

**L'Archivio elettronico delle *Opere* di C.E. Gadda:
come è stato costruito, come si consulta.**

Maria Luigia Ceccotti, Manuela Sassi
ILC-CNR
Pisa

nota interna , novembre 1997

- La casa editrice Garzanti ha fornito all'ILC, Istituto di Linguistica Computazionale del Consiglio Nazionale delle Ricerche, il materiale, codificato per la fotocomposizione, delle *Opere* di C.E. Gadda, edizione diretta da Dante Isella, collana 'I Libri della Spiga', Garzanti 1988-93, per la realizzazione di un Archivio elettronico;

- Il CNR ha brevettato il DBT, Data Base Testuale, progettato e sviluppato da Eugenio Picchi, dell'Istituto di Linguistica Computazionale, un sistema di analisi testuale che è la recente ed evoluta versione di una procedura di trattamento dei testi progettata nei primi anni '70.

Premessa

Nell'estate del 1993, - ricorreva il centenario della nascita e il ventennale della morte di C. E. Gadda - si formò presso l'Università Roma Tor Vergata, per iniziativa del prof. Antonio Paoletti, un gruppo di lavoro impegnato nella definizione di una metodologia da applicare allo studio del linguaggio tecnico-scientifico nelle opere di Gadda.

La constatazione che un simile progetto avrebbe avuto una partenza ottimale se fosse stato possibile avere a disposizione i testi gaddiani consultabili con strumenti informatici, una massa di dati esaustiva, dinamica, piuttosto che una raccolta, selezione lessicografica 'tradizionale', ha fatto sì che realizzare ciò divenisse un obiettivo prioritario con precedenza 'temporale'.

Nel gennaio del 1994 alla Garzanti Editore s.p.a. è stata inviata la richiesta di permettere all'ILC di utilizzare la loro memorizzazione su supporto magnetico dei testi gaddiani, da pochissimo pubblicati (1988-1993) nella collana 'I Libri della Spiga'.

- L'invio da parte della Garzanti, nel luglio '94, del materiale elettronico da loro approntato per la fotocomposizione e dell'autorizzazione a realizzare un Archivio elettronico delle Opere di Gadda con finalità didattiche,
- l'esistenza del DBT, un sistema di analisi testuale, brevetto CNR di Eugenio Picchi,
- l'impegno delle autrici a svolgere il lavoro redazionale, sia per quanto riguarda le scelte sia per la loro memorizzazione,
tutto ciò ha fatto sì che in tempi relativamente brevi venisse raggiunto l'importante traguardo della generazione di 33 archivi-Gadda-DBT e dell'Archivio (Corpus in DBT) che li comprende tutti, permettendo quindi la lettura e l'analisi delle *Opere* di Gadda mediante il computer.

In questa nota si descriveranno brevemente il Data Base Testuale e le fasi di lavoro che hanno prodotto i file-input per la generazione di archivi elettronici, secondo gli standard previsti dal sistema informatico di riferimento. Si evidenzieranno poi le possibilità di consultazione, tramite il DBT, di questi archivi, singolarmente, per sottoinsiemi, tutti quanti insieme.

Se nella presentazione di questo lavoro è stato definito 'Archivio' lo strumento informatico che permette la lettura e la consultazione delle opere di Gadda, in questo report si utilizzerà d'ora in poi il termine archivio in relazione a un'opera di Gadda (*La Madonna dei filosofi* o le *Traduzioni*, che comprendono tre testi autonomi o le *Schede autobiografiche*, in tutto sei pagine), il termine corpus per l'insieme dei 33 archivi o per un loro sottoinsieme.

1) Il Data Base Testuale

Il sistema DBT è un sistema interattivo testuale, che, tramite funzioni proposte da appositi menu, permette la lettura e la ricerca testuale in un

archivio, che può corrispondere ad un unico testo o a più testi, o in un **corpus**, un insieme di archivi.

