

# Predicati semantici e Lessico-grammatica

Annibale Elia, Simonetta Vietri<sup>1</sup>

## Introduzione: la semantica del Web Semantico

Con il termine "Web Semantico" si indica un insieme di programmi software che utilizzano etichette semantiche per il recupero di contenuti informativi a partire dai testi diffusi nel Web. Tim Berners-Lee, l'inventore del Web, ha lanciato questa provocazione informatico-filosofica da qualche anno<sup>2</sup>. Il Web tradizionale si basa ancora sull'uso di HTML (*Hyper Text Markup Language*), che è un linguaggio di marcatura dei testi di tipo essenzialmente "grafico-tipografico", il nuovo linguaggio di marcatura che permetterebbe interrogazioni di tipo semantico è l'XML (*eXtensible Markup Language*)<sup>3</sup>.

Il Web Semantico (WS), secondo Tim Berners-Lee e il W3C (consorzio di ricercatori che si muovono nella direzione del WS) non si sostituirebbe completamente al Web tradizionale, ma lo integrerebbe, prima attraverso nuove funzioni di ricerca e poi con una infrastruttura "intelligente", per facilitare l'accesso all'informazione attraverso "agenti" software in grado di risolvere anche specifici problemi di informazione quotidiana. L'obiettivo è quello di riuscire a rispondere a domande in linguaggio naturale o in un linguaggio formale molto facile da utilizzare, grazie a dei software autonomi, che vengono denominati appunto "agenti intelligenti". Il WS dovrebbe, quindi, rivoluzionare i motori di ricerca.

I motori di ricerca di adesso, anche i più evoluti, non permettono di cogliere la differenza fra il signor "Giacchino Rossini", che di professione fa il meccanico a Forlì, e il famoso compositore italiano di opere dell'Ottocento. Secondo i sostenitori del WS, se, al momento della ricerca, si potesse inserire a fianco della stringa "Giacchino Rossini" l'etichetta semantica ("tag") <COMPOSITORE>, verrebbero eliminati automaticamente tutti i documenti trovati nel Web che non contengono quella etichetta.

Evidentemente, ciò presuppone che tutti i documenti presenti nel Web e che parlino del compositore "Giacchino Rossini" dovrebbero essere stati già marcati *manualmente* con l'etichetta <COMPOSITORE> da coloro che li hanno immessi nel Web. Inoltre, dovrebbero essere presenti, per esempio, altre etichette relative ai contenuti musicali dei melodrammi, ad altri compositori e a tutte le informazioni connesse allo stesso dominio.

Secondo Berners-Lee, si tratta di accumulare nel tempo milioni e milioni di definizioni specifiche dei contenuti, organizzate in schemi di tipo logico-classificatorio. Questi schemi vengono denominati "ontologie", in onore della ben più complessa ontologia filosofica, che si occupa di definire l'essenza degli "enti". Per ritornare a Rossini, si tratterebbe di elaborare un sottoinsieme di una Ontologia Musicale, che potrebbe richiamare ad altre Ontologie per altri termini eventualmente presenti nei testi. Un testo esemplificativo come il seguente:

*Quando Giacchino Rossini scrisse il Barbiere di Siviglia, non immaginava che la prima di quell'opera a Napoli sarebbe stato un vero fiasco.*

potrebbe essere stato marcato in XML più o meno così:

```
Quando <compositore>Giacchino Rossini</compositore>
scrisse <opera>il Barbiere di Siviglia</opera>, non
immaginava che la prima di quell'opera a
<capoluogo_regione_campania>Napoli</capoluogo_regione_camp
ania> sarebbe stato un vero fiasco.
```

Solo i testi che contengono la stringa *Gioacchino Rossini* dotata dell'etichetta <COMPOSITORE> verranno dunque selezionati dal motore del WS. Con il sistema delle Ontologie, facendo interagire più domini, sarebbe possibile fare ulteriori ricerche che mettano in correlazione più tipi di testi e facciano riferimento a Siti (URL) in cui siano presenti, per esempio, tutti i compositori, tutte le Opere, tutti i libretti d'Opera, ma anche tutti i Teatri in cui le Opere sono state rappresentate, tutte le città in cui hanno sede i Teatri e così via, rendendo possibili risposte a domande del tipo "In quali città è stato rappresentato il Barbiere di Siviglia?", "Quali di queste città è Capoluogo di Regione?", "In quale negozio di Napoli trovo un CD musicale del Barbiere di Siviglia?", "Quanto costa il CD?", e così via.

La prospettiva del WS implica un grande spirito di collaborazione tra tutti i gestori dell'informazione e i navigatori del Web, l'adozione di un sistema comune di classificazione e uno sforzo costante di aggiornamento globale. Le sperimentazioni attuali sono ancora molto settoriali e legate a comunità di internauti fortemente sensibili allo spirito di collaborazione<sup>4</sup>, ma il WS va avanti.

Dal punto di vista linguistico, possiamo fare alcune considerazioni relative al tipo di semantica coinvolta nell'approccio del WS. Innanzitutto va osservato che le definizioni semantiche e le relative Ontologie riguardano essenzialmente dei nomi, siano essi propri o comuni. In secondo luogo, la semantica interviene essenzialmente come apparato logico-formale per disciplinare gerarchicamente la sistemazione concettuale di classi, sottoclassi e proprietà. Si potrebbe affermare che siamo in presenza di parole trattate come argomenti i cui predicati sono in definitiva esterni al mondo delle parole, in quanto rappresentati dagli operatori logici che provvedono a connettere, inferire, mettere in relazione gli argomenti.

In questo saggio, vogliamo contribuire al dibattito sul WS, aprendo una prospettiva diversa, fondata sulla possibilità di elaborare dei Predicati Semantici, costruiti a partire da concrete forme lessicali e dalle strutture sintattiche in cui esse trovano collocazione all'interno dei contesti informativi e comunicativi<sup>5</sup>. Tali Predicati verrebbero generati automaticamente nei testi da un "agente intelligente" lessico-grammaticale. I testi così etichettati si presterebbero ad interrogazioni in lingua naturale sui loro contenuti<sup>6</sup>.

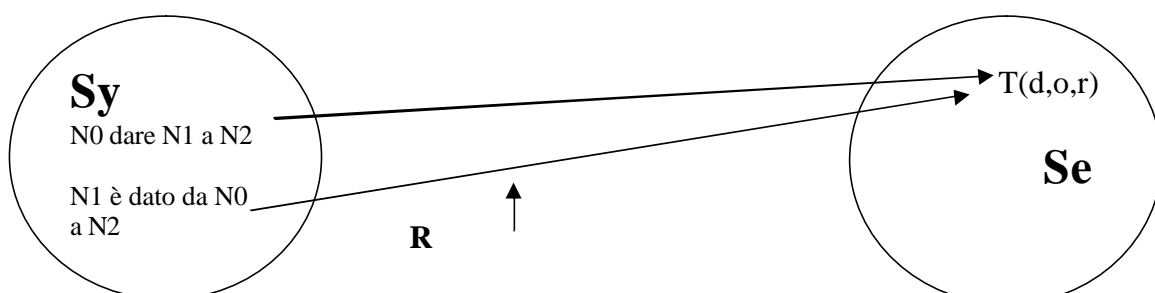
Faremo riferimento alla teoria del Lessico-Grammatica, che offre per le lingue trattate una descrizione potenzialmente esaustiva delle loro strutture lessico-sintattiche<sup>7</sup>.

## 1. Dalla nozione intuitiva di "scambio" ai Predicati di Trasferimento

Nel quadro della teoria lessico-grammaticale, si dispone dell'insieme delle strutture lessico-sintattiche di una lingua, il che comporta la possibilità, per esempio, di assegnare ad un verbo il valore di un Predicato semantico ed ai suoi attanti<sup>8</sup> (soggetto compreso) il valore di Argomenti selezionati da quel Predicato.

In questo saggio faremo riferimento a quei Predicati che formalizzano la nozione intuitiva di "scambio" tra persone e che possiamo definire Predicati di Trasferimento (T) che interessano tre Argomenti: un "datore" (d), un "oggetto" (o) e un "ricevente" (r), per cui abbiamo una formula generale del tipo T(d,o,r).

Sia **Sy** l'insieme delle forme lessico-sintattiche di una lingua e **Se** l'insieme degli elementi di senso o significato. Gli elementi di **Sy** possono essere associati a quelli di **Se** attraverso regole di interpretazione **R**.



