

L'utilizzo didattico di corpora: proposte metodologiche

Sassi Manuela, Ceccotti Maria Luigia,
Istituto di Linguistica Computazionale del CNR
Via Moruzzi 1
56100 PISA
manuela@ilc.pi.cnr.it, luigia@ilc.pi.cnr.it

ABSTRACT

A partire da alcune brevi considerazioni sull'uso di corpora, viene qui proposta una metodologia di recupero dell'informazione in essi contenuta a scopo didattico e, perché no, anche di curiosità. Ma soprattutto si vuole sottolineare le possibilità che oggi offre la rete affinché studenti ed insegnanti possano documentarsi e recuperare materiali interessanti per la didattica della scrittura.

1. Introduzione

Linguistica dei corpora, *Corpus Linguistics* in inglese, è il settore della linguistica computazionale che si pone l'obiettivo di ottenere da grandi quantità di testi *in machine readable form* le più svariate informazioni, utili ad un ampio spettro di utenti potenziali: studenti, docenti, studiosi, scrittori, lessicografi, linguisti, critici letterari, traduttori, sociologi, politologi, etc.

I primi corpora sono stati progettati e faticosamente costruiti, negli anni '60 (Juilland, LIF) e in centri di ricerca, ed i loro creatori, utilizzando, per metodologie ed obiettivi collaudati da secoli, la rivoluzionaria elaborazione automatica dei dati, si proponevano fundamentalmente lo studio quantitativo del lessico. Dopo quarant'anni il pubblico ha a disposizione intere annate di giornali, dizionari, centinaia di testi su CD-ROM dal giornalaio, nel supermercato, in libreria, può recuperare via Internet i più svariati testi ma per quanto riguarda il settore della ricerca i corpora elettronici continuano ad essere utilizzati, per lo più, nei centri in cui sono stati realizzati per una cerchia ristretta di utenti.

Manca, ci pare, una rassegna esaustiva sullo stato dell'arte, che permetta di fare conoscere i problemi di questo settore, in cui si potrebbe tentare una catalogazione per 'contenuto', 'caratteristiche tecniche', 'caratteristiche redazionali', 'finalità'.

Qui possiamo citare come esempio di pietre miliari di questo settore:

- il Brown Corpus, costituito da un milione di parole che è uno dei progenitori degli attuali corpora;
- il British National Corpus (ampio, 100milioni di parole);
- il Collins Cobuild (ampio, 300milioni di parole);
- l'ICLE, International Corpus of Learner English (piccolo, composto da saggi argomentativi prodotti da apprendenti avanzati di inglese di diverse nazionalità).

Inoltre per un aggiornamento sui siti che informano o rimandano a corpora disponibili in rete, si citano i seguenti indirizzi:

<http://ecate.itc.it:1024/bentivo/Corpora/repositories-ita.html>

<http://rmcisadu.let.uniroma1.it/camplani/cap2.html>

http://www.cilta.unibo.it/SITOCORIS_ITA.htm

<http://digilander.iol.it/elisae/Italiano/cortipo.htm>

2. Da cosa partiamo

Per questa proposta sono stati utilizzati alcuni dei Corpora che si possono consultare all'ILC, sempre a scopo didattico o di studio, e che sono:

- il Corpus di Riferimento dell'Italiano,
- il Corpus di quaderni di alunni del primo e secondo ciclo,
- l'Archivio Elettronico delle Opere di Carlo Emilio Gadda,
- il Corpus dell'Italiano Parlato.

Il Corpus di Riferimento dell'Italiano si basa su metodologie di creazione di corpora già ampiamente sperimentate nel corso degli ultimi 15 anni nell'Istituto [Bindi et alii, 1989 e 1991]. Si tratta di un corpus aperto, che tuttora viene regolarmente incrementato grazie alle versioni su CD delle annate di quotidiani a tiratura nazionale. Per questa proposta sono stati utilizzati 5 sottoinsiemi, uno per ogni testata, relativi al periodo che va dal 1992 al 1996, la versione più recente del Corpus, contributo dell'ILC al progetto PAROLE della Unione Europea [Bindi et alii, 1997]. Questi 5 sottoinsiemi vengono qui presentati come 5 corpora autonomi.

Il Corpus di quaderni di alunni del primo e secondo ciclo è un piccolo corpus chiuso, realizzato con materiali di alcune scuole elementari di Pisa e Padova negli anni 1983-1985. Sono stati trascritti quaderni di seconda e terza elementare con pensierini, temi, riassunti e esercizi, codificando anche le eventuali correzioni degli insegnanti [Ceccotti et alii, 1986] [Sassi et alii, 1990].

L'Archivio Elettronico delle Opere di Carlo Emilio Gadda, realizzato a partire dall'edizione dei testi nella collana "I libri della Spiga" [Garzanti, 1988-93], è stato presentato ufficialmente presso la sede di Roma del CNR il 14 novembre del '97 [Ceccotti, Sassi, 1997 e 1999]. La peculiarità di questo Corpus è che comprende tutte le opere di un autore contemporaneo, ad eccezione di pochi testi ancora inediti nel 1993.

Il Corpus dell'Italiano Parlato è un corpus in DBT creato presso l'ILC a partire dai testi su cui è stato elaborato il Lessico di frequenza dell'Italiano Parlato (LIP) [De Mauro et alii, 1993]. La decisione di inserirlo in questa proposta dipende dalla opportunità di offrire un ventaglio di Corpora di Riferimento, tenendo conto che il 'parlato' rispetto al testo scritto, a parità di estensione, contiene un numero inferiore di parole diverse.

Nella tabella al paragrafo 4. si elencano i dati quantitativi relativi a ciascun Corpus.

