

In corso di pubblicazione su Anna Cardinaletti(a cura di), "Intorno all'Italiano Contemporaneo", Franco Angeli, Milano.

STRUTTURE SINTATTICHE DALL'ANALISI COMPUTAZIONALE DI CORPORA DI ITALIANO

Rodolfo Delmonte
Università Ca' Foscari di Venezia

1. Introduzione

In questo articolo descriveremo il VIT, Treebank Sintattico dell'Italiano dell'Università di Venezia (Venice Italian Treebank) di 300.000 parole, creato dal Laboratorio di Linguistica Computazionale del Dipartimento di Scienze del Linguaggio. Il corpus è costituito da 50.000 parole di testo parlato e di 250.000 di testo scritto. Ci limiteremo qui a descrivere le caratteristiche della porzione di Treebank di testo scritto.¹

¹ Il primo treebank fu creato negli anni 1985-88 con il contributo di Laura Brugè, Anna Cardinaletti, Roberto Dolci, Giuliana Giusti, Paola Merlo, che collaborarono anche alla creazione del primo lessico di frequenza sottocategorizzato dell'italiano, per il quale furono scelte le prime 4000 parole nella lista di frequenza del LIF. Queste attività furono finanziate con il contratto di ricerca stipulato con la DIGITAL Equipment che era interessata a costruire una versione italiana del suo sintetizzatore della voce chiamato DECTALK, cioè un sistema di sintesi automatica della voce da testo scritto per l'italiano sulla base di quello realizzato per la lingua inglese americana. A questo scopo era necessario ricreare gli stessi strumenti linguistici presenti nella versione originale e cioè un parser sintattico robusto, un analizzatore morfologico e un lessico in grado di funzionare con qualsiasi testo italiano senza limitazioni di vocabolario. Il treebank che fu concepito all'epoca era solo cartaceo, per mancanza di altri esemplari disponibili a livello internazionale – quello della University of Pennsylvania era in fase di elaborazione – e per mancanza di software adeguato per

Il lavoro di costruzione del Treebank prevede diverse fasi che illustreremo brevemente anche con l'ausilio di figure. È infatti essenziale sapere che l'annotazione sintattica viene svolta attraverso una sequenza di operazioni automatiche, o semiautomatiche, seguite da operazioni manuali di verifica. In certi casi si è costretti a ciclare su una stessa sequenza allo scopo di ottimizzare il risultato finale. Le operazioni fondamentali sono indicate nelle Fig. 1 e 2 qui in basso.

I testi che verranno annotati sintatticamente devono essere digitalizzati, cioè disponibili su computer e, se necessario, corretti; se ci sono errori ortografici o di altra natura, questi vanno eliminati se non si vuole che nel risultato finale appaiano strutture sintattiche malformate. Successivamente, si passa alla fase di tokenizzazione dei testi, che trasforma il testo lineare in una forma verticale, assegnando a ciascuna parola una riga o record. In questa fase è necessario verificare che le parole composte da lettere e numeri, lettere e segni grafici, le date, le formule e altre strutture ortografiche che non siano semplicemente sequenze di lettere dell'alfabeto siano state trasformate adeguatamente e che non si sia persa alcuna parola del testo originario.

Dal file tokenizzato si procede alla creazione delle *Multiwords* o forme polirematiche, che utilizza un dizionario creato ad hoc, nel quale si possono aggiungere eventuali altre forme o espressioni idiomatiche che si intende analizzare come parola singola. In questa fase viene creato un eventuale lessico specializzato per il dominio in analisi: questo è quanto è avvenuto nell'analisi dei testi dei dialoghi spontanei all'interno dei progetti nazionali AVIP/API/IPAR (vedi Delmonte 2003; Delmonte *et al.* 2004), che hanno anche visto la codifica di

produrre l'annotazione in maniera interattiva e consistente. La documentazione cartacea – che ancora si trova nel Laboratorio di Linguistica Computazionale dove fu elaborata – venne utilizzata per creare una grammatica context-free dell'italiano probabilistica, cioè una lista di tutte le regole di riscrittura prodotte dall'annotazione manuale e per ogni regola diversa il numero di frequenze della regola stessa nel corpus. Il corpus prescelto era di 40.000 parole e composto da articoli di quotidiani e riviste di carattere politico, economico, di attualità e di linguaggio burocratico. I testi erano digitalizzati e disponibili su computer mainframe, ma non taggati in POS. Questo primo lavoro fu documentato in Delmonte - Dolci (1989).

Il lavoro di costruzione di Treebank fu poi continuato in maniera discontinua e sporadica, riutilizzando i testi di cui sopra e ampliando lentamente il campione, fino all'approvazione del progetto nazionale SI-TAL nel 1998, che costituì la spinta fondamentale per giungere anche a una normalizzazione della annotazione sintattica.

Il Treebank attuale utilizza quei testi e altri testi elaborati per il progetto nazionale SI-TAL e i progetti AVIP/API/IPAR, nonché testi annotati in progetti a contratto per incarico dell'IRST.

semiparole, non parole e altre forme di disfluenza. Ove possibile, il lessico specializzato deve riportare anche il lemma di riferimento. A questo punto, il file contiene *Multiwords* e parole singole che però, nel caso costituiscano delle forme amalgamate come i verbi con clitico vengono poi trasformate in due elementi separati.

Il programma successivo assegna i tag o le parti del discorso a ciascun token individuato precedentemente sulla base di un dizionario di parole invariabili e di un analizzatore morfologico (Delmonte - Pianta 1996), che nel caso in cui mancasse la radice corrispondente nel radiciario, può procedere al “guessing”. Questa operazione prevede una decomposizione della parola in affissi – flessivi e di derivazione – allo scopo di individuare una radice esistente, in mancanza della quale la parola viene classificata con l’etichetta provvisoria “npro” (nome proprio) se maiuscola, altrimenti “fw” (foreign word) se minuscola (Delmonte - Pianta 1998). In questa fase le parole amalgamate vengono decomposte e vengono create due parole; inoltre viene costruita una immagine del testo in forma di frasi, frasi che verranno poi utilizzate dall’analisi sintattica che assume appunto la frase come ambito della sua analisi.

Come accennato più sopra, le fasi di analisi morfologica e di lemmatizzazione, oltre a quelle di creazione di lessici specialistici, nonché quella di costruzione delle forme politematiche, hanno richiesto uno o più cicli di verifica manuale per la loro costruzione definitiva. Una volta costruita una versione verticalizzata dei testi da analizzare sintatticamente con ciascun “token” associato a uno o più “tag” o parti del discorso, abbiamo iniziato la fase, semi-automatica, della disambiguazione, cioè della assegnazione del significato associato a ciascuna etichetta come deciso dal contesto. Abbiamo elencato in Appendice 1. i tag che nel progetto SI-TAL (vedi Montemagni *et al.* 2000) hanno preso il nome di SFEATS.

I testi che abbiamo analizzato hanno un livello di ambiguità media del 1,9%, cioè ogni parola può avere una media di due significati, circa. A questo scopo sono stati creati degli algoritmi per la disambiguazione che sono in parte statistici e in parte sintattici (vedi Delmonte - Pianta 1999; Delmonte, Chiran, Bacalu 2000; Bianchi - Delmonte 2002), e che convergono in un programma che dispone di una interfaccia per l’annotatore, permettendogli di prendere decisioni in tempo reale sulla etichetta da assegnare nel contesto attuale, anche se diversa da quella indicata dall’analisi automatica. Le operazioni che vengono compiute sono indicate nella Fig. 2.

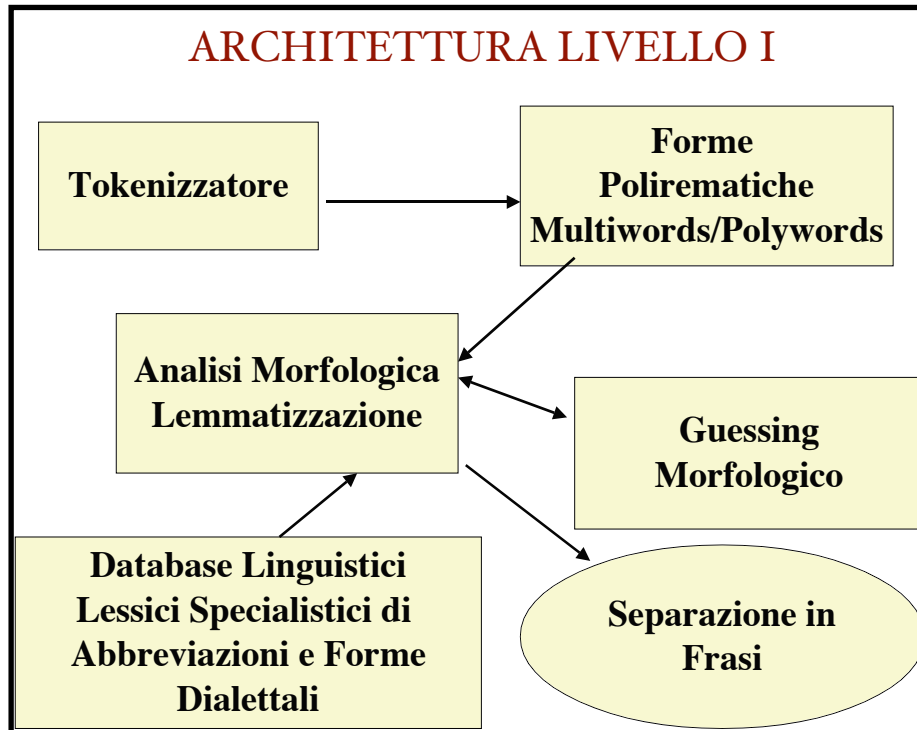


Fig.1 Suite di Annotazione – Primo Livello: Tokenizzazione, Lessico e Tagging

La creazione del treebank vero e proprio corrisponde alle due ultime operazioni di Fig. 2 – Shallow Parsing e Mapping Funzionale (Delmonte 2000) – ed è quindi l’ultimo passo di un processo di analisi abbastanza lungo ed elaborato, durante il quale il testo iniziale viene completamente trasformato. L’analisi automatica del parser viene poi passata alla verifica manuale e infine alla collazione da parte di un supervisore che è responsabile della eventuale unificazione delle “varianti” strutturali proposte dai diversi annotatori per uno stesso tipo strutturale. Operazione questa molto delicata e in alcuni casi che potrebbe richiedere la rielaborazione completa di porzioni non indifferenti del Treebank stesso, come è accaduto nel progetto SI-TAL per le strutture comparative e quantificate, di cui discuteremo in §##.

