

**Publicato come** Delmonte R.(1997), Rappresentazioni lessicali e linguistica computazionale, Atti SLI, Lessico e Grammatica - Teorie Linguistiche e applicazioni lessicografiche, Roma, Bulzoni, pp.431-462.

# **RAPPRESENTAZIONI LESSICALI E LINGUISTICA COMPUTAZIONALE**

*Rodolfo Delmonte*

*Università degli Studi di Venezia "Ca'Foscari"*

*Laboratorio di Linguistica Computazionale*

## **1. Il lavoro lessicale all'Università di Venezia**

In questa relazione viene presentato il lavoro svolto in campo lessicale presso l'Istituto di Linguistica dell'Università di Venezia, sotto la direzione del Prof. Rodolfo Delmonte. Le persone che hanno collaborato e che ancora collaborano a questa attività sono parecchie, ma meritano di essere citate anche se in ordine sparso: Anna Cardinaletti, Giuliana Giusti, Paola Merlo, Roberto Dolci, Laura Brugè, Monica Dan, Giovanni Alberti, Marialuisa Furlan, Analisa Sanson, Andrea Cacco, Emanuele Pianta, Gerhard Bruegger, Francesco Stiffoni. A partire dall'inizio degli anni '80 sono stati finanziati lavori in ambito lessicale da varie istituzioni e enti con scopo diversi. Il risultato dello spoglio di corpora è stato travasato in un unico database su cui sono state compiute una serie nutrita di analisi allo scopo di creare gli strumenti utili alla ricerca in campo morfologico computazionale e di descrizione automatica fonologica e prosodica della lingua italiana.

Il risultato è un corpus di circa 500.000 occorrenze di testi provenienti da campi non omogenei, di dimensioni non bilanciate. Lo scopo del lavoro non è quello di fornire uno strumento di comparazione per definire l'italiano standard in un certo periodo temporale - gli anni '80, ma c'è anche un testo di Svevo! - ma semmai ricercare termini usati poco frequentemente e comunque non in disuso, nè arcaici. A questo scopo, tutte le forme risultate incomprensibili all'analisi automatica sono state controllate manualmente.

La lista di frequenza o IWL (Italian Word List) si compone di 30.000 forme molto ricca di hapax legomena, cioè di forme con frequenza uguale o inferiore a 3 occorrenze. Se si confronta la nostra lista con le altre disponibili, cioè quella del LIF, quella del VELI e quella del LIP, la nostra è molto più estesa. La nostra lista contiene tutte le forme incluse nella zona alta delle liste precedenti di frequenza, cioè con frequenza uguale o superiore a 4 occorrenze, e tutte le forme del LIP.

Inoltre, le attività più cospicue sono state indirizzate ad esplorare la possibilità di generare da una serie di descrizioni codificate secondo criteri di completezza linguistica, le rappresentazioni lessicali necessarie all'analisi sintattica, semantica e concettuale. In altre parole, la prospettiva di creazione di un lessico computazionale della lingua italiana che contenesse perlomeno il nucleo centrale dei lemmi più frequenti utili all'analisi e all'interpretazione dei significati letterali e non di un testo. La grammatica in una prospettiva lessicalista prevede che il lessico svolga una funzione autonoma di deposito di informazioni in grado di interagire con tutti i livelli di rappresentazione linguistica.

1.1 In particolare il lessico interagisce con il componente morfologico nella creazione di parole nuove e nella creazione delle forme flesse di una lingua: questo aspetto è stato studiato nella creazione di IMMORTALE, un programma che permette di analizzare e di generare le forme flesse della lingua italiana. L'ipotesi di partenza è quella che si indirizza al parlante maturo, che ha già creato un

dizionario mentale stabile della lingua e che possiede padronanza completa delle regole di formazione di parole e delle regole di flessione. In questa prospettiva, prima di utilizzare le regole di derivazione si tentano tutte le possibili scomposizioni flessionali e se queste falliscono - non danno nessuna analisi accettabile o permessa - si tenta una decomposizione utilizzando gli affissi derivazionali.

1.2 Il lessico interagisce strettamente con il componente sintattico e semantico della grammatica nella analisi e generazione di testi. Sia nella scelta dei complementi obbligatori o opzionali che nella verifica della compatibilità semantica di aggiunti a vari livelli strutturali, l'informazione lessicale è essenziale nel guidare il parser.

GETA\_RUN è un programma che esegue l'analisi di testi costruendo una rappresentazione semantica completa nella forma di Discourse Model, cioè di un database di fatti e siti relativi a eventi e entità del mondo del discorso o testo in analisi (Delmonte, 1992; 1995). La rappresentazione lessicale che utilizziamo è una forma di sottocategorizzazione estesa che per i verbi include i seguenti elementi:

i. categoria lessicale (nome, verbo ecc.)

ii. categoria sintattica (transitivo, inaccusativo, psicologico, ecc.)

iii. categoria aspettuale (attività, risultato, stativo, ecc.)

iv. categoria semantica (riportivo, soggettivo, fattivo, ecc.)

v. lista di sottocategorizzazione con argomenti obbligatori, contenente per ogni argomento le seguenti informazioni:

1. costituenza sintattica; 2. funzione grammaticale; 3. ruolo semantico; 4. schema di controllo per argomenti predicativi, oppure restrizioni selettive per argomenti non predicativi.

Per i verbi con più possibili sottocategorizzazioni, vi è una lista di liste, a parità di significato. Casi particolari sono le forme idiomatiche. Nomi e aggettivi possono avere una rappresentazione simile.

1.3 Infine il lessico interagisce con la rappresentazione fonologica nella creazione delle forme di parola flesse nel contribuire alla assegnazione dell'accento primario e delle prominenze interne alla parola.

I due database IWL (Italian Word List) e SIWL (Spoken Italian Word List) sono stati creati allo scopo di studiare le relazioni che intercorrono tra le informazioni morfologiche, lessicali, e quelle fonologiche. In particolare essi contengono la rappresentazione fonologica completa di ciascuna parola flessa del database IWL consistente in 30.000 forme, compresa la sua suddivisione sillabica. Un database apposito è stato creato per le parole omografe e non omofone che sono circa 1500 per le quali si ha doppia rappresentazione fonologica.

La Spoken Italian Word List è stata elaborata dal Laboratorio di Linguistica Computazionale dell'Università di Venezia per rendere disponibile uno strumento di ricerca in campo fonologico a livello di parola su cui compiere esperimenti. La lista di frequenza è stata elaborata sommando materiali provenienti da lavori svariati compiuti dal Laboratorio per Enti e istituzioni.

Nel creare SIWL siamo passati dalla categorizzazione alla lemmatizzazione, e abbiamo ottenuto una lista composta di parole flesse, e di parole invariabili, di lunghezza superiore o uguale a tre grafemi eliminando cioè parole funzionali monosillabiche atone. Inoltre, SIWL non contiene nomi propri né parole straniere. Il risultato è un totale di forme uniche che ammonta a 26828 forme, ma una lista estesa di 43534 forme lemmatizzate. Questo significa che alle forme uniche sono state aggiunte ulteriori 16706 forme con lemma diverso.

Se ci limitiamo alla categoria verbale, abbiamo 24217 forme lemmatizzate su 15432 forme uniche, di cui 13310 sono costituite da verbi privi di clitico, e 2521 verbi cliticizzati. I lemmi verbali presi singolarmente sono 3797. Il livello di ambiguità risulta quindi essere di 1,6, corrispondente al livello generale di ambiguità

calcolato sul totale delle forme lemmatizzate. Abbiamo poi isolato le forme che hanno due o più pronunce e abbiamo ottenuto i seguenti dati: 3123 forme lemmatizzate con più pronunce; 1652 forme che si pronunciano perlomeno in due modi diversi su 821 forme uniche. In questo caso, il livello di ambiguità è molto più alto, 3,8. La SIWL viene descritta nel dettaglio nel lavoro Rodolfo Delmonte, Francesco Stiffoni(1995)(in corso di pubblicazione).

Il Laboratorio di Linguistica Computazionale dell'Università di Venezia rende disponibile gratuitamente una versione delle proprie applicazioni e database linguistici a tutti quelli che fossero interessati: è sufficiente accedere attraverso ftp, al site dell'Università che si trova su vega.unive.it e aprire la directory/folder "linguistic"; oppure dare direttamente il comando ftp.unive.it.

## **2. L'analisi morfologica e sintattica computazionale dei testi**

Concentreremo la nostra attenzione sull'analisi morfosintattica e in particolare sul l'analizzatore morfologico chiamato IMMORTALE (Italian Morphological Tagger and Lemmatizer).

Se teoricamente sembra utile partire da temi, radici con vocale tematica, e di postulare poi l'esistenza di una regola di cancellazione di vocale, nella nostra applicazione, siamo invece partiti da radici o da teste complete di affissi derivazionali, così come è possibile evincerle da un dizionario cartaceo. Infatti il nostro dizionario di radici è stato creato sulla base di più dizionari con l'aggiunta di informazioni provenienti dai vari lessici di frequenza disponibili sull'italiano, il LIF, il VELI e il LIP. Inoltre, sono stati aggiunte le radici ricavate dalla nostra lista di frequenza chiamata IWL. Complessivamente sono circa 90.000 radici.