3. Che cosa abbiamo sperimentato

I suddetti corpora sono stati omogeneamente trattati con il sistema DBT [Picchi, 1998], a partire dalla fase di creazione e gestione degli archivi elettronici, fino a quella di consultazione e di restituzione del risultato. Il DBT (Data Base Testuale) è stato concepito come un sistema di trattamento e interrogazione di testi che tiene in considerazione l'aspetto filologico e facilita lo studio dei particolari fenomeni linguistici. Si caratterizza anche per il rispetto totale delle qualità lessicografiche della risorsa linguistica a cui viene applicato e per la capacità di ottenere, rapidamente e in modo interattivo, funzioni di analisi testuale automatica, quali:

- la ricerca libera di parole nel testo (con possibilità molto estese di mascheramento e di combinazione attraverso gli operatori booleani);
- il calcolo delle frequenze in ordine alfabetico o decrescente;
- le concordanze sotto vari formati (direttamente pubblicabili via RTF) e in vari ordinamenti (secondo l'ordine di presenza nei testi, oppure secondo l'ordine alfabetico delle forme che seguono/precedono nella frase la parola oggetto di ricerca);
- gli index locorum, arricchiti da riferimenti puntuali;
- il calcolo dell'indice di leggibilità, basato sulle formule statistiche più accreditate.

Tutto ciò, unito ad altri ulteriori vantaggi (possibilità di analisi morfologica, integrazione di un tagger automatico della lingua, ecc.) ha permesso di definire e sviluppare, a partire dallo stesso strumento computazionale, una strategia di trattamento dell'informazione con fini

multidisciplinari. Il DBT si è dimostrato come uno strumento agile, duttile e, unito alla sua capacità di gestire grandi quantità di dati con risparmio di tempo e di spazio-disco, capace di adattarsi alle esigenze delle analisi più specifiche.

Il materiale esemplificativo proposto è stato prodotto utilizzando una delle funzioni statistiche più potenti di questo sistema di analisi testuale, la *co-occorrenza*: "Per co-occorrenza statistica (Mutual information) si intende il calcolo della probabilità che hanno le parole di un testo di essere associate, in quel testo, con una o più parole prestabilite", definizione tratta dal Manuale di uso on-line del DBT.

Secondo il LIF [Bortolini, 1972], il verbo più frequente, escludendo gli ausiliari, risulta il verbo '*fare*', dato confermato dal LIP. Tale spunto, che è stato l'ispiratore di un lavoro di ricerca svolto su quotidiani nazionali senza il sostegno della tecnologia [Ceccotti, 1990], è stato ripreso per una ricerca nel corpus dei quaderni, corpus usato come fonte-dati di partenza.

Dalla ricerca automatizzata del lemma *fare*¹, è stata ottenuta la mappa completa delle forme di tale verbo presenti nei testi; da questa sono state estrapolate le forme più frequenti, cioè *fa*, *far* e *fare*, che sono state utilizzate come 'parole prestabilite' nella consultazione degli altri corpora scelti. Per tutte le loro occorrenze selezionate, mediante la funzione DBT di co-occorrenze statistiche, sono stati ottenuti degli elenchi di parole, 'mutualmente connesse', di cui si propongono esemplificazioni nel paragrafo seguente.

4. Che cosa proponiamo

In questa presentazione sono stati riuniti alcuni esempi di videate che sono una rappresentazione, molto parziale, della potenzialità per scopi didattici degli strumenti precedentemente descritti (Corpora e DBT). E soprattutto si vuole promuovere l'aspetto stimolante che tali supporti didattici, con caratteristiche dinamiche, possano avere sugli insegnanti nel momento in cui utilizzano l'IT in classe, in alternativa a precisi percorsi didattici da seguire passivamente.

Nelle videate relative a quotidiani (n.2-5) e periodici (n.6) si evidenziano nella parte destra alcuni esempi di concordanze di sintagmi (es.: *fare chiarezza*, *fare concessioni*, *fare miracoli*), che appaiono in più corpora. Si dimostra così come la lista di parole, a sinistra, è un insieme dinamico di link ai contesti in cui si verificano le co-occorrenze, di alcune delle quali si visualizzano le concordanze.

Nella definizione dei parametri della funzione di co-occorrenza, si è indicato 4 come numero limite di parole a destra (ossia le seguenti) della terna data (*fa*, *far*, *fare*) e 0 a sinistra (ossia le precedenti).

Come soglia di frequenza minima si è fissato 7 nel caso di corpora più estesi, mentre per i rimanenti la soglia varia da 5 a 2; si tratta comunque di valori modificabili interattivamente sulla base dei risultati. Sono i risultati stessi, aggiornabili velocemente con un click di mouse che permette di variare tale soglia, che suggeriscono di volta in volta un valore piuttosto che un altro da assegnare a tale parametro.

Nella stessa maniera si può usufruire interattivamente del cambio di ordinamento: in ordine di importanza, secondo la formula statistica applicata (come appare nelle videate seguenti), oppure in ordine alfabetico.

¹ A partire dal DMI, il dizionario macchina dell'italiano [Gruppo di Pisa, 1979], poi incrementato nel corso degli anni, è stato sviluppato un componente di analisi morfologica, che inserito nel DBT in fase di interrogazione, permette, attraverso la generazione automatica delle forme di un lemma dato, di ottenerne tutte le occorrenze (frequenze e contesti) di un testo o corpus.

Una precisazione da fare è che si tratta di una funzione morfologica e non semantica, per cui non viene svolta nessuna distinzione fra omografi; sarà poi il contesto stesso, immediatamente raggiungibile, a disambiguare il termine. Nel nostro caso la forma *fa* sarebbe abbastanza ambigua, ma leggendo il risultato delle co-occorrenze si possono facilmente individuare i casi in cui si tratta del verbo fare.

Di seguito segnaliamo alcuni dati quantitativi relativi a ciascun corpus:

Quaderni:	58.700 parole, per un totale di 247 occorrenze di <i>fa, far, fare</i> .
Corsera :	circa 3.200.000 parole, 39 giorni ('92-'96) e 7540 occorrenze di <i>fa, far, fare</i> .
Sole24ore:	circa 4.150.000 parole, 51 giorni ('92-'96) e 5319 occorrenze di <i>fa, far, fare</i> .
LaRepubblica:	circa 3.485.000 parole, 39 giorni ('92-'96) e 8048 occorrenze di <i>fa, far, fare</i> .
La Stampa:	circa 3.300.000 parole, 45 giorni ('92-'96) e 7846 occorrenze di <i>fa, far, fare</i> .
Periodici:	circa 800.000 parole, da 9 testate, 68 giorni ('85-'88) e 2152 occorrenze di <i>fa, far, fare</i> .
Gadda:	circa 1.600.000 parole e 2214 occorrenze di <i>fa, far, fare</i> .
Corpus LIP:	circa 500.000 parole e 2204 occorrenze di <i>fa, far, fare</i> .

