

# Banche dati e strumenti di analisi testuale per il progetto “Donne e politica”

*Manuela Sassi*

Istituto di Linguistica  
Computazionale del CNR di Pisa

- ◆ L' Istituto di Linguistica Computazionale del CNR (ILC-CNR) opera nel settore sin dal 1967, anno in cui fu formata la Divisione Linguistica presso il Centro Nazionale Universitario di Calcolo Elettronico di Pisa (CNUCE oggi ISTI-CNR);
- ◆ nel 1978 viene trasformato in istituto della rete CNR.

- ◆ Nei 40 anni di attività l'ILC-CNR è stato uno dei maggiori promotori della nozione di *risorse linguistiche* come componente centrale della "infrastruttura linguistica" (consapevole anche delle sue implicazioni culturali, sociali, economiche e politiche),
- ◆ ha coordinato le più importanti iniziative relative alle risorse linguistiche ed alla standardizzazione,
- ◆ è stato spesso promotore di nuovi "paradigmi" nel settore.

# La collaborazione con ASDO

- ◆ Nell'ambito del progetto "Donne e politica" è stato attivato un osservatorio volto a studiare le dinamiche esistenti nel processo elettorale.
- ◆ Uno degli oggetti studiati dall'osservatorio è costituito dal corpus di 712 articoli selezionati nel periodo gen-mag 2006, per un totale di 442.796 parole (di cui 6477 sono rappresentate da numeri).
- ◆ Il corpus è stato organizzato in banca-dati on-line al seguente link:  
[serverdbt.ilc.cnr.it/singolo/start\\_don.html](http://serverdbt.ilc.cnr.it/singolo/start_don.html)

# Creazione del corpus

- ◆ Gli articoli selezionati con News di Google in formato elettronico sono stati marcati per essere indicizzati con il DBT (Brevetto CNR di E.Picchi) che permette di applicare ai testi una serie di strumenti computazionali ad-hoc.
- ◆ Quelli ottenuti dall'Eco in RTF sono stati pesantemente revisionati perché erano poco leggibili e hanno ancora molti refusi.
- ◆ Il corpus è stato poi "lemmatizzato" in modo automatico per un accesso ottimizzato alle informazioni contenute (lemmi e dati morfo-sintattici che permettono una sintesi migliore)

# Utilizzo del corpus

- ◆ Consultazione on-line dei testi con interfaccia web (vedi link) tramite funzioni mirate:
  - Ricerca per parole (e/o parti di esse)
  - Ricerca supportata da dizionario e/o thesaurus
  - Funzione di co-occorrenza
  - Possibilità di abbinamento di più parole (Famiglia di ricerca in AND e OR con indicazione del parametro della distanza fra i termini indicati)



**Param**

**HardFam**

**Guida**

### ● News "Donne e politica"

Occorrenze :	436319	Numero totale di parole nel testo
--------------	--------	-----------------------------------

N.Forme :	32353	Numero di forme diverse nel testo
-----------	-------	-----------------------------------