Tale sistema, che opera su personal computer IBM ed è compatibile con il sistema operativo MS/DOS, riconosce il testo memorizzato in formato ASCII ottenibile tramite i più diffusi programmi di word-processing ed è stato utilizzato, nella sua prima versione del 1987, sotto DOS con un'interfaccia, sviluppata in Borland Pascal, che metteva a disposizione i menu a formato finestra. L'utilizzo successivo dell'interfaccia MS-Windows ha permesso un'interazione più 'amichevole' e da pochi mesi è attiva anche una versione sperimentale che permette di interrogare archivi presso l'ILC tramite INTERNET e il WEB-Server (www.ilc.pi.cnr.it).

La lettura di un testo

Il DBT permette la lettura del testo per unità logiche (racconto, saggio, capitolo, paragrafo, etc.). E' una lettura dinamica in quanto ogni parola del testo può essere 'catturata' per mezzo del mouse ed essere utilizzata per eseguire ricerche nell'archivio o nel corpus.

La ricerca testuale

Il sistema propone vari livelli di analisi di un testo, livelli attivati dall'inserimento di codici. Per semplificare distinguiamo un **livello base** ed un **livello codificato**.

Livello base

Un file diventa un file-input-DBT con l'inserimento di un unico codice, il *codice di riferimento logico* (lo descriveremo in seguito) che, inserito nel primo record del file, è l'identificatore di un'unità logica del testo, unità logica che può coincidere con tutto il testo. Il sistema riconosce i segni di interpunzione e le stringhe alfabetiche delle quali fornisce le concordanze, l'*index locorum*, l'*incipitario*, l'*explicitario*, l'indice alfabetic, l'indice inverso, l'indice decrescente, l'indice di leggibilità; produce l'analisi di sequenze di parole, di sequenze di caratteri e di preposizioni; inoltre mette a disposizione la funzione che calcola le cooccorrenze statistiche (*mutual information*).

Livello codificato

Un testo non è costituito di semplici stringhe alfabetiche e di segni di interpunzione: ci sono i nomi propri, i numeri, i segni speciali quali il *trait d'union*, sono spesso presenti parole di altre lingue, disegni, formule, note che devono essere distinte dal testo e così via. E'

l'inserimento quindi di codici, segnalatori (tag) di fenomeni linguistici, che consente di trasformare dati in informazioni, da selezionare ed estrarre tramite funzioni di ricerca, che vengono attivate dal sistema solo in presenza nel testo input dei codici richiesti. Conviene precisare che tutte le funzioni di ricerca del livello base sono attive anche a questo livello, permettendo, ad esempio, l'index locorum delle parole francesi presenti nell'archivio, concordanze distinte per le forme (Guisa e guisa, Oliva e oliva) che il sistema a livello base riconoscerebbe come unica forma (guisa, oliva).

2) Gli archivi-Gadda-DBT

Le fasi di lavoro che producono un testo consultabile mediante questo sistema di analisi testuale sono quindi:

- la creazione, mediante la digitazione o l'utilizzo del lettore ottico, di un testo in 'machine readable form',
- l'inserimento nel 'testo' di codici che permetteranno al sistema la 'manipolazione' dei dati-testo,
- l'analisi da parte del sistema del 'testo codificato' mediante la scomposizione degli elementi che sono
 - le parole, i numeri, i segni speciali che compongono il 'testo',
 - i codici inseriti dal redattore, che attiveranno funzioni del sistema,
- la generazione di un archivio elettronico da consultare mediante il sistema.

2.1) Il materiale Garzanti e la preparazione dei file-Gadda-input

Dal materiale Garzanti sono stati 'recuperati' in machine readable form i testi di Gadda, da un insieme di file, corrispondenti a più di 22 Mb, file-testo che complessivamente contavano circa 8 Mb. I file-Garzanti, 'ripuliti' dai codici predisposti per la fotocomposizione, hanno avuto una vita brevissima perchè hanno subito delle modificazioni relative all'impaginazione per vincoli imposti dal DBT.