Ricapitolando, in ogni videata vi è un elenco di parole, ad ognuna delle quali (es.: *conoscenza* nella videata n.1) sono associati quattro numeri, che indicano rispettivamente:

- 1 quante volte ogni parola è associata, nella stringa di testo di lunghezza predefinita, alle forme su cui è stata calcolato il valore di co-occorrenza (es.: 6 volte);
- 2 quante volte la parola co-occorrente è presente nel testo in esame (es.: 6 volte);
- 3 il valore statistico di "Mutual information" (es.: 7,893);
- 4 il valore medio della distanza, espressa in unità di elementi, tra la parola co-occorrente e quelle in esponente (es.: 2,000 cioè due parole in media).

The screenshot shows the DBT per Windows application. The main window displays two panes of co-occurrence statistics. The left pane, titled 'Co-Occorrenze statistiche', lists 36 items. Each item consists of a number in parentheses, followed by four columns of numerical data, and then a word in red. The words include: conoscenza, ginnastica, spesa, campioni, cartine, faccende, militare, pasegiata, {c}capricci, {c}trecce, passeggiata, buco, barnabò, pensiero, pupazzi, {c}passeggiata, lezione, minestra, cuocere, tictac, lunghe, fretta, proteggere, ritrovarono, corsa, capire, cerchi, colazione, giro, tornare, guerra, boschi, natura, nitrito, and rumore. The right pane, also titled 'Co-Occorrenze statistiche', lists 35 items with similar numerical data and words in red: nitrito, rumore, bagno, feste, foglioline, voce, farfalla, merenda, marito, venire, grossa, forte, perche, pianta, cucina, visita, ore, male, fattore, nei, festa, tre, acqua, prati, bella, vento, tutto, essere, pissi, cane, due, cosa, io, and scuola. The status bar at the bottom indicates '[3] | 66524424 | DBT - Copyright © 1995 CNR Italia - Eugenio Picchi'.

Videata n.1: dal Corpus dei Quaderni (Le parole precedute da {c} fanno parte del testo corretto dall'insegnante).

WinDBT - Il Corriere della Sera del 21/8/1992

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

lenze statistiche

23	93	9.424	1.130	notare
9	38	9.362	1.111	sorridere
7	30	9.340	2.000	piega
18	83	9.235	1.000	scattare
7	34	9.160	1.143	shopping
9	44	9.151	1.556	affidamento
17	84	9.135	1.118	ridere
11	55	9.118	1.000	esplodere
7	36	9.077	1.143	prevalere
14	75	9.019	1.000	funzionare
24	136	8.938	1.292	rispettare
17	97	8.928	1.000	nascere
28	163	8.899	1.464	piacere
10	60	8.855	1.000	riflettere
33	199	8.848	1.121	sentire
20	126	8.785	1.150	crescere
18	120	8.703	1.056	cadere
7	47	8.693	1.143	miracoli
13	88	8.681	1.154	eco
39	281	8.591	1.231	referimento
7	54	8.492	1.000	comodo
11	86	8.473	1.000	apparire
8	64	8.440	1.500	concessioni
71	578	8.415	2.127	conti
57	491	8.333	1.140	capire
47	406	8.329	1.255	pensare
15	134	8.281	1.000	dimenticare
20	179	8.278	1.150	discutere
11	100	8.256	1.545	chiarezza
51	474	8.224	1.137	sapere
7	69	8.139	1.000	comprendere
22	225	8.086	1.136	conoscere

[Cnt]-[Co_occ]: MIRACOLI

- 1) in piazza e ha bisogno di santi patroni pronti a **far miracoli**". Proprio alle "imprese ed dimentica il "takuan daikon", una radice cinese che **fa miracoli** contro i raffreddori. L
- 2) come il Carl Lewis compagno di sponsor. Non **fa miracoli** la nuova Inter, ma si sta al
- 3) Fatebenefratelli. "Il professor Marzetti - esulta Zeudi - **fa miracoli**. Questo è un osped
- 4) Con l'organico che ha, la Polizia Stradale **fa** già **miracoli**. Ma il vertice della Polizia se
- 5) non capirlo. Forse pensa che si possa continuare a **fare miracoli** come abbiamo fat
- 6) la bacchetta magica", anticipa che "non si possono **fare miracoli**" e delega la moltip
- 7)

[Cnt]-[Co_occ]: CONCESSIONI

i sono lontane. Nessuna è disposta a **far concessioni**: non la Francia, che teme il voto agr

zione di Andreotti) non è disposta a **fare concessioni** ai referendari, tanto meno ad annac

d Arafat, rendere più difficile a Rabin di **fare concessioni** che lo espongano all'accusa di r

lla lotta alla mafia An non intende **fare concessioni** né sconti. Ne è dimostrazione la mani

/rate in modo assoluto e su quali è possibile **fare concessioni**. È una battaglia continua su

erché alcuni consiglieri federali non vogliono saperne di **fare** altre **concessioni** al sindacato

. Mastella apre agli autonomi **fare** ulteriori **concessioni** ai sindacati" **fare**

nde sui loro territori, e perciò devono **fare** loro grosse **concessioni**". Il ceto medio, formato

[Cnt]-[Co_occ]: CHIAREZZA

- 1) interne, ben poco chiare: adesso vorremmo che a **far chiarezza** sia la magistratur
- 2) ". Bossi non lo dice esplicitamente, ma lo **fa capire** con grande **chiarezza**: due ar
- 3) carte sui politici mafiosi. Un'accusa gravissima. Bisognava **fare chiarezza** subito, r
- 4) stiamo cercando di fare pulizia nella pubblica amministrazione, di **fare chiarezza** s
- 5) ha "auspicato" che la magistratura possa "**fare chiarezza**". - (12)CS24-04-93 Eco bus
- 6) Aaroi, spiega: "Riteniamo indispensabile questo incontro per **fare chiarezza** e per
- 7) aveva paura. Una testimonianza quindi "importantissima" per **fare chiarezza** perc
- 8) chiare, semplici e comprensibili: l'unico modo per **fare chiarezza** è tornare alle un
- 9) comunque, sui cellulari in aereo è ora di **fare chiarezza**. Esiste un Regolamento c
- 10) - sostiene il settimanale - che senza governi capaci di **fare chiarezza** sull'Unione e
- 11) si decide ad usare il pugno di ferro e a **fare** un po' di **chiarezza**? la cessione di Br

[3] | 66524652 | DBT - Copyright © 1995 CNR Italia - Eugenio Picchi

Videata n. 2: da "Il Corriere della Sera".