L'accesso al dizionario viene limitato alle sole radici, mentre per la generazione di tutti gli affissi il parser morfologico utilizza delle sottoliste interne che hanno una propria classificazione. Per i suffissi derivazionali l'informazione di categoria di ingresso e di uscita è in forma di lista, per generare le forme omografe. Per i suffissi valutativi o alterazioni più frequenti e per i suffissi di grado superlativo ugualmente frequenti si è preferito elencare le forme già flesse con l'indicazione di categoria, senza attivare meccanismi di tipo derivazionale. Infine i suffissi flessivi costituiti da un unico grafema, come ad es. "a", "e", "i", "o", sono utilizzabili indifferentemente per generare aggettivi, nomi e verbi: la loro interpretazione dipende quindi dalla radice. Ovviamente, in questo modo si possono produrre anche più interpretazioni per la stessa forma, cioè omografi.

Nel caso in cui la parola non venga analizzata con successo sulla base delle regole di flessione - con o senza le regole di segmentazione dei clitici - scattano prima le regole di prefissazione e quindi quelle di derivazione. Nel caso della prefissazione vengono elencati i prefissi possibili anche in forma composta - i prefissi composti sono quelli creati dalla concatenazione di più prefissi - e si cerca il prefisso più lungo. Questo permette di risparmiare sulle possibili ipotesi utilizzando solo i prefissi di base, in quanto non tutte le combinazioni tra prefissi sono possibili.

Se questa operazione fallisce, si tenta di separare tutti i possibili suffissi derivazionali e si assume come primo tentativo che la parte rimanente costituisca una radice possibile della lingua: questa viene poi ricercata nel lessico, partendo dal suffisso più lungo e se la ricerca fallisce si tenta con quella successiva. I suffissi derivazionali allo stato attuale sono accompagnati da una informazione relativa alla categoria di ingresso e da quella di uscita: la prima viene confrontata con la categoria lessicale della radice e la seconda serve come categoria di uscita della parola analizzata. Perché questo avvenga, deve essere confrontata con la categoria fornita indipendentemente dall'analisi del suffisso flettivo. La procedura è abbastanza interessante: dopo aver segmentato la parte relativa alle regole di flessione, poiché alcuni suffissi derivazionali possono essere contenuti

in altri, si procede cercando tutti i suffissi possibili e l'ultimo della lista risulterà essere quello più lungo. Il segmento che resta viene cercato nel dizionario, e questa procedura si esegue ricorsivamente: cioè, se non esiste una radice che è uguale al segmento in analisi si prende il suffisso più corto rispetto a quello precedente e disponibile nella lista dei suffissi derivazionali analizzati, e si restituisce una possibile radice che viene cercata nel dizionario. Nel momento in cui si trova una radice che è uguale al segmento in analisi, si confronta la categoria di ingresso del suffisso derivazionale con quella della radice e quindi si procede con l'ipotesi di parola a verificare se la flessione è compatibile con essa.

La parte più laboriosa dell'analisi morfologica è costituita dai verbi irregolari e difettivi dell'italiano che vengono generati internamente sfruttando al massimo le loro regolarità interne (per una discussione più dettagliata vedi Delmonte(1988)). In pratica, il dizionario registra solo il fatto che il verbo appartiene alla classe dei verbi irregolari demarcata dalla sigla -iv- che assieme all'indicazione della coniugazione è sufficiente al parser ad interpretare la forma flessa corrispondente. A sua volta la radice viene ricercata dopo aver tentato una decomposizione del prefisso eventuale e comunque dopo aver confrontato il segmento con una lista di segmenti o stems relativi alle radici dei verbi irregolari. Questa lista contiene indicazioni precise relative al tipo di paradigma irregolare di riferimento e al tipo di flessione irregolare o regolare da utilizzare per verificare l'analisi. La scelta che è stata operata a livello teorico è stata quella di considerare la forma residente nel lessico come forma primaria e di registrare tutte le altre forme suppletive direttamente nell'analizzatore morfologico come forme secondarie collegate alla forma primaria. La chiamata principale è quella seguente, a cui abbiamo unito una serie nutrita di esempi di classificazione.

irv(root, class, subClass, formNumber, primaryForm, coniug). ••

irv(evol, 1,		a,	1,	evol, 2).
irv(evolut, 1,		a,	2,	evol, 2).
irv(accend, 1,		b,	1,	accend, 2).
irv(acces, 1,		b,	2,	accend, 2).
irv(cad, 1,	1,	c,	1,	cad, 2).
irv(cadd, 1,		c,	2,	cad, 2).
irv(cogli, 3,		a,	1,	cogli, 2).
irv(colg, 3,		a,	2,	cogli, 2).
irv(cols, 3,		a,	3,	cogli, 2).
irv(colt, 3,		a,	4,	cogli, 2).
irv(dic, 3,	3,	b,	1,	d, 3).
irv(d, 3,		b,	2,	d, 3).
irv(diss, 3,		b,	3,	d, 3).
irv(dett, 3,		b,	4,	d, 3).
irv(ess, 7,		a,	1,	ess, 2).
irv(son, 7,		a,	2,	ess, 2).
irv(se, 7,		a,	3,	ess, 2).
irv(si, 7,		a,	4,	ess, 2).
irv(er, 7,		a,	6,	ess, 2).
irv(fu, 7,		a,	7,	ess, 2).
irv(fo, 7,		a,	8,	ess, 2).
irv(stat, 7,		a,	9,	ess, 2).
irv(sa, 7,	7,	a,	10,	ess, 2).
irv(è, 7,		a,	11,	ess, 2).

Vi è una differenziazione fondamentale tra le forme suppletive di una radice di verbo irregolare: la forma primaria, e quelle secondarie. La forma primaria, quella demarcata dal numero 1, appunto, è quella più produttiva e più regolare, mentre le forme secondarie sono quelle più idiosincratice. La forma primaria corrisponde di norma con la radice utilizzata per produrre l'infinito e permette di solito di generare l'imperfetto indicativo e congiuntivo.

Le forme primarie sono composte dall'indicazione di Classe, Sottoclasse, Desinenza, Modo, Tempo, Persona, Numero. Includiamo qui di seguito la chiamata principale per le forme primarie e un esempio di descrizione relativa alla classe 6, sottoclasse 'a', che nel nostro esempio si riferisce al verbo "fare".

pfd(Cl, SubCl, Des, Mode, Tempo, Pers/Gen, Num).

pfd(6, a, arò, indic, fut, 1, s).  
 pfd(6, a, arai, indic, fut, 2, s).  
 pfd(6, a, arà, indic, fut, 3, s).  
 pfd(6, a, aremo, indic, fut, 1, p).  
 pfd(6, a, arete, indic, fut, 2, p).  
 pfd(6, a, aranno, indic, fut, 3, p).

Qui di seguito diamo la chiamata principale per le forme secondarie e alcuni esempi di descrizione flessionale

fd(Cl, SubCl, FormNum, Des, Modo, Tempo, Pers/Gen, Num).

fd(1, a, 2, a, part, pass, f, s).  
 fd(1, a, 2, e, part, pass, f, p).  
 fd(1, a, 2, i, part, pass, m, p).  
 fd(1, a, 2, o, part, pass, m, s).  
 fd(1, b, 2, a, part, pass, f, s).  
 fd(1, b, 2, e, indic, pass, 3, s).  
 fd(1, b, 2, e, part, pass, f, p).  
 fd(1, b, 2, ero, indic, pass, 3, p).  
 fd(1, b, 2, i, indic, pass, 1, s).  
 fd(1, b, 2, i, part, pass, m, p).  
 fd(1, b, 2, o, part, pass, m, s).

La parte più interessante del lavoro morfologico riguarda la generazione ed analisi dei verbi cliticizzati. Uno degli strumenti formali più potenti formulati nella letteratura linguistica computazionale in campo morfologico negli ultimi dieci anni, è rappresentato dalle two-level rules, cioè regole a due livelli che indirizzano contesti di applicabilità a livello lessicale e a livello superficiale contemporaneamente. Queste regole in genere sostituiscono un segmento con un'altro, inseriscono oppure eliminano un certo segmento in uno dei due livelli. Nel caso di parole del tipo di "fatti" in italiano sono necessarie regole più potenti, che siano in grado contemporaneamente di eliminare e di aggiungere segmenti a più livelli: in particolare, si deve operare sull'insieme limitato di clitici che permettono il raddoppiamento fonotattico, nel contesto di flessioni verbali che vanno reinterpretate - da flessione della terza persona singolare dell'indicativo presente, a quella apocopata dell'imperativo che viene rappresentata con la presenza del corrispondente segno grafico di apostrofo. Inoltre, è necessario verificare che la radice verbale sia monosillabica. Riportiamo qui di seguito l'analisi eseguita da IMMORTALE per la parola "fatti" con i tempi di elaborazione.

fatti 101/60 sec.

