completamento per ‘essere ciò che sono’; il loro completarsi, tuttavia, non deve necessariamente attuarsi nella realtà di cui parliamo, ma è inerentemente compreso nel significato lessicale del verbo. Si noti ad esempio che se un verbo telico viene coniugato secondo l’imperfetto – o la perifrasi imperfettiva –, l’enunciato non permette alcuna implicazione circa la conclusione dell’evento.

(1.9) Giovanni **stava scrivendo un libro** NON IMPLICA Giovanni **ha scritto un libro**.

Infatti se diciamo che ‘Giovanni sta scrivendo un libro’, non presupponiamo che il libro venga effettivamente completato da Giovanni, ma presupponiamo un mondo possibile *w*, continuazione naturale del mondo di cui stiamo parlando, in cui il libro viene completato.

³Bertinetto 1986 [2]

Alcuni criteri di verifica sintattica possono essere l'espressione '*in x tempo*' (che indica in quanto tempo l'evento è giunto a compimento), la presenza di espressioni numeriche precise nell'oggetto o, nel caso di verbi di movimento, l'uso di complementi di moto a luogo:

(1.10) Finora **abbiamo venduto** 114 lotti..

(1.11) Chiara **ha completato** la relazione in mezz'ora.

1.3 Classi azionali

1.3.1 States (stativi)

La classe di verbi che Vendler (1967)[16] chiama *states* (stativi⁴) è caratterizzata dai tratti [+omogeneo], [+durativo], [-dinamico] (cfr. tabella 1.1).

Gli stativi indicano qualità, abitudini, abilità e altre caratteristiche, permanenti o transitorie, possedute dal soggetto.

(1.12) **Si trovava** in Calabria in compagnia della moglie.

1.3.2 Activities (processi)

Come abbiamo già accennato, la 'condizione di densità' è una prerogativa degli stativi. Ammettono invece soluzioni di continuità i verbi appartenenti alla classe che Vendler (1967)[16] chiama *activities* (processi o continuativi⁵), i cui tratti caratteristici sono [+omogeneo], [+durativo], [+dinamico] (cfr. tabella 1.1).

Il tratto [+omogeneo] deve quindi essere inteso in termini di una 'densità' più ridotta rispetto agli stativi. L'omogeneità inoltre distingue le *activities* dai verbi di *accomplishment*. In alcuni casi per 'telicizzare' un verbo basta aggiungervi una 'finalizzazione', come può essere quella introdotta da un complemento oggetto:

(1.13) Di notte si può star seduti e si può leggere, studiare, **scrivere** a casa. Io studio l'inglese e **scrivo** a Claudia, la mia fidanzata.
[*activity*]

(1.14) Giorgio Celli **ha scritto** un romanzo giallo.
[*accomplishment*]

1.3.3 Accomplishments (realizzazioni)

Come abbiamo appena sottolineato, le *activities* e gli *accomplishments*⁶ (realizzazioni o risultativi⁷) si distinguono per il tratto [+/-omogeneo]. Gli *accomplishment*

⁴terminologia di Bertinetto (1986)[2]

⁵terminologia di Bertinetto (1986)[2]

⁶terminologia di Vendler (1967)[16]

⁷terminologia di Bertinetto (1986)[2]

sono infatti caratterizzati dai tratti [−omogeneo], [+durativo], [+dinamico] (cfr. tabella 1.1).

Gli *accomplishments* sono quindi verbi ‘telici’ [−omogenei] e [+durativi], e spesso si tratta di *activities* che vengono ‘telicizzate’ dalla presenza di un complemento oggetto (*mangiare* ≠ *mangiare una mela*).

Sono validi per gli accomplishments test di telicità come ‘*in x tempo*’ che, in questo caso, indicano il tempo necessario per il totale compimento dell’azione.

1.3.4 Achievements (conseguimenti)

La classe di verbi che Vendler (1967)[16] chiama *achievements* (conseguimenti) è caratterizzata dai tratti [−omogeneo], [−durativo], [+dinamico] (cfr. tabella 1.1).

Anche in questo caso si tratta di telici ([−omogenei]), ma si differenziano dagli *accomplishments* per il tratto [−durativo]. Gli *achievements* indicano in genere un mutamento di stato del soggetto e quindi sono detti anche ‘trasformativi’⁸.

Una caratteristica interessante degli *achievements* è che, al progressivo, sottintendono due possibili interpretazioni semantiche. Ad esempio, una frase come ‘Il treno **sta partendo**’ può essere interpretata in due modi: ‘il treno sta per partire’ [interpretazione imminente], oppure ‘il treno sta iniziando a muoversi’ [interpretazione imperfettiva].

L’espressione ‘*in x tempo*’, tipicamente usata come test di telicità, implica un’accezione particolare nel caso dei trasformativi. L’espressione proietta l’attenzione su una fase ‘preparatoria’ precedente al momento ‘culminante’ (quest’ultimo, infatti, è tipicamente puntuale) e indica il tempo necessario per la ‘preparazione’ dell’evento.

1.3.5 Gli incrementativi

Le categorie vendleriane, come viene fatto notare da Bertinetto e Squartini (1995)[3], non soddisfano ogni esigenza tassonomica riguardo alle classi azionali. Si osservino ad esempio le seguenti frasi:

- (1.15) Nel decennio ‘74–‘84 **sono aumentati** i detenuti per terrorismo e per tossicodipendenza.
- (1.16) Quando il pesce **creseva** troppo veniva abbandonato nel fiume più vicino.
- (1.17) Nei prossimi giorni, le presenze **si ridurranno** ulteriormente, fino a toccare un minimo di 350 mila a Ferragosto.

Verbi come *crescere*, *ridurre*, *ingrassare*, *ingiallire*, che Bertinetto e Squartini definiscono *gradual completion verbs* (=‘incrementativi’), esprimono un approccio graduale verso un obiettivo più o meno definito – hanno quindi una forte caratterizzazione semantica che li rende già una classe coerente.

⁸terminologia di Bertinetto (1986)[2]

Il fatto stesso che tali verbi esprimano un approccio graduale li identifica come [+dinamici] e [+durativi] e come compatibili con il progressivo: si potrebbero quindi assimilare agli *achievements* e alle *activities*. Gli incrementativi tuttavia si distinguono da queste ultime classi per alcune caratteristiche peculiari:

- mentre i verbi di *accomplishment* sono poco compatibili con ‘*poco*’, ‘*molto*’, ma sono compatibili con ‘*gradualmente*’ (‘Chiara **ha completato gradualmente** il ricamo della tovaglia’ ma non ‘*Chiara **ha scritto molto** una lettera’) e i verbi di *activity* si comportano in modo diametralmente opposto (‘**Ha corso molto**’ ma non ‘***Ha corso gradualmente**’), gli incrementativi sono compatibili con avverbi di entrambi i tipi (‘La situazione è **gradualmente migliorata**’, ‘Chiara è **molto dimagrita**’);
- i verbi incrementativi sono inoltre compatibili con una serie di avverbi di significato intrinsecamente comparativo, come *di poco*, *di molto*, *di parecchio*, *ulteriormente*, *sensibilmente*, *appena*, *apprezzabilmente*;
- se coniugati secondo il paradigma di un tempo perfettivo, gli incrementativi sono ambigui: ammettono due possibili interpretazioni semantiche. Ad esempio, una frase come ‘la pianta è **cresciuta**’ può significare che è stato raggiunto l’obiettivo finale [*crescere* = ‘diventare grande’], oppure che è stato raggiunto uno stadio intermedio [*crescere* = ‘diventare più grande’]. In entrambe le accezioni, il verbo è da considerare telico [+omogeneo] in quanto compatibile con i classici test di telicità, come l’avverbiale ‘*in x tempo*’.

Riassumendo, quindi, i tratti con cui abbiamo contraddistinto tale classe di verbi sono i seguenti:

[omogeneo]	[durativo]	[dinamico]	[incrementativo]
–	+	+	+

Tabella 1.2: tabella riassuntiva dei tratti caratterizzanti i verbi incrementativi

1.4 Le ‘opposizioni infralessicali’

Alcuni verbi presentano una naturale polisemia tra due interpretazioni azionali, che può emergere nel contrasto tra frasi con tempi verbali diversi o nell’opposizione tra soggetto animato e soggetto inanimato (‘opposizioni infralessicali’ - Lucchesi 1971 [8]). Il fenomeno si manifesta come un vero e proprio ibridismo azionale: un verbo come *ricordarsi*, ad esempio, può equivalere ad ‘arrivare a ricordarsi’ (interpretazione culminativa o ingressiva) o ad ‘avere un ricordo’ (interpretazione durativa o risultativa), dove l’uno indica ‘l’attuarsi dell’azione indicata dall’altro’[8]. Altri verbi di questo genere sono ad esempio *chiudere* (*giungere a chiudere* ⇔ *tener chiuso*),

conoscere (venire a conoscenza \Leftrightarrow essere a conoscenza), *indicare* (puntare, mostrare, rivelare \Leftrightarrow stare nella posizione di chi indica).

- aspetto perfetto \Leftrightarrow aspetto imperfetto:

(1.18) a. **Calzò** un elegantissimo paio di sandali.
[interpretazione culminativa o ingressiva][8]

(1.18) b. **Calzava** un elegantissimo paio di sandali.
[interpretazione durativa o risultativa][8]

- soggetto dinamico \Leftrightarrow soggetto statico:

(1.19) a. Il ministro del Tesoro **ha indicato** anche l'obiettivo del prossimo anno.
[interpretazione culminativa o ingressiva]

(1.19) b. Le previsioni **indicano** che ci vorranno una decina di giorni almeno affinché la commissione Lavoro raccolga i vari pareri.
[interpretazione durativa o risultativa]

1.5 Un approccio computazionale

L'idea di un approccio computazionale alla classificazione azionale viene suggerita da un articolo di Siegel e McKeown [14]. Il loro approccio non si basa sulle categorie vendleriane, ma propone due distinzioni di carattere più generale: la prima (tra *events* e *states*) riguarda il tratto [+/- dinamico], la seconda (tra *culminated events* e *nonculminated events*) riguarda il tratto [+/- omogeneo]. Gli autori propongono l'uso di diversi sistemi di apprendimento automatico (alberi decisionali, *Logistic Regression*, *Genetic Programming*): ognuno di questi sistemi sfrutta i cosiddetti "indicatori linguistici" (elementi contestuali) e gli effetti di co-occorrenza tra tali indicatori e il verbo da analizzare per determinare la forma verbale di ogni verbo. Siegel e McKeown, tuttavia, non tengono conto dell'ibridismo azionale: parlano di una *fundamental aspectual class*⁹, ovvero di un significato azionale prototipico che ogni verbo avrebbe *a priori*, e di possibili mutamenti (*aspectual coercion*) che possono intervenire a modificarlo.