WinDBT - Il Sole 24 Ore del 12/12/1996

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

lenze statistiche

7	7	11.943	1.000	quadrare
9	15	11.206	1.111	capolino
33	74	10.778	1.000	scattare
7	17	10.663	1.000	crollare
23	56	10.660	1.000	decollare
51	131	10.582	1.000	valere
16	43	10.517	1.063	saltare
11	33	10.358	1.000	cessare
7	21	10.358	1.000	fallire
11	36	10.233	1.000	slittare
29	96	10.216	1.069	emergere
46	153	10.210	1.130	notare
21	70	10.206	1.238	nascere
20	70	10.136	1.000	funzionare
14	54	9.996	1.000	pervenire
11	45	9.911	1.000	confluire
10	53	9.537	1.000	riflettere
12	65	9.506	1.333	ripartire
8	45	9.452	1.250	sorgere
20	118	9.383	1.150	sentire
11	65	9.380	1.000	eco
10	63	9.288	1.500	piacere
185	1300	9.131	1.124	referimento
10	78	8.980	4.000	leone
30	238	8.955	1.000	conoscere
9	73	8.924	1.000	cadere
18	148	8.904	1.000	chiarezza
10	83	8.890	1.400	credere
110	928	8.867	1.100	capo
14	121	8.832	1.143	segnare
31	294	8.698	1.129	capire

[Cnt]-[Co_occ]: CHIAREZZA

Il cardinale Tisserand, che pure non è contrario a **far chiarezza** tra cattolici e comunisti, dich

ΠΠΠ Persa l'occasione per **far chiarezza** sulla giurisdizione. È nota da tempo la teoria - (55)S

controllare le enormi sofferenze dell'Istituto. È necessario **far chiarezza** e che venga fuori, i

che sottolineava come le indagini Consob fossero adeguate a **far chiarezza** e dove non c'è

piuttosto divaricate e molto, molto confuse. Proviamo a **fare chiarezza**. La proposta Cee. R

che caratterizzano il mercato dei pacchetti non aiutano a **fare chiarezza** nella testa degli ute

politiche che aggravava ampiamente la legge. Una parola per **fare chiarezza** l'hanno spesa

dell'accensione di fari da parte del magistrato, per **fare chiarezza** in situazioni in cui quella c

con intenti persecutori, ma non foss'altro che per **fare chiarezza** _ con relativi interrogatori, c

oggi>. E certamente abbiamo più che mai bisogno di **fare chiarezza** sui miti, soprattutto per

ieri a sorpresa dal Papa, che ha così voluto **fare chiarezza** definitiva sull'annosa questione.

ieri a sorpresa dal Papa, che ha così voluto **fare chiarezza** definitiva sull'annosa questione.

revisore contabile è spesso nell'occhio del ciclone. A **fare chiarezza** arriva adesso un libro l

fa automaticamente ripartire i lavori, però è necessaria per **fare chiarezza**. Mangani ha fatto

principali), non sono stati, infatti, sufficienti a **fare chiarezza** su questa tassa comunale sull'oc

conclude il rapporto - hanno fatto del loro meglio per **fare chiarezza** sulla vicenda ma, come

Consob non ha mancato di rinnovare la sua disponibilità a **fare chiarezza**. A che cosa è d

vigore. Anche su questi aspetti il Testo unico potrà **fare chiarezza**. **vedere** Sicurezza lavoro: il C

[Cnt]-[Co_occ]: CONCESSIONI

Governo e le industrie del Giappone tratteranno duramente prima di **fare concessioni** su

ΠΠΠ Il presidente americano non sembra disposto a **fare concessioni** ΠΠΠ Il vertice di Ric

nuscito ad ottenere un aumento retributivo attorno al 4% senza **fare concessioni** di alcun

degli Usa. Il Governo di Tokio si prepara a **fare** altre **concessioni**, ma per ora non s

ΠΠΠ... per il dumping turco ΠΠΠIntanto Bruxelles si prepara a **fare** altre **concessioni** ad Ar

la pressione sta crescendo e il Governo ha dovuto **fare** le prime **concessioni**, aumentan

La Francia non avrebbe più la forza politica per **fare** le ultime **concessioni** agricole indis

delle Finanze, Theo Waigel, ha dovuto già ieri **fare** le prime **concessioni** di princip

Videata n.3: da "Il Sole 24 ore".

WinDBT - la Repubblica del 11/03/1994

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

lenze statistiche

10	15	10.761	1.000	quadrare
11	18	10.635	1.000	rivivere
15	34	10.165	1.200	valere
8	20	10.024	1.000	slittare
7	18	9.983	1.000	brillare
30	81	9.913	1.000	finta
7	25	9.509	1.429	salti
30	109	9.485	1.100	saltare
29	112	9.397	1.138	notare
9	35	9.387	1.556	comodo
7	28	9.346	1.286	cenno
8	33	9.302	1.750	autocritica
10	43	9.242	1.100	esplodere
11	50	9.162	1.000	west
7	34	9.066	1.000	shopping
8	39	9.061	1.375	sparire
14	69	9.045	1.214	scattare
11	55	9.024	1.000	emergere
15	77	8.986	1.000	funzionare
10	52	8.967	1.500	ridere
7	37	8.944	1.000	decollare
17	92	8.910	1.059	riflettere
46	252	8.892	1.348	sentire
17	99	8.804	1.000	eco
7	41	8.796	2.000	piega
19	117	8.724	1.000	rispettare
19	120	8.687	1.474	chiarezza
15	95	8.683	1.133	approvare
88	562	8.671	1.227	sapere
22	153	8.548	2.727	mestiere
99	692	8.541	2.152	conti