[cat=verbo, pred=f+are, scat=intr, modo=part, tempo=pass, gen=m, num=p]

[cat=verbo, pred=f+are, scat=intr:imp, modo=part, tempo=pass, gen=m, num=p]

[cat=verbo, pred=f+are, scat=intr:pron, modo=part, tempo=pass, gen=m, num=p]

[cat=verbo, pred=f+are, scat=rifl, modo=part, tempo=pass, gen=m, num=p]

[cat=verbo, pred=f+are, scat=tr, modo=part, tempo=pass, gen=m, num=p]

[cat=ppas, pred=fatt+o, gen=m, num=p]

[cat=nome, type=com, pred=fatt+o, gen=m, num=p]

[cat=verbo, pred=f+are, scat=intr:pron, encl=ti, case=dat, vf=fa', modo=imp, tempo=pres, pers=2, num=s]

[cat=verbo, pred=f+are, scat=rifl, encl=ti, case=dat, vf=fa', modo=imp, tempo=pres, pers=2, num=s]

[cat=verbo, pred=f+are, scat=tr, encl=ti, case=dat, vf=fa', modo=imp, tempo=pres, pers=2, num=s]

### **3. Il Temario**

L'analizzatore Immortale utilizza un temario di 90.000 radici in cui sono presenti le seguenti categorie principali:

1. AGG            aggettivo
2. AGGPP        aggettivo participio passato
3. AGGPR        aggettivo participio presente
4. AN            aggettivo+nome attributivo/predicativo
5. ART           articolo
6. AVV           avverbio
7. CONG         congiunzione
8. COSU         congiunzione subordinativa
9. CONGF        congiunzione coordinativa frasale
10. EL           elemento
11. INTER       interiezione
12. LOC         locuzione (avverbio, congiunzione, preposizione)
13. N            nome
14. NA           nome+aggettivo predicativo
15. NAPR        nome+aggettivo participio presente
16. PRE         prefisso
17. PREP        preposizione
18. PRON        pronome
19. SUFF        suffisso
20. V            verbo

L'insieme di categorie che il nostro dizionario possiede ha come caratteristica peculiare quella di utilizzare una etichetta per parole che svolgono più funzioni contemporaneamente, specializzandone l'uso in senso linguistico.

Il funzionamento dell'analizzatore morfosintattico IMMORTALE è tutto fondato sulla buona classificazione contenuta nel Temario che è composta da una indicazione di categoria, una indicazione di classe e una descrizione di tratti se necessario. In genere, la indicazione di classe permette di generalizzare sul comportamento della flessione, ovviamente in concomitanza con l'informazione derivante dalla categoria. Ad esempio, gli aggettivi di classe "e" avranno solo il plurale in "i" mentre quelli di classe "o" avranno anche la possibilità di variare in base al genere. Le restrizioni sulle possibili flessioni associate a ciascuna categoria e classe sono espresse nell'analizzatore e costituiscono le regole di buona formazione di parola, che permettono ad IMMORTALE di stabilire con esattezza quale lemma associare ad una particolare forma e quindi di costruire una lemmatizzazione automatica.

<b>1. AGG</b>	<b>aggettivo</b>	
agg:a	adj classe -a	focaia, cotogna
agg:alt	adj "other"	altro
agg:co	adj classe -co	antico
agg:c	adj colore	giallo
agg:comp	adj comparativo	miglior

agg:dim	adj dimostrativo		quello
agg:e	adj classe -e	temibile	
agg:escl	adj esclamativo		quale
agg:go	adj classe -go		sacrilego
agg:ico	adj classe -ico		sarcastico
agg:ind	adj indef		troppo
agg:ind:min	adj indef minoranza	pochino	
agg:ind:sup	adj indef superlativo	pochissimo	
agg:int	adj interrogativo		quanto
agg:io	adj classe -io	refrattario	
agg:num:car	adj numerale cardinale	dodici	
agg:num:ord	adj numerale ordinale	ventesimo	
agg:num:ord:ind	adj numerale ordinale indef	ennesimo	
agg:o	adj classe -o	bello	
agg:ore	adj classe -ore		fotorecettore
agg:poss	adj possessivo		mio
agg:rel	adj relativo	quale	
agg:sup	adj superlativo		ottimo
agg:z	adj invar	video	
agg:z:def	adj def quantificatore	tutto	
agg:z:min	adj minoranza	minor	
agg:z:neg	adj negativo	nessun	
agg:z:qd	adj distrib quantificatore	ogni	
agg:z:rel:ind	adj relativo indef	qualsiasi	

Le classi in senso tradizionale dell'aggettivo sono quindi le seguenti: -O, -ORE, -IO, -E, -GO, -CO, -ICO, -A. Otto classi, molte di più di quelle normalmente trattate nei manuali di grammatica, solitamente due. Gli aggettivi e nomi in -ORE fanno regolarmente il femminile in -RICE e quindi i plurali in -I.

## 2. AGGPP *participio passato aggettivo*

aggpp:o	classe -o	moderato
---------	-----------	----------

## 3. AGGPR *participio presente aggettivo*

aggpr:e	classe -e	mortificante
---------	-----------	--------------

## 4. AN *aggettivo+nome attributivo/predicativo*

an:comp	adj+nome maggioranza	maggiore
an:e	adj+nome classe -e	folle
an:ico	adj+nome classe -ico	generico
an:ind	adj+nome indef	diverso
an:io	adj+nome classe -io	medio
an:min	adj+nome minoranza	minore
an:num:car	adj+nome numerale cardinale	novanta
an:num:ord	adj+nome numerale ordinale	novantesimo
an:o	adj+nome classe -o	novello
an:sup	adj+nome superlativo	massimo
an:sup:min	adj+nome superlativo minoranzaminimo	
an:sup:ord	adj+nome superlativo ordinale	ultimo
an:sup:t	adj+nome superlativo temporale	postumo
an:z	adj+nome invar	pop

E' interessante notare che le classi in questo caso siano un sottoinsieme di quelle nominali e di quelle aggettivale, ma globalmente molto inferiori di numero.

## 5. ART *articolo*

art:def	articolo def	il
art:ind	articolo indef	una

## 6. AVV *avverbio (modificatore del significato del verbo)*

Le parole invariabili come l'avverbio o la congiunzione portano solo informazione semantica:

avv:l	avverbio locativo	qua
avv:mn	avverbio maniera	piano
avv:neg	avverbio negativo	affatto
avv:q	avverbio quantificato	poco
avv:r	avverbio relazionale	simile
avv:mn:min	avverbio maniera minoranza	adagino
avv:mn:sup	avverbio maniera superlativo	adagissimo
avv:t	avverbio temporale	presto
avv:t:min	avverbio temporale minoranza	prestino
avv:t:sup	avverbio temporale superlativo	prestissimo
avv:td	avverbio temporale deittivo	stanotte
avv:ti	avverbio temporale iterativo	spesso

**7. CONG congiunzione (coordina due sintagmi o frasi)**

cong:av	conj avversativa	bensi
cong:caus	conj causale	perché
cong:comp	conj comparativa	come
cong:conc	conj concessiva	comunque
cong:dis	conj disgiuntiva	o
cong:ecc	conj eccettuativa	fuorché
cong:exp	conj esplicativa	ossia
cong:l	conj locativa	dove
cong:neg	conj negativa	nemmeno
cong:ott	conj ottativa	almeno
cong:par	conj parallela	mentre
cong:pk	conj k-	che
cong:pri	conj privativa	senza
cong:qt	conj quantificata temporale	ogniqualevolta
cong:sum	conj sommativa	oltreché
cong:t	conj temporale	finché
cong:tdur	conj temporale durativa	sinché
cong:ts	conj temporale successione	dopodiché
cong:tt	conj temporale telica	appena

**8. CONGF congiunzione frasale**

congf:av	conj avversativa	viceversa
congf:caus	conj causale	stanteché
congf:conc	conj concessiva	tuttavia
congf:cons	conj consecutiva	così
congf:ecc	conj eccettuativa	tranneché
congf:exp	conj esplicativa	infatti
congf:ott	conj ottativa	magari
congf:par	conj parallela	intanto
congf:pres	conj presentativa	ecco
congf:ra	conj finale	perciò
congf:sum	conj sommativa	inoltre
congf:t	conj temporale	frattanto
congf:tp	conj temporale precedenza	prima
congf:ts	conj temporale successione	poi