Si veda al punto (1.20):

(1.20) a. He **hiccupped** [-durativo]

b. He was **hiccupping** [+durativo]

Gli autori presuppongono che un verbo (come l'inglese *hiccup*) abbia una sua classe aspettuale fondamentale (nel nostro caso, [-durativo]), e notano come l'uso del progressivo operi una commutazione azionale (la lettura abituale, [+durativa], di

⁹secondo una convenzione diffusa che denomina l'Aspetto "Aspetto soggettivo" e l'Azionalità "Aspetto oggettivo"

hiccup).¹⁰ Tuttavia non considerano il caso di verbi come *considerare* o *costituire* che, come si può vedere dalla tabella 6.2, presentano un’alternanza lessicalmente rilevante tra due o più significati, senza che emerga una “classe azionale fondamentale”, come si vede al punto (1.21):

- (1.21) a. L’Associazione nazionale editoria elettronica, che **abbiamo** di recente **costituito** in Italia, parla di una vera e propria guerra tra produttori internazionali [–durativo]
- b. E proprio l’assenza di regole **costituisce** il fascino che accompagna taluni titoli bellissimi [+durativo]

Casi del genere rispecchiano una reale polisemia intrinseca al verbo stesso, non riconducibile a una semplice (e occasionale) commutazione azionale: in questo caso è più corretto parlare di ibridismo azionale piuttosto che di commutazione.

Il lavoro di Siegel e McKeown invece mette sullo stesso piano la commutazione azionale e l’ibridismo vero e proprio: un simile approccio può dare buoni risultati per l’obiettivo che i due autori si propongono, ovvero la determinazione della “classe azionale fondamentale” di ogni verbo, soprattutto con verbi fortemente prototipici (per i quali potrebbe aver senso parlare di classe azionale fondamentale) e su distinzioni categoriali ampie come quella tra *states* ed *events*; tuttavia lo stesso approccio non riesce a modellizzare adeguatamente l’ibridismo azionale.

Il lavoro esposto in questa relazione si propone, al contrario, di dare rilevanza a un fenomeno come l’ibridismo, e di raggiungere un obiettivo ben diverso: determinare la categoria azionale non di un lemma, ma di una singola occorrenza nel suo contesto, e non solo per verbi altamente prototipici ma, come vedremo, anche con lemmi caratterizzati da una forte polisemia.

¹⁰ Altri elementi contestuali, come la struttura argomentale, possono operare commutazioni azionali

Capitolo 2

L'annotazione e il modello computazionale

2.1 L'annotazione del corpus e il Gold Standard

Il corpus TreSSI¹ comprende frasi dell'italiano scritto contemporaneo per circa 300.000 parole ed è annotato ai livelli ortografico, morfo-sintattico, sintattico (a costituenti e funzionale) e semantico-lessicale. È strutturato in un database relazionale ed è suddiviso in due partizioni: un corpus “bilanciato”, costituito da una selezione di diversi tipi di testi italiani e un corpus specializzato, costituito da testi appartenenti al dominio economico-finanziario.

Dal corpus sono stati selezionati 33 lemmi particolarmente frequenti, per un totale di 3429 occorrenze, con i relativi contesti (i lemmi sono elencati in appendice alla tabella 6.2). Per ogni frase del corpus, ogni occorrenza dei 33 verbi scelti è stata annotata secondo le quattro categorie di *state*, *activity*, *accomplishment*, *achievement* e *incrementativo*². Ogni classe azionale è descritta secondo la presenza o assenza di quattro tratti: [+/-omogeneo], [+/-durativo], [+/-dinamico], [+/-incrementativo], come in tabella 1.1. Si veda la tabella 6.1 in appendice per un esempio del corpus di partenza, in cui ogni token è abbinato al proprio lemma e alla corrispondente analisi morfologica, con annotazione. Il corpus così annotato è stato considerato il nostro Gold Standard. Nella tabella 6.2 in appendice sono indicati i 33 verbi selezionati, con il numero totale di occorrenze, e, per ogni verbo, quante volte ricorre come *state*, *activity*, *accomplishment*, *achievement*, *incrementativo*. Dal corpus così ottenuto sono stati inizialmente messi da parte cinque verbi (*cambiare*, *crescere*, *ridurre*, *salire*, *scendere* - per un totale di 300 occorrenze), che avevano talvolta l'interpretazione *incrementativo*: questo gruppo di verbi sarà utilizzato come test set in una

¹Trebank Sintattico Semantica dell'Italiano del progetto SI-TAL (progetto finanziato dal MUR-ST nell'ambito della legge 46/82 art.10 e realizzato sotto la direzione scientifica del Prof. Zampolli) - Montemagni et al. 2003, [10]

²I dati sono stati organizzati in un database MySQL e l'estrazione delle frasi è stata gestita con script in php.

fase successiva del lavoro. Le occorrenze rimanenti sono quindi 3129, suddivise in 28 verbi. La loro distribuzione per le quattro categorie vendleriane è indicata alla tabella 2.1.

verbi	tot.occorrenze	ACT	STA	ACH	ACC
28	3129	430	583	1294	822

Tabella 2.1: il training

2.2 Apprendimento automatico: la Maximum Entropy

L'approccio adottato presuppone un agente in grado di apprendere, ovvero di costruirsi delle generalizzazioni a partire da contesti noti (fase di *training*, o di addestramento) e di applicarle su contesti nuovi (fase di *test*). Un sistema di apprendimento sarà definito da un particolare *task* e dalla tecnica di apprendimento utilizzata. Nel nostro caso, il *task*, dato un determinato contesto, sarà l'assegnazione della categoria azionale corretta (ovvero la più probabile in quel contesto), e il modello sarà quello della cosiddetta *Maximum Entropy* (massimizzazione dell'entropia, Berger *et al.* 1996 [1]). Si tratta di un approccio di tipo supervisionato: durante la fase di addestramento i contesti sono corredati dalla risposta corretta (*training set*).

Secondo il modello della *Maximum Entropy*, dato un contesto linguistico c , e una risposta $r \in R$ dipendente da c , la probabilità condizionata $p(r | c)$ viene calcolata utilizzando come unici vincoli probabilistiche soltanto le distribuzioni di un insieme di marche rilevanti $m_i(r, c)$ di c (tali distribuzioni vengono estratte dai dati forniti in fase di *training*). È stato dimostrato che la probabilità p che soddisfa tale definizione è unica ed è la seguente:

(2.1)

$$p(r | c) = \frac{1}{Z_c} \prod_{i=1}^k r_i^{m_i(r,c)}$$

Dove Z_c sarà un fattore di normalizzazione, e $m_i(r, c)$ i parametri di ogni marca m per (r, c) (i pesi assegnati alle marche: ovvero quanto la marca m del contesto c è rilevante per la risposta r).

Ad esempio, se un tipo di marca è particolarmente frequente sui contesti di tipo *ACT* il peso di quella marca per la risposta *ACT* sarà particolarmente elevato. Si vedano gli esempi in tabella 2.2: se i pesi che il sistema assegna alla marca “esprnu-mogg” (ovvero alla presenza di un'indicazione numerica precisa nell'oggetto diretto), relativamente alle quattro categorie azionali, sono quelli riportati, vorrà dire che la presenza di questa marca è rilevante per l'individuazione della telicità, dal momento che i pesi più alti sono quelli relativi agli *ACC* e, in misura minore, agli *ACH*. Per stimare i pesi di ogni marca è stato utilizzato l'algoritmo GIS implementato nel

esprnumogg_ACT	0,155694
esprnumogg_STA	0,734207
esprnumogg_ACC	1,740197
esprnumogg_ACH	0,847124

Tabella 2.2: pesi di una marca

programma AMIS (Miyao and Tsujii 2005 [9]).

Nella fase di *test*, per ogni nuovo contesto il sistema utilizzerà i pesi calcolati in fase di *training* e li combinerà in modo da calcolare le probabilità $p(r | c)$ per ogni risposta $r \in R$. Quindi calcolerà la risposta più probabile in quel determinato contesto, ovvero g (quella che massimizzerà $p(r | c)$) e la fornirà come risposta:

(2.2)

$$g = \operatorname{argmax}(p(r | c))$$

2.2.1 La k-fold cross-validation

Il sistema di apprendimento automatico, una volta addestrato sul training set, andrà ovviamente valutato sul test set: una tecnica tipica dell'apprendimento automatico è la cosiddetta “k-fold cross-validation”. Il corpus viene diviso in k parti: la prima porzione ($\frac{1}{k}$) viene messa da parte come test set, e la parte rimanente ($\frac{k-1}{k}$) viene utilizzata come training set. In questo modo i dati su cui il sistema viene addestrato sono diversi da quelli su cui viene validato. L'operazione può essere quindi ripetuta k volte, utilizzando ogni volta una frazione diversa come test: alla fine si farà una media dei risultati ottenuti.

Ad esempio, se decidiamo di dividere il nostro corpus di 3129 frasi in 10 parti ($k = 10$), ne estrarremo un decimo, addestreremo il sistema sui $\frac{9}{10}$ rimanenti, e useremo il decimo messo da parte come test set. Quindi estrarremo il secondo decimo, utilizzeremo come training set i $\frac{9}{10}$ rimanenti, e così via, per dieci volte, ovvero fino a quando ognuna delle 10 parti in cui avevamo suddiviso il corpus sarà, a turno, utilizzata come test corpus.

Per la valutazione di questo modello di apprendimento automatico è stata usata la 30-fold cross validation: i dati così ottenuti permettono di valutare la capacità predittiva del nostro sistema su dati non visti, e allo stesso tempo il fatto di poter ripetere il test più volte (utilizzando sempre test set diversi) permette di ovviare alla relativa scarsità dei dati.

2.2.2 Precisione, accuratezza, f-measure

Nella presente relazione (in particolare, per le tabelle più dettagliate presenti in appendice) i risultati sono generalmente presentati in “tabelle di confusione”, che riportano in orizzontale le “risposte corrette”, in verticale le risposte fornite dal sistema, sulla diagonale le risposte che il sistema riesce a “indovinare”. Inoltre per

ogni tabella di risultati vengono calcolati due parametri importanti per la valutazione del sistema; si tratta di due termini tipici dell'Information Retrieval: la precisione (*precision*) e l'accuratezza (*recall*). Per ogni evento il sistema fornisce una categoria tra quattro categorie possibili, quindi, per ogni categoria, calcoleremo la precisione e l'accuratezza.

Definiamo ad esempio l'accuratezza per la categoria STA (A_{STA}) come la percentuale di STA corretti trovati dal sistema rispetto a tutti gli STA presenti nel test corpus e la precisione per la categoria STA (P_{STA}) come la percentuale di STA corretti trovati dal sistema su tutti gli STA individuati dal sistema. Precisione e accuratezza saranno indicate con valori compresi tra 0 e 1. Calcoleremo anche la cosiddetta *f-measure*, ovvero la media armonica tra precisione e accuratezza, per vedere quanto il sistema è bilanciato tra i due parametri. (A_{STA})

(2.3)

$$\frac{2(A_{STA} \cdot P_{STA})}{A_{STA} + P_{STA}}$$

Un sistema perfetto avrà precisione e accuratezza uguali a 1, e quindi anche *f-measure* uguale a 1.