[Cnt]-[Co_occ]: CHIAREZZA

o". Alla fine il ministero degli Affari sociali, **fa** un po' di **chiarezza**: arrivano martedì. Le voci si...
 er la procedura di separazione legale. Due settimane **fa** Massimo dimostrò con **chiarezza** c...
 etto di Palazzo Chigi. Dico però che bisogna intanto **far chiarezza** su alcune promesse che...
 Intanto, per **far chiarezza** sull'affare, la Commissione Finanze della Camera convoca in - (1...
 Cisl e Uil hanno chiesto al governo di **far chiarezza** sull'operazione di fusione tra Stet e Tel...
 estremo rigore. Ora è davvero il momento per **fare chiarezza**. Per intenderci, i record in vigo...
 dei campi di sterminio questa iniziativa ha il pregio di **fare chiarezza**". Al termine dell'ir...
 issortite dei mille gusti offerti in pasticceria. A **fare chiarezza** è stato il ministero delle Finanze...
 dendo atto della decisione, insiste sulla necessità di **fare chiarezza** "anche al fine di evitare...
 rà il coraggio di farlo collettivo si riuscirà a **fare chiarezza**". DescrCommento sulla vicenda d...
 gno su cui neanche la Chiesa riesce evidentemente a **fare chiarezza**". Gettonare - Se i ter...
 questo caso deve essere studiato a fondo, bisogna **fare chiarezza** e capire come sono and...
 sto film" dice Melvin, "serve a **fare chiarezza** sulla storia, che è sempre scritta dai vincitori...
 strage di via Fani non erano mai riusciti a **fare chiarezza** su questo punto, che adesso è fir...
 comizi e di incontri con militanti e simpatizzanti "per **fare chiarezza**". In attesa di vedere più...
 Sono sereno - replica Arezzo di Trifletti - Ho voluto **fare chiarezza** nella gestione degli affar...
 \Box\... l'abbiamo voluto per costringerlo finalmente a schierarsi, a **fare chiarezza**". "Ma voi...
 [richiede] richiede una risoluzione urgente". "Bisogna **fare chiarezza**", insiste, l' [as;v]

[Cnt]-[Co_occ]: COMODO

architettura moderna internazionale, su cui in casi del genere **fa comodo** scaricare tutte le re...
 E Alberto di Monaco? Anche quella una montatura che **fa comodo** al Principato (tutta p...
 Dal punto di vista pratico agli americani **fa comodo** che la loro moneta si svaluti: non lo fanno...
 mercato, ma anche del protezionismo più bieco quando gli **fa comodo**, nei light trucks spadi...
 quelli che si sono opposti maggiormente. Adesso che gli **fa comodo** vorrebbero stare aperti...
 L'80% degli zionisti sono dei mentitori. A me **fa comodo** questa rivalità, così ora passo per un...
 attacchi massicci in proprio. Sono demoni ma di certo **fa** anche **comodo** demoniz...
 adesso è forte anche 'dentro'. E **fa** tanto **comodo** alla causa comune. A...
 l'hanno fatta, eccoli di nuovo accusarsi di voler **fare** gli uni il **comodo** proprio e gli altri di spreg...

Videata n.4: da "La Repubblica".

WinDBT - La Stampa del 2/01/1994

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

lenze statistiche

7	7	11.383	2.000	grinza
12	19	10.720	1.000	quadrare
13	27	10.328	1.154	tremare
7	17	10.103	1.000	scoppiare
10	25	10.061	1.100	rivivere
8	22	9.923	1.000	slittare
21	58	9.917	1.048	scattare
13	36	9.913	1.154	sorridere
33	93	9.888	1.091	saltare
9	26	9.852	1.000	decollare
17	51	9.798	1.000	valere
20	62	9.750	1.000	finta
30	94	9.735	1.067	notare
10	33	9.660	1.700	invidia
8	29	9.525	1.250	cenno
10	40	9.383	1.000	sparire
18	84	9.160	1.222	ridere
40	193	9.112	1.225	piacere
13	66	9.039	1.077	miracoli
7	36	9.020	1.286	emergere
26	135	9.006	1.308	eco
17	89	8.994	1.176	rispettare
19	100	8.987	1.000	nascere
16	87	8.940	1.000	riflettere
7	39	8.905	2.286	valigie
85	506	8.809	2.176	conti
33	202	8.769	1.061	sentire
8	49	8.768	1.375	esplodere
19	117	8.760	1.263	chiarezza
9	58	8.695	1.333	soffrire
8	53	8.655	1.375	temere
15	101	8.631	2.533	salto

[Cnt]-[Co_occ]: CHIAREZZA

La pubblicazione del documento, se da un lato **fa chiarezza** su quelle che erano state l...
 ma da un punto di vista politico questa decisione **fa chiarezza**. Magistratura Democratica...
 destini di Marchegiani e molte di più su come si **fa** ad usare **chiarezza** e trasparenza. Si...
 convinzione che non ha alcun fondamento e sulla quale dovremo **far chiarezza**. Ma il m...
 su La Stampa il 5 e 12 gennaio 1995 desideriamo **far chiarezza** sulla situazione venuta...
 sentenza è in arrivo), l'occasione è opportuna per **far chiarezza** sui vari tipi di impianti el...
 inchiesta parlamentare, molti saggi non sembrano sufficienti ancora a **far chiarezza** su c...
 una novità che ogni tanto emergano episodi su cui occorre **fare chiarezza**. Ma non cred...
 sull'incauta propalazione di notizie>, nel tentativo di <**fare chiarezza** su questa assurda...
 accusare Maccari? <Ho fatto quel nome proprio per **fare chiarezza**, per riempire l'ultimo...
 non ha maturato la mia convinzione. Forse, a **fare chiarezza** definitiva dovrebbe essere...
 Voci, tante voci sulle quali il magistrato deve **fare chiarezza**. L'avviso di garanzia, va rico...
 presso l'Unità di crisi si susseguono i messaggi per **fare chiarezza** sull'accaduto. L'amb...
 presidente della giunta con un'interrogazione urgente, di <**fare chiarezza**> sulla posizio...
 abbiamo fatto una proposta molto limpida, che aiuterebbe a **fare chiarezza**, se ne giove...
 Luigi Abete, ha esortato le forze politiche a **fare chiarezza** sui programmi economici. «S...
 Colonne del Comune di Torino. «L'intento è **fare chiarezza** su un tema che solleva molt...
 già concretamente dimostrata) a fornire ogni informazione necessaria a **fare chiarezza**>...
 alla poltrona, nemico che la mia decisione contribuisca a **fare** un po' di **chiarezza**>»