Nello schema vediamo che la struttura di frase semplice attiva costruita a partire dal verbo *dare* (N0 V N1 a N2) e la corrispondente struttura di frase semplice passiva N1 essere Vpp da N0 a N2 sono entrambe associate a un solo Predicato semantico T(d,o,r), secondo una regola di interpretazione che associa N0 a “d”, N1 a “o” e N2 a “r”<sup>9</sup>.

Questi Predicati rappresentano un caso particolare di una nozione intuitiva più generale, quella che denota il trasferimento di un oggetto (inanimato o animato) da un luogo (L1) ad un altro (L2). Molti verbi che esprimono un trasferimento verso un luogo possono rappresentare anche il trasferimento tra due persone, come nelle frasi

- (1) *Max porta un topolino dal negozio (a casa di Ale + da Ale)*
- (2) *Max porta un topolino ad Ale*

Possiamo osservare, infatti, che *Ale* è la realizzazione del “ricevente” analogamente a (*a casa di Ale + da Ale*), che è la realizzazione del “luogo” L2, mentre *dal negozio* è il luogo L1.

Le strutture lessico-sintattiche e le Classi lessico-grammaticali sono le seguenti:

Classe 28    N0 *portare* N1 Loc N2 Loc N3  
Classe 23D   N0 *portare* N1 a N2

I Predicati di Trasferimento Locativi<sup>10</sup> sono del tipo TL(am, m, L1, L2), in cui am =: “agente del movimento”, m =: “oggetto mobile”, L1=: “luogo iniziale”, L2=: “luogo finale”, il che produce per la frase (1) uno schema come il seguente:

TL =: *portare*  
am =: N0 =: *Max*  
m =: N1 =: *un topolino*  
L1 =: Loc N2 =: *dal negozio*  
L2 =: Loc N3 =: *a casa di Ale*

Lo scambio può riguardare anche “messaggi”, oltre che oggetti, e in questo caso abbiamo delle realizzazioni di Predicati di Trasferimento di Messaggi in strutture lessico-sintattiche con verbi che reggono una frase completiva e attanti umani:

- (3) *Max dice ad Ale che parte*
- (4) *Max informa Ale che parte*
- (5) *Max chiacchiera con Ale del fatto che partirà*

In questo caso, i Predicati saranno della forma TM (dm, m, rm), in cui dm =: datore del messaggio, m =: messaggio, rm =: ricevente del messaggio)

TM =: *dire*  
dm =: N0 =: *Max*  
m =: N1 =: Ch F =: *che parte (Max parte)*  
rm =: N2 =: *Ale*

TM =: *informare*  
dm =: N0 =: *Max*  
rm =: N1 =: *Ale*

m =: di N2 =: \*di Ch F =: *che parte (Max parte)*

TM =: *chiacchierare*

dm =: N0 =: *Max*

rm =: N1 =: *Ale*

m =: di N2 =: del fatto Ch F =: *del fatto che partirà (Max o Ale parte)*

Le strutture lessico-sintattiche e gli schemi dei Predicati in ordine lineare degli argomenti sono così caratterizzate:

TM (dm, m, rm)

Classe 47 N0 V Ch F N2 =: *Max dice che parte ad Ale*

TM (dm, rm, m,), ma anche TM (dm, m, rm)

Classe 50 N0 V N1 \*di Ch F =: *Max informa Ale che parte*

Classe 53 N0 V \*di Ch F Prep N2 =: *Max chiacchiera con Ale del fatto che partirà*

Tornando ai Predicati di Trasferimento più semplici (quelli di oggetti tra persone), un esempio naturale è quello del verbo *dare*, di cui conosciamo la struttura lessico-sintattica appartenente alla Classe 23D e che ci permette di assegnare i valori argomentali ai singoli attanti:

N0 *dare* N1 a N2 N0 =: “datore”, N1 =: “oggetto”, N2 =: “ricevente”

per cui, data la frase

(6) *Max ha dato un libro ad Ale*

abbiamo

T =: *dare*

d =: N0 =: *Max*

o =: N1 =: *un libro*

r =: N2 =: *Ale*

lo stesso schema T(d,o,r) è valido per le frasi trasformate passive

(7) *Un libro è stato dato da Max ad Ale*

(8) *Un libro è stato dato ad Ale da Max*

La frase

(9) *Ale riceve un libro da Max*

è semanticamente simile a (7) e (8), ma la corrispondenza tra attanti e argomenti è diversa:

CLASSE 23R N0 *ricevere* N1 da N2 N0 =: “ricevente”, N1 =: “oggetto”, N2 =: “datore”

tuttavia, lo schema resta inalterato:

T =: *ricevere*  
d =: N0 =: *Max*  
o =: N1 =: *un libro*  
r =: N2 =: *Ale*

In verità, se teniamo conto dell'ordine lineare di realizzazione degli Argomenti, osserveremo che lo schema di *dare* è del tipo T(d,o,r), mentre quello di *ricevere* è T(r,o,d)

T =: <i>dare</i>	T =: <i>ricevere</i>
d =: N0 =: <i>Max</i>	r =: N0 =: <i>Ale</i>
o =: N1 =: <i>un libro</i>	o =: N1 =: <i>un libro</i>
r =: N2 =: <i>Ale</i>	d =: N2 =: <i>Max</i>

Analogamente, nel caso delle frasi (7) e (8), le trasformate passive, gli schemi, secondo l'ordine lineare, sono del tipo T(o,d,r) oppure T(o,r,d).

Passando in rassegna le Classi lessico-grammaticali abbiamo la seguente situazione:

T (d, o, r)  
CLASSE 23D      N0 V N1 a N2 =: *Max ha dato un libro ad Ale*  
Esempi di usi verbali<sup>11</sup>: dare, passare1, vendere, cedere, regalare, offrire, fittare1, trasferire1

T (r, o, d)  
CLASSE 23Rda      N0 V N1 da N2 =: *Ale riceve un libro da Max*  
Esempi di usi verbali: ricevere, avere, ottenere, comprare, accettare, acquistare, fittare2

CLASSE 23Ra      N0 V N1 a N2 =: *Ale ruba un libro a Max*  
Esempi di usi verbali: rubare, strappare, sfilare, sottrarre, togliere

T (r, d, o)  
CLASSE 24      N0 V N1 di N2 =: *Ale priva Max di un libro*  
Esempi di usi verbali: derubare, privare

T (o, d, r)  
CLASSE 7DP      N0 V da N1 a N2 =: *il libro passa da Max ad Ale*  
Esempi di usi verbali: passare2, transitare

CLASSE 30      N0 V N1 da N2 a N3 =: *Lia trasferisce un libro da Max ad Ale*  
Esempi di usi verbali: trasferire2

### 1.1. I Predicati correlati

A partire dal Predicato T, è possibile individuare altri predicati che entrino in relazione con esso. Per esempio, possiamo immaginare un Predicato di Possesso, non a livello intuitivo, ma sulla base di nozioni giuridiche. Il Codice Civile (c.c.) distingue nettamente tra "possesso" e "proprietà".

La **proprietà**, infatti, è una situazione di diritto, più specificamente è il diritto sulla cosa. L'art. 832 c.c afferma: "il proprietario ha diritto di godere [949] e disporre delle cose in modo pieno ed esclusivo, entro i limiti e con l'osservanza degli obblighi stabiliti dall'ordinamento giuridico".

Il **possesso**, invece, è una situazione di fatto. L'art. 1140 c.c. lo definisce come "il potere sulla cosa che si manifesta in una attività corrispondente all'esercizio del diritto di proprietà o di

altro diritto reale". La differenza tra le due figure si fonda, quindi, sulla diversità tra titolarità ed esercizio del diritto: fra l'essere proprietari di una cosa e il comportarsi come proprietari di essa.