2.3 Classi azionali fondamentali?

Nella tabella 6.2 in appendice per ogni lemma analizzato è indicata la frequenza relativa dell'interpretazione azionale più frequente: chiameremo questa frequenza "grado di prototipicità": si può notare come il corpus di riferimento comprenda sia verbi fortemente prototipici (*lavorare, vincere*), sia verbi ibridi (*chiamare, capire*), che presentano un'alternanza lessicalmente rilevante tra due o tre categorie azionali. Proviamo invece, con Siegel e McKeown [14], a dare legittimità al concetto di "classe azionale fondamentale" (*fundamental aspectual class*). Abbiamo simulato un sistema che, data un'occorrenza di uno dei 28 verbi in esame, assegna semplicemente la "classe azionale fondamentale" del lemma, ovvero la più frequente.

tutto il corpus					
	ACT	STA	ACC	ACH	
ACT	254	23	101	52	⇒ tot. act: 430
STA	9	295	111	168	⇒ tot. sta: 583
ACC	2	3	785	32	⇒ tot. acc: 822
ACH	0	51	81	1162	⇒ tot. ach: 1294
precisione:	0,96	0,79	0,73	0,82	occorrenze: 3129
accuratezza:	0,59	0,51	0,95	0,9	errori: 20,2%
f-measure:	0,73	0,62	0,83	0,86	corrette: 79,8%

Tabella 2.3: il sistema assegna il tratto più frequente per ogni lemma: tutto il corpus

Se il test viene condotto su tutto il corpus, la percentuale di errore è comunque abbastanza alta (come si può vedere in tabella 2.3).³

Per un verbo fortemente prototipico come *vincere* il sistema assegnerà la categoria più frequente (*ACH*) e non sbaglierà mai. Un test come questo risulterà invece interessante solo se condotto sui verbi più polisemici (banco di prova decisivo per un modello automatico di classificazione azionale), come ad esempio *capire*:

(2.4) Forse questa volta l'hanno capito. (*ACH*) \iff Dice di **capire** le emozioni che il fatto può suscitare (*STA*)

Il verbo *capire* presenta 81 occorrenze, di cui 34 (ovvero il 42%) come *STA* nel senso di “riuscire a comprendere”, e 47 (ovvero il 58%) come *ACH* nel senso di “arrivare a capire”. Assegnando sempre la categoria più frequente (*ACH*), il sistema fornirà la risposta corretta per ogni occorrenza di tipo *ACH* ma sbaglierà per ogni occorrenza di tipo *STA*, ovvero il 42% delle volte. Abbiamo selezionato quattro gruppi (tabella 6.7 in appendice) di verbi polisemici: il primo gruppo comprende verbi la cui interpretazione più frequente ricorre meno del 60% delle volte, il secondo comprende invece verbi la cui interpretazione più frequente ricorre meno del 70% delle volte (quindi gli stessi verbi del primo gruppo con l'aggiunta di *conoscere* - 62% - *prendere* - 69% - *presentare* - 68%), mentre con il terzo la soglia viene alzata all'80%, e con il quarto al 90%. Ogni test che preveda la decisione sulle quattro categorie vendleriane sarà condotto anche su questi quattro gruppi.

Proviamo quindi a ripetere il test dell'assegnazione della categoria più probabile sui verbi più polisemici: la percentuale di errore sale al 44% per il primo gruppo, al 40% per il secondo, al 35% per il terzo e al 30% per il quarto.

Ne concludiamo che l'assegnazione di una categoria prototipica può quindi risultare fuorviante per la modellizzazione dell'azionalità. Utilizzeremo comunque questi valori come riferimento, per calcolare una baseline di precisione e per verificare quanto il sistema riesca a migliorare con l'introduzione di tratti contestuali (tabella 2.4).

Gruppi	Errori	Baseline
Primo gruppo	43,9%	56,1%
Secondo gruppo	40%	60%
Terzo gruppo	35,4%	64,6%
Quarto gruppo	30,4%	69,6%
Tutto il corpus	20,2%	79,8%

Tabella 2.4: valori di riferimento per la precisione del sistema per la decisione su classi

³le tabelle pubblicate sono il risultato della 30-fold cross validation

2.4 Il problema dell'Inter-Tagger Agreement

Come già descritto, annotazioni azionali condotte da annotatori umani diversi potrebbero differire tra loro: il problema dell'Inter Tagger Agreement (ITA) è tipico di ogni approccio automatico alla semantica. Forse per l'annotazione azionale possiamo aspettarci che il problema dell'accordo tra annotatori sia ancora più complesso, dal momento che le stesse teorie dell'azionalità sono spesso in disaccordo tra loro. Nel paragrafo precedente abbiamo definito una baseline per il modello, ma sembra opportuno chiedersi che performance possiamo aspettarci dal modello su un compito semantico complesso come quello dell'annotazione azionale. L'accordo che degli annotatori umani riescono a raggiungere può quindi diventare un valore di riferimento, se non proprio un *upper bound*, per valutare quanto siano significativi i livelli di accuratezza raggiunti dal sistema. Dal secondo gruppo di verbi ibridi, abbiamo quindi selezionato 100 frasi e le abbiamo sottoposte ad altri tre annotatori umani, con il compito di annotare ogni frase secondo una delle quattro categorie vendleriane. Per valutare i risultati ottenuti, è stato utilizzato il calcolo della statistica k [17], che, data una coppia di annotatori A e B , combina l'accordo osservato con il cosiddetto "accordo atteso". L'accordo osservato (p_o) è definito come la percentuale di risposte su cui gli annotatori sono d'accordo, mentre l'accordo atteso (p_a) è invece calcolato come la probabilità che su nuove frasi gli annotatori continuino a essere d'accordo e tiene conto del fatto che gli annotatori potrebbero essere d'accordo anche per puro caso (si veda 2.5)

(2.5)

$$p_o = \frac{\text{(risposte concordi)}}{\text{(tot. risposte)}}$$

$$p_a = \frac{(ACT_A)}{\text{(tot. risp.)}} \cdot \frac{(ACT_B)}{\text{(tot. risp.)}} + \frac{(STA_A)}{\text{(tot. risp.)}} \cdot \frac{(STA_B)}{\text{(tot. risp.)}} +$$

$$+ \frac{(ACC_A)}{\text{(tot. risp.)}} \cdot \frac{(ACC_B)}{\text{(tot. risp.)}} + \frac{(ACH_A)}{\text{(tot. risp.)}} \cdot \frac{(ACH_B)}{\text{(tot. risp.)}}$$

La k -statistic tiene conto della differenza tra l'accordo osservato e l'accordo atteso, quindi fornisce una misura più soddisfacente di quanto i diversi annotatori siano d'accordo.

(2.6)

$$k = \frac{p_o - p_a}{1 - p_a}$$

I valori di k -statistic per tutte le possibili coppie di annotatori, compreso il gold standard e i risultati forniti dal sistema, sono riportati in tabella 2.5:

	gold standard	annot.A	annot.B	annot.C
gold standard	1	0,61	0,27	0,51
annot.A	0,61	1	0,29	0,53
annot.B	0,27	0,29	1	0,26
annot.C	0,51	0,53	0,26	1

Tabella 2.5: valori di k-statistic per ogni coppia di annotatori

2.5 Una baseline per i tratti

Il sistema sarà anche sottoposto a un altro tipo di decisione, non su quattro categorie, ma su un tratto binario ([+/-omogeneo], [+/-durativo], [+/-dinamico]). Trattandosi di un compito di tipo diverso, abbiamo bisogno di una diversa baseline, e saranno diversi anche i gruppi di ibridi: il verbo *chiamare* ad esempio, non è polisemico per il tratto [+/-durativo], dal momento che presenta un'alternanza tra le interpretazioni di *accomplishment* e di *state*, entrambe [+durative]. Elenchiamo alle tabelle 6.11, 6.12, 6.13 in appendice i verbi relativi ai quattro gruppi per ognuno dei tratti (come per la decisione su classi, il secondo gruppo comprende anche i verbi del primo, e il terzo i verbi del secondo); i livelli di soglia sono stati leggermente aumentati, poichè la decisione su due è più facile della decisione su quattro (i risultati in dettaglio in 6.14, 6.15, 6.16).

Riassumiamo brevemente la baseline per la decisione su tratti:

Gruppi	DUR	DIN	OMG
Primo gruppo	60,9%	63,9%	56,9%
Secondo gruppo	62,2%	68,3%	66,8%
Terzo gruppo	70%	75,5%	71,9%
Tutto il corpus	87,7%	88,3%	84,4%

Tabella 2.6: valori di riferimento per la precisione del sistema sulla decisione per tratti

Capitolo 3

Marche contestuali

Completata l’annotazione del corpus, è stato selezionato un insieme di marche rilevanti per modellare la caratterizzazione azionale di ogni occorrenza nel suo contesto. Le marche, prese singolarmente, non sono sempre altamente predittive, e d’altra parte marche fortemente predittive (come ad esempio gli avverbi “deliberativi”) sono molto rare. Il modello della Maximum Entropy (in questo, analogo con i sistemi utilizzati da Siegel e McKeown [14]) riesce a ovviare al problema della sparsità dei dati combinando tra loro i “pesi” delle marche. Una marca quindi non è un vincolo forte, ma tutte le marche presenti nel contesto contribuiscono al calcolo della categoria azionale di appartenenza di ogni occorrenza.

In appendice (tabella 6.3) le 14 marche utilizzate da Siegel e McKeown [14]: alcune di esse (gruppo α) sono state scelte in base a considerazioni linguistiche, e altre (gruppo β) sono state individuate nel corso della ricerca come statisticamente rilevanti. È da notare tuttavia l’assenza di molte marche che in un approccio totalmente automatico quale quello di Siegel e McKeown (in cui le marche non sono annotate a mano) si sarebbero potute comunque estrarre senza difficoltà: ad esempio, *already*, *not..yet*, avverbi di frequenza (*once*, *twice*, *three times*), la presenza di espressioni numeriche nell’oggetto diretto.

Questi ed altri tipi di elementi contestuali dalla ben nota rilevanza linguistica sulla categorizzazione azionale ([2], [3], [8]) sono stati invece utilizzati nel presente lavoro, insieme ad alcune marche già utilizzate da Siegel e McKeown (altre invece, come avverbi “valutativi”, avverbi “di modo”, avverbi “continui”, non sono state considerate rilevanti).

Ne risulta un insieme sostanzialmente più numeroso, così ripartito:

- tratti morfologici tempo-aspettuali e diatesi della testa verbale;
- presenza di argomenti e di complementi, retti dalla testa verbale e loro tratti morfologici, sintattici e semantici;
- presenza di modificatori della testa verbale (di varie tipologie).

L'annotazione delle marche è stata in parte estratta semi-automaticamente tramite apposite query sul database del corpus e in parte condotta a mano durante la fase di annotazione delle categorie azionali (tabelle 6.4, 6.5 e 6.6 in appendice).