Nel proseguio di questo articolo ci interesseremo solo ed esclusivamente dell'annotazione sintattica e dei criteri linguistici adottati, sottostanti all'annotazione stessa. In particolare, si partirà dai fondamenti teorici che hanno ispirato la definizione dello schema per giungere ad una dettagliata descrizione dei costituenti sintattici e dei relativi criteri di annotazione, per poi terminare in

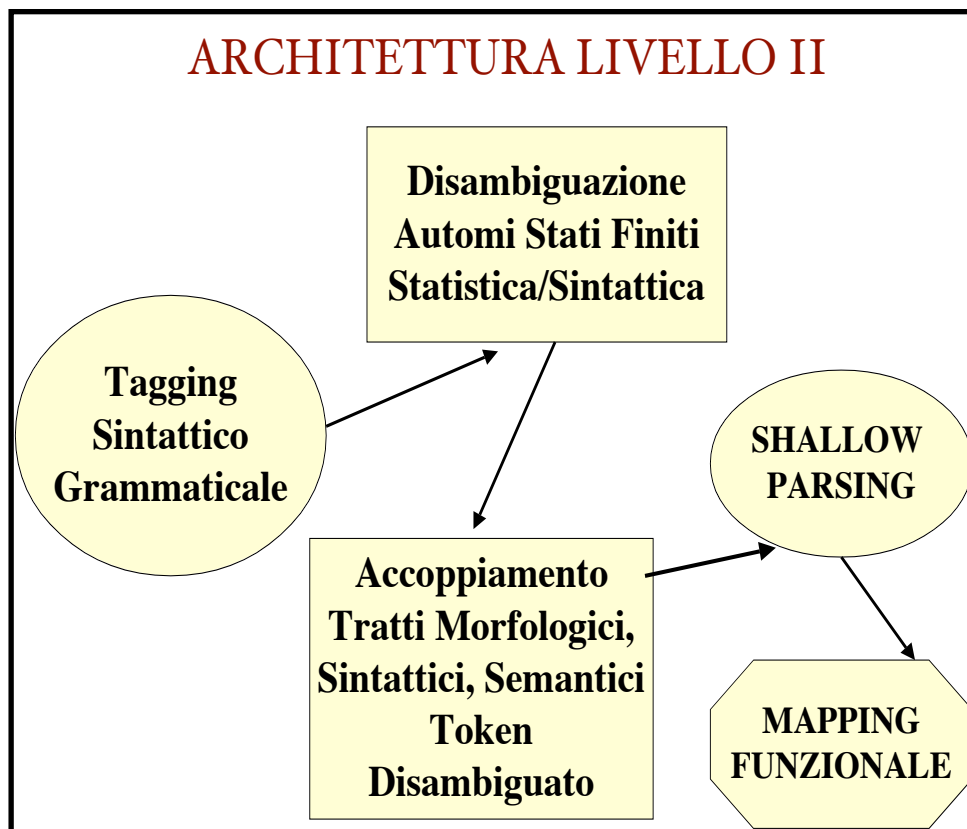


Fig.2 Suite di Annotazione – Secondo Livello: Disambiguazione e Parsing

§§6-8 con una esemplificazione dei casi strutturali più interessanti dal punto di vista sintattico e stilistico. La teoria ispiratrice per eccellenza è stata la Teoria X-barra alla quale abbiamo però operato delle variazioni ispirate dalla teoria LFG

(Bresnan 2001) e dalla necessità di rendere più piatta la rappresentazione, di ridurre, cioè, ove possibile, il numero di parentesi da dover poi controllare.

2. Teoria X-barra e costituenza sintattica

Schematicamente, la teoria X-barra prevede una organizzazione a teste e proiezioni di teste ognuna delle quali dotata di una barra, in ordine gerarchico: per cui il nodo da cui pende la testa è denotato da “Ø” e i successivi nodi dominanti hanno una barra, due barre e eventuali ulteriori barre – anche se due barre sono di solito il massimo livello di proiezione. L’organizzazione gerarchica della teoria prevede le seguenti regole astratte di riscrittura:

CP	-->	Spec, Cbarra
Cbarra	-->	C0, XP
XP	-->	Spec, Xbarra
Xbarra	-->	X, Complementi
CØ	-->	Complementatore
X	-->	Verbo, Aggettivo, Nome, Avverbio, Preposizione

Una prima scelta che abbiamo compiuto è l’organizzazione interna dello specificatore che, nel caso di costituenti non frasali, può essere costituito da uno o più elementi linguistici appartenenti a diverse categorie sintattiche minori:

Spec	-->	Determinanti, Quantificatori, Intensifiers
------	-----	--

Lo schema adattato delle regole X-barra è il seguente:

XP	-->	spec(atomico), Xbarra
Xbarra	-->	X, Complementi
X	-->	Verbo, Nome, Aggettivo, Avverbio, Preposizione

La scelta di avere una struttura di specificatore era troppo onerosa e abbiamo quindi lasciato in forma atomica il materiale linguistico che si trovasse prima della testa, altro che non richieda esso stesso una strutturazione propria##. Ovviamente, elementi anteposti, come gli aggettivi o gli avverbi, avranno una loro struttura propria di costituente. Lo stesso varrà per i quantificatori. Ma di tutti questi casi specifici parleremo nel dettaglio più in basso.

Per il livello di frase si preferisce quindi il seguente dettaglio dello schema X-barra:

CP	-->	Spec(Costituenti Dislocati), Cbarra
Cbarra	-->	C0, IP
IP	-->	Spec (SN<SOGGetto>), Ibarra
Ibarra	-->	I0, Complementi
CØ	-->	Complementatore
IØ	-->	Verbo flesso a tempo finito

Il verbo flesso assume a sua volta una struttura separata, che chiamiamo Ibar, e che può essere composta da più elementi aggiunti al livello di costituenza del verbo flesso:

Ibar	-->	X
X	-->	Verbo - ausiliari, modali, clitici, negazione, avverbiali

3. Costituenti sintattici nello schema di annotazione del VIT

Nella definizione dello schema di annotazione non adotteremo integralmente la teoria X-barra e i nomi che da essa derivano: nel dettaglio, utilizzeremo lo schema generale e le etichette IBAR e CP. Per il resto utilizzeremo sempre etichette di costituenti sintattici all'italiana: S(intagma) e non P(hrase) per non provocare confusione.

Le etichette adottate sono commentate nel dettaglio in quanto segue. Presentiamo qui in basso l'organizzazione astratta della nostra rappresentazione sintattica per il livello di frase, con le modifiche che introdurremo per la treebank dell'italiano. È importante ricordare che la stessa organizzazione si potrebbe applicare anche agli altri costituenti maggiori con teste semantiche – SA, SN, SP, ecc. –, ma poiché questi non posseggono argomenti obbligatori, verranno posti sullo stesso livello degli aggiunti e seguiranno quindi tutti la testa nell'ordine lineare in cui verranno analizzati in superficie.

Come detto più sopra, per il livello di frase utilizzeremo CP, per contenere i costituenti Dislocati o semplicemente Anteposti e eventuali Aggiunti, frasali e non. Per il resto utilizzeremo Ibarra come previsto nella teoria, per riscrivere il verbo flesso e il costituente SV che contiene i complementi; ognuno dei nodi di SV è specializzato per un tipo di sottocategorizzazione. Invece di IP utilizzeremo

F, e questo nodo verrà dominato da nodi di F specializzati. Non utilizzeremo quindi C0 e nemmeno I0. Lo schema adottato risulta dunque come segue:

CP --> (Aggiunti), Spec(Costituenti Dislocati), Cbarra
 Cbarra --> FC/FS/FInt, F
 F --> Spec (SN<SOGGetto>), Ibarra, (Aggiunti)
 Ibarra --> (Ausil,Avverb,Neg, Clitici), VerboFlesso, Complementi
 Complementi --> COMPT/ COMPC/COMPIN, (Aggiunti)

Contemporaneamente, ai fini del mapping tra il livello di annotazione sintattico a costituenti e quello sintattico-funzionale, verranno isolati introduttori semantici quali i subordinatori e il nucleo verbale, inserito nel costituente IBAR.

Per quanto riguarda poi la questione dei tratti di accordo, ci baseremo su suggerimenti di LFG che non prevedono nessun accordo verbale nella rappresentazione, ma tutti i tratti nei SN, che nel nostro caso sono quelli del livello morfosintattico. Riportiamo qui in basso esempi di strutture per i vari costituenti maggiori allo scopo di mostrare l'organizzazione dei componenti sulla base dello schema X-barra. Come si potrà notare, la testa è sempre l'ultimo elemento che precede i complementi e/o gli aggiunti, ed è a sua volta preceduta da eventuali costituenti minori che costituiscono il contenuto di SPECificatore, nodo che non riportiamo per non appesantire la rappresentazione stessa.

Ogni paragrafo che segue tratterà quei costituenti sintattici ai quali abbiamo dedicato successivamente una analisi linguistica dettagliata. Ogni costituente verrà discusso separatamente mettendo in grassetto l'elemento testa del costituente preso in considerazione.

3.1. F - Frase

Con questo simbolo si intende una frase semplice con un predicato principale e complementi che possono anche includere altre frasi. Sono anche considerate frasi semplici le frasi ellittiche.

- (1) a. Gullit ha dichiarato guerra alla stampa sportiva
 f-[sn-[npro-Gullit], ibar-[ausa-ha, **vppt-dichiarato**], compt-[sn-[n-guerra], sp-[part-alla, sn-[n-stampa], sa-[ag-sportiva]]]]
- b. Dobbiamo vincere domenica per interrompere questo brutto momento
 f-[ibar-[vsup-dobbiamo, **viin-vincere**], sn-[nt-domenica], sp-[p-per, sv2-[vit-interrompere, sn-[dim-questo, ag-brutto, nt-momento]]]]

3.2 IBAR – I Barra

Con questo simbolo viene riscritto il gruppo verbale che può includere al proprio interno gli ausiliari *essere* o *avere*, la negazione, eventuali clitici e avverbi che si possono interporre tra l'ausiliare e il verbo lessicale. Oltre agli ausiliari fanno parte di questo costituente i modali e i verbi aspettuali, e anche la forma perifrastica *stare per*. Analogamente ad altre treebank, il problema dei verbi con enclitico è stato risolto utilizzando la forma di parola con l'enclitico e ripetendo successivamente il clitico come parola separata. In questo caso, quindi, il clitico si troverebbe sia all'interno di IBAR che nel complemento (COMPT/IN/C) dove riceverebbe una interpretazione.

Segue una tipologia di costituenti IBAR:

- (5) a. non ha
ibar-[neg-non, **vc-ha**]
- b. è diventata
ibar-[ause-è, **vc-diventata**]
- c. si vende
ibar-[clit-si, **vt-vende**]
- d. le ha dato
ibar-[clitdat-le, ausa-ha, **vppt-dato**]
- e. è arrivata
ibar-[ause-è, **vppin-arrivata**]
- f. non devono guardare
ibar-[neg-non, vsup-devono, **viin-guardare**]
- g. è stato assunto
ibar-[ause-è, ausep-stato, **vppt-assunto**]
- h. sono stati spesso segnalati
ibar-[ause-sono, ausep-stati, avv-spesso, **vppt-segnalati**]

3.3. SN – Sintagma Nominale

Il Sintagma Nominale si sviluppa con una serie di specificatori, una testa ed eventuali complementi o aggiunti. Tra i costituenti specificatori ci sono gli articoli, i quantificatori, i numerali che, se non hanno struttura quantificata, vengono semplicemente elencati prima della testa; sempre prima del nome si trovano gli aggettivi e i possessivi che ricevono una propria struttura di

costituente. Gli aggettivi che seguono la testa del SN possono avere a loro volta una struttura interna con complementi ed aggiunti. Sempre all'interno del SN ci sono i costituenti SPD e gli eventuali complementi frasali nel caso si tratti di un nome fattivo o una forma polirematica. Le frasi relative sono tutte incluse nel SN.