**9. COSU congiunzione subordinante**

cosu:av	conj subord avversativa	anziché
cosu:caus	conj subord causale	perché
cosu:conc	conj subord concessiva	sebbene
cosu:cond	conj subord condizionale	se

cosu:par	conj subord parallela	allorquando
cosu:ra	conj subord finale	poiché
<b>10. EL</b>	<b>elemento</b>	
el:1	elemento primo	cloro
el:2	elemento secondo	algia
<b>11. INTER</b>	<b>interiezione (può essere usata per costruire una ellissi)</b>	
inter	interiezione	diamine
<b>12. LOC</b>	<b>locuzione</b>	
LOC AVV	locuzione avverbio	
loc:avv	locuzione avverbio	inintermediari
LOC AVV PREP	locuzione avverbiale con una data preposizione	
loc:avv:P	locuzione avverbio	insaputa (all)
LOC PREP	locuzione preposizione	
loc:prep	locuzione preposizione	attorno
<b>LOC CONG</b>	<b>locuzione congiunzione</b>	
loc:cong:av	locuzione conj avversativa	piuttosto
loc:cong:caus	locuzione conj causale	supposto
loc:cong:cond	locuzione conj condizionale	patto
loc:cong:cons	locuzione conj consecutiva	tale
loc:cong:ecc	locuzione conj eccettuativa	tranne
loc:cong:par	locuzione conj parallela	intanto
loc:cong:pres	locuzione conj presentativa	ecco
loc:cong:pri	locuzione conj privativa	senza
loc:cong:sum	locuzione conj summativa	pur
loc:cong:t	locuzione conj temporale	fintanto
loc:cong:tdur	locuzione conj temporale durativa	sintanto
loc:cong:tp	locuzione conj temporale precedenza	prima
loc:cong:ts	locuzione conj temporale successione	dopo
loc:cosu:pre	locuzione conj subord preconditione	ammesso

### **13. N nome**

Le classi nominali incluse nel nostro temario sono notevolmente superiori a quelle previste nei manuali tradizionali; in particolare è importante notare che abbiamo duplicazioni dovute alla variazione di genere. Un nome di classe -A con genere FEM, farà il plurale in -E, mentre un nome della stessa classe con genere MASC farà il plurale in -I. Globalmente abbiamo le seguenti 18 classi: -A2, -A FEM, -A MASC, -A MASC FEM, -CA FEM, -CA MASC, -CO, -E FEM, -E MASC, -E MASC FEM, -GA FEM, -GA MASC FEM, -GO, -IA, -IE, -IO, -O, -ORE. Prese singolarmente abbiamo -A2, -A, -CA, -CO, -E, -GA, -GO, -IA, -IE, -IO, -O, -ORE, cioè 12 classi, le stesse riportate nei manuali(vedi Serianni, 1988).

n:a2:f	nome fem classe -a2	ala
n:a:f	nome fem classe -a	alabarda
n:a:m	nome masc classe -a	automa
n:a:m:f	nome masc fem classe -a	borsista
n:ca:f	nome fem classe -ca	amica
n:ca:m	nome masc classe -ca	arciduca
n:co:m	nome masc classe -co	mammalucco
n:e:f	nome fem classe -e	bronchite
n:e:m	nome masc classe -e	clone
n:e:m:f	nome masc fem classe -e	commediante
n:ga:f	nome fem classe -ga	carlinga
n:ga:m:f	nome masc fem classe -ga	collega
n:go:m	nome masc classe -go	castigo
n:ia:f	nome fem classe -ia	coscia
n:ie:f	nome fem classe -ie	moglie

n:io:m	nome masc classe -io	monopolio
n:o:m	nome masc classe -o	monoplano
n:ore:m	nome masc classe -ore	motocoltivatore

Riportiamo a parte i nomi invariabili di cui notiamo una classe semantica e il genere e numero se necessario. La notazione del numero serve a restringere l'uso del nome all'interno di un sintagma nominale che contenga quel numero particolare. La mancanza di questa indicazione significa che il nome invariabile si potrà usare sia in sintagmi al singolare che al plurale. La classe "z" è riservata per i nomi comuni.

n:c:f	nome fem classe c(olore)	ambra
n:c:m	nome masc classe c(olore)	arancio
n:geo:f	nome fem classe geo(grafico)	carrara
n:geo:f:pl	nome fem plurale classe geo(grafico)	alpi
n:geo:f:sg	nome fem plurale classe geo(grafico)	murgia
n:geo:m	nome masc classe geo(grafico)	nord
n:p:f	nome fem classe p(ropio)	raffaella
n:p:m	nome masc classe p(ropio)	poseidone
n:p:m:pl	nome masc plurale classe p(ropio)	dioscuri
n:ab:f	nome fem classe abbreviazione	agip
n:ab:m	nome masc classe abbreviazione	centotredici
n:tg:f:pl	nome fem plurale classe t(empo)-giorno	ceneri
n:tg:f:sg	nome fem singol classe t(empo)-giorno	pasqua
n:tg:m	nome masc classe t(empo)-giorno	capodanno
n:tg:m:pl	nome masc plurale classe t(empo)-giorno	sabati
n:tg:m:sg	nome masc singol classe t(empo)-giorno	ferragosto
n:tm:m	nome masc classe t(empo)-mese	dicembre
n:ts:f:pl	nome fem plurale classe t(empo)-stagione	autunni
n:ts:f:sg	nome fem singol classe t(empo)-stagione	autunno
n:ts:m:pl	nome masc plurale classe t(empo)-stagione	estati
n:ts:m:sg	nome masc singol classe t(empo)-stagione	estate
n:z:f	nome fem invar	estraneità
n:z:f:pl	nome fem plurale invar	fauci
n:z:f:sg	nome fem singolare invar	atrazina
n:z:m	nome masc invar	iogurt, iter
n:z:m:f	nome masc fem invar	groviera
n:z:m:f:pl	nome masc fem plurale invar	idi
n:z:m:f:sg	nome masc fem singol invar	capobanda
n:z:m:pl	nome masc plurale invar	convenevoli
n:z:m:sg	nome masc singol invar	fiele

E' in corso un lavoro di classificazione più dettagliata che riportiamo qui di seguito e che per ora copre i primi 30 mila lemmi nominali del nostro dizionario AZ. Si tratta di classi concettuali generali che includono diverse sottoclassi, per le quali supponiamo di intervenire successivamente:

1. Esseri viventi - Organismi

Persone - Esseri umani Es.: bambino, ragazzo, sorella, madre, ecc.; Corpo Es.: mano, cistifellea, flessore, ala, ecc.; Animali - Fauna; Piante - Flora

2. Esseri non viventi - Oggetti; Oggetti naturali; Sostanze; Artefatti; Cibo Es.: frutti

3. Luoghi - Tempo

Esempi di "luoghi": casolare, agrumeto, aeroporto, anticamera

4. Eventi - Stati

Fenomeni naturali Es.: vento, brina; Atti, Azioni, Attività; Attributi, Proprietà, Ruoli Es.: avvocato, nunzio, fabbro; Eventi, Accadimenti; Sentimenti, Emozioni;

Processi, Risultati di processi Es.: componimenti letterari o musicali; Stati, condizioni Es.: malattie; Comunicazioni

#### 5. Astratti

Motivi; Quantità, Ammontare; Forma; Relazioni, Rapporti; Possesso, Mezzi di scambio Es.: monete, danaro, ecc.; Conoscenze, discipline Es.: docimologia, divisionismo (movimento pittorico), ; docetismo (dottrina eretica), ecc.; Titoli (militari, cavallereschi, ecc.), uffici e cariche Es.: censorato, centena; Lingue e linguistica Es.: causale (prop. subordinata); Note musicali; Unità di misura; Concetti matematici e geometrici Es.: cicloide, angolo, ascissa, cilindro, ecc.; Forme retoriche e stilistiche Es.: anafora, distico, ecc.;

#### 6. Gruppi - Collezioni; Istituzioni

Altri esempi: azienda, impresa, copisteria, cooperativa, ; insiemi di persone (Es.: brigata) od oggetti (Es.: bottame, bestiame)

#### 7. Parole straniere

Non sono compresi: 1) le parole entrate nell'uso comune. 2) gli adattamenti di parole straniere: (Es.: karacul = caracul = tipo di pecora).

#### 8. Voci idiomatiche e metaforiche

Es.: "cappotto" nella locuzione "fare cappotto"

#### 9. Colori - Entità intensionali

Es.: Babbo Natale, ciclope, angelo, chimera, ecc.

L. Leggi - Giurisdizione Es.: concetti giuridici, documenti, ecc.

Vengono inoltre attribuiti dei sottocodici che hanno i seguenti valori:

“.A” - nomi alterati (Es.: accrescitivi, diminutivi, vezzeggiativi);

“.I” - iterazioni;

“.N” - negazioni.

Quando un termine ha più valori, questi vengono tutti indicati.