3.1 Tratti morfologici tempo-aspettuali e diatesi della testa verbale

Categorie tempo-aspettuali - La categoria tempo-aspettuale è spesso rilevante per la scelta dell'interpretazione azionale corretta: il perfetto è considerato come preferenziale per gli eventi ([−omogeneo], [−stativo]), il progressivo è incompatibile con letture di tipo [−durativo] e [+stativo]. In alcuni casi di alternanze infralessicali, ad esempio, si può notare come un'interpretazione culminativa o ingressiva (compatibile con l'aspetto imperfettivo) si opponga a un'interpretazione durativa o risultativa (compatibile con l'aspetto perfettivo) (Lucchesi 1971 [8]), come già sottolineato al paragrafo 4.1. Gli “effetti di compatibilità” mostrati dagli esempi di Lucchesi dimostrano una correlazione non coercitiva, ma agevolmente sfruttabile da un sistema come quello proposto, in cui le marche non sono vincoli forti, ma indicatori probabilistici “deboli” da combinare tra loro. Le marche annotate sono le seguenti: **presente** (comprende i presenti dei modi indicativo, congiuntivo e condizionale), **imperfetto** (comprende gli imperfetti dei modi indicativo e congiuntivo), **passato remoto** (comprende i passati remoti del modo indicativo), **perfetto** (comprende i passati prossimi e i trapassati del congiuntivo e dell'indicativo, i futuri anteriori dell'indicativo, i condizionali passati, gli infiniti passati, i participi passati), **futuro**, **perifrasi progressiva**.

Diatesi - Per quanto riguarda la diatesi invece, gli attivi sono considerati il caso non-marcato, e sono indicati i passivi e la presenza del “si”.

3.2 Argomenti e complementi

Assenza dell'oggetto diretto - L'assenza dell'oggetto diretto può spesso aiutare a distinguere due significati diversi di un unico lemma polisemico.

- (3.1) a. Sono quindici anni che **lavoro** nel Dubai con la speranza di mettere via una somma sufficiente che mi permetta di tornare a casa.
- b. Le ore di credito sono correlate ai periodi trascorsi a terra corrispondenti ad **ore lavorate** per ragioni di addestramento.

Nell'esempio si può notare la distinzione tra due diversi significati di *lavorare* a valenze diverse: il primo uso, intransitivo, è un tipico esempio di *ACT*, il secondo è un uso un po' più colloquiale di *lavorare* come transitivo (*ACC*).

Marche di definitezza - L'uso di espressioni numeriche nell'oggetto diretto e nell'oggetto indiretto sono marche di telicità ancora più marcate:

(3.2) In tasca i militari gli **hanno trovato 950 mila lire**.

Nel caso in cui essi non siano preceduti da articoli o da espressioni numeriche, fanno propendere per un'interpretazione [+durativa] e [+omogenea]:

(3.3) Il nuovo consorzio **venderà** sistemi integrati.

La definitezza degli argomenti è invece una buona spia di telicità.

Animatezza - Il contrasto tra frasi con soggetti e oggetti animati può far emergere altre "opposizioni infralessicali" [8] (cfr. paragrafo 1.4).

3.3 Modificatori della testa verbale

Negazione - La negazione può rendere un verbo [+durativo], qualora indichi il persistere di una situazione in cui non accade nessun evento. Si veda l'esempio (3.4):

(3.4) Sono anni che qui **non** cambia nulla.

Modificatori avverbiali - L'avverbiale "*in x tempo*" è un tipico rivelatore di telicità, anche come test sintattico per individuare la categoria azionale di una occorrenza, mentre espressioni come "*per x tempo*", "*fino a*" (temporale), i decorrenziali, oltre a essere rivelatori di atelicità e di duratività, possono operare commutazioni azionali su predicati annullandone la telicità, come si vede al punto (3.5):

(3.5) a. Lidia **ha scritto** la tesi *per due ore* NON IMPLICA b. Lidia ha finito di scrivere la tesi.

Si noti inoltre che i decorrenziali possono essere di due tipi:

(3.6) a. Marco **sta lavorando** *da quattro ore* \iff b. Marco **è arrivato** *da quattro ore*.

Nella prima frase l'espressione *da quattro ore* indica che il processo si sta svolgendo *da quattro ore* (lettura più compatibile con l'Aspetto imperfettivo), nella seconda invece *da quattro ore* indica che l'evento telico si è concluso *da quattro ore* (lettura più compatibile con l'Aspetto perfettivo). Un'espressione del tipo "da x tempo" è stata annotata in modo diverso nei due casi, a seconda che fosse dipendente da un tempo perfettivo o da un tempo imperfettivo. Avverbi "deliberativi" ("*volentieri*", "*deliberatamente*") sono incompatibili con gli stativi, mentre avverbi di frequenza ("*spesso*", "*talvolta*") e avverbi iterativi ("*x volte*") introducono accezioni abituali e iterative compatibili con la categoria *ACT* (anche in commutazione azionale).

3.4 Lessicalizzare alcune marche contestuali

Un unico lemma può avere diverse strutture argomentali, che possono esprimere significati differenti e valori azionali differenti.

- (3.7) a. È necessario constatare gli eventuali tentativi delatori dei pentiti **trovando** riscontri (*ACT*)
- b. Leonardo Sciascia **si trova** a Milano da qualche settimana (*STA*)
- c. **Trovo** che la sua interpretazione sia plausibile (*STA*)
- d. Lavoravano per un mondo giusto, hanno **trovato** un destino infame (*ACH*)

Durante il calcolo della baseline comprendevamo in una marca relativa al lemma le tendenze semantiche di un verbo. Abbiamo quindi provato a fare qualcosa di più preciso, e a lessicalizzare alcune marche relative alla struttura argomentale. Così dunque non avremo più la marca “trovare” ma le marche “trovareOgg” (trovare + oggetto diretto) e “trovareCompl” (trovare + completiva), distinte come per rendere conto di significati distinti. Il sistema, calcolando il peso delle marche lessicalizzate (indicate in tabella 3.1), sfrutterà le correlazioni specifiche tra una delle strutture argomentali di un verbo e le categorie azionali di appartenenza. In appendice (tabella 6.6) i pesi delle marche lessicalizzate per *scrivere*.

lemma +	marca contestuale	⇒ nuova marca
	oggetto diretto	⇒ lemmaOgg
	oggetto indiretto	⇒ lemmaOggI
	moto a luogo	⇒ lemmaMal
	completiva	⇒ lemmaComple
	passivo	⇒ lemmaPass
	no arg.	⇒ lemmaNoArg
	si	⇒ lemmaSi
	sogg. animato	⇒ lemmaSoggAnim
	sogg. inanimato	⇒ lemmaSoggInanim
	ogg. animato	⇒ lemmaOggAnim
	ogg. inanimato	⇒ lemmaOggInanim

Tabella 3.1: le marche lessicalizzate

Capitolo 4

Primi esperimenti computazionali

4.1 La decisione su classi

Il sistema è stato sottoposto a una serie di cicli di valutazione e revisione, con una valutazione degli errori più frequenti e una continua consultazione della letteratura sulla classificazione azionale. Sono stati criteri di questo tipo che hanno fatto escludere marche come la presenza del soggetto, o che hanno portato a lessicalizzare alcune marche, come l'animatezza del soggetto. Questo processo ha quindi contribuito alla definizione graduale dell'insieme di marche contestuali considerate adatte per migliorare la performance del sistema (si tratta delle marche elencate nelle tabelle 6.4, 6.5, 6.6 in appendice).

Applicando questo insieme di marche, si ottengono dei risultati che, confrontati con la baseline, mostrano un decisivo miglioramento. Si vedano le tabelle 6.9 e 6.10 in appendice per i risultati, e 4.1 per il confronto con la baseline, che si rivela particolarmente interessante sui gruppi di ibridi: il sistema si rivela quindi capace di modellare, almeno in parte, l'ibridismo azionale.

Gruppi	Baseline	Conf. migliore
Primo gruppo	56,1%	69,34%
Secondo gruppo	60%	72,37%
Terzo gruppo	64,6%	75,47%
Quarto gruppo	69,6%	78,43%
Tutto il corpus	79,8%	85,39%

Tabella 4.1: confronto con la baseline per la decisione su classi

4.1.1 Confronto con altri annotatori

Proviamo ad addestrare il modello così ottenuto su tutto il corpus, escluse le 100 frasi che avevamo utilizzato per verificare l'Inter Tagger Agreement, e poi a testarlo su quelle frasi. I risultati ottenuti sono in tabella 4.2, espressi sia in termini di *k-statistic* sia in termini di percentuale di accordo (che avevamo chiamato p_o): possiamo osservare come l'accuratezza raggiunta del sistema su verbi "mediamente ibridi" sia in linea con l'accordo tra annotatori umani.

	annot.A	annot.B	annot.C	sistema
k-statistic	0,61	0,27	0,51	0,65
percentuale di accordo (p_o)	73%	44%	67%	76%

Tabella 4.2: confronto di ogni annotatore e del modello con il gold standard

4.1.2 Separazione delle marche

Il modello utilizza un insieme di marche abbastanza eterogeneo: come abbiamo visto, le marche utilizzate sono di tipo avverbiale, morfologico e sintattico-argomentale. Proviamo (tabella 4.3) a usarle distintamente in tre cicli separati: si noterà che le marche avverbiali sono sicuramente utili per la distinzione tra categorie azionali, soprattutto sui tratti di omogeneità e di dinamicità, meno sul tratto di duratività (si veda la distinzione tra ACC e ACH), ma spesso (colonna NC) il modello, basandosi solo su di esse, non è in grado di rispondere, per la frequente mancanza di queste marche. Tali marche sono infatti tanto predittive quanto rare. Le marche morfologiche riducono il numero delle occorrenze non classificate, ma rimangono problematiche per la duratività, tratto per cui le marche sintattico-argomentali (e ancora di più la combinazione di tutte le marche individuate, tabella 6.10) riescono a essere decisive.

4.1.3 Errori comuni del sistema

Da un'analisi più dettagliata degli errori emerge che il modello ha particolari difficoltà con i verbi *mettere*, *prendere*, *portare*. Si tratta di verbi spesso soggetti a uso idiomatico: espressioni come *mettere piede*, *prendere parte*, *portare sfortuna* sono molto frequenti, e benché gli argomenti di un verbo idiomatico non siano stati marcati (la marca sulla presenza dell'oggetto non è stata considerata in un contesto idiomatico come *mettere piede*), il modello continua ad avere problemi con usi così eterogenei di un singolo verbo.

- (4.1) Umberto Eco ha potuto divertirsi a prendere un po' tutti per il bavero, sia gli autori gialli sia i lettori di quelli, ma non solo i giallomani.
- (4.2) Questa è una fetta di Croazia dove fino a ieri sera noi non potevamo neppure mettere piede.

marche avverbiali						
	ACT	STA	ACC	ACH	NC	
ACT	22	37	0	14	357	⇒ tot. act: 430
STA	8	58	1	11	505	⇒ tot. sta: 583
ACC	8	26	2	70	716	⇒ tot. acc: 822
ACH	7	47	4	185	1051	⇒ tot. ach: 1294
precisione:	0,49	0,35	0,29	0,66		occorrenze: 610
accuratezza:	0,05	0,1	0	0,14		errori: 91,5%
f-measure:	0,09	0,15	0	0,24		corrette: 8,5%
marche morfologiche						
	ACT	STA	ACC	ACH	NC	
ACT	34	165	9	58	164	⇒ tot. act: 430
STA	30	363	0	117	73	⇒ tot. acc: 583
ACC	18	213	2	380	209	⇒ tot. acc: 822
ACH	13	223	2	635	421	⇒ tot. ach: 1294
precisione:	0,36	0,38	0,15	0,53		occorrenze: 3129
accuratezza:	0,08	0,62	0	0,49		errori: 66,9%
f-measure:	0,13	0,47	0	0,51		corrette: 33,1%
marche sintattico-argomentali						
	ACT	STA	ACC	ACH	NC	
ACT	282	17	89	42	0	⇒ tot. act: 430
STA	13	410	56	104	0	⇒ tot. acc: 583
ACC	13	14	755	40	0	⇒ tot. acc: 822
ACH	9	75	68	1142	0	⇒ tot. ach: 1294
precisione:	0,89	0,79	0,78	0,86		occorrenze: 3129
accuratezza:	0,66	0,7	0,92	0,88		errori: 17,3%
f-measure:	0,76	0,75	0,84	0,87		corrette: 82,7%

Tabella 4.3: training su gruppi separati di marche

Un altro problema è stato notato quando il verbo la cui categoria azionale ci interessa è all'infinito, o per forme participiali e gerundiali: in particolare gli infiniti dipendenti da causativi o da modali perdono l'informazione morfologica del verbo reggente.