Un commento a parte va fatto per i nominali che fungono da modificatori di altri nominali. Alcuni di questi sono nomi propri (*i fratelli Karamazov, la commissione Sangalli*); altri sono semplici nomi comuni che hanno una funzione di specificazione (*il settore cucine, la politica suicida*, ecc). In tutti questi casi non si individua esplicitamente la testa a livello di costituente, in quanto i due o più nominali si troverebbero allo stesso livello. Vi sono poi casi in cui la testa è una *Multiword* o una forma polirematica. Segue una tipologia di costituenti di tipo SN:

- (14)a. La Bimex
sn-[art-La, **npro-Bimex**]
- b. una bella ragazza
sn-[art-una, sa-[ag-bella], **n-ragazza**]
- c. molti libri
sn-[ind-molti, **n-libri**]
- d. dieci programmatori
sn-[num-dieci, **n-programmatori**]
- e. infondate speculazioni
sn-[sa-[ag-infondate], **n-speculazioni**]
- f. il loro strumento di ricerca
sn-[art-il, sa-[poss-loro], **n-strumento**, spd-[pd-di, sn-[n-ricerca]]]
- g. cosiddetta commissione sangalli
sn-[sa-[ag-cosiddetta], n-commissione, sn-[nh-sangalli]]
- h. il fatturato complessivo delle società
sn-[art-Il, **n-fatturato**, sa-[ag-complessivo], spd-[partd-delle, sn-[n-società]]]

3.4. SA – Sintagma Aggettivale

Oltre che all'interno del SN, questo costituente viene creato solo per la posizione predicativa degli aggettivi e in questo caso può avere una struttura interna, come in (b):

- (25)a. infelice
sa-[**ag-infelice**]

- b. inferta alla nostra penisola
sa-[**agn-inferta**, sp-[part-alla, sn-[poss-nostra, n-penisola]]]]

3.5. F2 – Frase 2 – Frasi Relative

Questo costituente ha la stessa funzione di CP, serve cioè a contenere costituenti che sono anteposti e che hanno una relazione semantica con qualche elemento nella frase sottostante. Possono contenere un SN che come testa abbia un elemento funzionale, il complementatore *che*, oppure un pronome relativo preceduto oppure no da un articolo. Potrà contenere anche un SP per i pronomi relativi obliqui, preceduto da una preposizione semplice o articolata, oppure individuati da *cui*. Il costituente F2 domina sempre invariabilmente un nodo F. Non contiene la traccia o l'elemento vuoto del costituente legato al pronome relativo (v. §4.2):

- (27)a. il ragazzo al quale Gullit ha parlato
sn-[art-il, n-ragazzo, f2-[sp-[part-al, **rel-quale**], f-[sn-[npro-Gullit], ibar-[ausa-ha, vppin-parlato]]]]
- b. le imprese italiane che producono e importano energia
sn-[art-le, n-imprese, ag-italiane, f2-[sn-[**rel-che**], f-[coord-[ibar-[vt-producono], cong-e, ibar-[vt-importano]], compt-[sn-[n-energia]]]]]]

3.6. FINT – Frase Interrogativa

Questo costituente ha la stessa funzione di CP, serve cioè a contenere costituenti che sono anteposti e che hanno una relazione semantica con qualche elemento nella frase sottostante. In particolare, come nel caso di F2 contiene un SN o un SP che hanno come testa un pronome interrogativo, preceduto o no da una preposizione. Il pronome interrogativo a sua volta può essere seguito o preceduto dal complementatore *che*.

- (29)a. chi abbia incontrato Maria
fint-[sn-[**int-chi**], f-[ibar-[auair-abbia, vppt-incontrato], compt-[sn-[nh-Maria]]]]
- b. di che cosa stai parlando?
fint-[spd-[pd-di, sn-[int-che_cosa]], f-[ibar-[vprog-stai, vgin-parlando]], puntint-?]

3.7. FAC – Frase Complemento

Questo costituente ha la stessa funzione di CP, serve però a contenere soltanto la testa di CP, CØ, cioè il complementatore *che*, che in certi contesti può anche essere vuoto. In questi casi però il nodo FAC verrà eliminato. Il nodo FAC può apparire nel contesto di un complemento verbale come COMPT oppure in posizione estraposta, all’inizio frase, sotto CP:

- (31) a. che c’era Maria con Luigi
compt-[fac-[pk-**che**, f-[ibar-[clit-c, vc-era], compc-[sn-[nh-Maria], sp-[p-con, sn-[npro-Luigi]]]]]]
- b. Che Maria sia partita significa che si è arresa
cp-[fac-[pk-**che**, f-[sn-[nh-Maria], ir_infl-[aueir-sia, vppin-partita]],
ibar-[vt-significa], compt-[fac-[pk-**che**, f-[ibar-[clit-si, ause-è, vppt-arresa]]]]]

3.8. FP – Frase Parentetica

Con FP si indica la presenza di punteggiatura nel testo che può servire successivamente ad individuare una eventuale frase parentetica o appositiva, o semplicemente a marcare la spezzatura del testo in parti informativamente differenti. Con FP sono designati gli aggiunti aperti del tipo participiale posti dopo la testa che li governa. Le strutture appositive possono essere ovviamente anche aggiunti di tipo chiuso, modificatori o attributi di vario tipo. Le parentetiche possono essere frasi di commento, in forma anche ellittica, racchiuse tra due simboli di punteggiatura.

- (33) Slitta la riunione, prevista per oggi, della cosiddetta commissione Sangalli
f-[ibar-[vin-slitta], sn-[art-la, n-riunione, fp-[punt-., f-[sv3-[vppt-prevista], sp-[p-per, sn-[n-oggi]], punt-.,], spd-[partd-della, sn-[ag-cosiddetta, n-commissione, nh-sangalli]]]]]

Per parentetica si intende anche la frase reggente di enunciati a discorso diretto, che riporta solitamente l’identità del parlante, e che si trova all’interno o alla fine del discorso diretto stesso (vedi §3.9):

- (34) I concerti – spiegano gli organizzatori – perché sono stati...

compt-[sn-[art-i, n-concerti]], fp-[par--, ibar-[vt-spiegano], compt-[sn[art-gli, n-organizzatori], par--], fs-[cosu-perché, cp-[f-[ibar-[ause-sono, ausep-stati,

3.9. DIRSP - discorso diretto

Questo costituente viene utilizzato soprattutto per frasi che contengono discorso diretto virgolettato, come in (a), e serve per contenere due frasi separate dal segno di punteggiatura “:”, come in (b). Questo segno è però ambiguo tra discorso diretto, discorso riportato, commento, ecc., e sarà discusso nel dettaglio in §8.2. Riportiamo qui solo alcuni esempi:

- (36)a. “Abbiamo deciso di confermare i concerti – spiegano gli organizzatori – perché sono stati superati tutti gli impedimenti.
dirsp-[par-", ibar-[ausa-abbiamo, vppt-deciso], compt-[sv2-[pt-di, vit-confermare, compt-[sn-[art-i, n-concerti]], fp-[par--, ibar-[vt-spiegano], compt-[sn-[art-gli, n-organizzatori], par--], fs-[cosu-perché, cp-[f-[ibar-[ause-sono, ausep-stati, vppt-superati]], sn-[qc-tutti, art-gli, n-impedimenti, punto-.]]]]]]]
- b. Il traino più consistente viene sempre dalle vendite all'estero: il +9,5% tendenziale di maggio, contro il 6% di marzo e il 7,4% di aprile.
dirsp-[f-[sn-[art-il, n-traino, sa-[in-più, ag-consistente]], ibar-[vin-viene], compin-[savv-[avv-sempre], spda-[partda-dalle, sn-[n-vendite, sp-[part-all, sn-[n-estero]]]]]], dirs-:, f-[sn-[art-il, num-plus9/5%, n-tendenziale, spd-[pd-di, sn-[nt-maggio]]], punt-, sp-[p-contro, coord-[sn-[art-il, num-6%, spd-[pd-di, sn-[nt-marzo]]], cong-e, sn-[art-il, num-7/4%, spd-[pd-di, sn-[nt-aprile]]]]], punto-.]

3.10. F3 - Frase frammento

Questo costituente contiene frasi che non hanno un IBAR al loro interno, ovvero sono semplicemente dei costituenti di tipo nominale, avverbiale, preposizionale, ecc. privi completamente di verbo. Casi tipici di strutture F3 sono i titoli. Includiamo alcuni esempi:

- (37)a. Intersind pronta al suo ingresso in Confindustria.
f3-[sn-[npro-intersind, sa-[ag-pronta, sp-[part-al, sn-[poss-suo, n-ingresso, sp-[p-in, sn-[npro-confindustria]]]]]], punto-.]
- b. Poi l'approvazione del programma d'attività presentato alla Giunta il 4 giugno.

f3-[cong-f-poi, sn-[art-l, n-approvazione, spd-[partd-del, sn-[n-programma, sp-[p-d, sn-[n-attività]], sv3-[ppas-presentato, compt-[sp-[part-alla, sn-[n-giunta]], sn-[art-il, num-4, nt-giugno]]]]], punto-.]

3.11. CP - Frase con aggiunti in inizio frase e frasi complesse

Anche in questo caso si indicano tutti i costituenti frasali che iniziano con elementi aggiunti a cui deve poi fare seguito la frase semplice. Si utilizza questa marca di costituente anche quando ci sono frasi complesse, cioè una frase semplice con frasi subordinate o coordinate, per indicare dei nessi frasali in genere:

- (39) Nello stesso periodo sulle reti Fininvest sono andati in onda oltre 10mila spot in più rispetto al gennaio-marzo '93 (8%).
 cp-[sp-[part-Nello, sn-[dim-stesso, nt-periodo]], sp-[part-sulle, sn-[n-reti, sn-[npro-Fininvest]]], f-[ibar-[ause-sono, vppin-andati], compin-[sp-[p-in, sn-[n-onda]], sn-[sq-[in-oltre, num-10, num-mila], n-spot, savv-[p-in, avv-più, fc-[ccong-rispetto_all, sn-[nt-gennaio, par-^c, nt-marzo, date-novantatre]]], par-(, sn-[num-8%], par-)]], punto-.]

3.12. SQ - Sintagma Quantificato e Strutture Comparative

Il Sintagma Quantificato che ha come testa un quantificatore ha costituito un problema notevole per la sua rappresentazione, soprattutto per il fatto di una carenza di bibliografia adeguata – quella della lingua inglese è inadeguata e insufficiente. Il SQ si trova spesso all'interno di strutture comparative come quelle riportate in (39), (40) e (49) e che discuteremo in §6.4. Il SQ può assumere diversi ruoli sintattici:

- può essere uno specificatore del Sintagma Nominale (SN), del Sintagma Aggettivale (SA), oppure del Sintagma Avverbiale (SAVV):

- (40) livelli dello 0,6% superiori a quelli del bimestre precedente e dell'1,1% più alti rispetto all'analogo periodo del 1993
 sn-[n-livelli, coord-[sa-[sq-[q-dello, num-0/6%], ag-superiori, sp-[p-a, sn-[deit-quelli, spd-[partd-del, sn-[nt-bimestre, ag-precedente]]]]], cong-e, sa-[sq-[q-dell, num-1/1%, in-più], ag-alti, fc-[ccong-rispetto_all, sn-[ag-analogo, nt-periodo, spd-[partd-del, sn-[num-1993]

- (41)a. una delle più importanti zone costiere del Paese
 sn-[sq-[q-una, qd-delle, in-più], ag-importanti, n-zone, sa-[ag-costiere], fc-
 [ccom-del, sn-[n-Paese]]
- b. qualunque altro impiegato statale
 sn-[sq-[ind-qualunque, q-altro], n-impiegato, sa-[ag-statale]]
- (43)a. oltre 10mila spot in più
 sn-[sq-[in-oltre, num-10, num-mila], n-spot, savv-[p-in, avv-più]]
- b. un pô prematuramente
 savv-[sq-[art-un, in-pô], avv-prematuramente]

- Il Sintagma Quantificato può essere un complemento di verbi che reggono SN quantificati:

- (45)a. è sceso nel mese corrente dell'1,1%
 ibar-[ause-è, vppin-scuso], compin-[sp-[part-nel, sn-[nt-mese, sa-[ag-
 corrente]]], sn-[q-dell, num-1/1%]]
- b. il PIL tedesco-occidentale risulta del 3,7% ed è stato del 4% in aprile
 sn-[art-il, npro-PIL, sa-[ag-tedesco_occidentale]], ibar-[vc-risulta], compc-[sq-
 [q-del, num-3/7%]]], fc-[cong-ed, f-[ibar-[ause-è, vppc-stato], compc-[sn-[q-
 del, num-4%]], sp-[p-in, sn-[nt-aprile]]]
- c. sono oggi più di 700
 ibar-[vc-sono, avv-oggi], compc-[sn-[sq-[in-più, q-di], num-700]]
- d. (49) risulta aumentato dell'1,9% rispetto ai primi tre mesi del 1993
 ibar-[vc-risulta], compc-[sa-[ppas-aumentato, sn-[q-dell, num-1/9%]], fc-
 [ccong- rispetto_ai, sn-[sq-[num-primi, num-tre], nt-mesi, sn-[num-1993]]]

- Il SQ può essere un modificatore o aggiunto nelle strutture comparative come mostrato nei seguenti esempi, in cui è strettamente in relazione al sintagma chiamato DegP in ambito chomskiano, essendo quindi espressione di un sintagma di grado:

- (53) a. maggiore di più di un quinto del voto
 sa-[ag-maggiore, spd-[pd-di, sn-[sq-[in-più, cong-di], num-un, num-quinto, fc-
 [ccom-del, sn-[n-voto]]]
- b. è aumentata ben di più di quella teorica
 ibar-[ause-è, vppin-aumentata], compin-[sq-[in-ben, cong-di, avv-più], fc-
 [ccom-di, sn-[dim-quella, sa-[ag-teorica]]]

3.13. CP_INT - Frase interrogativa con aggiunti in inizio frase

Questo costituente serve per marcare frasi interrogative che iniziano con aggiunti oppure congiunzioni frasali subordinanti o coordinanti, o ancora frasi interrogative frammento.

- (55)a. Effetto Multiplex – Quali sono gli effetti di questa concentrazione tra distribuzione ed esercizio?
cp_int-[sn-[n-Effetto, sn-[npro-Multiplex]], par-(-), fint-[sn-[int-Quali], f-[ibar-[vc-sono], compc-[sn-[art-gli, n-effetti, spd-[pd-di, sn-[dim-questa, n-concentrazione, sp-[p-tra, sn-[coord-[n-distribuzione, cong-ed, n-esercizio]]]]]]]]], puntint-?]
- b. Ma oggi, con i CD Rom e la multimedialità, è davvero lo stesso mestiere di 500 anni fa?
- c. Fine delle ambizioni o ennesimo voltafaccia su una vicenda sulla quale Mosca negli ultimi mesi non ha cessato di cambiare posizioni, di dire tutto e il contrario di tutto?

4. Dati quantitativi

In quanto segue presenteremo e discuteremo i dati quantitativi relativi ai costituenti contenuti nelle 10.000 frasi del nostro Treebank. In particolare focalizzeremo la nostra attenzione su alcune strutture interessanti dal punto di vista stilistico. Presenteremo poi alcuni esempi di elevata complessità strutturale (v. §6 e §7).

Tipo di Costituente	Simbolo	Numero di occorrenze
Enunciati Totali		10.000
Tokens Totali		250.000
Sintagmi Nominali	SN	67.667
Sintagmi Preposizionali	SP	20.615
Sintagmi Preposizionali DI/DA	SPD/SPDA	20.571
Sintagmi Aggettivali	SA	18.183
Sintagmi Avverbiali	SAVV	3.830
Sintagmi Quantificati	SQ	621
Complesso Verbale	IBAR	14.067

Complesso Verbale Irreale	IR_INFL	1.410
Strutture Coordinate	COORD	2.217
Frase Semplice	F	15.599
Frase Subordinata	FS	1.037
Frase Coordinata/Comparativa	FC	2.922
Frase Parentetica	FP	7.214
Frase Interrogativa Semplice	FINT	546
Frase Complessa	CP	5.404
Frase Interrogativa Complessa	CP_INT	199
Frase Complemento	FAC	901
Frase a Tempo Indefinito	SV2/SV3/SV5	7.355
Frase Relativa	F2	3.288
Frase a Discorso Diretto	DIRSP	716
Frase Frammento	F3	2.290
Struttura Complemento Transitivo	COMPT	11.311
Struttura Complemento Intransitivo	COMPIN	3.850
Struttura Complemento Copulativo	COMPC	3.350
TOTALE COSTITUENTI NON FRASALI		165.475
TOTALE COSTITUENTI FRASALI		47.471
TOTALE COSTITUENTI		212.946

Tab. 2 Dati Quantitativi Costituenti VIT

Come si può capire dai dati di Tab. 5, gli enunciati del nostro Treebank hanno in media 25 tokens. Ogni enunciato contiene in media 21 costituenti, 6,7 SN, 2 SP, 2 SPD/SPDA, 2 SA, 1,5 IBAR e 1/2 SV, 1,6 COMP. La media ovviamente non ci dice nulla della struttura reale degli enunciati anche se si può utilizzare come possibile approssimazione. Un primo dato interessante che non appare nella Tabella è il numero di frasi semplici prive di SOGGetto espresso lessicalmente: sono 6.227, su un totale di 15.599, dunque circa il 40%.

Abbiamo quindi guardato all'interno del corpus e delle strutture alte di costituenza per capire come sono organizzati gli enunciati e i costituenti al loro interno.

La prima struttura che abbiamo studiato è quella del Sintagma Aggettivale che è molto interessante in una lingua come l'italiano. Sono state individuate le seguenti caratteristiche generali:

17.372 Aggettivi
18.183 SA

1.225 possessivi
14.287 SA PostNominali
2.347 SA PreNominali Assoluti
1.431 SA PreNominali Interni
1.169 SA Complemento

Dei 18mila SA, una buona parte, più di un terzo, 5.000, è una struttura con SA coordinati. La grande maggioranza, 6.880 sono in posizione predicativa, e solo una minoranza – 2.321 – è in posizione prenominale di inizio costituente SN. Una porzione ancora inferiore è costituita da SA con determinanti o quantificatori di SN che li precedono. E infine, 577 sono gli SA utilizzati come complemento. Riporteremo più in basso alcuni esempi interessanti.

Nei paragrafi che seguono analizzeremo le strutture di SA dal punto di vista delle loro caratteristiche organizzative astratte.

5. Analisi di Sintagmi Aggettivali

5.1. Coordinazioni

Negli esempi riportati in basso abbiamo cercato di raccogliere una gamma di tipologie di coordinazione più estesa possibile, ma certo non esaustiva di tutti i casi presenti nel treebank. Il criterio di annotazione è improntato a non produrre parentesizzazioni sovrabbondanti, quindi di norma, se si hanno coordinazioni di sole teste, queste vengono lasciate libere all'interno del costituente generico COORD. Come si può notare, ci sono coordinazioni di 2, 3, 4 e anche 5 aggettivi di seguito: in (70), la coordinazione avviene tra coppie di aggettivi e nell'esempio (69), si alterna una struttura di SPDI, un SN, e un SA in forma ellittica, mostrando in questo modo come il modificatore aggettivale è semanticamente equivalente agli altri modificatori. Gli aggettivi sono a loro volta modificati da avverbi, da negazione e dall'operatore quantificato *più*, da strutture parentetiche, e in alcuni casi da altri SA, come in (60), dove *internazionali* viene ulteriormente specificato dal SA coordinato. Di particolare interesse sono gli aggettivi in posizione attributiva, che come si può vedere possono avere struttura ed essere coordinati, come mostrato da (63) e (64).

(67) a. aspetti privatistici, amministrativi, penali tributari e comunitari

- b. (70) grandi risorse umane e materiali, produttive e tecnico-scientifiche, intellettuali e culturali
- c. (69) la progettazione, la fornitura e il montaggio degli impianti elettrici, di termoventilazione, antincendio, idrosanitari
- d. (59) un governo efficace e democraticamente controllato
- e. (61) soggetto generale, ma non esclusivo dello sviluppo
- f. (60) una cordata d'operatori internazionali, arabi o australiani
- g. (63) più alti e più diffusi bisogni culturali e formativi.
- h. (64) un funesto e insieme affascinante miraggio

5.2. Pronominalizzazioni

In questa sezione presentiamo alcuni casi di pronominalizzazione aggettivale: per rappresentare il fenomeno sintattico abbiamo inserito il SA all'interno di un SN, intendendo con questo motivare la possibile presenza di un pronome vuoto, replica della testa nominale reggente. I casi esemplificati vanno dalla ellissi del nominale in (72) alla sua totale assenza negli altri casi. Bisogna poi diversificare casi di aggettivi come il deittico *medesimi* di (74), che hanno uno statuto diverso dagli altri; un aggettivo quantificato come quello di (75); l'uso nominale di *bello* per intendere "l'aspetto bello, interessante" di (76), e infine gli aggettivi che vengono usati liberamente anche in forma nominale, come *fresco* di (77)

- (72)a. Per le immobilizzazioni finanziarie e le immateriali si rende necessario...
- b. (74) i medesimi dovranno...
- c. (75) il massimo cui l'alleanza era...
- d. (76) il bello è che Berlusconi e Fini la pensano in...
- e. (77) il fresco c'è solo in Scandinavia, dove le...

5.3. Dipendenze interne

Come tutti gli altri costituenti, i Sintagmi Aggettivali possono avere a loro volta dei modificatori interni come mostrato in (81)

- (81) di orientamenti politici e culturali diversi

5.4. Strutture quantificate e comparative

Abbiamo già incontrato strutture comparative negli esempi (40,49) che riportiamo qui in basso:

- (40) a. livelli dello 0,6% superiori a quelli del bimestre precedente e dell'1,1% più alti rispetto all'analogo periodo del 1993
 b. (49) il PIL tedesco occidentale risulta aumentato dell'1,9% rispetto ai primi tre mesi del 1993

In queste frasi troviamo un modello di struttura comparata di due tipi determinati dal predicato reggente: in (a) il Verbo regge la struttura comparativa, in (b) è l'aggettivo che regge la struttura comparativa:

- () a. A [Operatore/Verbo] di Quant rispetto a B
 b. A di Quant [Operatore/Aggettivo a] B
 A di Quant [Operatore+Aggettivo rispetto a] B

Il caso (a) si riferisce all'esempio (49) dove A è seguito da un predicato verbale Operatore/Verbo che contiene all'interno del suo significato lessicale un operatore semantico (più o meno) e che regge la struttura "di Quant" alla sua destra dove Quant è una quantità numerica introdotta da DI (e le sue forme articolate) che funge da misura della comparazione che si effettua tra A e B; ed è seguita da B, il secondo elemento di comparazione. Il cambiamento del tipo di predicato da Verbo ad Aggettivo, come in (b), provoca uno spostamento dell'elemento "di Quant" a sinistra e Operatore/Aggettivo è un aggettivo di grado come *superiore*, *inferiore* che contiene un operatore (più o meno) nel suo significato lessicale, mentre Operatore+Aggettivo indica una coppia di elementi costituita di nuovo da un operatore lessicalmente espresso e un aggettivo di grado. Queste strutture sono diverse da quelle che riportiamo qui in basso, più semplici da analizzare in quanto strutture predicative di tipo copulativo.