Es.: "charter" a) noleggio a tempo di una nave (4); b) aereo non di linea (2)

### 14. NA

#### *nome+aggettivo predicativo*

na:a:f	nome+adj fem classe -a	femmina
na:a:m	nome+adj masc classe -a	erbicida
na:a:m:f	nome+adj masc/fem classe -a	estremista
na:ca:f	nome+adj fem classe -ca	scarica
na:co:m	nome+adj masc classe -co	sciocco
na:e:f	nome+adj fem classe -e	sciupone
na:e:m	nome+adj masc classe -e	plantare
na:e:m:f	nome+adj masc fem	possidente
na:ga:f	nome+adj fem classe -ga	centrifuga
na:ga:m:f	nome+adj masc fem classe -ga	belga
na:go:m	nome+adj masc classe -go	coprofago
na:go:m:sg	nome+adj masc only/sing class/go	gallego
na:ia:f	nome+adj fem classe -ia	greggia
na:io:m	nome+adj masc classe -io	grigio
na:io:m:sg	nome+adj masc solo/sing class/io	romancio
na:o:m	nome+adj masc classe -o	guasto
na:o:m:sg	nome+adj masc solo/sing class/o	ridicolo
na:ore	nome+adj classe -ore	roditore
na:z:m	nome+adj masc invar	standard
na:z:m:f	nome+adj masc fem invar	snob

### 15. NAPR

#### *nome+aggettivo participio presente*

napr:e:f	nome+adj fem classe -e	stimolante
napr:e:m	nome+adj masc classe -e	spiovente
napr:e:m:f	nome+adj masc fem classe -e	richiedente

### 16. PRE

#### *prefisso*

pre	prefisso	ri	
pre:d	prefisso raddoppiamento	fra	
pre:n	prefisso nome		arci
pre:v	prefisso verbo		oltre

**17. PREP preposizione**

prep	preposizione	di	
prep:art	preposizione articolo		dal

A differenza di quanto avviene in progetti europei quali MULTEXT, la classificazione delle parole funzionali è legata al ruolo linguistico che il tag assegnato assume all'interno di un parser morfologico, sintattico, o semantico. L'etichetta riflette l'analisi linguistica contestuale che ad essa si associa in ambiti teorici moderni, che si differenziano da ambiti tradizionali presenti tuttora in manuali scolastici. Riportiamo qui di seguito la classificazione riportata in MULTEXT (Specifications and Notation for Lexicon Encoding from MULTEXT LRE 62-050), per alcune classi di parole funzionali,

**PRONOME**

Tipo [personale/p, dimostrativo/d, indefinito/i, possessivo/s, interrogativo/t relativo/r, esclamativo/e, riflessivo/x, reciproco/l]; Persona [1, 2, 3]; Genere [masc/m, fem/f, neutro/n]; Numero [sing/s, plur/p]; Case [nominativo/n, genitivo/g, dativo/d, accusativo/a, obliquo/o, oggetto/j]; Possessore [singul/s, plural/p]

**DETERMINATORE**

Tipo [dimostrativo/d, indefinito/i, possessivo/s, interrogativo/i]; Persona [1, 2, 3]; Genere [masc/m, fem/f, neutro/n]; Numero [sing/s, plur/p]; Caso [nominativo/n, genitivo/g, dativo/d, accusativo/a, obliquo/o]; Possessor [sing/s, plur/p]

**ARTICOLO**

Tipo [definite/d, indefinite/i]; Genere [masc/m, fem/f, neutro/n]; Numero [[sing/s, plur/p]; Caso [nominativo/n, genitivo/g, dativo/d, accusativo/a]

**AVVERBIO**

Tipo [generale/g, particella/p]; Grado [positivo/p, comparativo/c, superlativo/s]

**ADPOSIZIONE**

Tipo [preposizione/p, postposizione/t, circumposizione/c]; Formazione [semplice/s, clitico/c]

**CONGIUNZIONE**

Tipo [coordinante/c, subordinante/s]

**NUMERALE**

Tipo [cardinale/c, ordinale/o]; Genere [masc/m, fem/f, neutro/n]; Numero [sing/s, plur/p]; Caso [nominativo/n, genitivo/g, dativo/d, accusativo/a]

Nel nostro caso, le parole funzionali sono viste direttamente dall'analizzatore IMMORTALE e in particolare i clitici ricevono il caso direttamente durante l'analisi in quanto sono in buona parte ambigui.

**18. PRON pronome**

pron:an	pron anaforico	stesso	
pron:alt	pron "other"	altro	
pron:cl	pron clitico	li	
pron:dim	pron dimostrativo	quegli	
pron:escl	pron esclamativo	quanto	
pron:ind	pron indef	troppo	
pron:int	pron interrogativo	quale	
pron:loc:cong	pron locuzione congiunzione	cui	
pron:neg	pron negazione	nessuno	
pron:pers	pron personale	io	
pron:poss	pron possessivo	tuo	

pron:qd	pron quantificatore distributive	ciascuno
pron:qp	pron quantificatore partitive	molto
pron:rel	pron relativo	quale
pron:z:an	pron invar anaforico	sé
pron:z:def	pron invar def	tutto
pron:z:ind	pron invar indef	chicchessia
pron:z:int	pron invar interrogativo	chi
pron:z:escl	pron invar esclamativo	che
pron:z:min	pron invar minoranza	minore
pron:z:num	pron invar numerale	entrambi
pron:z:qd	pron quantificatore distributivo	ognuno
pron:z:qp	pron quantificatore partitivo	molto
pron:z:rel:ind	pron relativo indef	qualunque

### 19. SUFF *suffisso*

suff:a	suffisso adj	oica/o/che/ci
suff:a:a	suffisso adj adj	igna/e/o/i
suff:a:n	suffisso adj nome	ismi/o
suff:ind	suffisso indef	unque
suff:n	suffisso nome	ene
suff:n:a	suffisso nome adj	ico/i/ica/iche
suff:n:avv	suffisso nome avverbio	oni
suff:n:n	suffisso nome nome	onzoli/o
suff:n:v	suffisso nome verbo	eggiare
suff:num	suffisso numerale	pla/ple
suff:v:a	suffisso verbo adj	ibile/ibili
suff:v:n	suffisso verbo nome	ibilità
suff:v:v	suffisso verbo verbo	icchiare

### 20. V *verbo*

v:1:cop	verbo copulativo 1.	sembrare
v:1:intr	verbo intrans. 1.	serpeggiare
v:1:intr:imp	verbo intrans. impersonale 1.	sgelare
v:1:intr:pron	verbo intrans. pronominale 1.	sgolare
v:1:rifl	verbo riflessivo 1.	slacciare
v:1:rifl:rec	verbo riflessivo reciproco 1.	somigliare
v:1:tr	verbo trans. 1.	somministrare
v:1:tr:erg	verbo trans. ergativo 1.	ammosciare
v:2:intr	verbo intrans. 2.	ripetere
v:2:intr:imp	verbo intrans. impersonale 2.	calere
v:2:intr:pron	verbo intrans. pronominale 2.	compiere
v:2:rifl	verbo riflessivo 2.	credere
v:2:rifl:rec	verbo riflessivo reciproco 2.	sfottere
v:2:tr	verbo trans. 2.	spremere
v:2:tr:erg	verbo trans. ergativo 2.	scotere
v:3:intr	verbo intrans. 3.	applaudire
v:3:intr:pron	verbo intrans. pronominale 3.	scucire
v:3:rifl	verbo riflessivo 3.	sentire
v:3:rifl:rec	verbo riflessivo reciproco 3.	susseguire
v:3:tr	verbo trans. 3.	vestire
v:3:tr:erg	verbo trans. ergativo 3.	adempire
v:a3:intr	verbo intrans. 3. incoativo	avvizzare
v:a3:intr:imp	verbo intrans. impers. 3. incoat.	imbrunire
v:a3:intr:pron	verbo intrans. pronom. 3. incoat.	avvilire
v:a3:rifl	verbo riflessivo 3. incoat.	istruire
v:a3:rifl:rec	verbo riflessivo recipr. 3. incoativo	riunire

v:a3:tr	verbo trans. 3. incoativo	riverire
v:a3:tr:erg	verbo trans. ergativo 3. incoat.	aggrinzire
v:i-v:1:cop	verbo copulative 1. irreg.	stare
v:i-v:1:intr	verbo intrans. 1. irreg.	andare
v:i-v:1:intr:imp	verbo intrans. impers. 1. irreg.	dare
v:i-v:1:intr:pron	verbo intrans. pronom. 1. irreg.	liquefare
v:i-v:1:rifl	verbo riflessivo 1. irreg.	disfare
v:i-v:1:tr	verbo trans. 1. irreg.	contraffare
v:i-v:2:aux	verbo ausiliare 2. irreg.	essere
v:i-v:2:cop	verbo copulativo 2. irreg.	corrispondere
v:i-v:2:intr	verbo intrans. 2. irreg.	compiacere
v:i-v:2:intr:imp	verbo intrans. impers. 2. irreg.	parere
v:i-v:2:intr:pron	verbo intrans. pronom. 2. irreg.	perdere
v:i-v:2:rifl	verbo riflessivo 2. irreg.	ritrarre
v:i-v:2:rifl:rec	verbo riflessivo recipr 2. irreg.	reggere
v:i-v:2:tr	verbo trans. 2. irreg.	rendere
v:i-v:3:cop	verbo copulativo 2. irreg.	apparire
v:i-v:3:intr	verbo intrans. 2. irreg.	aprire
v:i-v:3:intr:imp	verbo intrans. impers 2. irreg.	convenire
v:i-v:3:intr:pron	verbo intrans. pronom. 2. irreg.	imbiondire
v:i-v:3:rifl	verbo riflessivo 3. irreg.	rioffrire
v:i-v:3:rifl:rec	verbo riflessivo reciproco 2. irreg.	contraddire
v:i-v:3:tr	verbo trans. 2. irreg.	dire
v:i-v:a3:intr	verbo intrans. 3. incoat.. irreg.	impallidire
v:i-v:a3:intr:pron	verbo intr pronom. 3. incoat. irreg.	inacidire
v:i-v:a3:rifl	verbo riflessivo 3. incoat. irreg.	profferire
v:i-v:a3:tr	verbo trans. 3. incoat.. irreg.	inaridire