(4.3) Basta chiedere a Poulidor, infaustamente chiamato Poupou, che segue il Tour con un'aria da clown.

(4.4) Tutto ciò lascia capire come il gruppo Lazard abbia già mangiato la foglia.

Una possibile direzione di miglioramento potrebbe essere il recupero e l'integrazione di questa informazione morfologica, che, come abbiamo visto, può rivelarsi fondamentale per la disambiguazione del valore azionale di una forma verbale.

4.2 La decisione su tratti

Analizziamo ora brevemente anche l'esperimento condotto per la decisione su tratti. Come abbiamo già visto, anche in questo caso abbiamo calcolato una baseline (tabella 2.6) e abbiamo definito tre gruppi di ibridi (tabelle 6.11, 6.12, 6.13) su cui valutare il sistema.

Gruppi	Baseline	Conf. migliore
DUR		
Primo gruppo	60,9%	72,8%
Secondo gruppo	62,2%	74,3%
Terzo gruppo	70%	79,1%
Tutto il corpus	87,7%	90,6%
DIN		
Primo gruppo	63,9%	79,9%
Secondo gruppo	68,3%	84,9%
Terzo gruppo	75,5%	85,4%
Tutto il corpus	88,3%	92%
OMG		
Primo gruppo	56,9%	79,5%
Secondo gruppo	66,8%	81,7%
Terzo gruppo	71,9%	83,2%
Tutto il corpus	84,4%	89,9%

Tabella 4.4: confronto con la baseline sulla decisione per tratti

La tabella 4.4 riassume i risultati ottenuti (in dettaglio in appendice, tabelle 6.17, 6.18, 6.19). La baseline è in questo caso notevolmente più alta (si tratta di una decisione su due categorie e non su quattro), ma notiamo che il modello ha sempre particolari problemi con il tratto [+/-durativo]. È interessante notare anche che anche gli altri annotatori umani avevano notato una particolare difficoltà

con tale tratto, e segnatamente nella scelta tra le due categorie di *accomplishment* e *achievement*.

4.3 Correlazioni tra marche e classi azionali

L'interesse di uno studio del genere non è soltanto nel verificare quanto un sistema di apprendimento automatico riesca a “far bene” per un compito semantico difficile come la classificazione azionale. Un modo per analizzare diversi risvolti del progetto può essere l'analisi dei pesi che il modello assegna alle marche contestuali per ogni classe azionale: i grafici 4.1 e 4.2 mostrano le marche, divise per le quattro categorie vendleriane e ordinate per peso. Più una marca ha peso, più favorirà quella categoria: le marche in neretto sono quelle che favoriscono la corrispettiva classe azionale, le marche in grigio chiaro hanno peso 1 (e quindi non influiscono il computo della categoria azionale), le marche in bianco sono marche inibitorie di quella classe azionale.

Ad esempio, tra le marche che favoriscono la categoria di *activity* si notino gli avverbiali iterativi e di frequenza, il modificatore ‘*per x tempo*’, l'oggetto plurale senza articolo, mentre la presenza di espressioni numeriche nell'oggetto contribuisce a inibire la stessa categoria (infatti si configura come classico indicatore di telicità). Per le categorie durative (*activity*, *state*, *accomplishment*) i temporali puntuali (‘alle 5 in punto’) sono marche inibitorie, mentre hanno un peso molto maggiore per gli *achievement*, tipicamente non durativi.

Un modello stocastico come quello proposto può quindi offrire l'occasione di indagare la natura stessa di categorie semantiche complesse come le classi azionali: l'analisi degli errori e l'individuazione degli elementi contestuali che aiutano il modello a migliorare ci confermano alcune intuizioni e alcuni assunti tradizionali della letteratura sulla classificazione azionale, e forniscono loro un supporto sperimentale.