- (82)a. il tutto sarebbe certo molto più efficace, controllabile e democratico di quanto oggi non sia.
 f-[sn-[art-il, qc-tutto], ibar-[vcir-sarebbe, avv-certo], compc-[sq-[in-molto, q-più, coord-[sa-[ag-efficace, punt-., ag-controllabile, cong-e, ag-democratico]], fc-[ccom-di, q-quanto, f-[ibar-[avv-oggi, neg-non, vcir-sia]]]]]]], punto-.]
 b. Più buono di così non poteva essere.
 cp-[sq-[in-Più, sa-[ag-buono], fc-[ccom-di, savv-[avv-così]]], f-[ibar-[neg-non, vsupp-poteva, vci-essere]], punto-.]

5.5. Tipologie di spec di SA

Abbiamo già visto in §6.1 come i SA possano contenere nello specificatore un avverbiale o una negazione. In questa sezione riportiamo altri casi di specificatore, tra cui *allora* di (86a), *più* di (b), a sua volta modificato da *sempre* in (c):

- (86)a. ricordando la moglie Monica allora incinta
- b. più alti e diffusi bisogni culturali e formativi.
- c. di una sempre più libera e democratica vita interna

5.6. Focalizzate inverse (in Frasi F)

In questa sezione e in quella successiva analizziamo strutture di complementi aggettivali che hanno subito una dislocazione a sinistra perché focalizzate e la frase ha quindi una struttura inversa con il soggetto in posizione postverbale.

- (90) Diversi sono anche i livelli di concentrazione del mercato
dirsp-[cp-[sa-[ag-Diversi], f-[ibar-[vc-sono], compc-[sn-[in-anche, art-i, n-
livelli, spd-[pd-di, sn-[n-concentrazione, spd-[partd-del, sn-[n-mercato]]]]]]]]]]

5.7. Focalizzate inverse (in Frammenti F3)

Le strutture di complementi aggettivali che hanno subito una dislocazione a sinistra si possono anche trovare all'interno di Frammenti di frase in cui il verbo è stato omesso: in tutti i frammenti è però possibile ricostruire un verbo ESSERE.

- (92)a. Buono anche l'andamento delle vendite negli USA
f3-[sa-[ag-Buono], cong-anche, sn-[art-l, n-andamento, spd-[partd-delle, sn-[n-
vendite, sp-[part-negli, sn-[np-Usa]]]],
- b. Probabile il ricorso al blocco del turn-over
f3-[sa-[ag-Probabile], sn-[art-il, n-ricorso, sp-[part-al, sn-[n-blocco, spd-
[partd-del, sn-[npro-turn_over]]]]],

5.8. Inversione nel complemento

In (a) e (b) riportiamo alcuni esempi di inversione del SOGGetto in posizione postverbale dopo un complemento predicativo di tipo SA, con verbi copulativi. In (c) abbiamo un caso di inversione dell'OGGetto nel complemento posizionandosi dopo il complemento predicativo di tipo SA, questa volta con un verbo transitivo a predicazione secondaria.

- (98) Diventa così più acuta la contraddizione
 f-[ibar-[vc-diventa], compc-[savv-[avv-così], sa-[in-più, ag-acuta], sn-[art-la, n-contraddizione]

5.9. Loro dativo

È interessante notare come in tutto il Treebank ci sia un solo caso di *loro* dativo, dipendente da un aggettivo, e precisamente in una frase di italiano burocratico:

- (101) ...presente ordinanza e con il punteggio loro spettante in base alla tabella Allegato A)
 sn-[ag-presente, sn-[n-ordinanza]]], cong-e, sp-[p-con, sn-[art-il, n-punteggio]], sv3-[sn-[clitdat-loro], ppre-spettante, comp-[p-in, sn-[n-base, sp-[part-alla, sn-[n-tabella, sn-[n-Allegato, sn-[abbr-A], par-),

6. Altre strutture complesse

6.1. SN dislocati a sinistra

In (a) riportiamo un caso di dislocazione a sinistra con clitico di ripresa, cioè di OGGetto preposto in inizio frase, prima del SOGGetto e quindi all'interno di CP, che viene poi ripreso da un clitico nell'IBAR; in (b) un caso di HANGING TOPIC, *il concorso*, di nuovo posizionato in CP, e ripreso da un clitico nella frase interrogativa successiva; anche (c) è un caso di Hanging Topic, di forma idiomatica:

- (102)a. Una decisione importante Ghitti l'ha riservata all'imputato numero uno del processo:

- dirsp-[cp-[sn-[art-una, n-decisione, sa-[ag-importante]], f-[sn-[nh-Ghitti], ibar-[clitac-l, ausa-ha, vppt-riservata], compt-[sp-[part-all, sn-[n-imputato, sn-[n-numero, num-uno], spd-[partd-del, sn-[n-processo]]]]]]];
- b. Il concorso l'ha vinto Francesco o Giovanni?
cp_int-[sn-[art-il, n-concorso], f-[ibar-[clitac-l, ausa-ha, vppt-vinto], compt-[sn-[sn-[nh-Francesco], cong-o, sn-[nh-Giovanni]]]], puntint-?]
- c. Brutta faccenda, i sudditi si ribellano.
cp-[f3-[sn-[sa-[ag-brutta], n-faccenda]], punt-, f-[sn-[art-i, n-sudditi], ibar-[clit-si, vt-ribellano], punto-.]]

6.2. Relative come enunciati

Riportiamo in questa sezione alcuni esempi di Frasi Relative utilizzate come Enunciati indipendenti. È abbastanza evidente che leggendole separatamente dal contesto, le frasi risultino tutte incomplete e solo parzialmente comprensibili. L'espedito retorico di riportare in una frase separata una porzione di un enunciato in forma di Frase Relativa potrebbe essere interpretato come uno stilema per enfatizzare maggiormente un concetto.

- (105) Che, sostengono alla Farnesina, non ha per niente gradito l'operazione by_pass.
cp-[f2-[rel-Che, cp-[fp-[punt-, f-[ibar-[vt-sostengono], compt-[sp-[part-alla, sn [npro-Farnesina]]]], punt-,], f-[ibar-[neg-non, ausa-ha, sp-[p-per, avv1-niente], vppt-gradito], compt-[sn-[art-l, n-operazione, n-by_pass]], punto-.]]]]

6.3. Strutture ad Aux-to-Comp e ad inversione dell'italiano letterario

Le strutture che presentiamo in questa sezione, tratte da testi di Dino Buzzati, sono tipiche del linguaggio letterario; non ne abbiamo rintracciate in testi di altri generi. I quattro esempi che riportiamo qui in basso hanno strutture in parte simili e comunque riconducibili al fenomeno definito nella letteratura linguistica come Aux-to-Comp (Rizzi 1982), in cui cioè l'Ausiliare si trova posizionato nel Comp di CP, precedendo in questo modo il SN SOGGetto. Nell'esempio (a) si tratta di *L'aver*, in (b) di *si fosse*.

- (114)a. L'aver il figlio abbandonato il mare per la città le era sempre sembrato, in cuor suo, un tradimento alle tradizioni di famiglia.
cp-[topf-[auxtoc-[sn-[art-l, ausai-avere]], sn-[art-il, n-figlio], sv3-[vppt-abbandonato, compt-[sn-[art-il, n-mare], sp-[p-per, sn-[art-la, n-città]]]]]],

- f-[ibar-[clitdat-le, ause-era, avv-sempre, vppt-sembrato], fp-[punt-., sp-[p-in, sn-[n-cuor, sa-[poss-suo]]], punt-., compt-[sn-[art-un, n-tradimento, sp-[part-alle, sn-[n-tradizioni, spd-[pd-di, sn-[n-famiglia]]]]]], punto-.]
- b. E, si fosse egli trasferito pure nel più remoto continente, ancora il colombre si sarebbe appostato nello specchio di mare più vicino
 fc-[congf-e, punt-., cp-[topf-[auxtoc-[clit-si, aueir-fosse], sn-[pron-egli], sv3-[vppin-trasferito, cong-pure, compin-[sp-[part-nel, sn-[sa-[in-più, ag-remoto], n-continente]]]], punt-., cp-[savv-[avv-ancora], f-[sn-[art-il, n-colombre], ibar-[clit-si, ause-sarebbe, vppin-appostato], compin-[sp-[part-nello, sn-[n-specchio, spd-[pd-di, sn-[n-mare, sa-[in-più, ag-vicino]]]]]]],

Negli esempi (116) e (117) sia l'ausiliare sia il verbo al participio passato sono in CP. La loro posizione è individuata dalla presenza del *che* complementatore che si trova nel Comp di CP, ed entrambi precedono il SN SOGGetto, che in (116) è *sei mesi*, in (117) il Pronome inespresso *io*.

- (116)a. Trascorsi che furono sei mesi
 cp-[topf-[sv3-[vppin-trascorsi, cong-che, ause-furono, compin-[sn-[num-sei, nt-mesi]]]],
- b. E, ritornato che fui in città, la mia sontuosa villa era sparita;
 fc-[congf-E, punt-., cp-[topf-[sv3-[vppin-ritornato, cong-che, ause-fui], compin-[sp-[p-in, sn-[n-città]]]], punt-., f-[sn-[art-la, sa-[poss-mia], sa-[ag-sontuosa], n-villa], ibar-[ause-era, vppin-sparita]], punt-;

Come è possibile notare, si è scelto di trattare come strutture topicalizzate gli SN di (114) e gli SV3 di (116) con una marca specifica di costituente TOPF, che non abbiamo neppure conteggiato in quanto appare solo le 4 volte che mostriamo qui sopra. Il senso di questa annotazione è di rendere la spezzatura prosodica (Ronat 1982; Delmonte 1985, 1987a, 1987b) che accompagna il fenomeno dell'Aux-to-Comp, che non permette in alcun modo di gerarchizzare gli elementi sintattici che si trovano in due posizioni separate, come schematicamente indicato in (), o invertite, come in ():

- () a. L'aver / il figlio / abbandonato
 b. E si fosse / egli / trasferito
- () a. Trascorsi / che / furono / sei mesi
 b. E ritornato / che / fui / in città

Si potrebbe scegliere di frammentare il costituente TOPF e separarlo dagli altri che si troverebbero a galleggiare in CP allo stesso livello, come in (), ma in questo modo si dovrebbe recuperare la semantica dei costituenti, una parte della quale è di natura puramente pragmatica:

- () cp-[topf-[auxtoc-[sn-[art-l, ausai-avere]], sn-[art-il, n-figlio], sv3-[vppt-
abbandonato, compt-[sn-[art-il, n-mare], sp-[p-per, sn-[art-la, n-città]]]]]

Oppure si potrebbe decidere di posizionare i costituenti sotto CP assegnando la marca di infinitiva, inframezzando il SN soggetto, come in (). Il SN viene poi ripreso dal clitico *lo* nella frase principale seguente, rendendo la struttura una dislocazione a sinistra con clitico:

- () cp-[sn-[art-l, sv3-[ausai-avere, sn-[art-il, n-figlio], vppt-
abbandonato, compt-[sn-[art-il, n-mare], sp-[p-per, sn-[art-la, n-città]]]]]]]

Lo stesso argomento si potrebbe sviluppare nei confronti della frase (114b), in cui di nuovo abbiamo scelto la rappresentazione con i costituenti posizionati allo stesso livello di CP piuttosto che quella suggerita dalla letteratura al riguardo, come in (), in cui l'ausiliare viene lasciato appeso a CP assieme al clitico e il participio passato viene ad occupare la posizione all'interno di IBAR che si trova ora sotto F.