Anche in questo le classi sono un sottoinsieme di quelle che abbiamo elaborato a parte per il nostro dizionario AZ. Si tratta di 27 classi che uniscono informazioni sintattiche e semantiche di cui parliamo in modo più dettagliato in Delmonte, 1995. I verbi che abbiamo classificato sono complessivamente circa 13.000.

tr	transitivo	[6700]	
tr_cop	transitivo+predicazione secondaria		[112]
tr_perc	transitivo_percettivo		[24]
ditr(+preps)	ditransitivo	[386]	
psych1	psicologico 1		[59]
psych2	psicologico 2		[251]
psych3	psicologico 3		[19]
inac	inaccusativo	[764]	
inerg	inergativo	[1586]	
inac/inerg	inaccusativo/inergativo		[161]
inerg/inac	inergativo/inaccusativo		[15]
[inac,inerg]	inaccusativo e inergativo		[10]
[inerg,inac]	inergativo e inaccusativo		[11]
rifl	riflessivo	[890]	
rifl_rec	riflessivo reciproco		[203]
rifl_in	riflessivo inerente		[304]
erg_rifl	ergativo e riflessivo		[1742]
imp	impersonale	[30]	
imp_atm	impersonale atmosferico		[32]
cop	copulativo	[8]	
mod	modali		[5]
C_mov	di movimento		[255]
C_prop	proposizionale		[210]

#### **4. Lemmatizzazione e analisi sintattica**

##### *4.1 La lemmatizzazione*

Elenchiamo qui di seguito la lista di strutture di tratti o di coppie attributi-valore generati da IMMORTALE quando lavora in modalità di “lemmatizzazione automatica” oppure nell’analisi sintattica:

pred = predicato semantico (coincidente con il lemma)

gen = genere (mas, fem)

num = numero (sing, plur)

pers = persona (1, 2, 3)

tempo = (pres, past, imp, fut)

modo = (ind, cong, imper, cond, part, ger, inf)

scat = classe di sottocategorizzazione solo per i verbi (trans, intr, cop, refl\_in, erg, refl\_rec, refl, imp, aux)

tipo = un tipo semantico per tutte le categorie ad eccezione del verbo

encl = un enclitico in verbi cliticizzati (uno dei clitici)

caso = caso/i per clitici (nom, dat, acc, gen, obl, part)

pref = prefisso (uno dei prefissi)

suff = suffisso (uno dei suffissi)

deg = grado (comp, min, sup)

cat/cat1/cat2 = categorie contenute in parole sincretiche (preposizione articolata)

pred1/pred2 = predicati addizionali contenuti in parole composte

vform = solo per verbi cliticizzati, è la corrispondente forma verbale

##### *4.2 Il “Tagging” sintattico*

Elenchiamo qui di seguito le categorie sintattiche generative automaticamente da IMMORTALE ed utilizzate dal parser in esso contenuto. Sono state ottenute unendo all’etichetta di base informazioni lessicali e semantiche contenute nella lemmatizzazione. Solo la parte rilevante per l’analisi sintattica è stata inclusa: ad esempio, i tratti di accordo non appaiono. I “tag” sono in tutto 88.

- |           |   |                            |
|-----------|---|----------------------------|
| 1. ag     | = | aggettivo                  |
| 2. agn    | = | aggettivo nominale         |
| 3. art    | = | articolo def/indef         |
| 4. clit   | = | clitico pronome            |
| 5. clitac | = | clitico pronome accusativo |
| 6. clidat | = | clitico pronome dativo     |
| 7. deit   | = | deittico pronome           |
| 8. dim    | = | dimostrativo aggettivo     |
| 9. fw     | = | parola straniera = nome    |
| 10. in    | = | intensificatore            |
| 11. ind   | = | aggettivo indef            |
| 12. int   | = | interrogativo              |
| 13. n     | = | nome                       |
| 14. nc    | = | nome colore                |
| 15. nf    | = | nome fattivo               |
| 16. nh    | = | nome umano                 |
| 17. nt    | = | nome temporale             |
| 18. num   | = | numerale                   |
| 19. nnum  | = | numerale nome              |
| 20. poss  | = | possessivo aggettivo       |
| 21. pron  | = | pronome                    |
| 22. pers  | = | pronome personale          |

23. q	=	quantificatore
24. qc	=	quantificatore collettivo
25. qd	=	quantificatore distributivo
26. rel	=	relativo
27. w	=	parola sconosciuta = nome
28. avv	=	avverbio
29. avv1	=	avverbio locuzione
30. ccom	=	congiunzione "come" comparativo
31. cong	=	congiunzione
32. congf	=	congiunzione frasale
33. conjl	=	congiunzione locuzione
34. cosu	=	congiunzione subordinata
35. cosul	=	congiunzione subordinata locuzione
36. dirs	=	discorso diretto subordinazione
37. intj	=	interiezione
38. neg	=	negazione
39. p	=	preposizione
40. part	=	preposizione più articolo
41. partd	=	preposizione "di" più articolo
42. partda	=	preposizione "da" più articolo
43. pd	=	preposizione "di"
44. pda	=	preposizione "da"
45. php	=	preposizione locuzione
46. ausai	=	ausiliare avere infinito
47. ause	=	ausiliare essere flesso
48. ausei	=	ausiliare essere infinito
49. ausep	=	ausiliare essere participio passato
50. part_pass	=	participio passato assoluto
51. part_pres	=	participio presente assoluto
52. pt	=	particella verbale
53. vc	=	verbo copulativo
54. vcir	=	verbo copulativo modo irreale
55. vgin	=	verbo intrans gerundio
56. vgprog	=	verbo intrans progressivo
57. vgs	=	verbo modal gerundio
57. vgsf	=	verbo a ricostruzione gerundio
58. vgt	=	verbo trans gerundio
59. viin	=	verbo intrans infinito
60. vin	=	verbo intrans flesso
61. virin	=	verbo intrans modo irreale
62. virt	=	verbo trans modo irreale
63. vit	=	verbo trans infinito
64. vppin	=	verbo intrans participio passato
65. vppt	=	verbo trans participio passato
66. vprin	=	verbo intrans participio presente
67. vprog	=	verbo progressive
68. vprogir	=	verbo progressive modo irreale
69. vprt	=	verbo trans participio presente
70. vmodir	=	verbo modale modo irreale
71. vmod	=	verbo modale
72. vt	=	verbo trans flesso
73. punt	=	interna alla frase ( , - = § )
74. punto	=	alla fine di frase ( ! ? . ; )
75. puntint	=	punteggiatura non dichiarativa ( ? ! )

76. dirs	=	introduttore discorso diretto/indiretto ( : )
77. par	=	parentetiche ( '(' ')' ' "' )
78. pk	=	parola k-
79. auag	=	ausiliare avere gerundio
80. auair	=	ausiliare avere modo irreal
81. aueg	=	ausiliare essere gerundio
82. aueir	=	ausiliare essere modo irreal
83. ausa	=	ausiliare avere fless
84. vsfa	=	verbo a ricostruzione gerundio
85. rel	=	relativo
86. relin	=	relativo indef
87. relob	=	relativo obliquo
88. relq	=	relativo quantificatore

I tags sintagmatici e radice (di frase) sono 33. In totale, I simboli terminali e non terminali sono 121. Ciascuna rete di transizione corrisponde al lato sinistro di una regola di grammatica context-free e viene trattata come tale dal parser. La descrizione delle regole è composta da una etichetta di rete, da un nodo di ingresso, da una etichetta indicante uno dei quattro tipi di mosse permesse: cat, push, jmp, pop; e infine da uno o da N nodi di uscita. Con CAT si accetta in ingresso una parola che appartiene a quella particolare categoria sintattica, o tag; con PUSH si chiama una sottorete con quella etichetta, o meglio una regola che abbia quel particolare costituente da riscrivere; con JMP, o "jump" si salta a un nodo più lontano nella stessa rete in cui si sta operando; e con POP si esce dalla regola e si chiude il costituente. I costituenti sintattici non terminali generati da IMMORTALE sono quelli elencati di seguito, per ciascuno dei quali indichiamo le mosse e le categorie di ingresso, in altre parole il primo simbolo alla destra della freccia di riscrittura di una corrispondente regola context-free:

1. PARAG = paragrafo

Questa rete serve a chiamare un paragrafo, cioè una sequenza di frasi. A sua volta chiama il nodo CP

arc(parag, 1, cat, inizio, 2).

arc(parag, 2, push, cp, 2).