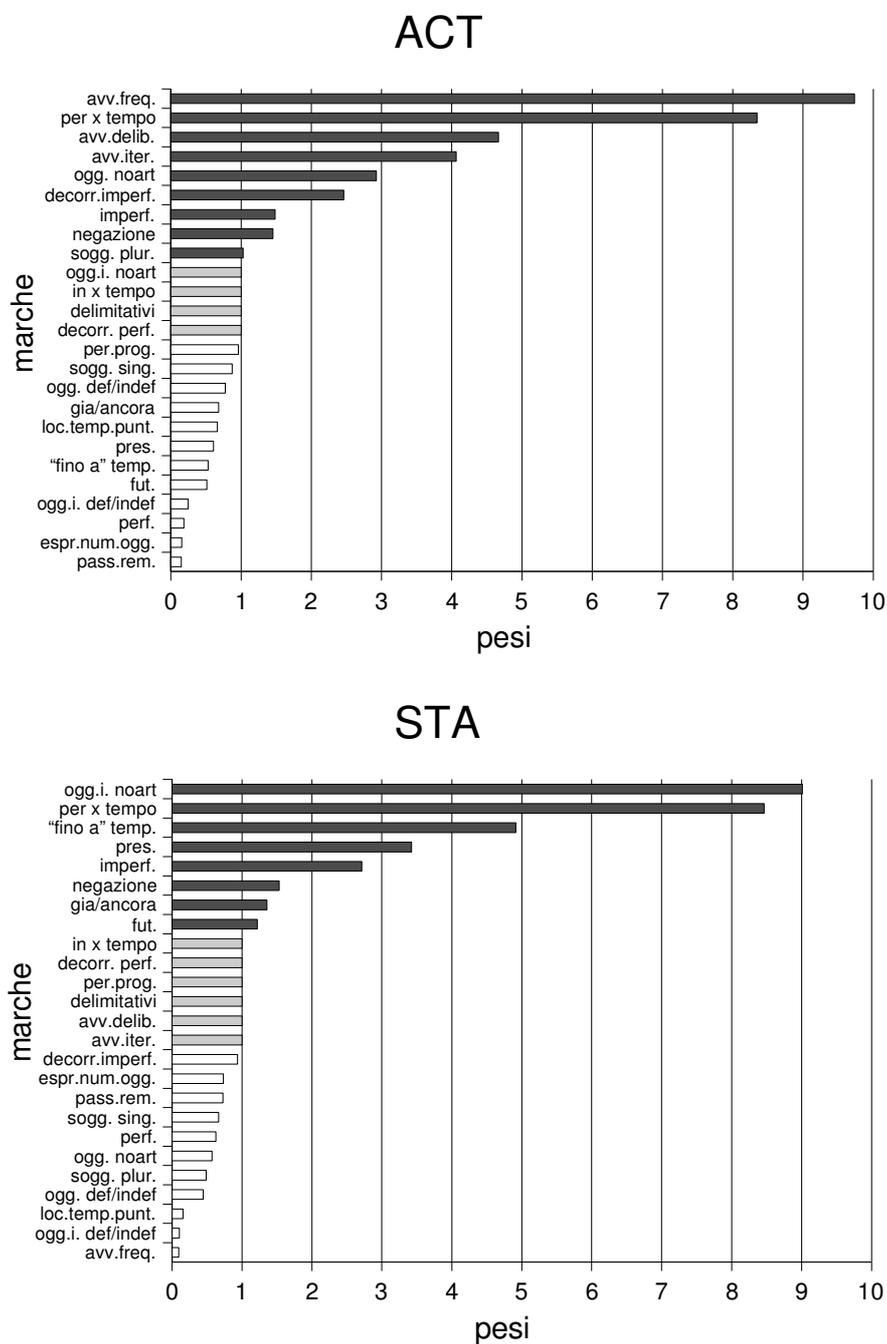


Figura 4.1: Pesì delle marche per le categorie ACT e STA

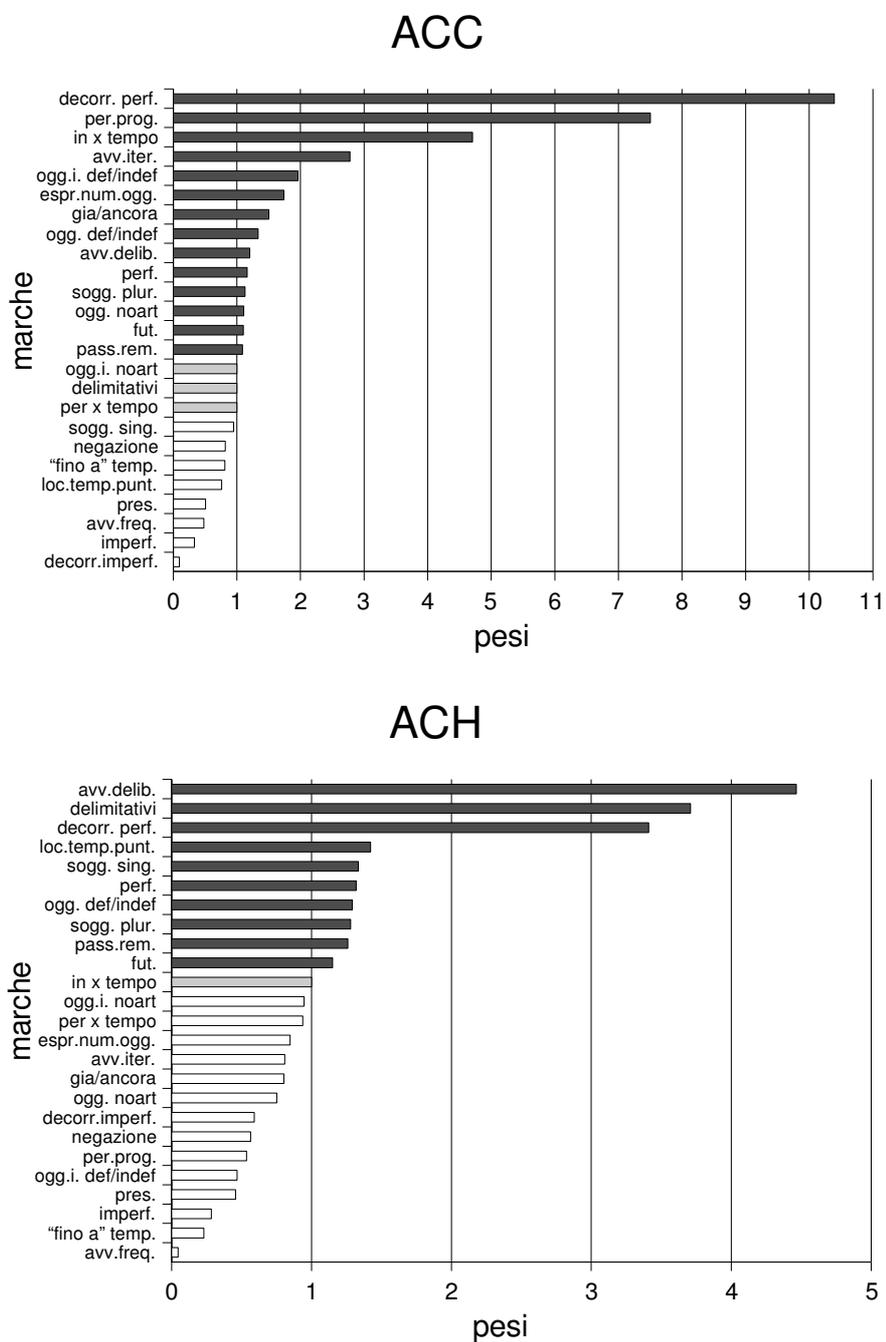


Figura 4.2: Pesì delle marche per le categorie ACC e ACH

Capitolo 5

Conclusioni

Il progetto presentato offre diverse possibilità di sviluppo, in diverse direzioni. Come ricordato, abbiamo messo da parte i cinque verbi che comprendevano la possibilità di un'interpretazione incrementativa (*cambiare, crescere, salire, ridurre, salire, scendere* - per un totale di 300 occorrenze), riservandoci la possibilità di utilizzarli in seguito: potrebbe essere interessante condurre gli stessi test descritti nella presente relazione sui lemmi esclusi, e analizzare il comportamento del sistema in rapporto agli incrementali; in questo caso, più che le “risposte giuste” ci interesserebbe vedere come il modello colloca tali verbi secondo le quattro categorie vendleriane, o come li analizza per tratti.

Il corpus annotato per lo svolgimento del progetto si presta a diverse nuove elaborazioni computazionali, quali ad esempio la realizzazione di un nuovo modello di apprendimento automatico, ma non supervisionato, o la costruzione di un sistema più “robusto”, che lavori non solo sui 28 verbi discussi in questa relazione, ma su un gruppo più ampio di lemmi, se non su qualunque frase (debitamente annotata), indipendentemente dal verbo presente.

L'esperimento dell'Inter Tagger Agreement fornisce diversi spunti per un'analisi più approfondita in ambito psicolinguistico sulle percezioni del parlante nativo riguardo alle categorie azionali del verbo.

Capitolo 6

Appendice

Un esempio di matrice di annotazione, con parte degli indicatori del corpus di partenza e la nuova annotazione azionale.

phrase_id	id	spec_lemma	lemma	pos	mfeats	[omg]	[dur]	[din]	[incr]
24	mw_541	Michelangelo	michelangelo	SP	NN				
24	mw_542	Lamberti	lamberti	SP	NN				
24	mw_543	,	,	PU	NULL				
24	mw_547	geometra	geometra	S	NS				
24	mw_548	lecchese	lecchese	A	NS				
24	mw_549	:	:	PU	NULL				
24	mw_550	nello	in	E	MS				
24	mw_551	Zaire	zaire	SP	NN				
24	mw_552	era	essere	V	S3II				
24	mw_553	arrivato	arrivare	V	MSPR	-	-	+	-
24	mw_554	per	per	E	NULL				
24	mw_555	la	lo	RD	FS				
24	mw_556	prima	primo	NO	FS				
24	mw_557	volta	volta	S	FS				
24	mw_558	martedi'	martedi'	S	MN				
24	mw_559	scorso	scorso	A	MS				
24	mw_560	.	.	PU	NULL				

Tabella 6.1: l'annotazione del corpus

lemma	freq.	STA	ACT	ACH	ACC	INC	prot.
arrivare	185	6	8	171	0	0	92%
cambiare	72	4	5	28	0	35	49%
capire	81	34	0	47	0	0	58%
chiamare	93	36	13	0	44	0	47%
chiedere	152	0	11	0	141	0	93%
chiudere	100	4	0	96	0	0	96%
comprendere	64	53	0	11	0	0	83%
conoscere	68	42	0	26	0	0	62%
controllare	65	58	6	0	1	0	89%
costituire	62	48	0	14	0	0	77%
crescere	50	0	0	0	0	50	100%
entrare	90	8	0	82	0	0	91%
indicare	97	37	1	58	1	0	60%
lasciare	121	14	2	104	1	0	86%
lavorare	84	0	82	0	2	0	98%
mettere	166	1	5	66	94	0	57%
morire	87	0	0	83	4	0	95%
parlare	181	9	172	0	0	0	95%
partire	78	4	0	74	0	0	95%
passare	129	7	30	0	92	0	71%
portare	174	29	18	0	127	0	73%
prendere	131	8	9	90	24	0	69%
presentare	143	28	3	15	97	0	68%
ridurre	51	0	0	0	12	39	75%
salire	67	1	1	11	2	52	78%
scendere	60	0	0	23	4	33	55%
scrivere	106	3	21	0	82	0	77%
spiegare	115	7	0	0	108	0	94%
tornare	117	1	1	113	2	0	97%
trattare	113	94	17	0	2	0	83%
trovare	173	52	22	99	0	0	57%
vendere	73	0	9	64	0	0	88%
vincere	81	0	0	81	0	0	100 %
33	3429	588	436	1356	840	209	

Tabella 6.2: occorrenze per ogni classe azionale

Marca contestuale	Esempio
gruppo α	
perfect	<i>They have landed.</i>
progressive	<i>I am behaving myself.</i>
duration <i>in</i> -PP	<i>She built it in an hour.</i>
duration <i>for</i> -PP	<i>I sang for ten minutes.</i>
temporal adverb	<i>I saw to it then.</i>
manner adverb	<i>She studied diligently.</i>
evaluation adverb	<i>They performe horribly.</i>
continuous adverb	<i>She will live indefinitely.</i>
gruppo β	
not or never	<i>She can not explain why.</i>
no subject	<i>He was admitted to the hospital</i>
past/pres participle	<i>...blood pressure going up</i>
present tense	<i>I am happy.</i>
past tense	<i>I was happy.</i>

Tabella 6.3: Le marche utilizzate da Siegel e McKeown [14]

Marca contestuale	ACT	STA	ACC	ACH	favorita	inibita
avverbiali						
negazione	1.453370	1.530657	0.819275	0.565232	STA	ACH
[non] già / ancora	0.680708	1.357458	1.502010	0.803345	ACC	ACT
in x tempo	1.00000	1.00000	4.704783	1.00000	ACC	-
per x tempo	8.346233	8.465071	1.00000	0.936987	STA	ACH
decorr. perfetti	1.000000	1.000000	10.394530	3.409577	ACC	ACT
decorr. imperfetti	2.461314	0.936084	0.097192	0.591363	ACT	ACC
delimitativi	1.00000	1.00000	1.00000	3.709383	ACH	-
fino a (temporale)	0.529972	4.917146	0.813037	0.230849	STA	ACH
loc.temp. punt.	0.662383	0.157643	0.761981	1.421949	ACH	STA
argomentali						
espr.num. nell'ogg.	0.155694	0.734207	1.740197	0.847124	ACC	ACT

Tabella 6.4: Le marche contestuali annotate a mano

Marca contestuale	ACT	STA	ACC	ACH	Favorita	Inibita
avverbiali						
avv. deliberativi	4.664793	1.00000	1.203594	4.465059	ACT	STA
avv. di frequenza	9.733636	0.096907	0.479991	0.046111	ACT	ACH
avv. iterativi	4.064083	1.00000	2.777080	0.808716	ACT	ACH
morfologiche						
pres.	0.606318	3.425157	0.506779	0.457735	STA	ACH
imperf.	1.483545	2.713320	0.330211	0.284292	STA	ACH
ind. rem.	0.05164	0.52859	0.98487	1.46935	ACH	ACT
pass. rem.	0.149773	0.728024	1.088681	1.259437	ACH	ACT
perfetti	0.185689	0.628172	1.163407	1.320138	ACH	ACT
futuri	0.511835	1.217487	1.100161	1.151105	STA	ACT
perif.prog.	0.962752	1.000000	7.504087	0.535936	ACC	ACH
sintattico-argomentali						
sogg. singolare	0.872792	0.666794	0.949006	1.335535	ACH	STA
sogg. plurale.	1.030458	0.489138	1.130233	1.279751	ACH	STA
espr.num.ogg.	0.155694	0.734207	1.740197	0.847124	ACC	ACT
ogg. def/indef	0.775064	0.445000	1.332720	1.291717	ACC	STA
ogg. noart	2.925374	0.572128	1.108255	0.752471	ACT	STA
ogg.i. def/indef	0.247932	0.104030	1.962390	0.468215	ACC	STA
ogg.i. noart	1.000000	9.009894	1.000000	0.947445	STA	ACH

Tabella 6.5: Le marche contestuali non lessicalizzate estratte semi-automaticamente

Marca contestuale	ACT	STA	ACC	ACH	Favorita	Inibita
Ogg. dir.	14.684090	1.000000	12.221620	1.000000	ACT	STA
Ogg. indir.	5.520467	1.000000	3.399329	1.000000	ACT	STA
M.a luogo	1.000000	1.000000	1.000000	1.000000	ACT	ACT
Completive	1.000000	0.831049	37.129950	1.000000	ACC	STA
Passivo	1.000000	6.841561	27.886500	1.000000	ACC	ACT
NoArg	29.028790	1.000000	18.529110	1.000000	ACT	STA
Sogg. Anim.	8.299703	1.000000	3.345199	1.000000	ACT	STA
Sogg. Inanim.	1.000000	1.909118	7.755223	1.000000	ACC	ACT
Ogg. Anim.	1.000000	1.000000	11.981640	1.000000	ACC	ACT
Ogg. Inanim.	1.304770	3.725482	4.123706	1.000000	ACC	ACH
Si	1.000000	416.0874	1.000000	1.000000	STA	ACT

Tabella 6.6: Le marche contestuali lessicalizzate per *scrivere*

lemma	freq.	STA	ACT	ACH	ACC	prot.
primo gruppo						
chiamare	93	36	13	0	44	47%
mettere	166	1	5	66	94	57%
trovare	173	52	22	99	0	57%
indicare	97	37	1	58	1	58%
capire	81	34	0	47	0	58%
secondo gruppo						
conoscere	68	42	0	26	0	62%
presentare	143	28	3	15	97	68%
prendere	131	8	9	90	24	69%
terzo gruppo						
passare	129	7	30	0	92	71%
portare	174	29	18	0	127	73%
costituire	62	48	0	14	0	77%
scrivere	106	3	21	0	82	77%
quarto gruppo						
comprendere	64	53	0	11	0	83%
trattare	113	94	17	0	2	83%
lasciare	121	14	2	104	1	86%
vendere	73	0	9	64	0	88%
controllare	65	58	6	0	1	89%