- () fc-[cong-f-e, punt-., cp-[clit-si, aueir-fosse, f-[sn-[pron-egli], ibar-[vppin-
trasferito], compin-[cong-pure, sp-[part-nel, sn-[sa-[in-più, ag-remoto], n-
continente]]]]]

Il problema si ripresenterebbe con l'inversione degli altri due esempi di (116), in cui il participio passato precede l'ausiliare che a sua volta viene preceduto da un *che* complementatore. Anche qui abbiamo scelto una rappresentazione a Sintagma Verbale a tempo indefinito con il SN soggetto posizionato all'interno del complemento, piuttosto della semanticamente scomoda ():

- () cp-[vppin-trascorsi, pk-che, f-[ibar-[ause-furono], compin-[sn-[num-sei, nt-
mesi]]]]]

Nella sintassi superficiale che abbiamo prescelto non c'è spazio per il movimento di costituenti minori o maggiori, né abbiamo indicato categorie vuote

(v. §4.2). Le marche funzionali introdotte servono quindi a supplire la caratterizzazione semantica del fenomeno, che ha valenze stilistiche molto forti.

7. Il Discorso Diretto

Per finire presenteremo strutture di Discorso Diretto (DD), che costituiscono un elemento molto interessante dal punto di vista teorico e descrittivo. Gli enunciati contenenti DD sono generalmente composti da due sotto enunciati legati tra di loro, uno dei quali ha la funzione di individuare il soggetto parlante del discorso diretto, e quindi l'enunciato virgolettato contenente quanto detto. L'enunciato riportante l'identità del parlante a cui va ascritto il contenuto del discorso diretto è prodotto da un soggetto diverso da quello dell'altro enunciato: ci sono quindi due enunciati appartenenti a due soggetti diversi, la cui semantica è quindi indipendente l'una dall'altra. Così viene trattato il fenomeno da Banfield (1982), che parla di Expressions.

Per semplificare il problema, si potrebbe pensare che l'enunciato di DD sia come quello di Discorso Riportato, cioè un complemento frasale e quindi in quanto tale una proposizione indipendente, ma la stessa Banfield dimostra come gli enunciati di DD non possano in nessun modo essere assimilati a quelli di Discorso Riportato o Indiretto. D'altro canto, i verbi di comunicazione reggenti sono pur sempre dei verbi transitivi che prendono come loro complemento un oggetto che può essere una frase: in tutti gli esempi che riportiamo in basso è difatti possibile cliticizzare l'enunciato DD.

A noi interessa in questa sede mettere in luce le difficoltà insite nella rappresentazione sintattica del DD e della frase reggente parentetica, che si può trovare in varie posizioni. Le più ovvie e "convenienti" sono all'inizio o alla fine del DD, ma più spesso, la frase parentetica si trova inserita nel DD stesso, rendendo impossibile, o di non facile trattamento, una soluzione a frase complemento.

7.1 La posizione della parentetica

A. parentetica inserita alla fine del discorso diretto:

(118) "Che l'eccesso di coordinamento e di pianificazione può anche andare a detrimento della creatività del ricercatore", ha commentato il vicepresidente del

CNR.

dirsp-[par-", fac-[pk-Che, f-[sn-[art-l, n-eccesso, coord-[spd-[pd-di, sn-[n-coordinamento]], cong-e, spd-[pd-di, sn-[n-pianificazione]]]], ibar-[vsup-può, in-anche, viin-andare], compin-[sp-[p-a, sn-[n-detrimento, spd-[partd-della, sn-[n-creatività, spd-[partd-del, sn-[n-ricercatore]]]]]]], par-", *punt-*, *f-[ibar-[ausa-ha, vppt-commentato], compt-[sn-[art-il, n-vicepresidente, spd-[partd-del, sn-[npro-Cnr]]]]], punto-*.]

B. parentetica inserita tra più enunciati di discorso diretto

- (121) “Ad ogni metro troviamo cadaveri – ha raccontato un impiegato dell’ufficio del governatore – ci sono case devastate dalla frana
dirsp-[par-", cp-[sp-[p-ad, sn-[qd-ogni, n-metro]], f-[ibar-[vt-troviamo], compt-[sn-[n-cadaveri]]], *fp-[par--], f-[ibar-[ausa-ha, vppt-raccontato], compt-[sn-[art-un, n-impiegato, spd-[partd-dell_, sn-[n-ufficio, spd-[partd-del, sn-[n-governatore]]]]]]], par-*, f-[ibar-[clitabl-ci, vc-sono], compc-[sn-[n-case, sv3-[ppas-devastate, comppas-[spda-[partda-dalla, sn-[n-frana]]]]], punt-,

C. parentetica inserita tra SOGG e IBAR

- (122) “A questo punto la data”, ha detto D’Alema ieri sera, “dipende da Berlusconi”.
dirsp-[par-", cp-[sp-[p-a, sn-[dim-questo, n-punto], f-[sn-[art-la, n-data], par-", *fp-[punt-*, *f-[ibar-[ausa-ha, vppt-detto], compt-[sn-[npro-D_, npro-Alema, savv-[avv-ieri], nt-sera]]], punt-*, par-", ibar-[vin-dipende], compin-[spda-[pda-da, sn-[nh-Berlusconi]], par-", punto-.]]]]]

D. parentetica inserita tra materiale in CP e F

- (125) “In ogni caso” puntualizza lui, “comincerò ad occuparmene dalla settimana prossima”.
dirsp-[par-", cp-[sp-[p-in, sn-[qd-ogni, n-caso]], par-", *f-[ibar-[vt-puntualizza], compt-[sn-[pron-lui], punt-*, par-", f-[ibar-[vin-comincerò], compin-[sv2-[pt-ad, vcl-occuparmene, compt-[clit-me, clit-ne], spda-[partda-dalla, sn-[nt-settimana, sa-[ag-prossima]], par-", punto-.]]]]]]]

E. parentetica che interrompe una frase complemento nel CP

- (127) “Se oggi assume un nuovo impiegato”, spiega Agostinucci, “l’azienda dà per scontato che di un telefono: lo deve dotare”.
dirsp-[par-", fs-[cosu-Se, cp-[savv-[avv-oggi], ibar-[vt-assume], compt-[sn-[art-

un, sa-[ag-nuovo], n-impiegato]]], par-", fp-[punt-, f-[ibar-[vt-spiega], compt-[sn-[nh-Agostinucci]]], punt-,], par-", f-[sn-[art-l, n-azienda], ibar-[vin-dà], compin-[sp-[p-per, sn-[agn-scontato, compc-[fac-[pk-che, cp-[spd-[pd-di, sn-[art-un, n-telefono]]], dirs-:, f-[ibar-[clitac-lo, vsup-deve, vit-dotare]]]]]]], par-", punto-.]

F. discorso diretto proprio, privo di parentetica

- (129) “Senza PAR-CONDICIO niente elezioni”
 dirsp-[par-", cp-[sp-[p-senza, sn-[n-PAR_CONDICIO]], f3-[savv-[avv-niente], sn-[n-elezioni]], par-", punto-.]

7.2. I due punti

I due punti “:” sono introduttori generici del DD, ma in realtà servono a varie strategie retoriche: oltre a quella canonica di precedere il DD virgolettato, come in G., servono come in H. a introdurre una o più strutture sintattiche non necessariamente frasali, che svolgono la funzione retorica di chiarificatori o esplicatori di quanto detto nella porzione di enunciato, la tesi, che precede i due punti. Emblematico il caso rappresentato in N., in cui prima dei due punti il “narratore” o soggetto della narrazione fa un proprio commento, poi dopo i due punti viene riportato DD virgolettato che viene interrotto dalla frase parentetica. Quindi in questo caso siamo in presenza di tre enunciazioni con tre funzioni retoriche a livello di discorso e con semantiche separate. Nell’esempio contenuto in O. addirittura abbiamo un uso interno al DD dei due punti: anche qui esiste la frase parentetica, e ciò che segue i due punti è interno al DD, quindi interpretabile dal suo contenuto semantico come lista di SP come chiarificatori di quanto espresso nella frase precedente.

G. i due punti hanno la funzione canonica di precedere il discorso diretto virgolettato

- (133) E Keys ha risposto subito: <in questo programma il mondo degli affari è pronto a svolgere un ruolo positivo e attivo>.
 dirsp-[fc-[cong-e, f-[sn-[nh-Keys], ibar-[ausa-ha, vppt-risposto], compt-[savv-[avv-subito]]], dirs-:, cp-[par-<, sp-[p-in, sn-[dim-questo, n-programma]], f-[sn-[art-il, n-mondo, spd-[partd-degli, sn-[n-affari]], ibar-[vc-è], compc-[sa-[ag-pronto, sv2-[pt-a, vit-svolgere, compt-[sn-[art-un, n-ruolo, sa-[coord-[ag-positivo, cong-e, ag-attivo]]]]]]], par->, punto-.]

H. i due punti introducono il DD, ma la parentetica che rivela l'identità del riportato viene omessa

- (130) E poi stringe i cordoni della borsa sul ripiano dei 18mila miliardi di debiti delle USL, mentre frena ancora sul ruolo dei privati nell'assistenza sanitaria: <sarebbe come appaltare la giustizia ai privati>.
dirsp-[fc-[cong-e, cong-poi, f-[ibar-[vt-stringe], compt-[sn-[art-i, n-cordoni, spd-[partd-della, sn-[n-borsa]]], sp-[part-sul, sn-[n-ripiano, spd-[partd-dei, sn-[num-18, num-mila, num-miliardi, spd-[pd-di, sn-[n-debiti, spd-[partd-delle, sn-[npro-USL]]]]]]]]], punt-, fs-[cosu-mentre, f-[ibar-[vt-frena, avv-ancora], compt-[sp-[part-sul, sn-[n-ruolo, spd-[partd-dei, sn-[n-privati]], sp-[part-nell_, sn-[n-assistenza, sa-[ag-sanitaria]]]]]]]]], dirs-:, f-[par-<, ibar-[vcir-sarebbe, savv-[avv-come]], compc-[sv2-[vit-appaltare, compt-[sn-[art-la, n-giustizia], sp-[part-ai, sn-[n-privati]]]]], par->, punto-.]

I. i due punti introducono una lista di SN come informazione chiarificatrice

- (131) E ci saranno anche i vecchi vertici e gli ex manager della Ferruzzi: l'ex amministratore delegato della Montedison Carlo Sama;
dirsp-[fc-[cong-e, f-[ibar-[clit-ci, vcir-saranno], compc-[coord-[sn-[in-anche, art-i, sa-[ag-vecchi], n-vertici], cong-e, sn-[art-gli, sa-[ag-ex], n-manager, spd-[partd-della, sn-[npro-Ferruzzi]]]]]]], dirs-:, f-[coord-[sn-[art-l, sa-[ag-ex], n-amministratore, sa-[ag-delegato], spd-[partd-della, sn-[npro-Montedison]], nh-Carlo, nh-Sama], punt

L. dopo i due punti, segue un discorso diretto riportato libero e poi un discorso diretto virgolettato

- (132) E non ha evitato di dare consigli: il servizio segreto civile andrebbe sciolto e riformato, ha detto, aggiungendo che sono <irresponsabili le parole del Ministro dell'Interno che riguardano me>.
dirsp-[fc-[cong-e, f-[ibar-[neg-non, ausa-ha, vppt-evitato], compt-[sv2-[pt-di, vit-dare, compt-[sn-[n-consigli]]]]]]], dirs-:, f-[sn-[art-il, n-servizio, sa-[ag-segreto], sa-[ag-civile]], ibar-[vprogir-andrebbe, coord-[vppt-sciolto, cong-e, vppt-riformato]], fp-[punt-, f-[ibar-[ausa-ha, vppt-detto], punt-,], sv5-[vgt-aggiungendo, compt-[fac-[pk-che, f-[ibar-[vc-sono], par-<, compc-[sa-[ag-irresponsabili], sn-[art-le, n-parole, spd-[partd-del, sn-[n-Ministro, spd-[partd-dell, sn-[n-Interno]]]]], f2-[rel-che, f-[ibar-[vt-riguardano], compt-[sn-[pron-me]]]]]]]]], par->, punto-.]