[Cnt]-[Co_occ]: MIRACOLI

La banda-Scala spazza l'Ascoli...L'Onore **fa miracoli**, poi cede a Melli e M...
 si convince che l'altro cambierà, perché l'amore **fa miracoli**. È vero? Assolutamente...
 e collo. La radice del ginseng è nota, **fa miracoli** (o quasi): l'estratto ha poteri tonifican...
 vede, è la prova vivente che Dio esiste e **fa miracoli**. Ecco le più famose visioni della G...
 questi granata in formazione rimaneggiata mettono paura, l'emergenza **fa miracoli**>...
 con la troupe regalo se ne esce con un <**Fa miracoli**, vero?>. Dieci miliardi in tutto il t...
 Poi Pasquale Prunas scese a Roma, continuando a **far miracoli** grafici piccoli e grand...
 una sconfitta di Bush. <Dico che non poteva **fare miracoli**. E che è stato sfortunato per...
 ritornando ad allenarsi solo da una settimana, non poteva **fare miracoli**. Viene comu...
 parlare del professor Marcellino, hanno saputo che riesce a **fare miracoli**. Lo contatta...
 sembra smarrito e Williams, da solo, non può **fare miracoli**. Dopo la sfiurata iniziale di...
 aggressione di chi è impaurito. Ovviamente, non può **fare miracoli**, appena le grandi...
 chiunque esso sia, non è messo in condizione di **fare i miracoli** richiesti. Forse tutta e...

[3] | 66601484 | DBT - Copyright © 1995 CNR Italia - Eugenio Picchi

Videata n.5: "La Stampa".

WinDBT - Panorama Mondadori 1985-1988

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

cooccorrenze statistiche

5	5	13.249	1.000	capolino
5	5	13.249	1.400	fruttare
6	12	12.249	1.167	scattare
5	10	12.249	1.000	rivivere
11	30	11.801	1.273	ridere
5	14	11.763	1.000	invidia
10	30	11.664	2.000	salto
12	44	11.374	1.000	crescere
22	93	11.169	2.136	conti
13	59	11.067	1.154	sentire
6	28	11.027	1.167	funzionare
6	30	10.927	1.000	nascere
7	40	10.734	1.286	credere
7	48	10.471	1.000	ricorso
5	37	10.361	1.000	aumentare
9	70	10.290	1.000	referimento
18	151	10.180	1.222	capire
6	56	10.027	2.333	mestiere
15	148	9.946	1.533	male
12	136	9.746	1.500	paura
5	57	9.738	2.400	spesa
6	70	9.705	1.833	freddo
13	156	9.664	1.154	pensare
8	106	9.521	1.500	passare
6	89	9.358	2.333	bagno
7	106	9.328	1.429	arrivare
9	138	9.310	2.222	figura
6	92	9.310	1.500	tendenza
5	79	9.267	1.000	domande
5	80	9.249	2.400	diagnosi
12	204	9.161	1.167	capo
5	86	9.145	1.000	conoscere

[Cnt]-[Co_occ]: CAPOLINO

Tra loro serpeggia l'impazienza. Quando il cameriere Gassman **fa capolino** dalla porta certi secrétaires: una forma veramente eccentrica, in cui **fa capolino** quella vena di preletterario tuttora avvolto dal mistero; tutto bene, ma **fa capolino**, di quando in quando, qu di Bruce, il pescecane de Lo squalo, che **fa capolino** dal pozzo dei desideri in cortile. A che da qualche tempo in qua hanno incominciato a **far capolino** nelle maggiori città. R

[Cnt]-[Co_occ]: RIDERE

po' un nome passepartout, una specie di attrezzo che **fa ridere**, piangere ecc.. Ecco co Le donne si giudicano dalle loro misure? Il grasso **fa ridere**? E soprattutto: è ragionevol democratica" - e se questo è un termine che **fa ridere**, viste le fortune che si sprecano c solo il pettegolezzo, l'indiscrezione, la volgarità. **Fa** sempre **ridere** uno che invoca: "Mie lo svolgimento. Quella famigliola perfetta, stereotipata! Mi **fa un po' ridere** l'ide stupisce apprendere che egli si diletta di astrologia solo per **far ridere** le nobili non politico" $\lambda\pi\lambda\lambda$ Un artigiano felice di saper **far ridere**. E la storia di Elisabetta come io lavoro come un artigiano e sono felice di saper **far ridere**. - (185)GZ13-01-85 UNA REGIN questa vocazione del comico, cioè già da ragazzino voleva **far ridere**? R. Veramente i Rascel, Macario, Tognazzi e Vianello, ma per **far ridere** bisognava inventarsi un modo la testa il muro di casa. Io invece volevo **far ridere** dicendo per esempio: "Buonasera c

[Cnt]-[Co_occ]: TENDENZA

- della Castellana, vicinissimo alla zona che in questo momento **fa tendenza**, qu
- Anzi essere vergini adesso fa moda, **fa tendenza**, è un optional di qualità. Me
- ΑΤι3... nuova tecnica inglese per lanciare il ballerino-coreografo post punk che
- l'eleganza personale. Guida fast per sapere cosa **fa tendenza** fra i modaioli d
- Snob o NS, estrema frontiera della beautiful people che **fa tendenza**. Incalzan
- Anzi essere vergini adesso **fa moda**, **fa tendenza**, è un optional di qualità. Me

[3] | 66523416 | DBT - Istituto di Linguistica Computazionale - Pisa

Videata n.6: Periodici.