**Parole**

Ricerca parole nel testo

**Morfologia**

Ricerca nel testo con la morfologia

**Thesaurus**

Ricerca nel testo con thesaurus

**Punteggiatura**

Ricerca della punteggiatura nel testo





# PI SYSTEM

## DBT Data Base Testuale

**Quadro**

**Param**

**Guida**

Nr	- Forma -	Freq.		
<a href="#">1</a>	<input checked="" type="checkbox"/> femmina	1	<b>In Quadro</b>	Trasferisci in quadro
<a href="#">2</a>	<input checked="" type="checkbox"/> femmine	15	<b>FaiFam</b>	Creazione automatica di famiglia in OR
<a href="#">3</a>	<input checked="" type="checkbox"/> femminile	596	<b>Tutte</b>	Seleziona tutte le parole
<a href="#">4</a>	<input checked="" type="checkbox"/> femminili	118	<b>Nessuna</b>	Toglie la selezione a tutte le parole
<a href="#">5</a>	<input checked="" type="checkbox"/> femminilità	3	<b>Contesti</b>	Contesti di tutte le parole della pagina
<a href="#">6</a>	<input checked="" type="checkbox"/> femminilita'	2		
<a href="#">7</a>	<input checked="" type="checkbox"/> femminilizza	1		
<a href="#">8</a>	<input checked="" type="checkbox"/> femminilizzazione	4		
<a href="#">9</a>	<input checked="" type="checkbox"/> femminismo	67		
<a href="#">10</a>	<input checked="" type="checkbox"/> femminista	66		
<a href="#">11</a>	<input checked="" type="checkbox"/> femministe	37		
<a href="#">12</a>	<input checked="" type="checkbox"/> femministi	11		
<a href="#">13</a>	<input checked="" type="checkbox"/> femmis	3		



● **Contesti**

Parola corrente : **femminile ( 596 )**

**Chiudi**

**Continua**

- |    |  |           |  |
|----|--|-----------|--|
| 1  | dato dalla forte rappresentanza al     | femminile | , che mi fa particolarmente - <a href="#">20060120 arezzone.it.15</a>                |
| 2  | per sensibilizzarle sulla presenza     | femminile | nelle liste che vorranno formate - <a href="#">20060119 provincia.grosseto.it.14</a> |
| 3  | Da una maggiore presenza               | femminile | nei ruoli di responsabilità, - <a href="#">20060119 provincia.grosseto.it.38</a>     |
| 4  | regime fascista, l'impegno             | femminile | nella guerra e nella militanza - <a href="#">20060118 La Nazione/Lucca.16</a>        |
| 5  | teso alla ricostruzione dell'identità  | femminile | spesso offuscata nell'analisi storica - <a href="#">20060118 La Nazione/Lucca.19</a> |
| 6  | di vita e di impegno                   | femminile | nella vita di fabbrica, - <a href="#">20060118 La Nazione/Lucca.20</a>               |
| 7  | Lucchese Storia e Storie al            | femminile | . Ha pubblicato: La - <a href="#">20060118 La Nazione/Lucca.48</a>                   |
| 8  | il primo presidente di sesso           | femminile | in uno dei continenti più - <a href="#">20060118 L'Avanti!.18</a>                    |
| 9  | penultimo in Europa per rappresentanza | femminile | nelle Assemblee parlamentari. C' - <a href="#">20060118 L'Avanti!.29</a>             |
| 10 | tendenza? "La rappresentanza           | femminile | nelle istituzioni italiane è crollata - <a href="#">20060118 L'Avanti!.31</a>        |
| 11 | ha avuto una rappresentanza            | femminile | molto bassa. A ciò - <a href="#">20060118 L'Avanti!.39</a>                           |
| 12 | owero che la rappresentanza            | femminile | sarà cospicua come sarebbe accaduto - <a href="#">20060118 L'Avanti!.46</a>          |
| 13 | modo specifico, alla condizione        | femminile | ". Dopo l'elezione della - <a href="#">20060118 L'Avanti!.62</a>                     |
| 14 | con un governo al                      | femminile | ed un programma tagliato sulle - <a href="#">20060118 Libero.6</a>                   |
| 15 | una elezione agevole nel mondo         | femminile | . Il Movimento Difesa del - <a href="#">20060118 sudnews.it.6</a>                    |
| 16 | primo un partito tutto al              | femminile | . Si chiama Pde: - <a href="#">20060118 lavocedimilano.it.6</a>                      |
| 17 | trascurare l'importanza del ruolo      | femminile | nelle amministrazioni locali. Il - <a href="#">20060118 lavocedimilano.it.18</a>     |
| 18 | Pde, che si definisce                  | femminile | - etico -trasversale, - <a href="#">20060118 lavocedimilano.it.19</a>                |
| 19 | Partito d'Europa a LEADERSHIP          | FEMMINILE | 2. stravolge le regole - <a href="#">20060118 lavocedimilano.it.16</a>               |
| 20 | è l'esempio di civiltà                 | FEMMINILE | italiana che ha scritto una - <a href="#">20060118 lavocedimilano.it.21</a>          |
| 21 | un modello di società al               | femminile | ed in questo caso si - <a href="#">20060118 lavocedimilano.it.30</a>                 |