M. la frase che segue i due punti è usata come fatto probante della tesi o ipotesi della prima parte dell'enunciato.²

- (135) E, in effetti, le ultime elezioni politiche hanno provocato una vera e propria rivoluzione: nessun leader politico che aveva tenuto le redini del potere amministrativo della provincia negli ultimi decenni è stato riconfermato.
 dirsp-[fc-[cong-e, punt-., savv-[avvl-in_effetti], punt-., f-[sn-[art-le, sa-[ag-ultime], n-elezioni, sa-[ag-politiche]], ibar-[ausa-hanno, vppt-provocato], compt-[sn-[art-una, sa-[coord-[ag-vera, cong-e, sa-[poss-propria]]], n-rivoluzione]]], dirs-:, f-[sn-[ind-nessun, n-leader, sa-[ag-politico], f2-[rel-che, f-[ibar-[ausa-aveva, vppt-tenuto], compt-[sn-[art-le, n-redini, spd-[partd-del, sn-[n-potere, sa-[ag-amministrativo], spd-[partd-della, sn-[n-provincia]]]]], sp-[part-negli, sn-[sa-[ag-ultimi], nt-decenni]]]]], ibar-[ause-è, ausep-stato, vppt-riconfermato], punto-.]

N. il discorso diretto viene ascritto a un anonimo *qualcuno* comunque virgolettato

- (137) Ma, a volte, i rischi possono trasformarsi in opportunità: <Perché – deve aver pensato qualcuno – non cedere il credito in sofferenza verso Arganti a Sasea per ridurre l'esposizione del gruppo Novara?>.
 dirsp-[fc-[congf-Ma, punt-., savv-[p-a, avvl-volte], punt-., f-[sn-[art-i, n-rischi], ibar-[vsup-possono, vcl-trasformarsi], compc-[clit-si, sp-[p-in, sn-[n-opportunità]]]], dirs-:, par-<, fint-[int-Perché, fp-[par-(-), f-[ibar-[vsup-deve, ausai-aver, vppt-pensato], compt-[sn-[ind-qualcuno]]], par-(-)], f-[ibar-[neg-non, vit-cedere], compt-[sn-[art-il, n-credito, sp-[p-in, sn-[n-sofferenza]], sp-[p-verso, sn-[nh-Arganti]]], sp-[p-a, sn-[npro-Sasea]], sp-[p-per, sv2-[vit-ridurre, compt-[sn-[art-l_, n-esposizione, spd-[partd-del, sn-[n-gruppo, sn-[np-Novara]]]]]]], puntint-?, par->], punto-].

O. il discorso diretto inizia subito e i due punti vengono comunque adottati da chi parla con funzione fattiva

- (138) <In questi anni – osserva Monsignor Carlo Mazza, direttore dell'Ufficio

² Questo può anche essere esplicitato dalla presenza della congiunzione *infatti*:

- (i) Infine, c'è da risolvere il problema del Totta, di cui il Banesto ha il 50% ma in modo ritenuto irregolare dal Governo portoghese: infatti, metà del pacchetto fu acquistato attraverso società di comodo.

Nazionale del Turismo della CEI – il nostro settore ha registrato un confortante aumento della qualità: dall'accoglienza, ai trasporti, alla cultura delle persone (accompagnatori e visitatori) coinvolti nei viaggi>.

dirsp-[cp-[par-<, sp-[p-In, sn-[dim-questi, nt-anni]], fp-[par-(-), f-[ibar-[vt-osserva], compt-[sn-[n-Monsignor, sn-[nh-Carlo, nh-Mazza], fp-[punt-., sn-[n-direttore, spd-[partd-dell, sn-[n-Ufficio, sa-[ag-Nazionale], spd-[partd-del, sn-[n-Turismo]], spd-[partd-della, sn-[npro-Cei]]]]]]], par-(-), f-[sn-[art-il, sa-[poss-nostro], n-settore], ibar-[ausa-ha, vppt-registrato], compt-[sn-[art-un, sa-[ag-confortante], n-aumento, spd-[partd-della, sn-[n-qualità]]]]], dirs-:, f3-[spda-[partda-dall, sn-[n-accoglienza]], punt-., sp-[part-ai, sn-[n-trasporti]], punt-., sp-[part-alla, sn-[n-cultura, spd-[partd-delle, sn-[n-persone, fp-[par-(, sn-[coord-[n-accompagnatori, cong-e, n-visitatori]], par-)], sv3-[ppas-coinvolti, compt-[sp-[part-nei, sn-[n-viaggi]]]]]]], par->, punto-.]

P. la tesi è costituita da un SN frammento a cui fa seguito la spiegazione, che può a sua volta essere un frammento o una frase completa

(139) Unica cura: uno choc collettivo, che lui pensò di fornire appunto con Psycho nel 1960.

dirsp-[f3-[sn-[sa-[ag-unica], n-cura], dirs-:, f3-[sn-[art-uno, n-choc, sa-[ag-collettivo]], punt-., f2-[rel-che, f-[sn-[pron-lui], ibar-[vt-pensò], compt-[sv2-[pt-di, vit-fornire, avv-appunto, compt-[sp-[p-con, sn-[npro-Psycho]]], sp-[part-nel, date-1960], punto-.]]]]]

Q. uso generico per discorso riportato senza frase parentetica:

(142) Intanto il personale incalza: c'è un contratto in arrivo.

dirsp-[fc-[cong-intanto, f-[sn-[art-il, n-personale], ibar-[vin-incalza]]], dirs-:, f-[ibar-[clit-c, vc-è], compc-[sn-[art-un, n-contratto], sp-[p-in, sn-[n-arrivo]]], punto-.]

8. Riferimenti bibliografici

Banfield A. (1982), *Unspeakable Sentences. Narration and Representation in the Language of Fiction*, Boston, Routledge & Kegan Paul, 23-63.

Bianchi D., Delmonte R. (2002), "Tecniche di apprendimento applicate al problema del tagging: una prima valutazione per l'Italiano", Workshop "NLP e Web: La Sfida della Multimodalità tra Approcci Simbolici e Approcci Statistici", Convegno Nazionale

- AI*IA, Siena, 20-34.
- Bresnan J. (2001), *Lexical-Functional Syntax*, Blackwells.
- Delmonte R. (1985), "Parsing Difficulties and Phonological Processing in Italian", *Proceedings of the 2nd Conference of the European Chapter of ACL*, Geneva, 136-145.
- Delmonte, R. (1987a), "Focus and the Semantic Component", *Rivista di Grammatica Generativa* 12, 81-121.
- Delmonte R. (1987b), "The Realization of Semantic Focus and Language Modeling", in *Proceedings XIth ICPHS*, 1987, Tallinn (Estonia), Vol. 2, 101-104.
- Delmonte R. (2000), "Shallow Parsing And Functional Structure In Italian Corpora", LREC, Atene, 113-119.
- Delmonte R. (2003), "Parsing Spontaneous Speech", in *Proceedings of EUROSPEECH2003*, Pallotta V., Popescu-Belis A., Rajman M. (eds.#) *Robust Methods in Processing of Natural Language Dialogues*, Genève, 1-6.
- Delmonte R., Bristol A., Chiran L., Bacalu C., Tonelli S., (2004), "Parsing the Oral Corpus AVIP/API", in *Atti del Convegno Internazionale "Il Parlato Italiano"*, Napoli, Università di Napoli, 20-40.
- Delmonte R., Chiran L., Bacalu C. (2000), "Elementary Trees For Syntactic and Statistical Disambiguation", *TAG+5*, Paris, 237-240.
- Delmonte R., Dolci R. (1989), "Parsing Italian with a Context-Free Recognizer", *Annali di Ca' Foscari* XXVIII, 1-2, 123-161.
- Delmonte R., Pianta E. (1996), "IMMORTALE - Analizzatore Morfologico, Tagger e Lemmatizzatore per l'Italiano", in *Atti V Convegno AI*IA "Cibernetica e Machine Learning"*, Napoli, 19-22
- Delmonte R., Pianta E. (1998), "Immortal: How to Detect Misspelled from Unknown Words", in *BULAG*, PCUF, Besançon, 193-218.
- Delmonte R., Pianta E. (1999), "Tag Disambiguation in Italian", in *Proceedings Treebank Workshop ATALA*, Paris, 43-49.
- Montemagni S. *et al.* (2000), "The Italian Syntactic-Semantic Treebank: Architecture, Annotation, Tools and Evaluation", *LINC*, ACL, Luxembourg, 18-27.
- Rizzi L. (1982), *Issues in Italian Syntax*, Foris Publications, Dordrecht.
- Ronat M. (1982), Logical Form and Prosodic Islands, *Journal of Linguistic Research* 3, 33-48.

APPENDICE 1. *Le SFEATS*

Per permettere la annotazione in costituenti sintattici in maniera tale che ciascun costituente abbia una propria identità semantica chiara, e facilmente relazionabile alle annotazioni funzionali a teste introdotte nel livello di analisi superiore, è stato necessario

utilizzare un insieme di TAGs che permettessero una facile individuazione di relazioni semantiche importanti. La presenza di questi tags permette di diversificare in maniera utile anche le varie interpretazioni che una stessa parola può assumere nei vari contesti. La lista delle SFEATS viene presentata qui in basso con un commento per ogni tag.

abbr	= abbreviazione	in	= <i>intensifier</i>
ag	= aggettivo	ind	= quantificatore indefinito
agn	= agg. nominale predicativo	int	= pronome interrogativo
art	= articolo definito/indefinito	intj	= interiezione
auag	= ausiliare <i>avere</i> gerundio	n	= nome
auair	= aus. <i>avere</i> modo/tempo irreali	nc	= nome di colore
auæg	= ausiliare <i>essere</i> gerundio	neg	= negazione
aueir	= aus. <i>essere</i> modo/tempo irreali	nf	= nome fattivo
ausa	= ausiliare <i>avere</i> tempo definito	nh	= nome umano - nome propri
ausai	= ausiliare <i>avere</i> infinito	nt	= nome temporale
ause	= ausiliare <i>essere</i> tempo definito	np	= nomi propri geografici
ausei	= ausiliare <i>essere</i> infinito	npro	= nomi propri per organizzazioni, società e altro
ausep	= ausiliare <i>essere</i> participio passato	num	= numeri, numerali, quantificatori
avv	= avverbio	p	= preposizione
avvl	= locuzione avverbiale	par	= segno di parentetica: - () “
ccom	= congiunzione comparativa ³	part	= preposizione articolata
ccong	= congiunzione di coordinazione ⁴	partd	= preposizione articolata <i>di</i>
clit	= pronome clitico generico	partda	= preposizione articolata <i>da</i>
clitac	= pronome clitico accusativo	pd	= preposizione <i>di</i>
clitabl	= pronome clitico ablativo	pda	= preposizione <i>da</i>
clitdat	= pronome clitico dativo	php	= preposizione in locuzione
cong	= congiunzione	pk	= complementatore <i>che</i>
cong f	= congiunzione frasale	poss	= aggettivo e pronome possessivo
conjl	= congiunzione locuzione	ppas	= participio passato assoluto
cosu	= congiunzione subordinante	ppre	= participio presente assoluto
cosul	= cong. subordinante locuzione	pron	= pronome
da_riempire	= segno di punteggiatura per sospensione “...”	pt	= particella verbale (<i>a, ad, di, da</i>)
date	= numero di data	punt	= segno di virgola
deit	= pronome deittico	puntint	= segni di punteggiatura non dichiarativa (? !)
dim	= aggettivo dimostrativo	punto	= segno di punto e punto e virgola
dirs	= segno di punteggiatura (:)	q	= quantificatore
equal	= segno di uguale	qc	= quantificatore collettivo
escl	= pronome esclamativo	qd	= quantificatore distributivo
fw	= parola straniera	rel	= pronome relativo ⁵

³ A livello di costituente.