WinDBT - Gadda, La Madonna dei Filosofi, RRI, Garzanti

File Ricerca Contesti Famiglie Opzioni Varie Finestre Guida

Co-Occorrenze statistiche

6)	4	4	9.526	1.000	imbizzare	38)	3	30	6.204	1.667	legna
7)	5	9	8.678	2.400	tonto	39)	5	52	6.148	1.800	istruzione
8)	4	8	8.526	1.500	cin	40)	3	37	5.902	1.667	discreta
9)	3	7	8.304	3.000	acquolina	41)	11	138	5.877	2.545	giro
10)	3	7	8.304	1.667	assegnamento	42)	5	64	5.848	2.400	asino
11)	4	10	8.204	2.250	passeggiatina	43)	8	104	5.826	2.250	bagno
12)	11	32	7.986	1.455	pipi	44)	9	121	5.777	1.000	colazione
13)	3	10	7.789	1.667	congetture	45)	11	148	5.776	1.455	presto
14)	3	12	7.526	2.333	concessione	46)	3	41	5.754	2.000	ovo
15)	4	19	7.278	2.500	do	47)	5	70	5.719	1.000	capire
16)	4	19	7.278	2.500	sparire	48)	3	43	5.685	1.000	chiudere
17)	4	19	7.278	2.000	uopo	49)	8	117	5.656	1.750	venire
18)	9	44	7.237	2.111	spia	50)	5	74	5.639	2.000	star
19)	7	36	7.164	2.143	fretta	51)	3	46	5.588	1.333	dispetto
20)	4	22	7.067	1.000	diventar	52)	4	62	5.572	1.000	dimenticare
21)	4	22	7.067	1.250	testimonio	53)	4	65	5.504	1.250	notare
22)	3	17	7.024	1.667	ritrovare	54)	3	49	5.497	1.000	motto
23)	3	18	6.941	2.333	esplosivi	55)	3	50	5.467	1.333	difetto
24)	11	70	6.856	1.727	piangere	56)	5	84	5.456	2.000	affari
25)	3	22	6.652	1.333	funzionare	57)	4	68	5.439	1.000	perdere
26)	3	22	6.652	2.333	riflettere	58)	35	611	5.401	1.600	male
27)	5	38	6.600	1.400	seguire	59)	3	53	5.383	2.000	copia
28)	6	49	6.497	1.000	comodo	60)	17	323	5.278	1.176	paura
29)	13	112	6.419	1.308	piacere	61)	7	133	5.278	1.286	niente
30)	22	191	6.408	1.364	pensare	62)	4	76	5.278	1.250	obbligo
31)	3	27	6.356	1.000	vela	63)	3	57	5.278	1.667	invidia
32)	3	27	6.356	2.333	viste	64)	7	135	5.257	1.714	scale
33)	7	67	6.268	2.143	conti	65)	3	58	5.253	2.667	dio
34)	3	29	6.253	2.000	palo	66)	3	58	5.253	1.000	storie
35)	3	29	6.253	2.667	pupi	67)	9	176	5.237	1.222	onore
36)	3	29	6.253	2.667	regali	68)	3	59	5.229	1.333	ridere
37)	5	49	6.234	1.800	cenno	69)	7	139	5.215	1.000	passare
38)	3	30	6.204	1.667	legna	70)	4	80	5.204	1.250	entrare

[4] | 66525232 | DBT - Istituto di Linguistica Computazionale - Pisa

Videata n.7: Il Corpus Gaddiano.

Le-Uccorrenze statistiche				Le-Uccorrenze statistiche							
5)	4	4	7.816	2.500	plin	44)	6	57	4.568	2.333	differenza
6)	4	6	7.231	1.000	sparire	45)	4	38	4.568	2.500	esami
7)	17	28	7.096	1.294	schifo	46)	4	38	4.568	1.750	riunione
8)	5	9	6.968	2.400	paragone	47)	5	48	4.553	2.400	fotografia
9)	18	33	6.942	1.778	caldo	48)	5	49	4.523	2.200	leggi
10)	4	8	6.816	2.000	voltura	49)	4	40	4.494	2.250	telefonata
11)	7	15	6.717	1.571	fisioterapia	50)	6	64	4.401	2.333	certe
12)	5	11	6.679	2.400	laser	51)	6	64	4.401	2.000	ricerca
13)	33	95	6.291	1.455	piacere	52)	4	43	4.390	1.000	diventare
14)	5	16	6.138	3.400	recupero	53)	7	78	4.338	2.143	mangiare
15)	4	13	6.116	1.250	riflettere	54)	6	67	4.335	1.000	presto
16)	4	14	6.009	2.000	affari	55)	6	68	4.314	1.667	sentire
17)	4	15	5.909	3.250	riforma	56)	5	59	4.256	2.000	domande
18)	5	21	5.746	2.000	seminario	57)	6	75	4.172	2.000	passo
19)	4	17	5.729	2.000	saggio	58)	5	64	4.138	1.600	servizio
20)	9	42	5.594	1.333	causa	59)	4	52	4.116	2.000	tardi
21)	8	38	5.568	1.375	conoscere	60)	4	54	4.061	2.750	atto
22)	4	19	5.568	1.250	comodo	61)	4	54	4.061	1.000	pagare
23)	13	65	5.494	1.077	riferimento	62)	4	54	4.061	2.500	relazione
24)	4	20	5.494	2.250	impressione	63)	4	55	4.035	3.250	capitolo
25)	5	26	5.438	1.600	foto	64)	13	188	3.962	1.308	male
26)	4	22	5.357	1.500	ridere	65)	4	58	3.958	3.250	dibattito
27)	9	51	5.314	2.333	tesi	66)	5	75	3.909	2.000	capo
28)	4	23	5.293	2.000	passi	67)	5	76	3.890	2.000	grosso
29)	4	24	5.231	2.000	fotografie	68)	4	62	3.862	1.250	paura
30)	4	24	5.231	3.000	offerta	69)	8	125	3.850	2.500	domanda
31)	6	38	5.153	2.333	analisi	70)	5	79	3.834	1.600	fronte
32)	7	45	5.132	2.143	nomi	71)	6	95	3.831	3.167	bambini
33)	4	26	5.116	2.250	visita	72)	6	100	3.757	2.167	roba
34)	4	27	5.061	1.500	discutere	73)	6	103	3.715	3.167	sull'
35)	4	27	5.061	3.000	rai	74)	18	328	3.629	2.278	lavoro
36)	4	28	5.009	2.750	contrario	75)	4	73	3.626	2.750	attivit�
37)	8	61	4.885	1.125	passare	76)	4	73	3.626	1.250	nu
38)	6	46	4.878	2.333	esame	77)	6	112	3.594	2.333	carta
39)	14	122	4.693	1.000	capire	78)	4	77	3.549	3.250	domenica
40)	4	35	4.687	2.250	chissa'	79)	18	349	3.539	1.222	vedere
41)	5	45	4.646	2.800	compito	80)	4	78	3.531	2.500	proposta
42)	4	36	4.646	3.000	controllo	81)	6	119	3.506	2.500	corso
43)	7	65	4.601	2.571	lavori	82)	4	80	3.494	3.000	arrivare

Videata n.8: Corpus dei testi del LIP.