2. CP = aggiunti frasali

Questa rete chiama le singole frasi, gli aggiunti di frase e le frasi aggiunte, nonché frasi interrogative e frasi subordinate. Infine nell'arco 2, se queste chiamate vanno a vuoto, chiama il nodo di frase.

arc(cp, 1, cat, [fine,punto,puntint], N).

arc(cp, 1, push, [sp, savv,f2,f4,fint,fs,fc,sv2,sv3,sv5,fp,saa,f], N).

3. F = frase

Questa regola chiama tipicamente il sintagma nominale soggetto e il sintagma verbale predicato.

arc(f, 1, push, [sn1, ibar, ir\_infl], N).

4. FINT = frase interrogativa

L'ingresso a questo tipo di frasi è ammesso solo attraverso una particolare categoria, tipicamente un pronome interrogativo o una preposizione e un pronome interrogativo.

arc(fint, 1, cat, [p, int, intj, int], N).

5. FA = frase appositiva

arc(fa, 1, cat, [ag, avv, num, cong, q, n], N).

6. F1 = frase parentetica

arc(f1, 1, cat, punt, 1).

arc(f1, 1, push, [sn1, sa1, savv, sp1, sv3], N).

7. F2 = frase relativa prepositiva

arc(f2, 1, cat, [p, rel, relob, rel], N).

8. F3 = frase relativa ellittica  
arc(f3, 1, cat, [p, part, rel, relob, relin], N).

9. F4 = frase completiva prepositiva  
arc(f4, 1, cat, [p, pk, relin], N).

10. FC = frase coordinata  
arc(fc, 1, cat, cong, 2).

11. FAC = frase completiva  
arc(fac, 1, cat, pk, 2).

arc(fac, 2, push, cp, 3).

12. FS = frase subordinata  
arc(fs, 1, cat, [cosu, conjl], N).

13. DIRSP = frase discorso diretto  
arc(dirsp, 1, cat, dirs, 2).

14. FP = frase parentetica  
arc(fp, 1, cat, [punt, par], N).

15. IBAR = sintagma verbo flesso

Questo costituente chiama tutti i costituenti funzionali che si aggiungono al verbo flesso e lo stesso verbo flesso sia che si tratti di ausiliare sia che si tratti di verbo lessicale. Questo costituente chiamerà per ciascun verbo lessicale un diverso tipo di sintagma verbale, realizzando così la cosiddetta “verb guidance” nel parsing: il verbo a seconda della propria classe sintattica, - transitivo, copulativo, intransitivo - realizzerà una diversa struttura di complementi. Le strutture passive sono individuate a questo livello di costituenza.

arc(ibar, 1, cat, [neg, clit, clitac, clitdat, vprog, ausa, ause, vsupp, vt, vin, vc, in], N).

16. IR\_INFL = sintagma verbale tempo irreal

Lo stesso costituente precedente in cui però il modo è irreal, cioè congiuntivo o condizionale.

arc(ir\_infl, 1, cat, [neg, clit, clitac, vprogir, ausair, auseir, vsupir, virt, virin, vcir], N).

17. SVT = sintagma verbale transitivo

arc(svt, 1, push, [dirsp, savv, fp, sp1, sn1, fac, sv2, fint, sq], N).

18. SVIN = sintagma verbale intransitivo

arc(svin, 1, push, [sp1, sv2], 2).

arc(svin, 1, jmp, [], 2).

arc(svin, 1, pop, [], []).

19. SVC = sintagma verbale copulativo

arc(svc, 1, push, [sa1, sn, sp1, fp], N).

20. SV2 = frase infinitiva complemento/aggiunta

arc(sv2, 1, cat, [p, pt, ausei, ausai, vit, viin, avv], N).

21. SV3 = frase participiale aggiunta

arc(sv3, 1, cat, [vppin, vppt, vppt, vprin], N).

22. SV5 = frase gerundiva aggiunta

arc(sv5, 1, cat, [clit, auag, aueg, vgs, vgt, vgin, vsfa], N).

23. SQ = sintagma quantificato

arc(sq, 1, cat, [in, q, ind, qd], N).

24. SAA = sintagma aggettivale aggiunto

arc(saa, 1, push, sa1, 2).

arc(saa, 2, cat, punt, 3).

25. SA1 = sintagma aggettivale coordinato

arc(sa1, 1, push, sa, 2).

arc(sa1, 2, cat, cong, 3).

26. SA = sintagma aggettivale

arc(sa, 1, cat, [neg, ind, ccom, avv, in, ag], N).

27. SN1 = sintagma nominale con aggiunti sp  
 arc(sn1, 1, push, [sq, sn, f4, sp1], N).
28. SN = sintagma nominale  
 arc(sn, 1, cat, [in, neg, art, qc, poss, num, dim, pers, deit, pron, n, nf, nh, nt, ag, agn, vppt, ag], N).
29. SP1  
 arc(sp1, 1, push, sp, 2).  
 arc(sp1, 2, cat, cong, 3).
30. SP = sintagma preposizionale  
 arc(sp, 1, cat, [in, p, php, part], N).
31. SAVV = sintagma avverbiale  
 arc(savv, 1, cat, [neg, avv, avvl, in], N).
32. SPD1 = sintagma preposizionale (di/da) coordinato  
 arc(spd1, 1, push, spd, 2).  
 arc(spd1, 2, cat, cong, 3).
33. SPD = sintagma preposizionale (di/da)  
 arc(spd, 1, cat, [in, pd, partd, pda, partda], N).

### **5. Alcune analisi sintattiche**

Presentiamo qui di seguito alcune analisi sintattiche realizzate dal nostro parser sulla base dell'analizzatore morfologico e del temario di cui si è parlato più sopra. Il parser è stato presentato in un lavoro del 1989 (vedi Delmonte & Dolci), ed è un semplice attraversatore di reti basato su una grammatica context-free. Le regole della grammatica sono quelle elencate al punto precedente che utilizzano i tags prodotti dal tagger automatico e i simboli non terminali della grammatica. Le analisi forniscono la lista dei tags singoli associati a ciascuna parola del testo sulla base delle regole sintattiche. Inoltre, viene fornita la traccia dell'attraversamento di ciascun nodo principale, che qui non viene mostrata per mancanza di spazio. In questo modo i costituenti maggiori contengono le varie teste lessicali e i vari costituenti minori, ricevono tutti l'indicazione di un nodo astratto superiore normalmente quello di frase, permettendo quindi di ricostruire l'albero complessivo. I testi che riportiamo sono estratti da giornali o da circolari ministeriali.

1. I docenti immessi in ruolo per l'insegnamento su posti di tipo speciale possono presentare domanda di trasferimento per i posti che siano stati messi a disposizione ma solo i docenti titolari di posto potranno partecipare ai movimenti per compensazione.

Parse = [f-[subj-[sn-[art-[i], n-[docenti], sv3-[vppin-[immessi], sp1-[sp-[p-[in], sn-[n-[ruolo]]], sp-[p-[per], sn-[art-[l], n-[insegnamento], sv3-[savv-[avv-[su]], n-[posti], sp1-[sp-[p-[di], sn-[n-[tipo], n-[speciale]]]]]]]]]]], ibar-[vsupp-[possono], vit-[presentare], svt-[obj-[sn-[n-[domanda], spd1-[spd-[pd-[di], sn-[n-[trasferimento]]]]]]], sp1-[sp-[p-[per], sn-[art-[i], n-[posti], f2-[rel-[che], subj-[np-[e], ibar-[ause-[siano], ausep-[stati], vppt-[messi], svin-[sp1-[sp-[p-[a], sn-[n-[disposizione]]]]]]]]]]], fs-[cosu-[ma], cp-[f-[subj-[sn-[in-[solo], art-[i], ag-[docenti], n-[titolari], spd1-[spd-[pd-[di], sn-[n-[posto]]]]]]], ibar-[vsupp-[potranno], vit-[partecipare], svt-[obj-[np-[e], sp1-[sp-[part-[ai], sn-[n-[movimenti]]], sp-[p-[per], sn-[n-[compensazione]]]]]]], punto-[.]]]]