## ● Contesto largo

Parola corrente : **femminile( 596 )**

**Chiudi**

Contesti1

20060120 arezzone.it

impegno che li aspetta "Le decisioni che prenderete dovranno nascere dall'intera Assemblea. Avrete un compito importante, perché rappresenterete coloro che vi hanno eletto e che con la loro preferenza hanno permesso a tutti voi di essere qui. I vostri progetti andranno a vantaggio dell'intera comunità. Il consiglio Comunale dei ragazzi non va interpretato come un gioco, anche se spesso oggi la politica, quella "vera", assomiglia sempre più ad un teatrino. Vorrei infine sottolineare un fatto singolare, dato dalla forte rappresentanza al **femminile**, che mi fa particolarmente piacere". Si è poi proceduto all'elezione del Sindaco Junior: con tre preferenze è risultato eletto Matteo Tavanti della I E della Scuola Media "Dante Alighieri". "Ringrazio tutti coloro i quali mi hanno eletto, gli insegnanti, i miei amici, i miei genitori e tutti i presenti. La prima cosa che farò sarà quella di presentare al Sindaco Brandi il mio progetto relativo alla realizzazione di una Ludoteca - ha dichiarato appena eletto - sono contento che quasi la metà del Consiglio sia composto da ragazze,



**Quadro****Param****Guida****Famiglia Hard**

Ricerca nel testo con "Famiglia hard"

Elemento 1

rappresentanza

AND Elemento 2

femminile

AND Elemento 3

AND Elemento 4

AND Elemento 5

Le diverse celle riempite devono essere considerate in AND tra di loro mentre all'interno di ogni singola cella possono essere inserite più ricerche che saranno considerate in OR tra di loro. Le ricerche possono consistere in singole forme, in espressioni con \*, \$ e parentesi secondo la sintassi di DBT, di entrate lessicali per utilizzare la morfologia dell'italiano disponibile in linea (per indicarle scrivere "L=" seguita dal lemma desiderato) e da elementi del thesaurus (sinonimi) per i quali ricercare la catena sinonimale presente e ottenere la generazione morfologica di ogni sinonimo ottenuto (per indicarli scrivere "S=" seguita dal sinonimo desiderato)