Una volta applicato un Predicato della forma T (d,o,r), possiamo costruire due nuovi schemi di Predicato, uno per la Proprietà (Pr) e uno per il Possesso (Ps):

Pr (p1, o), in cui p1 =: "proprietario", o =: "oggetto"

Ps (p2, o), in cui p2 =: "possessore", o =: "oggetto"

a partire dai quali possiamo assegnare automaticamente gli argomenti p1 e p2 alle frasi seguenti

(10) *Ale ha acquistato un libro da Max*

(11) *Ale ha rubato un libro a Max*

che hanno le strutture lessico-sintattiche, le relative Classi lessico-grammaticali e gli schemi di Predicato seguenti

Classe 23Rda	N0 <i>acquistare</i> N1 da N2	T (r, o, d)
Classe 23Ra	N0 <i>rubare</i> N1 a N2	T (r, o, d)

N0 =: Ale (10) =: +p1,+ p2

N0 =: Ale (11) =: +p2, -p2

N2 =: Max (10) =: -p1, -p2<sup>12</sup>

N2 =: Max (11) =: +p1, -p2

per cui possiamo espandere lo schema precedente in quello seguente :

Classe 23Rda	N0 <i>acquistare</i> N1 da N2	T (r, o, d) Pr (r = p1, o) Pr (r = p2, o)
Classe 23Ra	N0 <i>rubare</i> N1 a N2	T (r, o, d) Pr (d = p1, o) Ps (r = p1,o)

## 1.2. Predicati di trasferimento in lingue diverse

Disponendo della descrizione lessico-grammaticale di una lingua è possibile compararla a quella di un'altra lingua e individuare le classi equivalenti. Inoltre, poiché in ogni classe sono rappresentati tutti gli usi verbali che entrano in determinate strutture lessico-sintattiche, è possibile, dato un Predicato semantico, assegnare ai complementi di ogni verbo il valore di Argomento semantico corrispondente.

Se prendiamo le strutture del francese, dell'italiano, dell'inglese e del tedesco, per esempio, possiamo mettere in relazione delle frasi concrete e assegnare gli Argomenti del Predicato T (d, o, r) anche in ordine di successione lineare.

Francese:

N0 V N1 à N2 = : *Max a donné un livre à son amie* **T (d, o, r)**

Italiano:

N0 V N1 a N2 = : *Max ha dato un libro alla sua amica* **T (d, o, r)**

Inglese:

N0 V N1 N2 =: *Max gave his friend a book* **T (d, r, o)**

Tedesco:

N0(nom) V N1(dat) N2(acc) =: *Max gab seiner Freundin ein Buch*      **T (d, r, o)**

Se prendiamo le classificazioni lessico-grammaticali del francese<sup>13</sup> e dell'italiano, lingue che dispongono di un livello di approfondimento descrittivo comparabile, possiamo avere degli schemi di questo tipo:

Francese	Italiano
<b>N0 V N1 à N2</b>	<b>N0 V N1 a N2</b>
36DT <i>donner</i> T(dor)	23D <i>dare</i> T(dor)
38LD <i>lancer</i> T(dor)	28 <i>lanciare</i> T(dor)
38LR <i>abandonner</i> T(dor)	28SC <i>abbandonare</i> T(dor)
36DT <i>voler</i> T (rod)	23D <i>rubare</i> T (rod)
38LS <i>enlever</i> T(rod)	23D <i>togliere</i> T(rod)
36R <i>abonner</i> T(dro)	23R <i>abbonare</i> T(dro)
<b>N0 V N1 de N2</b>	<b>N0 V N1 di/da N2</b>
36DT <i>dédommager</i> T(dro)	23D <i>dispensare da</i> T(dro)
37M1 <i>nantir</i> T(dro)	24 <i>dotare di</i> T(dro)
36DT <i>recevoir</i> T(rod)	23R <i>ricevere da</i> T(rod)
38LS <i>soutirer</i> T(rod)	29 <i>travasare da</i> T(rod)
36DT <i>arnaquer</i> T (rdo)	23D <i>scroccare</i> T(rod)
37E <i>gruger</i> T(rdo)	24 <i>spogliare</i> T(rod)

## 2. Il web semantico e il lessico-grammatica

Le descrizioni lessico-grammaticali sono state inserite da molti anni in programmi di analisi testuale automatica concepiti sotto la direzione scientifica di Maurice Gross<sup>14</sup>. Questi programmi permettono, attraverso un sistema integrato di dizionari morfo-grammaticali elettronici, tavole lessico-grammaticali e grammatiche locali, di estrarre informazioni e dati nei testi e di assegnare strutture, tag e classificazioni a partire dai dati.

Ciò significa che è possibile, per esempio, all'interno di qualsiasi insieme di testi in formato elettronico (offline o online, cioè sul web) assegnare automaticamente i valori degli Argomenti semantici a partire da Predicati che siano stati associati a strutture lessico-sintattiche e quindi alle concrete entrate lessicali. Estrazioni ed associazioni possono essere estese automaticamente anche all'interno di testi in lingue diverse, purché dispongano di una completa descrizione lessico-grammaticale.

La prima implicazione di questo processo di estrazione e di associazione è la possibilità di avere una omogenea e *automatica* pre-marcatura di Predicati e Argomenti semantici all'interno di svariati miliardi di testi presenti nel web e di poter accedere a queste informazioni non solo attraverso le forme nominali, come si procede attualmente, bensì anche attraverso i Predicati, cioè i verbi e le forme verbali suppletive.

Prendiamo un testo esemplificativo:

*Lia aveva lanciato dal finestrino del treno la borsa nera a Luciano, che, dopo una capriola, era riuscito a prenderla, rimettendosi in piedi a fatica. Luciano corse fino alla macchina parcheggiata poco lontano, si sedette alla guida e con il cuore in gola estrasse dalla borsa il pacchetto che gli interessava. Mise in moto e arrivò velocemente al baretto, dove incontrò Carlo, al quale diede il prezioso pacchetto, ricevendo da lui una*

*busta gialla stretta e lunga. Carlo lo pregò di andarsene via di corsa perché c'erano in giro personaggi pericolosi. Luciano ripartì in tutta fretta e arrivò in piazza Merli verso le 23.*

A partire dagli usi verbali presenti nel testo, elaboriamo ed applichiamo i Predicati Semantici che abbiamo illustrato finora.

T (d, o, r)	Predicato di Trasferimento
TL (am, m, lp, ld)	Predicato di Trasferimento Locativo
TM (dm, m, rm)	Predicato di Trasferimento di Messaggi
M (am, lp, ld)	Predicato di Movimento

Sulla base di alcune sperimentazioni effettuate con NOOJ, che utilizza dizionari elettronici morfosintattici dell'italiano e tavole informatiche per le classi lessico-grammaticali, è possibile assegnare automaticamente i Predicati e gli Argomenti correlati, come nell'output seguente, che individua gli attanti e i verbi ed etichetta Predicati e Argomenti:

Lia <d> aveva lanciato <T> dal finestrino del treno la borsa nera <o> a Luciano <r>, che, dopo una capriola, era riuscito a prenderla, rimettendosi in piedi a fatica. Luciano <am> corse <M> fino alla macchina parcheggiata poco lontano dalla stazione <ld>, si sedette alla guida e con il cuore in gola (Luciano) <am> estrasse <TL> dalla borsa <lp> il pacchetto <m> che gli interessava. Mise in moto e (Luciano) <am> arrivò <M> velocemente al baretto <ld>, dove incontrò Carlo al quale <r> (Luciano) <d> diede <T> il prezioso pacchetto <o> (Luciano) <r> ricevendo <T> da lui <d> una busta gialla stretta e lunga <o>. Carlo <dm> lo <rm> pregò <TM> di andarsene via <m> di corsa perché c'erano in giro personaggi pericolosi. Luciano <am> ripartì in tutta fretta e arrivò <M> in piazza Merli <ld> verso le 23.

Sulla base di quest'output, è possibile elaborare delle domande di natura semantica sul testo:

**Quali sono i soggetti di azione *datori di oggetti*?**

Lia  
Luciano  
Carlo

**Quali sono i soggetti di azione *datori e riceventi di messaggi*?**

Carlo, Luciano

**Quali sono i soggetti di azione *agenti di movimento*?**

Luciano, Luciano

**Quali sono i soggetti di azione *agenti di trasferimento locativo*?**

Luciano

**Qual è il luogo di destinazione dei movimenti di Luciano?**

la macchina parcheggiata poco lontano dalla stazione



il baretto  
piazza Merli

***Che cosa riceve Luciano e da chi?***

la borsa nera da Lia  
una busta gialla stretta e lunga da Carlo

***Che cosa dà Luciano a Carlo ?***

il prezioso pacchetto

***Che cosa trasferisce Luciano e da dove?***

il pacchetto dalla borsa

E' possibile creare molte altre domande e, sulla base del sistema di corrispondenza tra lessico-grammatiche di lingue diverse, è possibile interrogare simultaneamente più testi in più lingue.