⁴ e/o di comparazione a livello frasale.

⁵ Include anche il *che*.

relin	= relativo indefinito	virin	= verbo intrans. modo/tempo irreale
relob	= relativo obliquo	virt	= verbo trans. modo/tempo irreale
relq	= relativo quantificatore	vit	= verbo transitivo infinito
sect	= lettera o numero indicatore di sezione	vppc	= verbo copulativo part. passato
slash	= segno "/"	vppin	= verbo intransitivo part. passato
vc	= verbo copulativo tempo definito	vppt	= verbo transitivo participio passato
vci	= verbo copulativo infinito	vprc	= verbo copulativo part. presente
vcir	= verbo copul. modo/tempo irreale	vprin	= verbo intransitivo part. presente
veg	= verbo copulativo gerundio	vprog	= verbo progressivo
vcl	= verbo con enclitico	vprogir	= verbo progressivo modo/tempo irreale
vgin	= verbo intransitivo gerundio	vprt	= verbo transitivo part. presente
vgprog	= verbo intrans. progressivo gerundio	vsf	= verbo <i>fare</i> a ricostruzione
vgs	= verbo modale gerundio	vsup	= verbo modale
vgsf	= verbo <i>fare</i> a ricostruzione gerundio	vsupir	= verbo modale modo/tempo irreale
vgt	= verbo transitivo gerundio	vt	= verbo transitivo tempo definito
viin	= verbo intransitivo infinito		
vin	= verbo intransitivo tempo definito		

APPENDICE 2: *Costituenti non discussi nell'articolo*

SP– Sintagma Preposizionale

I costituenti SP sono normalmente composti obbligatoriamente da una testa di tipo P (preposizione) o PART (ovvero preposizione articolata) e da un costituente SN che ne costituisce un elemento essenziale. Sono trattati come SP anche gli aggiunti infinitivali che quindi dopo la testa hanno un costituente SV2. Va ricordato che si escludono da questo caso i sintagmi preposizionali introdotti da **di** e **da**, che hanno una marca sintattica specializzata, come chiarito più in basso.

- () a. sp-[**part-alla**, sn-[n-stampa, sa-[ag-sportiva]]]
 b. sp-[**p-senza**, sv2-[viin-chiedere]]
 c. sp-[**p-dopo**, sv2-[vcl-averlo, clitac-lo, vppin-chiesto]]

SPDA– Sintagma Preposizionale “da”

Il costituente SPDA è costituito da una testa preposizionale di tipo specializzato, il *da* o le preposizioni articolate di *da*, e un SN. Questo costituente può servire da Agente Obliquo nelle frasi passive, oppure può individuare complementi o circostanziali di vario tipo, tra cui un locativo con l'indicazione di Origine o Provenienza.

- () spda-[**partda-dalla**, sn-[npro-Union_Carbide]]

SPD – Sintagma Preposizionale “di”

Il costituente SPD è costituito da una testa preposizionale di tipo specializzato, il DI o le preposizioni articolate di DI, e un SN. Questo costituente può servire da complemento verbale, da aggiunto o complemento nominale e da secondo termine di paragone in una struttura comparativa: in questo caso viene trattato come congiunzione di una relazione quantificata (vedi più in basso il SQ). Se invece si tratta di un relazione di comparativo superlativo, viene analizzato come pd.

- () a. spd-[**partd-delle**, sn-[n-società]]
b. spd-[**pd-di**, sn-[n-ricerca]]

SAVV – Sintagma Avverbiale

Questo costituente ha come testa un avverbiale anche in forma locutiva e può avere una struttura interna.

- () a. savv-[avv-**bene**]
b. savv-[avvl-**alla_spicciolata**]

SV2 – Sintagma Verbale 2 – Frasi Infinitive

Questo costituente ha come testa un verbo all’infinito, che può essere preceduto oppure no da una particella verbale, limitatamente alle seguenti preposizioni *a, di, da*. Tutte le altre preposizioni introducono un SP e contribuiscono all’interpretazione semantica della frase infinitiva che assume quindi valore di aggiunto. Nel caso invece rappresentato dalle particelle verbali, queste ultime vengono semplicemente assorbite dalla struttura infinitiva e non contribuiscono all’interpretazione semantica. In genere, la presenza della particella indica la funzione di complemento dell’infinitiva. Il costituente SV2 viene anche utilizzato per l’infinito passato che ha quindi un ausiliare seguito dal verbo lessicale al participio passato. Inoltre, si costruisce come SV2 anche l’infinito con clitico. Esso contiene i propri complementi e aggiunti, intendendo con aggiunti quelli che modificano il significato del verbo che costituisce la testa lessicale del SV2.

- () a. sv2-[pt-di, **vt-fare**, sn-[n-soldi], savv-[avv-rapidamente]]
b. sv2-[**vcl-averlo**, clitac-lo, vppin-chiesto]

SV3 – Sintagma Verbale 3 – Frasi Participiali

Questo costituente ha come testa il participio passato in funzione di aggiunto modificatore nominale o frasale. Al suo interno ci sono i complementi e gli aggiunti della testa lessicale. Non vengono inclusi eventuali avverbiali modificatori del participio passato che possono precederlo e la testa risulta quindi essere sempre il primo elemento del costituente - vedi "quasi fatto" = [savv-[in-quasi], sv3-[vppt-**fatto**]]

- () a. sv3-[vppt-**controllate**]
 b. sv3-[vppt-**rivolte**, sp-[part-all, sn-[n-Italia]], spda-[partda-dalla, sn-[n-commissione, sa-[ag-europea]]]]

SV5 – Sintagma Verbale 5 – Frasi Gerundive

Questo costituente ha come testa il gerundio in funzione di aggiunto modificatore frasale. Al suo interno ci sono i complementi e gli aggiunti della testa lessicale. Non vengono inclusi eventuali avverbiali modificatori del gerundio che possono precederlo e la testa risulta quindi essere sempre il primo elemento del costituente. Il costituente SV5 viene anche utilizzato per il gerundio passato che ha quindi un ausiliare seguito dal verbo lessicale al participio passato. Inoltre, si costruisce come SV5 anche il gerundio con clitico.

- () a. sv5-[**vgin-parlando**]
 b. sv5-[**auveg-essendo**, vppin-arrivato, savv-[avv-presto]]

COORD - Coordinazione

COORD/Nome di costituente serve per coordinare costituenti tra di loro e nel caso di SN ad assegnare numero plurale all'insieme creato dai due o più SN coordinati. Questo costituente comprende due tipi di teste: la congiunzione coordinate individuata da “cong-e” oppure un segno di punteggiatura, “punt,-”. Le strutture coordinate possono contenere liste e frasi ellittiche che nel Penn Treebank vengono indicate da costituenti appositi LST e FRAG. Abbiamo scelto di non introdurre marche specializzate per questo tipo di strutture in quanto la nostra rappresentazione non include le categorie vuote. Abbiamo invece trattato a parte le frasi coordinate, che potrebbero oppure no prevedere strutture ellittiche, con un costituente separato FC (vedi sotto).

coord-[sn-[art-I, n-superstiti], **cong-e**, sn-[art-i, n-parenti, spd-[partd-delle, sn-[n-vittime]]]]
 coord-[ibar-[vt-producono], **cong-e**, ibar-[vt-importano]], compt-[sn-[n-energia]]]]]]
 coord-[ibar-[ausa-ha, vppt-confermato], compt-[sn-[npro-Eugenio, npro-Scalfari]], **cong-ed**, ibar-[ausa-ha, vppt-deliberato], compt-[sv2-[pt-di, vit-aumentare], sn-[art-il, n-numero, spd-[partd-dei, sn-[n-consiglieri]]]]]]

FC – Frase Congiunta/Coordinata/Comparativa

Un sottocaso di coordinazione è costituito dalla coordinazione di frasi, una delle quali ellittica, che viene demarcata da un costituente esplicito, FC.

- () fc-[f-[sn-[art-Gli, n-esempi], ibar-[ause-sono, vppt-scritti], compt-[sp-[p-in, sn-[n-corsivo]]]], **punt-**, f-[sn-[art-le, n-glosse], sp-[p-in, sn-[n-stampatello]]]]

Inoltre il costituente FC include anche le seguenti tipologie:

- frasi coordinate dalla punteggiatura come il punto e virgola;
- frasi o costituenti ellittici coordinati da congiunzione comparativa che esemplifichiamo qui in basso:

- () a. sa-[sq-[q-dell, num-1/1%, in-più], ag-alti, fc-[ccong-rispetto_all, sn-[ag-analogo, nt-periodo, spd-[partd-del, sn-[num-1993]]]
b. fc-[ccong-come, f-[ibar-[ausa-ha, vppt-spiegato], compt-[savv-[avv-ieri], sn-[nh-mandela]]

COMP-T/C/IN Complemento Verbali

Questo costituente ha la funzione di individuare tutti i complementi verbali, di verbi transitivi, intransitivi e copulativi. Contiene tutti i costituenti che servono da complementi del verbo e sono retti lessicalmente; oltre ai complementi obbligatori, conterrà i complementi opzionali. Questo costituente contiene inoltre anche gli aggiunti del verbo, tutti quei circostanziali che servono ad individuare delle modalità specifiche dell'azione connotata dal significato del verbo reggente. Gli aggiunti spaziotemporali sono inclusi a livello più alto. Questo costituente conterrà anche il SN soggetto posposto di frasi passive o invertite per motivi di carattere informativo, pragmatico o semantico.

Con verbi intransitivi si intendono perlomeno i seguenti tipi sintattici: verbi inaccusativi, verbi inergativi, verbi atmosferici, verbi impersonali. Gli SP aggiunti spaziotemporali sono inclusi a livello più alto, altro che non costituiscano essi stessi dei complementi spaziotemporali. Questo costituente conterrà anche il SN soggetto posposto di frasi invertite per motivi di carattere informativo, pragmatico o semantico. Casi particolari sono costituiti da frasi con verbo ergativo, frasi con il *si* impersonale, frasi con verbi riflessivi inerenti.

Nel caso dei verbi copulativi, il Complemento contiene anche gli SP che costituiscano una predicazione di Locativi ripresi oppure no dal *ci*.

COMPT – Complemento Transitivo

- () a. ibar-[vt-promise], compt-[sp-[p-a, sn-[nh-Maria]], sv2-[pt-di, viin-partire]]
b. ibar-[ausa-ha, vppt-promesso], compt-[sv2-[pt-di, vit-accettare, sn-[art-il, n-lavoro]]]

COMPIN – Complemento Intransitivo

- () a. ibar-[vin-esce], compin-[sp-[part-allo, sn-[n-scoperto]]]
b. ibar-[vin-arrivai], compin-[sp-[p-in, sn-[n-ritardo]]]

COMPC – Complemento Copulativo

- () a. ibar-[clit-c, vc-era], compc-[sn-[nh-Maria], sp-[p-con, sn-[npro-Luigi]]]
b. ibar-[clit-'C', vc-è], compc-[sn-[art-una, n-vespa], sp-[part-nella, sn-[poss-mia, n-camera]]]