Tags = [art-[i], n-[docenti], vppin-[immessi], p-[in], n-[ruolo], p-[per], art-[l], n-[insegnamento], avv-[su], n-[posti], p-[di], n-[tipo], n-[speciale], vsupp-[possono], vit-[presentare], n-[domanda], pd-[di], n-[trasferimento], p-[per], art-[i], n-[posti], rel-[che], ause-[siano], ausep-[stati], vppt-[messi], p-[a], n-[disposizione], cosu-



[sn-[n-[replica]], sp1-[sp-[part-[alle], sn-[n-[accuse]]]], fp-[par-['']], neg-[non], sn-[ag-[oneste], fp-[par-['']], sp1-[sp-[part-[sul], sn-[n-[buco], spd1-[spd-[pda-[da], sn-[num-['15mila'], n-[miliardi]]]]]]]]], punto-[.], sn-[nh-[ciampi]], ibar-[vt-[contrattacca], svt-[dirsp-[dirs-[ :], sn1-[sn-[nh-[berlusconi]]], ibar-[ausa-[ha], vppt-[mentito]]]], punto-[.], fp-[par-['']], ibar-[ausa-[abbiamo], vppt-[risanato], svt-[sn1-[sn-[art-[i], n-[conti], spd1-[spd-[partd-[dello], sn-[ag-[stato], fp-[par-['']]]]]]]]], punto-[.], sn-[n-[roma]], punto-[.], fp-[par-['']], sn-[n-[nessuno]], ibar-[vsupp-[può], vit-[parlare], svt-[fp-[punt-[','], savv-[avv-[onestamente]], sp1-[sp-[p-[di], sn-[n-[falsificazioni]]], par-['']]]]], punto-[.], fc-[cong-f-[e], sp-[p-[in], sn-[dim-[quell], fp-[par-['']], savv-[avv-[onestamente]], par-['']]]]], ibar-[clit-[c], vc-[è], svc-[sn-[qc-[tutta], art-[l], n-[irritazione], spd1-[spd-[partd-[del], sn-[n-[presidente], spd1-[spd-[partd-[del], sn-[n-[consiglio]]]]]]]]], sp1-[sp-[p-[per], sn-[art-[una], n-[accusa], f2-[rel-[che], sn1-[sn-[art-[il], poss-[suo], n-[governo]]], ibar-[neg-[non], vsupp-[può], vit-[tollerare], svt-[dirsp-[dirs-[ :], ibar-[vc-[aver], svc-[sa1-[sa-[ag-[detto]], sn-[art-[il], n-[falso]], sp1-[sp-[part-[sui], sn-[n-[conti], n-[pubblici]]]]]]]]]]], punto-[.], sn-[nh-[ciampi]], ibar-[ausa-[ha], savv-[avv-[così]], vppt-[risposto], svt-[sp1-[sp-[part-[all], sn-[n-[accusa], spd1-[spd-[pd-[di], sn-[ag-[falso], fp-[punt-[','], sv3-[vppt-[lanciata], sp1-[sp-[p-[da], sn-[nh-[berlusconi]], sp-[p-[durante], sn-[art-[il], n-[faccia]], sp-[p-[a], sn-[n-[faccia], sa1-[sa-[ag-[televivivo]]]], sp-[p-[con], sn-[art-[il], n-[segretario], spd1-[spd-[partd-[del], sn-[n-[pds]]]]]]], sn-[nh-[occhetto], fp-[punt-[','], sv5-[vgt-[rompendo], sn-[art-[la], ag-[stretta], n-[consegna], spd1-[spd-[partd-[del], sn-[n-[silenzio], fac-[pk-[che], ibar-[clitac-[si], ause-[era], vppin-[dato], svin-[sp1-[sp-[p-[per], sn-[qc-[tutta], art-[la], n-[campagna], sa1-[sa-[ag-[elettoriale]]]]]]]]], punto-[.], ibar-[ausa-[ha], vppt-[ribadito], svt-[sn1-[sn-[art-[le], n-[spiegazioni], sa1-[sa-[avv-[già], ag-[date], sp1-[sp-[part-[sulla], sn-[ag-[finanziaria], fp-[punt-[','], sp1-[sp-[part-[sul], sn-[ag-[maggior], n-[deficit], ag-[provocato]]]], savv-[avv-[essenzialmente]]], spd1-[spd-[partda-[dal], sn-[n-[peso], sa1-[sa-[neg-[non], ag-[previsto]]], spd1-[spd-[partd-[della], sn-[n-[recessione]]]]]]]]]]]]]]]]], punt-[',']], cong-[e]]]]]]]]], ibar-[ausa-[ha], savv-[avv-[orgogliosamente]], vppt-[precisato], svt-[sn1-[sn-[art-[i], n-[risultati], ag-[conseguiti]], sp1-[sp-[in-[proprio], part-[sulla], sn-[n-[finanza], sa1-[sa-[ag-[pubblica]]]]]]]]], punto-[.], sp-[p-[tra], sn-[art-[i], n-[risultati]], sn1-[sn-[dim-[quell], n-[accordo], spd1-[spd-[pd-[di], sn-[nt-[luglio]]]], sp1-[sp-[p-[tra], sn-[art-[le], n-[parti], ag-[sociali], f2-[rel-[che], ibar-[vt-[considera], svt-[sn1-[sq-[q-[una]]], spd-[partd-[delle], sn-[ag-[principali], n-[conquiste], spd1-[spd-[partd-[del], sn-[poss-[suo], n-[governo]]]]]]]]]]]]], punto-[.], fc-[cong-f-[e], sn-[pron-[proprio]], sp1-[sp-[part-[sull], sn-[n-[accordo]]]], par-['']], ibar-[vt-[fanno], svt-[sn1-[sn-[ag-[quadrato], fp-[par-['']], savv-[avv-[insieme]], sn-[nh-[confindustria], cong-[e], ag-[sindacati], fp-[punt-[','], f4-[pk-[che], ibar-[clitac-[si], ause-[sono], vppt-[rivolti], svt-[savv-[avv-[prima]], sp1-[sp-[part-[al], sn-[n-[presidente], spd1-[spd-[partd-[del], sn-[n-[consiglio]], cong-[e]]], cong-[poi], sp-[part-[al], sn-[n-[presidente], spd1-[spd-[partd-[della], sn-[n-[repubblica]]]], fint-[int-[perché], sn-[in-[soprattutto], dim-[quest], n-[ultimo]], ir\_infl-[clitac-[si], virt-[faccia], svt-[sn1-[sn-[n-[garante], spd1-[spd-[partd-[del], sn-[nf-[fatto], fac-[pk-[che], sp-[p-[dopo], sn-[art-[le], n-[elezioni]], sn1-[sn-[pron-[esso]]], ibar-[ause-[sia], vppt-[rispettato], svt-[fp-[punt-[','], f4-[relin-[chiunque], ibar-[vin-[risulti], svin-[sa1-[sa-[ag-[vincitore]]]]]]]]]]], punto-[.], ibar-[clitac-[si], vt-[tratta], svt-[fp-[punt-[','], savv-[avv-[come]], ibar-[ausa-[ha], vppt-[ricordato], svt-[sn1-[sn-[nh-[carlo], nh-[callieri], fp-[punt-[','], sn-[n-[vicepresidente], spd1-[spd-[partd-[degli], sn-[ag-[industriali], fp-[punt-[','], sv2-[p-[di], vit-[difendere], sn-[dim-[questo], n-[compromesso]], sp1-[sp-[p-[tra], sn-[art-[le], n-[parti], ag-[sociali], spd1-[spd-[partda-[dalle], sn-[n-[tentazioni], sv2-[p-[di], vit-[rimettere], sp1-[sp-[p-[in], sn-[n-[discussione]]]]]]]]]]]]], sn-[art-[le], n-[regole], f2-[rel-[che], ibar-[clitac-[lo], vt-[garantiscono], svt-[sp1-[sp-[p-[da], sn-



Lista di lemmi del LIFUV contenente le strutture predicato-argomentali relative ad aggettivi, nomi e verbi in forma codificata. Ciascun lemma viene ripetuto un numero di volte pari alle diverse accezioni semantiche e sintattiche. A parte sono classificate le 1375 parole funzionali invariabili, utilizzate da IMMORTALE per l'analisi morfosintattica di testi.

Il LIFUV viene utilizzata da GETARUN in cui si accede attraverso un'interfaccia creata da E. Pianta. Altrimenti può essere visualizzata attraverso un programma in C che permette di vedere separatamente la lista predicato-argomentale in forma di rappresentazione concettuale

oppure in forma sintattico-funzionale. 4932 nomi + 4844 verbi + 1623 aggettivi = 11399 entrate lessicali

## **7. Bibliografia**

Dardano M., P. Trifone, 1983, Grammatica Italiana, Zanichelli, Bologna.

Delmonte R., 1988, Computational Morphology for Italian, in AA.VV., Studi di Linguistica Computazionale, Unipress, Padova, Chapt. I, 1-20.

Delmonte R., 1989, From Subcategorization Frames to Thematic Roles, in R. Delmonte (ed.), 1989, Lessico, Strutture e Interpretazione - Studi Linguistici Applicati I, Unipress, Padova, Chapt. IV, 167-212.

Delmonte R. (1992), Linguistic and Inferential Processes in Text Analysis by Computer, Unipress, Padova.

Delmonte R., 1995, (in corso di pubblicazione), Lexical Representations: Syntax-Semantics interface and World Knowledge, in Notiziario AIIA (Associazione Italiana di Intelligenza Artificiale), Roma, pp. 8.

Delmonte R. (ed), 1989, Lessico, Strutture e Interpretazione, Unipress, Padova.

Delmonte R., R. Dolci (1989), Parsing Italian with a Context-Free Recognizer, Annali di Ca' Foscari XXVIII, 1-2, 123-161.

Delmonte R. F. Stiffoni, 1995, (in corso di pubblicazione), SIWL - Il Database Parlato della lingua Italiana, Convegno AIA - Gruppo di Fonetica Sperimentale, Trento.

Scalise S., 1984a, Morfologia lessicale, CLESP, Padova.

Scalise S., 1984b, Generative Morphology, Foris Pub., Dordrecht.

Serianni L. (1988), Grammatica Italiana, Utet, Torino.