Prima di concludere, vorremmo fare qualche esempio dei passaggi intermedi necessari, a livello di analisi morfosintattica dei testi, per arrivare ad individuare i concreti elementi lessicali cui associare gli Argomenti correlati ai diversi Predicati.

### **3. Strategie e tattiche di ricerca all'interno dei testi**

Una presentazione aggiornata del lessico-grammatica dell'italiano e degli strumenti linguistici e computazionali messi in campo presso il Dipartimento di Scienze della Comunicazione dell'Università di Salerno<sup>15</sup> è contenuta in Vietri 2004, a cui rinviamo per un inventario generale delle strategie e tattiche linguistiche di ricerca automatica all'interno di testi<sup>16</sup>. In questa sede ci sembra opportuno richiamare soltanto alcuni principi generali.

Gli strumenti necessari per la ricerca testuale sono principalmente di tre tipi:

- dizionari elettronici morfosintattici in forma canonica e flessa delle parole semplici (monorematiche), e delle parole composte (polirematiche);
- tavole lessico-grammaticali delle strutture possibili a partire da: usi verbali, usi nominali predicativi, usi aggettivali, usi polirematici (verbali, nominali, aggettivali, avverbiali e preposizionali);
- grammatiche locali in forma di automi a stati finiti (strutture semi-fisse, trasformazioni di frase, forme agglutinate, avverbi).

Per quanto riguarda le grammatiche elettroniche si vedano i paragrafi successivi. Qui di seguito diamo, invece, un esempio di stringhe del dizionario elettronico delle forme flesse:

amo,amare,V+t+a+PR+1+s  
amo,N+m+s  
cortese,A+f+s  
cortese,A+m+s  
di,PREP  
dottore,N+m+s  
dottoressa,dottore,N+f+s

dottorresse,dottore,N+f+p  
 dottori,dottore,N+m+p  
 lentamente,AVV

e un esempio di tavola lessico-grammaticale:

	N0 =: Num	N0 =: il fatto Ch F	N0 =: V1 Comp		V	N0 V che F a N2	F0 V che Fcong a N2	N0 V di V0 Comp a N2	N0 V di V2 Comp a N2	il fatto Ch F a N2	N1 =: se F o se F	N1 =: N1 V1-inf Comp	N0 V a N2 di N3	Passivo
-	+	+		apportare	+	-	-	+	+	-	-	-	-	+
+	-	-		articolare	+	+	+	+	-	+	-	-	-	+
+	-	-		assegnare	-	+	-	+	+	-	-	-	-	+
+	-	-		asserire	+	-	+	-	+	-	-	-	-	+
+	+	+		assicurare	+	-	+	-	+	-	-	-	-	+
+	-	-		attribuire	+	-	-	+	+	-	-	-	-	+

Nella tavola binarie il “+” e il “-“ indicano l’accettabilità o l’inaccettabilità delle proprietà indicate nelle colonne in alto. Le prime tre colonne a sinistra indicano proprietà relative al sintagma N0, le nove colonne a destra del verbo sono relative a diverse costruzioni (indicativo o congiuntivo nella completiva, altre forme di subordinazione possibili, passivo).

### 3.1. Le trasformazioni

I predicati di trasferimento possono presentarsi non solo in forma attiva, ma anche sotto forma di frasi trasformate, ad esempio le frasi risultanti dall’applicazione della passivizzazione e della pronominalizzazione. Così, se consideriamo una delle frasi presenti nel testo esemplificativo e cioè:

(1) *Luciano diede il pacchetto a Carlo*

vediamo che potrebbe presentarsi in forma passiva:

*Il pacchetto fu dato a Carlo da Luciano*

o in forma pronominalizzata:

*Luciano lo diede a Carlo*

Nei paragrafi che seguono prenderemo in esame la possibilità di mettere in correlazione automaticamente le frasi attive, passive e pronominalizzate prima di etichettare i predicati semantici.

#### 3.1.1. Passive

In un testo è possibile trovare forme di frasi passive in cui uno o più argomenti possono essere omessi. Così, dato l’esempio di frase (1) avremo :

- (2) *Il pacchetto fu dato a Carlo da Luciano*
- (3) *Il pacchetto fu dato da Luciano*
- (4) *Il pacchetto fu dato a Carlo*
- (5) *Il pacchetto fu dato*

Il verbo *dare* nella sua forma attiva di (1) ha una struttura predicativa del tipo T (d,o,r) che assegna al gruppo nominale a sinistra del predicato il ruolo di DATORE , al gruppo nominale immediatamente a destra il ruolo di OGGETTO e al gruppo preposizionale il ruolo di RICEVENTE. Nella forma passiva tale assegnazione non è corretta, infatti nelle frasi (2)-(5), tale predicato realizza tutti o solo alcuni dei suoi argomenti, secondo un ordine diverso da T (d, o, r):

- in (2): T (o, r, d)
- in (3): T (o, d)
- in (4): T (o, r)
- in (5): T (o)

E' necessario quindi individuare la forma di frase passiva e metterla in correlazione con la frase attiva. Solo dopo questa operazione sarà possibile etichettare correttamente i predicati. NOOJ permette di sviluppare delle grammatiche non solo flessive e morfologiche, ma anche sintattiche che sono in grado di mettere in correlazione frasi risultanti dall'applicazione di operazioni di sostituzione e spostamento come la passiva (si veda la nota 14). E' possibile costruire una grammatica che riconosce le frasi passive e genera le corrispondenti frasi attive.

La grammatica qui di seguito è semplificata e si limita, in questa sede, a DATORI e RICEVENTI indicati da nomi propri, ma è possibile prevedere gruppi nominali più complessi:

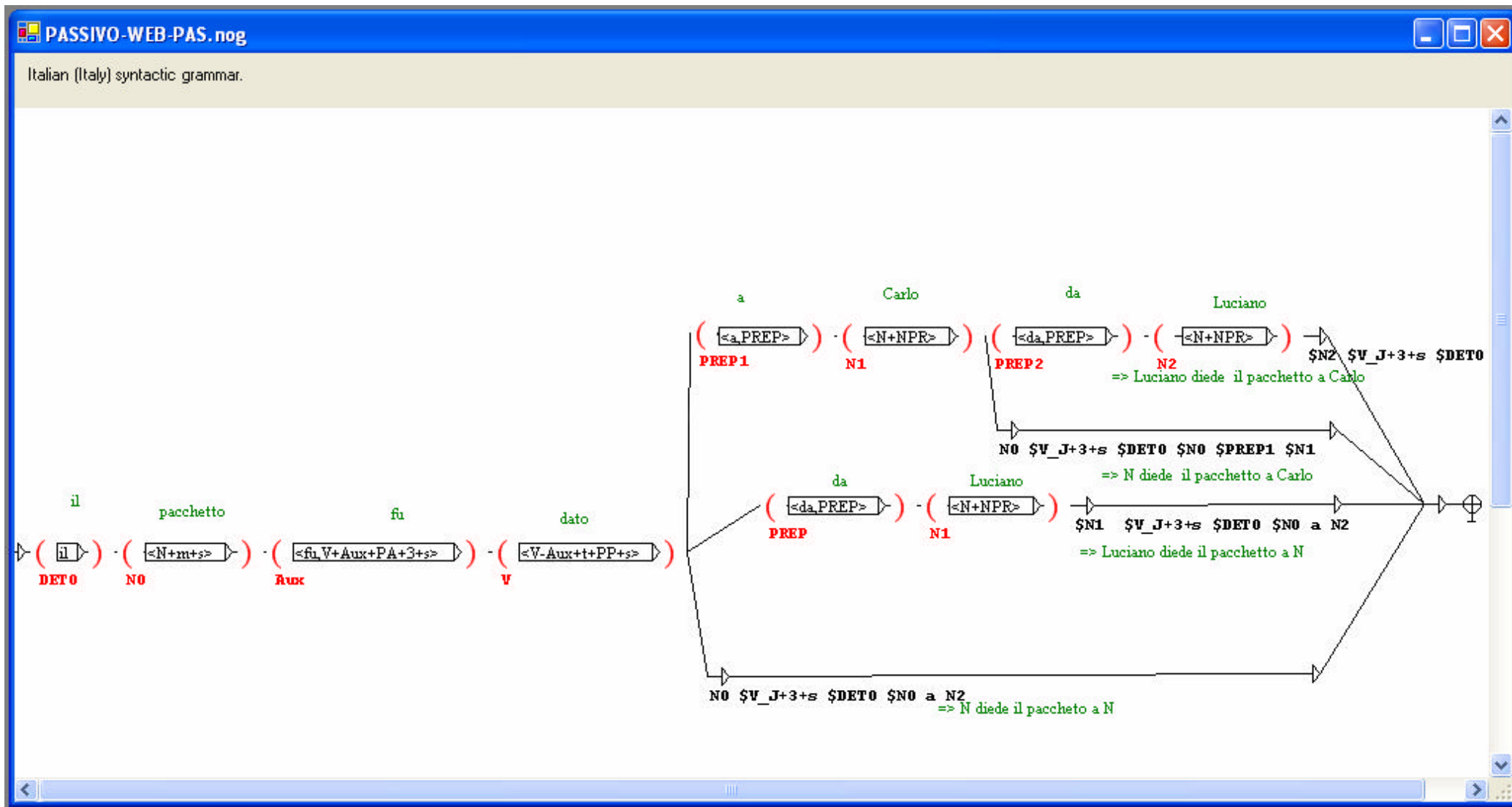


Figura 1: esempio di grammatica per il passivo

L'applicazione di questa grammatica agli esempi di frasi (2)-(5) dà in output le seguenti correlazioni tra frasi passive e frasi attive:

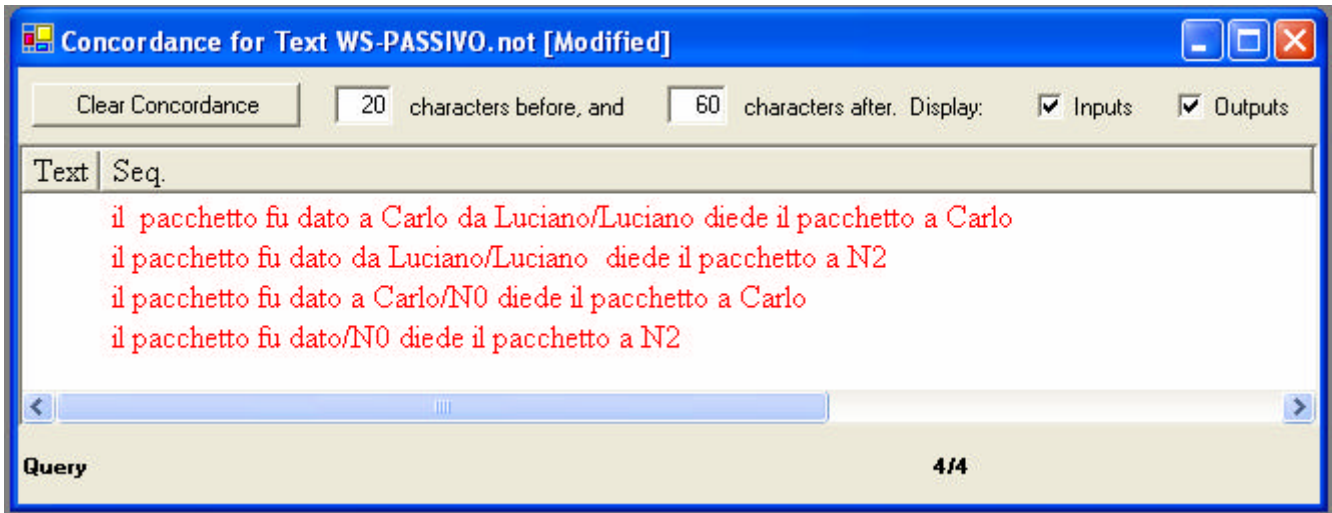


Figura 2: esempio di correlazione passivo-attivo

Nell'esempio (3), il RICEVENTE, che non è espresso, sarà recuperato nella sua forma attiva con la notazione N2; nell'esempio (4) sarà il DATORE, argomento vuoto, ad essere indicato come N0, mentre nell'esempio (5) saranno sia il DATORE sia il RICEVENTE ad essere indicati nella forma attiva. In base alla forma attiva sarà possibile etichettare i predicati in modo corretto.

### 3.1.2. Pronominalizzazioni

Le frasi possono presentarsi nei testi sotto diverse forme di frasi pronominalizzate, in questo caso considereremo la cliticizzazione. Data la frase (1), che qui riportiamo:

(1) *Luciano diede il pacchetto a Carlo*

possiamo applicare la cliticizzazione all'OGGETTO (=il pacchetto):

(1a) *Luciano lo diede a Carlo*

al RICEVENTE (= Carlo):

(1b) *Luciano gli diede il pacchetto*

sia all'OGGETTO che al RICEVENTE:

(1c) *Luciano glielo diede*

Negli esempi di cui sopra, l'ordine in cui il predicato *dare* realizza i suoi argomenti è il seguente:

- in (1a): T (d, o, r), in cui l'OGGETTO precede il verbo
- in (1b): T (d, r, o), in cui il RICEVENTE precede il verbo
- in (1c): T (d, r, o), in cui sia l'OGGETTO sia il ricevente precedono l'OGGETTO

Anche in questo caso, prima di etichettare il testo, sarà necessario mettere in correlazione le frasi che sono il risultato dell'applicazione della cliticizzazione con le frasi a cui non sono state applicate le operazioni di sostituzione e spostamento. La grammatica che permette tale operazione di riconoscimento e generazione di frasi è la seguente:

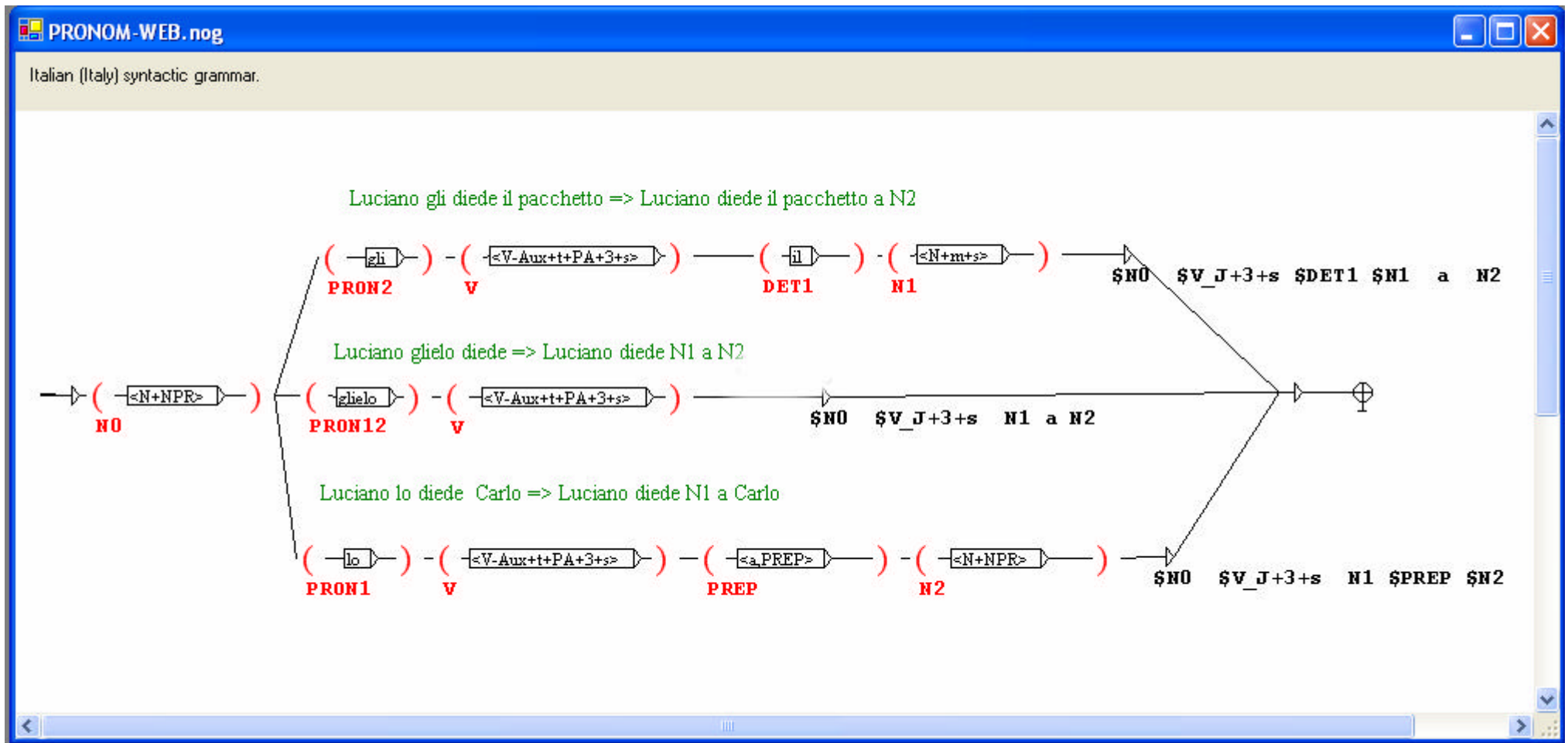


Figura 3: esempio di grammatica per la pronominalizzazione

L'applicazione di questa grammatica agli esempi di frasi (1a)-(1c) dà in output le seguenti correlazioni tra frasi pronominalizzate e frasi attive:

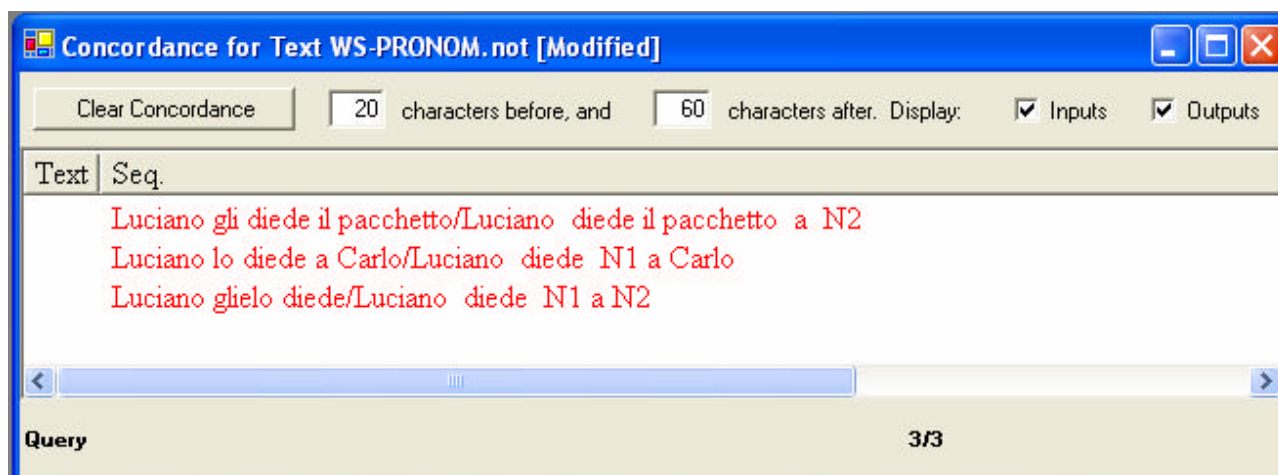


Figura 4: esempio di correlazione pronominalizzazione

In base alla forma non pronomi nalizzata sarà possibile etichettare i predicati in modo corretto.

Immaginiamo che in un testo occorran le frasi:

*Il pacchetto fu consegnato a Carlo. Luciano lo diede a Carlo*

Dopo aver applicato le grammatiche “Passivo” e “Pronominalizzazione”, il testo etichettato sarà:

*Il pacchetto (o) fu consegnato (T) a Carlo (r). Luciano (d) lo (o) diede (T) a Carlo (r)*

### 3.1.3. Le forme agglutinate

In italiano, la cliticizzazione applicata a forme verbali all’infinito, al gerundio, al participio passato e all’imperativo prevede che il clitico non sia preposto alla forma verbale, ma posposto e direttamente legato ad essa, dando origine ad un’unica parola, come negli esempi:

*Dargli il pacchetto, significava per Luciano....*

*Dandolo a Carlo, Luciano.....*

*Datoglielo, ....*

*Dallo a Carlo, disse.....*

In fase di analisi automatica, il riconoscimento di tali forme pone un problema in più rispetto a forme verbali come *significava* o *disse* perché queste ultime sono immediatamente riconducibili al lemma, in questo caso *significare* e *dire*; al contrario, parole come *datoglielo*, *dallo* o *dargli* possono essere associate al lemma *dare* solo dopo averle analizzate come sequenze di VERBO+PRONOME.

La grammatica che segue, ridotta e semplificata per ragioni di spazio, permette il riconoscimento delle forme agglutinate al gerundio; grafi simili possono essere costruiti per le altre forme verbali.

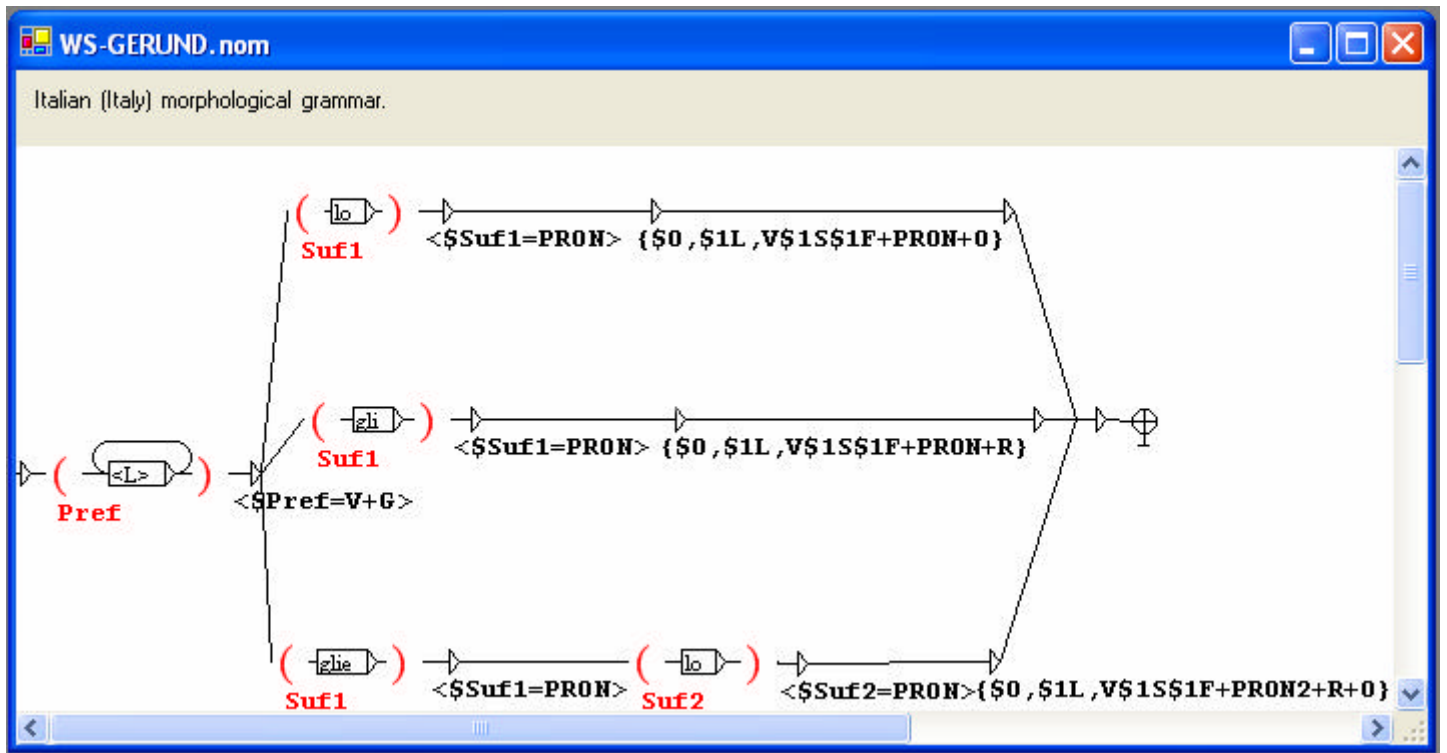


Figura 5: esempio di grammatica per le forme agglutinate

Se, in fase di analisi morfo-lessicale applichiamo tale grammatica a frasi del tipo:

- Dandoglielo, Luciano sorrise*
- Dandolo a Carlo, Luciano sorrise*
- Dandogli il pacchetto, Luciano sorrise*

NOOJ riconoscerà e lemmatizzerà le forme lessicali nel seguente modo:

```

Untitled [Modified]
Vocabulary contains 15 entries

# INFLECTIONAL AND DERIVATIONAL DESCRIPTION FILES I
#
# Special Features: +NW (non-word) +UNAMB (unambiguous)
#                   +FLX= (inflectional paradigm)
#
# Special Characters: '\ ' ' " ' + ' , ' # ' ' '
#
a, PREP
Carlo, N+NPR
Dandogli, dare, V+t+a+G+PRON+R
Dandoglielo, dare, V+t+a+G+PRON2+R+O
Dandolo, dare, V+t+a+G+PRON+O
il, DET+m+s
Luciano, N+NPR
pacchetto, N+m+s
sorrise, sorridere, V+i+a+PA+3+s

```

Figura 6: esempio di lemmatizzazione



Come si può vedere la forma *dandogli* è associata al verbo *dare*, un Verbo transitivo che richiede l'ausiliare *avere*, nella forma al Gerundio (V+t+a+G) seguito da un pronome con funzione di RICEVENTE (PRON+R). Allo stesso modo *dandoglielo* è associato al verbo *dare* (V+t+a+G) seguito da due pronomi che hanno, rispettivamente, funzione di RICEVENTE e OGGETTO (PRON2+R+O). Infine, la forma *dandolo* è un verbo al gerundio seguito da un Pronome con funzione OGGETTO.

Sulla base di questa analisi sarà possibile poi etichettare correttamente il testo in base ai predicati semantici.

## Conclusioni minime

Come abbiamo cercato di mostrare, la descrizione esaustiva degli usi lessico-grammaticali di una lingua, associata ad un dizionario elettronico morfosintattico ed a svariate grammatiche ad automi finiti per la soluzione di ambiguità o l'individuazione di trasformazioni linguistiche "locali", può permettere l'analisi in Predicati semantici, linguisticamente definiti, e il conseguente trattamento automatico del contenuto argomentativo di un testo non etichettato preventivamente.

Gli strumenti linguistici da mettere a punto sono molti e di notevole complessità<sup>17</sup>, ma essi presentano il vantaggio di essere validi per fenomeni che possono presentarsi in grande quantità all'interno di milioni di testi reali.

Noi crediamo che lo sviluppo di una semantica costruita su basi empiriche, cioè lessico-grammaticali, offra delle buone prospettive di evoluzione alle problematiche messe in campo dal movimento del Web Semantico.

## Bibliografia

Berners-Lee T., Hendler J., Lassila O. (2001), *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, "Scientific American", 284(5)

Boons J.P., Guillet A., Leclère C. (1976). *La structure des phrases simples en français. Constructions intransitives*, Droz: Genève

Courtois B. (1990a), *Un système de dictionnaires électroniques pour les mots simples du français*, "Langue Française" n.87, Paris: Larousse

Courtois B. (1990b), *Dictionnaire DELAS par classes flexionnelles*. Rapport Technique du LADL n. 26, Université Paris 7 : Paris.

Courtois B., Silberztein M. (1990a), *Développement d'outils de navigation dans les grandes documentations*, Rapport Technique du LADL n. 28, Université Paris 7 : Paris

Courtois B., Silberztein M. (eds.), (1990b), *Les dictionnaires électroniques du français*. "Langue Française" n. 87, pp. 11-22, Larousse: Paris.

D'Agostino E. (1992), *Analisi del discorso*, Loffredo: Napoli

D'Agostino, E. (1995), cura di, *Tra sintassi e semantica. Descrizione e metodi di elaborazione automatica della lingua d'uso*, ESI: Napoli

- D'Agostino E., Elia A. (1998), *Il significato delle frasi: un continuum dalle frasi semplici alle forme polirematiche*, in AAVV, "Ai limiti del linguaggio", Laterza: Bari
- Elia A. (1984), *Le verbe italien*. Schena-Nizet: Bari - Paris
- Elia A. (1995), *Per una disambiguazione semi-automatica di sintagmi composti: i dizionari elettronici lessico-grammaticali*, in Cipriani R., Bolasco S., a cura di, "Ricerca qualitativa e computer", Franco Angeli: Milano
- Elia, A. (1995b), *Standards informatici per la lessicografia computazionale*, in AAVV, "Lessico e grammatica", Bulzoni: Roma
- Elia A. (1995c), *Dizionari elettronici e applicazioni informatiche*, in Bolasco S., Lebart L., Salem A., a cura di "JADT 1995, III Giornate internazionali di Analisi Statistica dei Dati Testuali", CISU: Roma
- Elia A., Renzi L. (1995d), *Il problema delle reggenze*, in AAVV, "Lessico e grammatica", Bulzoni: Roma 1995
- Elia A. (1996), *Elogio dell'imperfezione*, in Gambarara D., Gensini S., Pennisi A., a cura di, "Language Philosophies and the Language Sciences", Nodus Publikationen: Münster.
- Elia A. (2002), *Discorso scientifico e linguaggio settoriale. Un esempio di analisi lessico-grammaticale di un testo neuro-biologico*, in "Quaderni del Dipartimento di Scienze della Comunicazione – Università di Salerno", Cicalese A., Landi A., a cura di, "Simboli, linguaggi e contesti", n. 2, Carocci: Roma
- Elia A., Martinelli M., D'Agostino E. (1981), *Lessico e Strutture sintattiche*, Liguori: Napoli
- Fuiano L. (2004), *Semantic Web e approccio lessico-grammaticale*, in "Studi Italiani di Linguistica Teorica e Applicata", n.1, Pacini: Roma
- Gross M. (1975), *Méthodes en syntaxe*, Hermann: Paris
- Gross M. (1978), *Observations on Semantic Theories*. "Theoretical Linguistics" 5:1, p. 3-17
- Gross M. (1979), *On the failure of Generative Grammar*, in "Language" 55 :4
- Gross M. (1981), *Les bases empiriques de la notion de prédicat sémantique*, in "Langages" n. 63, Larousse: Paris
- Gross M. (1990), *Sur la notion harissienne de transformation et son application au français*, in "Langages" n. 99, Larousse: Paris
- Gross M. (2002), *Le temps grammatical et le temps qui s'écoule*, LADL: Paris
- Guillet A., La Fauci N. (1984), eds, *Lexique-Grammaire des Langues Romanes*, John Benjamins: Amsterdam/Philadelphia
- Guillet A., Leclère C. (1992), *La structure des phrases simples en français. Constructions transitives locatives*, Droz: Genève

- Harris Z.S. (1976), *Notes du cours de syntaxe*, Editions du Seuil: Paris
- Harris Z.S. (1978), *Operator-grammar of English*, in "Linguisticae Investigationes", II, 1, John Benjamins: Amsterdam/Philadelphia.
- Harris Z.S. (1982), *A Grammar of English on Mathematical Principles*, Wiley: New York
- Harris Z.S. (1988-95), *Linguaggio e informazione*, Adelphi: Milano
- Leclère C. (2002), *Organization of the Lexicon-Grammar of French verbs*, "Linguisticae Investigationes" 25:1, John Benjamins: Amsterdam/Philadelphia.
- Leclère C. (2005), *The lexicon-grammar of French verbs: a syntactic database*, In "Linguistic Informatics - State of the Art and the Future", Kawaguchi Y. et alii (eds.), Tokyo University of Foreign Studies, UBLI 1, John Benjamins: Amsterdam/Philadelphia.
- Silberztein M. (1993.), *Dictionnaires électroniques et analyse automatique de textes*, Masson: Paris.
- Silberztein M. (1994), *INTEX: a corpus processing system*. In "Proceedings of COLING 1994": Kyoto.
- Silberztein M. eds. (2004), *INTEX pour la Linguistique et le Traitement Automatique des Langues*. "Proceedings of the 4th and 5th INTEX workshop", Bordeaux, May 2001, Marseille, May 2002, Presses Universitaires de Franche-Comté: Besançon.
- Silberztein M. (2004), *NooJ: a cooperative object oriented architecture for NLP*, "Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues", Muller C., Royauté et M. Silberztein (éds.), Presse Universitaire de Franche-Comté: Besançon.
- Silberztein M. (2005), *NooJ's dictionaries*, in Zygmunt Vetulani (ed.), "Proceedings of the 2nd Language & Technology Conference, April 21-23, 2005, Poznań, Poland", Wydawnictwo Poznanskie Sp. z o.o.
- Tesnière L. (1953), *Esquisse d'une syntaxe structurale*, Kincksieck: Paris
- Vietri S. (1984), *Lessico e sintassi delle espressioni idiomatiche. Una tipologia tassonomica in italiano e in inglese*, Liguori: Napoli
- Vietri S. (1990), *La sintassi delle frasi idiomatiche*, in "Studi italiani di linguistica teorica e applicata", XIX:1, Liviana Ed: Padova
- Vietri S. (1996), *The syntax of the Italian verb essere Prep.*, in "Linguisticae Investigationes", XX:2 John Benjamins: Amsterdam/Philadelphia
- Vietri S. (2001), *Navigare nei testi. Teorie e applicazioni informatiche per la linguistica testuale*. Editoriale Scientifica Italiana: Napoli.
- Vietri S., Elia A. (2001), *Analisi automatica dei testi e dizionari elettronici*, in Cordeschi R., Burattini E., ed. 2001. "Intelligenza Artificiale. Manuale per le discipline della comunicazione". Carocci: Roma.