22	storica, con la scarsa	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	nelle cariche elettive e i insufficiente - <a href="#">ZUUBUJUB womenews.net.49</a>
23	il suo mandato, la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	era di appena il 10%. - <a href="#">20060306 novionline.net.15</a>
24	la delusione per la scarsa	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	nelle liste elettorali per il Senato - <a href="#">20060308 Il Giornale.3</a>
25	posto nel mondo per la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	nelle assemblee elettive. Pieno sostegno - <a href="#">20060308 Corriere.19</a>
26	Nelle liste Ds la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	è del 35%, dentro Rifondazione - <a href="#">20060308 Il Tempo.25</a>
27	vostro voto il riequilibrio della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. Non voglio parlare qui della - <a href="#">20060308 Il Tempo.48</a>
28	più che triplicare la sua	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	in parlamento (oggi infatti ne - <a href="#">20060308 Corriere.27</a>
29	italiano e il ruolo della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	in politica. L'iniziativa, - <a href="#">20060310 INFORM.7</a>
30	blocco non riusciremo ad avere	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	forte. Sulle riforme se uno - <a href="#">20060315 L'Unita`.71</a>
31	rosa" per aumentare la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	in Parlamento si è quindi tradotta - <a href="#">20060315 quaderniradicali.it.11</a>
32	An si è rivolta alla	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. «La destra di Alleanza - <a href="#">20060318 liberta.it.12</a>
33	Europa non può mancare la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. Per questo sosteniamo il voto - <a href="#">20060320 INFORM.31</a>
34	oneroso compito di risollevere la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	, oggi a livelli mortificanti: - <a href="#">20060320 panorama.it.9</a>
35	"Cronaca della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	politica " Salerno - Grande successo di - <a href="#">20060321 casertanews.it.1</a>
36	in 14 legislature della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	politica ", promosso dalle Consigliere di Parità - <a href="#">20060321 casertanews.it.4</a>
37	accento sulle problematiche legate alla	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	in politica, utilizzando il forte - <a href="#">20060321 casertanews.it.7</a>
38	partiti sullo stato attuale della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	nelle istituzioni e nei partiti. - <a href="#">20060321 womenews.net.9</a>
39	sul tema della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	politica nei paesi della UE."Le - <a href="#">20060328 ilsannioquotidiano.it.4</a>
40	è a un aumento della	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. Un altro dato illuminante sulle - <a href="#">20060407 businessonline.it.20</a>
41	quale garanzia per una reale	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	". Anche l'ex ministra per - <a href="#">20060414 womenews.net.27</a>
42	maschile - ad allargare la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. Sul fatto che non abbiano - <a href="#">20060501 womenews.net.37</a>
43	che questa città avesse una	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	più ampia, oviamente, ma - <a href="#">20060501 womenews.net.132</a>
44	clii e garantita un'adeguata	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	si legge nel rapporto dell' - <a href="#">20060119 La Repubblica/PALERMO.43</a>
45	alle politiche sia garantita la	<a href="#">rappresentanza</a>	<a href="#">femminile</a>	. Stefania Presticiacomo, ministro per - <a href="#">20060120 Libero.49</a>



# PI SYSTEM

DBT Data Base Testuale

- [DefFam](#)
- [Param](#)
- [HardFam](#)
- [Guida](#)

## ● Quadro

Parola corrente : concilia ( 3 )

Nr.	Forma	Freq.
1	concilia	3
<a href="#">2</a>	conciliare	20
<a href="#">3</a>	conciliarle	1
<a href="#">4</a>	conciliazione	12

[Contesti](#)

[Sinistri](#)

[Destri](#)

[Parole](#)

Ricerca parole nel testo

[Morfologia](#)

Ricerca nel testo con la morfologia

[Thesaurus](#)

Ricerca nel testo con thesaurus

[Punteggiatura](#)

Ricerca della punteggiatura nel testo

[CoOccorrenze statistiche](#)

[Analisi preposizioni](#)

Dimensione contesti 10

Distanza x famiglie 10





**Quadro**

**Guida**

### ● Co-occorrenze

Nr	- Forma -	Freq.		
1	<input checked="" type="checkbox"/> concilia	3	<b>Tutte</b>	Seleziona tutte le parole
2	<input checked="" type="checkbox"/> conciliare	20	<b>Nessuna</b>	Toglie la selezione a tutte le parole
3	<input checked="" type="checkbox"/> conciliarle	1	ItalCorr.\$SW	Selezione di una lista di parole vuote
4	<input checked="" type="checkbox"/> conciliazione	12	<b>CoOccorrenze statistiche</b>	

Ricerca co-occorrenze

# PI SYSTEM

DBT  
Data  
Base

Quadro

Param

Guida

## CoOccorrenze

1	3	38	9.902	2.333	impegni
2	6	84	9.758	2.000	difficoltà
3	2	30	9.658	2.000	misure
4	4	113	8.745	2.000	tempi
5	2	69	8.457	3.000	maternità
6	2	79	8.261	2.000	persona
7	9	492	7.793	1.556	lavoro
8	2	111	7.771	1.000	devono
9	3	215	7.402	2.000	impegno
10	4	292	7.375	2.500	famiglia
11	2	239	6.664	2.000	tempo
12	2	271	6.483	1.500	casa
13	2	293	6.370	3.000	politico
14	2	573	5.403	3.000	fatto
15	2	3284	2.884	2.500	donne