Tabella 6.7: i quattro gruppi con maggior grado di ibridismo azionale per la decisione su classi

primo gruppo					
	ACT	STA	ACC	ACH	
ACT	0	0	18	23	⇒ tot. act: 41
STA	0	0	37	123	⇒ tot. sta: 160
ACC	0	0	138	1	⇒ tot. acc: 139
ACH	0	0	66	204	⇒ tot. ach: 270
precisione:	-	-	0,53	0,58	occorrenze: 610
accuratezza:	-	-	0,99	0,76	errori: 43,9%
f-measure:	-	-	0,69	0,66	corrette: 56,1%
secondo gruppo					
	ACT	STA	ACC	ACH	
ACT	0	0	21	32	⇒ tot. act: 53
STA	0	42	65	131	⇒ tot. sta: 238
ACC	0	0	235	25	⇒ tot. acc: 260
ACH	0	26	81	294	⇒ tot. ach: 401
precisione:	-	0,62	0,58	0,61	occorrenze: 952
accuratezza:	-	0,18	0,9	0,73	errori: 40%
f-measure:	-	0,27	0,71	0,67	corrette: 60%
terzo gruppo					
	ACT	STA	ACC	ACH	
ACT	0	0	90	32	⇒ tot. act: 122
STA	0	90	104	131	⇒ tot. sta: 325
ACC	0	0	536	25	⇒ tot. acc: 561
ACH	0	40	81	294	⇒ tot. ach: 415
precisione:	-	0,69	0,66	0,61	occorrenze: 1423
accuratezza:	-	0,28	0,96	0,71	errori: 35,4%
f-measure:	-	0,4	0,78	0,66	corrette: 64,6%
quarto gruppo					
	ACT	STA	ACC	ACH	
ACT	0	23	90	43	⇒ tot. act: 156
STA	0	295	104	145	⇒ tot. sta: 544
ACC	0	3	536	26	⇒ tot. acc: 565
ACH	0	51	81	462	⇒ tot. ach: 594
precisione:	-	0,79	0,66	0,68	occorrenze: 1859
accuratezza:	-	0,54	0,95	0,78	errori: 30,4%
f-measure:	-	0,64	0,78	0,73	corrette: 69,6%

Tabella 6.8: il sistema assegna il tratto più frequente per ogni lemma

primo gruppo					
	ACT	STA	ACC	ACH	
ACT	8	7	10	16	⇒ tot. act: 41
STA	6	110	7	37	⇒ tot. sta: 160
ACC	2	3	118	16	⇒ tot. acc: 139
ACH	8	27	48	187	⇒ tot. ach: 270
precisione:	0,33	0,75	0,64	0,73	occorrenze: 610
accuratezza:	0,2	0,69	0,85	0,69	errori: 30,7%
f-measure:	0,25	0,72	0,73	0,71	corrette: 69,3%
secondo gruppo					
	ACT	STA	ACC	ACH	
ACT	8	9	13	23	⇒ tot. act: 53
STA	4	174	9	51	⇒ tot. sta: 238
ACC	3	5	216	36	⇒ tot. acc: 260
ACH	9	40	61	291	⇒ tot. ach: 401
precisione:	0,33	0,76	0,72	0,73	occorrenze: 952
accuratezza:	0,15	0,73	0,83	0,73	errori: 27,6%
f-measure:	0,21	0,75	0,77	0,73	corrette: 72,4%
terzo gruppo					
	ACT	STA	ACC	ACH	
ACT	47	12	40	23	⇒ tot. act: 122
STA	12	232	31	50	⇒ tot. sta: 325
ACC	26	10	490	35	⇒ tot. acc: 561
ACH	8	41	61	305	⇒ tot. ach: 415
precisione:	0,51	0,79	0,79	0,74	occorrenze: 1423
accuratezza:	0,39	0,71	0,87	0,73	errori: 24,5%
f-measure:	0,44	0,75	0,83	0,74	corrette: 75,5%
quarto gruppo					
	ACT	STA	ACC	ACH	
ACT	65	19	40	32	⇒ tot. act: 156
STA	17	437	29	61	⇒ tot. sta: 544
ACC	29	13	487	36	⇒ tot. acc: 565
ACH	9	53	63	469	⇒ tot. ach: 594
precisione:	0,54	0,84	0,79	0,78	occorrenze: 1859
accuratezza:	0,42	0,8	0,86	0,79	errori: 21,6%
f-measure:	0,47	0,82	0,82	0,79	corrette: 78,4%

Tabella 6.9: configurazione migliore: i gruppi di ibridi

tutto il corpus					
	ACT	STA	ACC	ACH	
ACT	319	21	50	40	⇒ tot. act: 430
STA	20	454	30	79	⇒ tot. sta: 583
ACC	33	15	733	41	⇒ tot. acc: 822
ACH	10	55	63	1166	⇒ tot. ach: 1294
precisione:	0,84	0,83	0,84	0,88	occorrenze: 3129
accuratezza:	0,74	0,78	0,89	0,9	errori: 14,6%
f-measure:	0,79	0,8	0,86	0,89	corrette: 85,4%

Tabella 6.10: configurazione migliore: tutto il corpus

lemma	freq.	[+durativo]	[−durativo]	prot.
primo gruppo				
trovare	173	74	99	57%
capire	81	34	47	58%
mettere	166	100	66	60%
indicare	97	39	58	60%
conoscere	68	42	26	62%
prendere	131	41	90	69%
secondo gruppo				
costituire	62	48	14	77%
terzo gruppo				
lasciare	121	17	104	86%
chiudere	114	16	98	86%
vendere	73	9	64	88%
presentare	143	128	15	90%

Tabella 6.11: i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-durativo]

lemma	freq.	[+dinamico]	[−dinamico]	prot.
primo gruppo				
capire	81	47	34	58%
chiamare	93	57	39	61%
indicare	97	60	37	62%
conoscere	68	26	42	62%
trovare	173	121	52	70%
secondo gruppo				
costituire	62	14	48	77%
presentare	143	115	28	80%
terzo gruppo				
portare	174	145	29	83%
comprendere	64	11	53	83%
trattare	113	19	94	83%
lasciare	121	107	14	88%
controllare	65	7	58	89%

Tabella 6.12: i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-dinamico]

lemma	freq.	[+omogeneo]	[−omogeneo]	prot.
primo gruppo				
chiamare	93	49	44	53%
trovare	173	74	99	57%
capire	81	34	47	58%
indicare	97	38	59	61%
conoscere	68	42	26	62%
secondo gruppo				
passare	129	37	92	71%
portare	174	47	127	73%
scrivere	106	24	82	77%
costituire	62	48	14	77%
presentare	143	31	112	78%
terzo gruppo				
comprendere	64	53	11	83%
prendere	131	17	114	87%
lasciare	121	16	105	87%
vendere	73	9	64	88%

Tabella 6.13: i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-omogeneo]

tutto il corpus			
	+DUR	-DUR	
+DUR	1583	252	⇒ tot.+DUR: 1835
-DUR	132	1162	⇒ tot.-DUR: 1294
precisione:	0,92	0,82	occorrenze: 3129
accuratezza:	0,86	0,9	errori: 12,3%
f-measure:	0,89	0,86	corrette: 87,7%
primo gruppo			
	+DUR	-DUR	
+DUR	142	188	⇒ tot.+DUR: 330
-DUR	92	294	⇒ tot.-DUR: 386
precisione:	0,61	0,61	occorrenze: 716
accuratezza:	0,43	0,76	errori: 39,1%
f-measure:	0,5	0,68	corrette: 60,9%
secondo gruppo			
	+DUR	-DUR	
+DUR	190	188	⇒ tot.+DUR: 378
-DUR	106	294	⇒ tot.-DUR: 400
precisione:	0,64	0,61	occorrenze: 778
accuratezza:	0,5	0,74	errori: 37,8%
f-measure:	0,56	0,67	corrette: 62,2%
terzo gruppo			
	+DUR	-DUR	
+DUR	318	214	⇒ tot.+DUR: 532
-DUR	121	462	⇒ tot.-DUR: 583
precisione:	0,72	0,68	occorrenze: 1115
accuratezza:	0,6	0,79	errori: 30%
f-measure:	0,65	0,73	corrette: 70%

Tabella 6.14: il sistema assegna il tratto più frequente per ogni lemma (baseline per duratività)

tutto il corpus			
	+DIN	-DIN	
+DIN	2469	77	⇒ tot.+DIN: 2546
-DIN	288	295	⇒ tot.-DIN: 583
precisione:	0,9	0,79	occorrenze: 3129
accuratezza:	0,97	0,51	errori: 11,7%
f-measure:	0,93	0,62	corrette: 88,3%
primo gruppo			
	+DIN	-DIN	
+DIN	285	26	⇒ tot.+DIN: 311
-DIN	159	42	⇒ tot.-DIN: 201
precisione:	0,64	0,62	occorrenze: 512
accuratezza:	0,92	0,21	errori: 36,1%
f-measure:	0,75	0,31	corrette: 63,9%
secondo gruppo			
	+DIN	-DIN	
+DIN	400	40	⇒ tot.+DIN: 440
-DIN	187	90	⇒ tot.-DIN: 277
precisione:	0,68	0,69	occorrenze: 717
accuratezza:	0,91	0,32	errori: 31,7%
f-measure:	0,78	0,44	corrette: 68,3%
terzo gruppo			
	+DIN	-DIN	
+DIN	652	77	⇒ tot.+DIN: 729
-DIN	230	295	⇒ tot.-DIN: 525
precisione:	0,74	0,79	occorrenze: 1254
accuratezza:	0,89	0,56	errori: 24,5%
f-measure:	0,81	0,66	corrette: 75,5%

Tabella 6.15: il sistema assegna il tratto più frequente per ogni lemma (baseline per dinamicità)

tutto il corpus			
	+OMG	-OMG	
+OMG	581	432	⇒ tot.+OMG: 1013
-OMG	56	2060	⇒ tot.-OMG: 2116
precisione:	0,91	0,83	occorrenze: 3129
accuratezza:	0,57	0,97	errori: 15,6%
f-measure:	0,7	0,89	corrette: 84,4%
primo gruppo			
	+OMG	-OMG	
+OMG	42	195	⇒ tot.+DUR: 237
-OMG	26	249	⇒ tot.-DUR: 275
precisione:	0,62	0,56	occorrenze: 512
accuratezza:	0,18	0,91	errori: 43,2%
f-measure:	0,28	0,69	corrette: 56,9%
secondo gruppo			
	+OMG	-OMG	
+OMG	90	334	⇒ tot.+DUR: 424
-OMG	40	662	⇒ tot.-DUR: 702
precisione:	0,69	0,66	occorrenze: 1126
accuratezza:	0,21	0,94	errori: 33,2%
f-measure:	0,32	0,78	corrette: 66,8%
terzo gruppo			
	+OMG	-OMG	
+OMG	143	376	⇒ tot.+DUR: 519
-OMG	51	945	⇒ tot.-DUR: 996
precisione:	0,74	0,72	occorrenze: 1515
accuratezza:	0,28	0,95	errori: 28,2%
f-measure:	0,4	0,82	corrette: 71,9%