Vietri S., Elia A., D'Agostino E. (2004), *Lexicon-grammar, Electronic Dictionaries and Local Grammars in Italian*, in Laporte, E., Leclère, C., Piot, M., Silberztein M., eds.. "Syntaxe, Lexique et Lexique-Grammaire" Volume dédié à Maurice Gross. *Linguisticae Investigationes Supplementa* 24, John Benjamins: Amsterdam/Philadelphia

Vietri S. (2004), *Lessico-grammatica dell'italiano. Metodi, descrizioni, applicazioni*. UTET: Torino

Vietri S. (2004), *Lemmatization of Idioms in Italian*, in Muller C., Royauté J., Silberztein M., eds. 2004. "INTEX pour la Linguistique et le Traitement Automatique des Langues. Proceedings of the 4th and 5th INTEX workshop", Bordeaux, May 2001, Marseille, May 2002, Presses Universitaires de Franche-Comté: Beausson

Vasseux P. (1982), *Le système LEXSYN de classification des données du LADL*. Rapport interne du LADL, Université Paris 7: Paris

---

<sup>1</sup> Annibale Elia è l'autore dei paragrafi 1, 2 e dell'Introduzione, Simonetta Vietri è l'autore del paragrafo 3 e delle Conclusioni minime.

<sup>2</sup> Si veda Berners-Lee et Alii 2001.

<sup>3</sup> Si veda Fuiano 2004, in appendice, per un rapido confronto linguistico tra HTML e XML.

<sup>4</sup> All'Università di Stanford è stato elaborato un Portale dei Vini che sperimenta un "agente" specializzato nell'Ontologia dei Vini e che simula l'interazione con un' Ontologia delle Pietanze. Per un approfondimento si veda il sito <http://www.ksl.stanford.edu/people/dlm/webont/wineAgent/>.

<sup>5</sup> Si veda M. Gross, 1978, 1981, che per primo ha impostato il tema dei predicati semantici nell'ottica lessico-grammaticale.

<sup>6</sup> Si veda Fuiano 2004 per un primo approccio al problema.

<sup>7</sup> Per una versione recente del Lessico-grammatica dell'italiano si veda Vietri 2004. Per una visione d'insieme delle applicazioni lessico-grammaticali nel mondo si veda il sito Internet dell'Università di Marne la Vallée: <http://infolingu.univ-mlv.fr/>.

<sup>8</sup> Il termine attante, che si deve a Lucien Tesnière 1953, è usato largamente nella teoria lessico-grammaticale per evidenziare che il verbo intrattiene relazioni comparabili sia con il "soggetto" che con i suoi "complementi".

<sup>9</sup> Il sistema di notazione del Lessico-grammatica è illustrato in Elia A., Martinelli M., D'Agostino E. 1981. In questa sede, ci limitiamo a ricordare che *N* indica il gruppo nominale, i numeri 0, 1 e 2 si riferiscono alla posizione del gruppo nominale nella struttura di frase: *N0* =:"soggetto", *N1*, *N2*, *N3* =: "complemento", *Loc* indica una qualsiasi preposizione locativa, *Ch F* indica una completiva generica, *di \*Ch F* indica l'inesistenza di una sequenza preposizionale in *di* per le completive che sono correlate ad un gruppo nominale preposizionale in *di*. Le classi lessico-grammaticali sono numerate per macro-differenze: da 1 a 15 riguardano gli usi verbali cosiddetti intransitivi che non accettano una completiva o una infinitiva, da 16 a 40 riguardano gli usi verbali cosiddetti transitivi che non accettano una completiva o una infinitiva, da 41 a 60 riguardano gli usi verbali cosiddetti transitivi o intransitivi che accettano una completiva o una infinitiva, le classi finora elaborate sono circa 50, le proprietà combinatorie esaminate per ogni uso verbale sono circa 150.

<sup>10</sup> Esistono anche i Predicati di Movimento (M) semplici in cui l'Agente del Movimento (am) trasferisce se stesso da un luogo di provenienza (lp) ad un luogo di destinazione (ld) : *Max (am) corre (M) fino al ponte (ld)* con struttura lessico-sintattica *N0 V Loc N1*.

<sup>11</sup> I numeri accanto ai verbi stanno a indicare che esistono usi verbali con diversa struttura lessico-sintattica, come accade per *fittare*: *Mario fitta una casa a Lia*, *Lia fitta una casa da Mario* .

<sup>12</sup> Nella frase *Ale ha acquistato un libro da Max*, *Max* non è più il proprietario del libro dal momento che, con la vendita, ha trasferito il diritto di proprietà ad *Ale*.

<sup>13</sup> Si veda, per il francese, M.Gross 1975, Ch.Leclère 2002, 2005.

<sup>14</sup> I primi programmi automatici per la gestione dei dizionari elettronici e delle tavole lessico-grammaticali furono elaborati negli anni '70 da Philippe Vasseux e Nathalie Bely prima e, in un secondo momento, da Blandine Courtois al Laboratoire d'Automatique Documentaire et Linguistique del CNRS presso l'Université de Paris 7. Successivamente, per diversi anni, hanno lavorato all'ideazione di componenti nuove per l'elaborazione dei dati molti informatici e linguisti. Un primo programma integrato, INTEX, fu realizzato da Max Silberztein negli anni '90. Dopo la scomparsa di Maurice Gross, nel 2001, un gruppo di informatici, legato all'Université de Marne La Vallée, diretto da Eric Laporte, ha elaborato il programma UNITEX, dalle funzionalità simili a quelle di INTEX. Max Silberztein, che ha un suo gruppo di ricerca all'Università di Franche-Comté, ha elaborato NOOJ, un nuovo programma per l'analisi morfosintattica automatica dei testi che rappresenta un'evoluzione di INTEX. Il sito Internet <http://perso.orange.fr/rosavram/pages/noojpag.html> descrive NOOJ e permette di scaricare versioni demo in alcune lingue, tra cui l'italiano. Questi programmi, per realizzare applicazioni di analisi testuale hanno bisogno di dizionari

---

elettronici morfosintattici, di tavole lessico-grammaticali e di grammatiche locali, elaborati negli ultimi trent'anni da diverse équipes di linguisti per svariate lingue (dal francese, l'italiano e l'inglese al serbo e al coreano).

<sup>15</sup> Per una presentazione generale delle ricerche lessico-grammaticali nel mondo rinviamo al sito Internet <http://infolingu.univ-mlv.fr/>, in cui si trovano indicazioni esaurienti sia dal punto di vista bibliografico che da quello sitografico e computazionale. Inoltre nel sito c'è una lista dei maggiori laboratori che applicano la stessa metodologia per il trattamento linguistico computazionale delle lingue.

<sup>16</sup> Per gli aspetti più propriamente computazionali è utile anche Vietri 2002, Vietri, Elia 2001.

<sup>17</sup> Non abbiamo affrontato, per esempio, le procedure necessarie per costruire grammatiche locali efficaci per estrarre la presenza di un nome all'interno di un qualsiasi sintagma nominale reale. I grafi che sono stati realizzati per tale scopo, infatti, hanno presupposto un dettagliato studio descrittivo e classificatorio della sintassi e del lessico dei predeterminanti di un nome.