< [ ] >

News

Cooccorrenze  
Statistiche

Mutual  
Information



### ● CoOccorrenze

CoOccorrenze

**Chiudi**

N. contesti

9

- 1 maschili, non sono congeniali alle donne, che devono **conciliare lavoro** e maternità, oppure lavoro e cura delle persone - [20060107 Corriere.45](#)
- 2 che i figli crescono e diventano autonomi si possono benissimo **conciliare lavoro** e famiglia, a condizione naturalmente che il partner - [20060325 INFORM.73](#)
- 3 uno che nel suo governo non ha fatto leggi per **conciliare lavoro** e domesticita` o responsabilita` comuni di uomini e donne - [20060419 peacelink.37](#)
- 4 proprio le politiche pensate per aiutare le donne che vogliono **conciliare lavoro** e maternità. Diversi studi dell'Ilo e l' - [20060220 La Repubblica.31](#)
- 5 è la precarietà del lavoro, è la difficoltà di **conciliare lavoro** e famiglia, la carenza dei servizi sociali per - [20060114 CGIL.101](#)
- 6 è la precarietà del lavoro, è la difficoltà di **conciliare lavoro** e famiglia, la carenza dei servizi sociali per - [20060111 emilianet.it.185](#)
- 7 che le donne siano poche perché già troppo impegnate a **conciliare** casa e **lavoro** è diffusa soprattutto tra gli elettori di - [20060307 Corriere.180](#)
- 8 maggiore partecipazione delle donne alla vita pubblica (misure di **conciliazione** famiglia **lavoro**, più asili nido, maggiore ripartizione dei - [20060307 Corriere.270](#)
- 9 in merito a temi di rilevante importanza come la **conciliazione** dei tempi **Lavoro-Famiglia**, la questione della rappresentanza - [20060109 ilquotidiano.it.13](#)