5. Conclusioni

Questo contributo   scaturito da due constatazioni apparentemente in conflitto: la disponibilit  'illimitata' di risorse per acquisire conoscenza (siamo nel dominio del continuo) e l'opportunit  di strumenti didattici (siamo nel dominio del discreto) adeguati all'attuale societ  dell'informazione.

Riguardo alla seconda constatazione, Didamatica 2001 ci sembra una sede significativa per ribadire che negli anni passati molto   stato detto, poco   stato fatto per rendere concretamente disponibile e fruibile nelle scuole ci  che   stato prodotto dalla ricerca nel settore della linguistica. La situazione attuale non ci sembra cambiata² anche perch  per 'voltare pagina' sarebbe necessario un progetto nazionale di riferimento per tutte le scuole del territorio nazionale e non (basti pensare alle scuole italiane all'estero, alla societ  europea in fieri, all'insegnamento dell'italiano per stranieri). I tempi, ci chiediamo, non sono ancora pronti perch  nelle scuole il dubbio, lessicale o sintattico, di un alunno possa essere risolto anche con l'ausilio di una banca di dati testuali?

Auspichiamo allora che tra breve un alunno della scuola elementare, per esempio, avendo letto in un testo l'espressione *far capolino*, per impararne le sfumature di uso e le sue traduzioni nelle varie lingue, possa collegarsi alla rete scoprendo cos  che l'espressione pu  essere sostituita, a seconda del contesto:

² Nel lontano 1986 nell'ambito della Fiera del Levante fu presentato il corpus dei quaderni e le sue possibili applicazioni in ambito didattico; in quella occasione gli insegnanti presenti manifestarono notevole interesse per tale metodologia, che risultava ancora prematura per la scarsa diffusione nelle scuole della tecnologia telematica.

- in italiano, da *apparire, comparire, mostrarsi fugacemente, presentarsi, sbirciare, spiare, affacciarsi, saltar fuori, spuntare, cominciare a nascere, aprirsi un varco, sfondare, filtrare, trapelare, trasparire, tradire la propria presenza*;
- in inglese da *to peep in, to peep out, to rise, to sprout, to begin to grow, to appear*;
- in francese da *passer la tête (à travers), percer, pointer, paraître, se lever, éclore, pousser*;
- in spagnolo da *asomarse, atisbar, nacer, aparecer, echar una mirada furtiva, echar una ojeada furtiva, mirar furtivamente, ojear, fisgar*;
- in tedesco da *hervor-kommen, hervor-vorgucken, auftauchen, erscheinen, (heimlich) betrachten, (heimlich) mustern, spähen*;
- in portoghese da *mostrar-se apenas, espreitar furtivamente, aparecer, despontar, nascer, surgir, afirmar-se, espiar, dar uma olhada*.

Bibliografia

- Bindi R., Monachini M., Orsolini P., *Corpus di Riferimento della lingua italiana contemporanea*, ILC-TLN-1989-1, Pisa, 1989.
- Bindi R., Monachini M., Orsolini P., *Italian Reference Corpus. General Information and Key for Consultation*, ILC-TLN-1991-1, Pisa, 1991.
- Bindi R., Goggi S., Biagini L., Picchi E., Rossi S., *Italian Corpus Documentation, LE-PAROLE-WP2.11*, Pisa, 1997.
- Bortolini U., Tagliavini C., Zampolli A., *Lessico di frequenza della lingua italiana contemporanea, (LIF)*, Garzanti, Milano, 1972
- Ceccotti M.L., Pecchia L., Ruimy N., Sassi M., "Elaborazioni di dati linguistici e letterari: prospettive per la didattica", in *Giornata di Studio su Applicazioni dell'elaboratore nella didattica delle discipline umanistiche*, organizzata da Sassi M. per l'AICA, 22 gennaio 1986, Pisa, 1986.
- Ceccotti M.L., Pecchia L., Ruimy N., Sassi M., Sorrenti R., "Progetto Prisma: studio sulla produzione linguistica nella scuola dell'obbligo", in *Scuola e Comunicazione*, N.1, Pisa, 1986.
- Ceccotti M.L., *Una testimonianza dell'uso corrente di due verbi con le concordanze semantiche, dei verbi FARE e DIRE, relative a due quotidiani in lingua italiana dell'otto agosto millenovecentonovanta: Avanti! e Avvenire*, Rapporto tecnico, ILC-CS-1, Pisa, 1990.
- Ceccotti M.L., Sassi M., *L'Archivio elettronico delle Opere di C.E.Gadda: come è stato costruito, come si consulta*, ILC, Pisa, 1997
- Ceccotti M.L., Sassi M., *Alla ricerca dei termini gaddiani. Una pre-concordanza*, ILC-CNR, S.T.A.R., Pisa, 1999.
- De Mauro T., Mancini F., Vedovelli M., Voghera M., *Lessico di frequenza dell'italiano parlato, (LIP)*, Etaslibri, Roma, 1993.
- Gruppo di Pisa, "Il dizionario di macchina dell'italiano", in *Linguaggi e Formalizzazioni*, Gambarara D., Lo Piparo F., Ruggiero G. (eds.), Atti del Convegno Internazionale di Catania, 1976.
- Opere di Carlo Emilio Gadda*, edizione diretta da Dante Isella, collana "I Libri della Spiga", Garzanti Editore, Milano, 1988-93.
- Picchi E., "Informatica e scienze umane: Procedure di analisi testuale", in *Parola e Immagine* a cura di Maria Assunta Zanetti, (Pubblicazione della Facoltà di Lettere e Filosofia dell'Università di Pavia; 88), Firenze, La Nuova Italia Editrice, 1999, p.181-190.
- Sassi M., Caligaris C., Pecchia L., *Progetto Prisma: una proposta per l'utilizzo di sistemi informatici nella scuola*, DIDAMATICA 90, Catania, 1990.