Tabella 6.16: il sistema assegna il tratto più frequente per ogni lemma (baseline per omogeneità)

tutto il corpus			
	+DUR	-DUR	
+DUR	1682	153	⇒ tot.+DUR: 1835
-DUR	142	1152	⇒ tot.-DUR: 1294
precisione:	0,92	0,88	occorrenze: 3129
accuratezza:	0,92	0,89	errori: 9,4%
f-measure:	0,92	0,89	corrette: 90,6%
primo gruppo			
	+DUR	-DUR	
+DUR	233	97	⇒ tot.+DUR: 330
-DUR	98	288	⇒ tot.-DUR: 386
precisione:	0,7	0,75	occorrenze: 716
accuratezza:	0,71	0,75	errori: 27,2%
f-measure:	0,7	0,75	corrette: 72,8%
secondo gruppo			
	+DUR	-DUR	
+DUR	279	99	⇒ tot.+DUR: 378
-DUR	101	299	⇒ tot.-DUR: 400
precisione:	0,73	0,75	occorrenze: 778
accuratezza:	0,74	0,75	errori: 25,7%
f-measure:	0,74	0,75	corrette: 74,3%
terzo gruppo			
	+DUR	-DUR	
+DUR	419	113	⇒ tot.+DUR: 532
-DUR	120	463	⇒ tot.-DUR: 583
precisione:	0,78	0,8	occorrenze: 1115
accuratezza:	0,79	0,79	errori: 20,9%
f-measure:	0,78	0,8	corrette: 79,1%

Tabella 6.17: configurazione migliore (duratività)

tutto il corpus			
	+DIN	-DIN	
+DIN	2438	108	⇒ tot.+DIN: 2546
-DIN	141	442	⇒ tot.-DIN: 583
precisione:	0,95	0,8	occorrenze: 3129
accuratezza:	0,96	0,76	errori: 8%
f-measure:	0,95	0,78	corrette: 92%
primo gruppo			
	+DIN	-DIN	
+DIN	257	54	⇒ tot.+DIN: 311
-DIN	49	152	⇒ tot.-DIN: 201
precisione:	0,84	0,74	occorrenze: 512
accuratezza:	0,83	0,76	errori: 20,1%
f-measure:	0,83	0,75	corrette: 79,9%
secondo gruppo			
	+DIN	-DIN	
+DIN	391	49	⇒ tot.+DIN: 440
-DIN	59	218	⇒ tot.-DIN: 277
precisione:	0,87	0,82	occorrenze: 717
accuratezza:	0,89	0,79	errori: 15,1%
f-measure:	0,88	0,8	corrette: 84,9%
terzo gruppo			
	+DIN	-DIN	
+DIN	642	87	⇒ tot.+DIN: 729
-DIN	96	429	⇒ tot.-DIN: 525
precisione:	0,87	0,83	occorrenze: 1254
accuratezza:	0,88	0,82	errori: 14,6%
f-measure:	0,88	0,82	corrette: 85,4%

Tabella 6.18: configurazione migliore (dinamicità)

tutto il corpus			
	+OMG	-OMG	
+OMG	818	195	⇒ tot.+OMG: 1013
-OMG	121	1995	⇒ tot.-OMG: 2116
precisione:	0,87	0,91	occorrenze: 3129
accuratezza:	0,81	0,94	errori: 10,1%
f-measure:	0,84	0,93	corrette: 89,9%
primo gruppo			
	+OMG	-OMG	
+OMG	181	56	⇒ tot.+DUR: 237
-OMG	49	226	⇒ tot.-DUR: 275
precisione:	0,79	0,8	occorrenze: 512
accuratezza:	0,76	0,82	errori: 20,5%
f-measure:	0,78	0,81	corrette: 79,5%
secondo gruppo			
	+OMG	-OMG	
+OMG	303	121	⇒ tot.+DUR: 424
-OMG	85	617	⇒ tot.-DUR: 702
precisione:	0,78	0,84	occorrenze: 1126
accuratezza:	0,71	0,88	errori: 18,3%
f-measure:	0,75	0,86	corrette: 81,7%
terzo gruppo			
	+OMG	-OMG	
+OMG	370	149	⇒ tot.+DUR: 519
-OMG	106	890	⇒ tot.-DUR: 996
precisione:	0,78	0,86	occorrenze: 1515
accuratezza:	0,71	0,89	errori: 16,8%
f-measure:	0,74	0,87	corrette: 83,2%

Tabella 6.19: configurazione migliore (omogeneità)

Elenco delle tabelle

1.1	tabella riassuntiva dei tratti caratterizzanti di ogni classe	8
1.2	tabella riassuntiva dei tratti caratterizzanti i verbi incrementativi . .	12
2.1	il training	16
2.2	pesi di una marca	17
2.3	il sistema assegna il tratto più frequente per ogni lemma: tutto il corpus	18
2.4	valori di riferimento per la precisione del sistema per la decisione su classi	19
2.5	valori di k-statistic per ogni coppia di annotatori	21
2.6	valori di riferimento per la precisione del sistema sulla decisione per tratti	21
3.1	le marche lessicalizzate	26
4.1	confronto con la baseline per la decisione su classi	27
4.2	confronto di ogni annotatore e del modello con il gold standard . . .	28
4.3	training su gruppi separati di marche	29
4.4	confronto con la baseline sulla decisione per tratti	30
6.1	l'annotazione del corpus	37
6.2	occorrenze per ogni classe azionale	38
6.3	Le marche utilizzate da Siegel e McKeown [14]	39
6.4	Le marche contestuali annotate a mano	39
6.5	Le marche contestuali non lessicalizzate estratte semi-automaticamente	40
6.6	Le marche contestuali lessicalizzate per <i>scrivere</i>	40
6.7	i quattro gruppi con maggior grado di ibridismo azionale per la deci- sione su classi	41
6.8	il sistema assegna il tratto più frequente per ogni lemma	42
6.9	configurazione migliore: i gruppi di ibridi	43
6.10	configurazione migliore: tutto il corpus	44
6.11	i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-durativo]	44
6.12	i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-dinamico]	45

6.13	i tre gruppi con maggior grado di ibridismo azionale per la decisione sul tratto [+/-omogeneo]	45
6.14	il sistema assegna il tratto più frequente per ogni lemma (baseline per duratività)	46
6.15	il sistema assegna il tratto più frequente per ogni lemma (baseline per dinamicità)	47
6.16	il sistema assegna il tratto più frequente per ogni lemma (baseline per omogeneità)	48
6.17	configurazione migliore (duratività)	49
6.18	configurazione migliore (dinamicità)	50
6.19	configurazione migliore (omogeneità)	51

Ringraziamenti

Doverosi ringraziamenti vanno a Alessandro Lenci, Pier Marco Bertinetto, Simonetta Montemagni, Felice Dell'Orletta, senza i quali questo progetto non sarebbe potuto neanche nascere.

Ringrazio mia madre, mio padre e mia sorella, per la loro pazienza.

Ringrazio anche “i quattro moschettieri”, ovvero Vincenzo Mantova, Giovanni Pizzi, Leonardo Bartoloni e Giuseppe Ottaviano, per il sostegno tecnico (ma non solo), e in particolare per le miracolose resurrezioni del mio portatile.

Ringrazio gli amici siciliani e gli amici che vivono in Toscana, perché mi sono sempre accanto.

E ovviamente, last but not least, chi mi ha saputo dare entrambe le cose, con più calore di tutti, Jack.

Bibliografia

- [1] BERGER, A., DELLA PIETRA, S., DELLA PIETRA, V. [1996], *A maximum entropy approach to natural language processing*, in *Computational Linguistics*, 22(1): 39-71.
- [2] BERTINETTO, P.M. [1986], *Tempo, Aspetto e Azione nel verbo italiano. Il sistema dell'indicativo*, Firenze, Accademia della Crusca.
- [3] BERTINETTO, P.M. e SQUARTINI, M. [1995], *An attempt at defining the class of gradual completion verbs*, in P.M. Bertinetto, V. Bianchi, J. Higginbotham, M. Squartini (curr.), *Temporal Reference, Aspect and Actionality: Semantic and Syntactic Perspectives*, Torino, Rosenberg & Sellier: 11-26.
- [4] DELL'ORLETTA, F., LENCI, A., MONTEMAGNI, S., PIRRELLI, V. [2005], *Climbing the path to grammar: a maximum entropy model of subject/object learning*, (<http://www.aclweb.org/anthology/W/W05/W05-0509>) - *Proceedings of the ACL 2005 Workshop on Psychocomputational Models of Language Acquisition, Ann Arbor, USA*.
- [5] GRIES, S.TH., HAMPE, B., SHÖNEFELD D. [2005], *Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions*, in *Cognitive Linguistics*, 16.4:635-76.
- [6] GRIES, S.TH., HAMPE, B., SHÖNEFELD D. [2005], *Converging evidence II: More on the association of verbs and constructions*, (http://people.freenet.de/Stefan_Th_Gries/Research/Coll_vs_Freq_2.pdf).
- [7] JURAFSKY, D. e MARTIN, J.H. [2000], *Speech and Language Processing - an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, New Jersey, Prentice Hall.
- [8] LUCCHESI, V. [1971], *Fra grammatica e vocabolario. Studio sull'aspetto del verbo italiano*, in *Studi di grammatica italiana*, 1: 179-270.
- [9] MIYAO, Y. [2005], *Amis - A maximum entropy estimator for feature forests*, (<http://www-tsujii.is.s.u-tokyo.ac.jp/amis/index.html>), University of Tokyo.

- [10] MONTEMAGNI, S. et al. [2003], *Building the Italian syntactic-semantic treebank*, in A. Abeillé (ed.), *Treebanks. Building and Using Parsed Corpora*, Kluwer, Dordrecht: 189-210.
- [11] RATNAPARKHI, A. [1998], *Maximum Entropy Models for Natural Language Ambiguity Resolution*, Ph.D. Dissertation, University of Pennsylvania.
- [12] ROTHSTEIN, S. [2004], *Structuring Events - A Study in the Semantics of Lexical Aspect*, Blackwell Publishing.
- [13] RUSSELL, S. J. e NORVIG., P. [1998], *Intelligenza Artificiale: un approccio moderno*, UTET.
- [14] SIEGEL, E. V., MCKEOWN, K. R. [2000], *Learning Methods to Combine Linguistic Indicators: Improving Aspectual Classification and Revealing Linguistic Insights*, in *Computational Linguistics*, 26(4): 595-628.
- [15] SIEGEL, E. V., [1998], *Linguistic Indicators for Language Understanding: Improving using machine learning methods to combine corpus-based indicators for aspectual classification of clauses*, Ph.D. Dissertation, Columbia University.
- [16] VENDLER, Z. [1967], *Verbs and Times*, in *Linguistics in Philosophy*, Cornell University Press, Ithaca, NY, 97-121 (revised version of Vendler, Z., 1957, *Verbs and Times*, *Philosophical Review*, 66, 143-60).
- [17] VIERA, J. A., GARRETT, M. J. [2005], *Understanding Interobserver Agreement: The Kappa Statistic*, in *Family Medicine*, 37(5): 360-3.