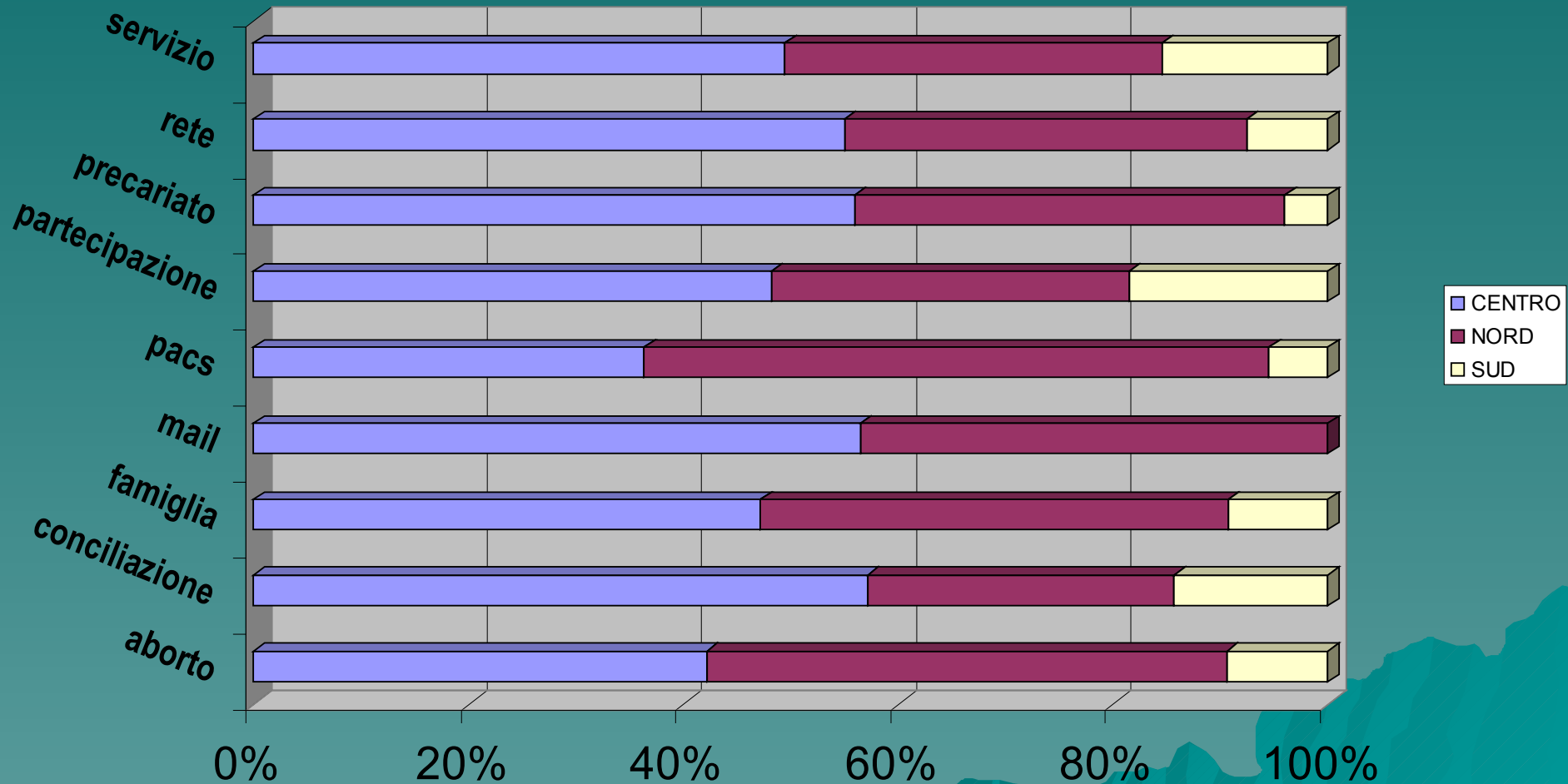


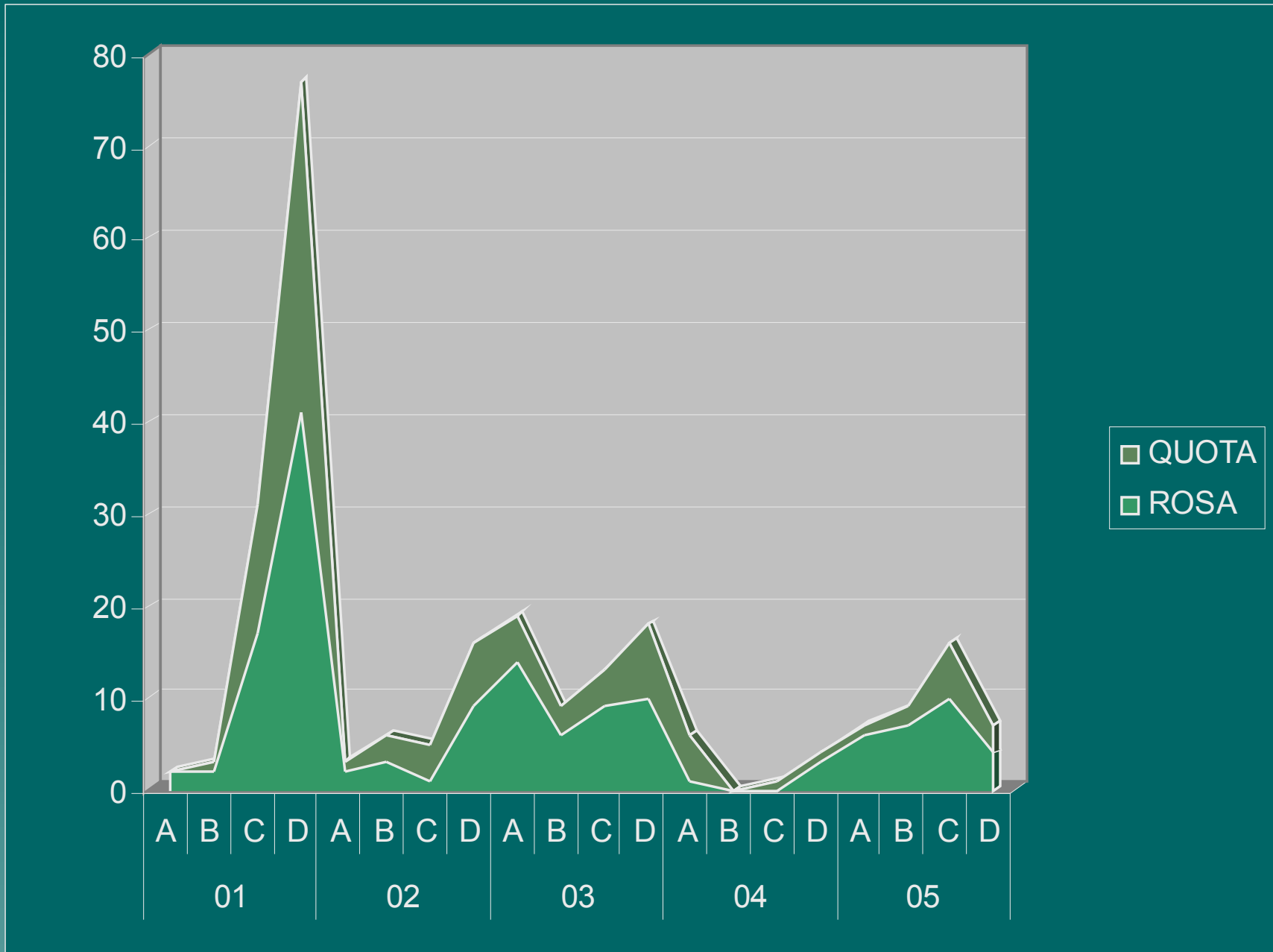


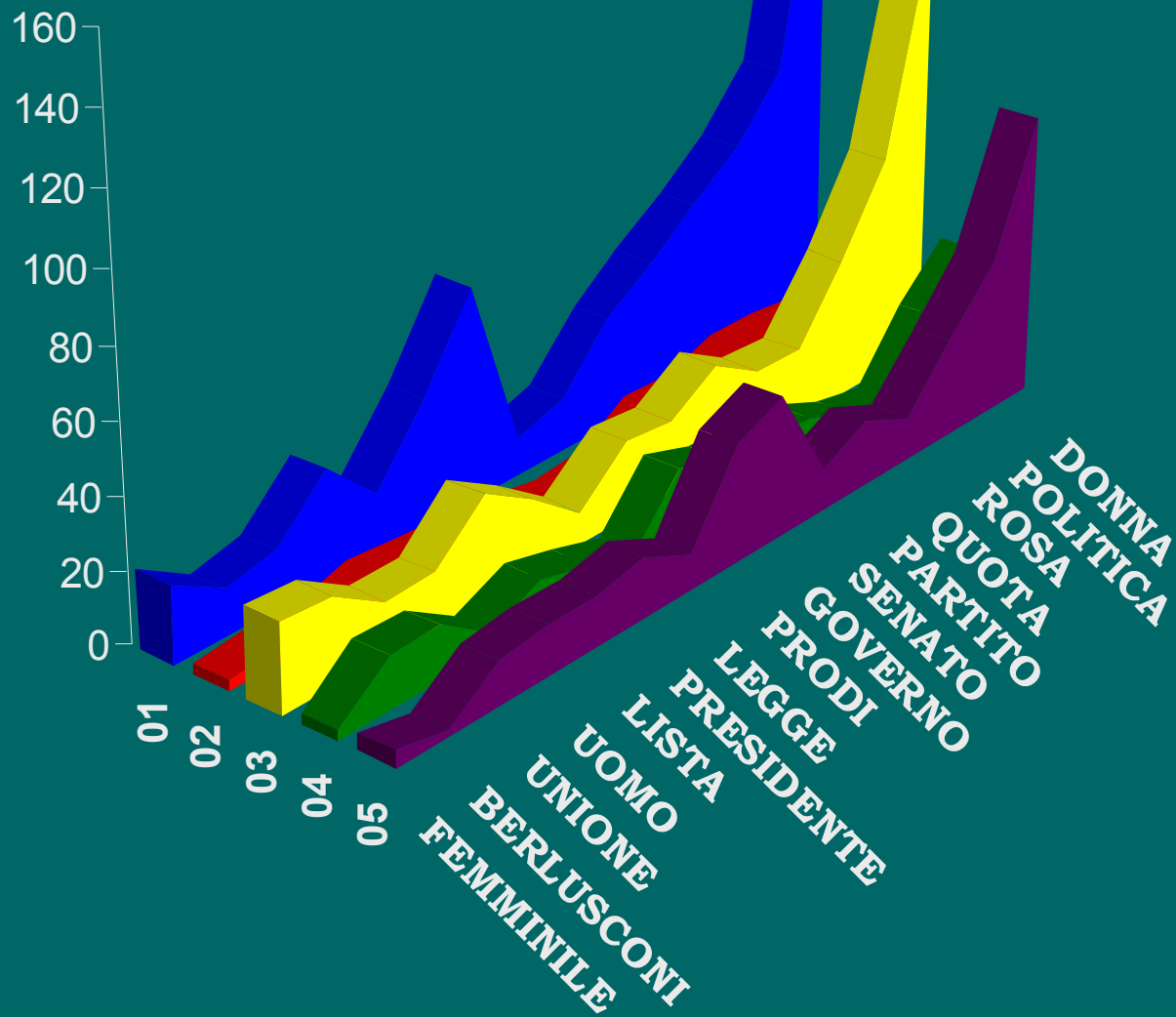
# Altre possibilità di uso del corpus

- ◆ Nella versione DBT-web è possibile copiare i risultati delle richieste in modalità testo.
- ◆ Nella versione su PC locale è disponibile la funzione di Salva-finestra che trasferisce i risultati in Word e su fogli di calcolo che possono essere usati per incrociare i dati numerici e generare tabelle e grafici illustrativi.

Esempio: i risultati quantitativi dei termini selezionati dallo studio effettuato sul corpus sono stati incrociati con la locazione delle testate







<p>PARTECIPAZIONE</p> <p>Questo è un esempio di elaborazione dei dati ottenuti dalla lemmatizzazione, tramite la quale è stato recuperato il sostantivo <i>partecipazione</i> con gli aggettivi ad esso riferiti.</p>	ATTIVO	ABANO.TV	1
		CASERTANEWS.IT	1
		CORRIERE	1
		GIORNAL.IT	1
		LIBERO	1
		MAREMMANEWS	1
		MARGHERITAONLINE.IT	1
		ORVIETOSI.IT	1
	BASSO	CORRIERE	1
	CONTINUATIVO	WW2.CARTA.ORG	1
	DIFFUSO	MARGHERITAONLINE.IT	1
		WW2.CARTA.ORG	1
	DIRETTO	GIORNALE	1
	DISTRETTO	LIBERTA.IT	1
EFFETTIVO	CORRIERE	4	
FEMMINILE	ASIANEWS.IT	1	
	CORRIERE	2	
	LIBERTA.IT	2	
	LUCACOSCIONI.IT	1	
	PANORAMA.IT	1	
	QN ECONOMIA E POLITICA	2	
	TODAY.REUTERS.IT	1	
	VOCEDIMILANO.IT	1	
WEB.RADICALPARTY.ORG	1		

# Conclusioni aperte

- ◆ Quanto esposto finora non è che una piccola parte delle possibilità offerte sia dal sistema presentato sia dalle competenze che si sono formate negli anni.
- ◆ Tali competenze rischiano di andare perdute se il nostro paese non attua una seria politica per la ricerca e per il passaggio delle consegne ai giovani.