Le parole spezzate a fine riga appaiono nella pagina elettronica ricomposte all'inizio del rigo contenente la seconda parte della parola. Se questo tipo di ricomposizione è prodotto dal programma di generazione dell'archivio-DBT, un intervento analogo è stato precedentemente applicato da tastiera nei file-input in due casi :

- quando una parola spezzata si trova a fine pagina. In tal caso la ricomposizione della parola è stata effettuata sul testo-input solitamente nella pagina che contiene la maggior parte dei caratteri della parola. La parola spezzata è stata ricomposta nella pagina successiva quando è corredata di una nota a piè della pagina contenente la seconda sottostringa.
- quando una data, una numerazione, un'abbreviazione continuano nel record (riga) successivo, dal momento che il codice relativo è applicato ogni volta su un record (nel quale può essere applicato più volte), anche in questo caso è obbligatorio un intervento nel testo-input, che viene leggermente modificato - la data, la numerazione, la sigla vengono ricomposti in un record - per permettere l'applicazione di questi codici.

Una caratteristica di molte pagine gaddiane, la frequente presenza dei puntini di sospensione, ha costretto all'inserimento di uno spazio bianco tra la fine o l'inizio di una parola da questi seguita o preceduta. Le eccezioni sono poche e riguardano parole 'sospese' quali per esempio

Gadda, *Quer pasticciaccio...* (Letteratura), RR2, Garzanti [1]
baciato la signora ... «Quanno je prendevano certe ***ff...urie*** ...». Don
Lorenzo riesci a salvar l'effe: - RR2-QPL.3.1668.Pag.0403.9

Gadda, *Quer pasticciaccio...*, RR2, Garzanti [1]
Bottafavi ch'aveva incontrato le venditrici de caciotta: «***n...o***,»
disse, con gran pena dei labbri che non - RR2-QP.4.230.Pag.0091.38

Gadda, *Quer pasticciaccio...* (Letteratura), RR2, Garzanti [1]
L'avete già ... licenziata? ...». «... ***Sss...i***, ... sì, sì ...». «...
Ebbene - RR2-QPL.4.908.Pag.0431.33

La pagina elettronica è modificata anche quando vi sono

- note a piè di pagina,
- disegni, formule.

Di questi due ultimi casi parleremo dettagliatamente quando saranno descritti i relativi codici: codice nota e codice immagine.

2.2) I codici DBT

Codice riferimento logico

%RR1-

Questo è l'unico codice che **deve** essere inserito in un archivio-input-DBT. Ne basta uno, inserito nel primo record qualora l'archivio sia considerato un'unica unità logica. Solitamente, tenendo conto dei capitoli, dei paragrafi di un testo, ne vengono decisi e immessi parecchi.

Quindi la prima importante decisione presa è stata quella relativa ai riferimenti logici da inserire e per i quali vincolanti sono stati ovviamente i dati dell'edizione garzantiana. Tranne che in poche eccezioni si è fatto riferimento ai dati di 'Bibliografia e Indici', pubblicato insieme con il volume quinto dell'edizione garzantiana.

Nella costruzione di tale codice è stato tenuto presente il seguente schema:

- % come primo carattere obbligatorio,
- la sigla del volume Garzanti:
 - RR1 per *Romanzi e Racconti I*,
 - RR2 per *Romanzi e Racconti II*,
 - SGF1 per *Saggi Giornali Favole I*,

- SGF2 per *Saggi Giornali Favole II*,
- SV(P) per *Scritti Vari*
- S(V)P per *Scritti Postumi*
- '-'
- la sigla dell'archivio DBT, sigla che non può essere più lunga di tre caratteri (vd. Tabella 1 a pag. 39). Unica eccezione riguarda le tre opere *Madonna dei filosofi*, *Il Castello di Udine*, *L'Adalgisa*, facenti parte de *I sogni e la folgore*, una raccolta pubblicata da Einaudi nel 1955 ed il cui titolo compare anche nel primo volume della Garzanti. Nel codice di queste tre opere dopo il trattino compare SF seguita dal punto e dalla sigla dell'opera.
- '.'
- [un codice alfanumerico che caratterizza un particolare testo]. Per alcuni archivi è stato inserito questo codice per individuare caratteristiche dei vari testi. L'esempio più importante è GGP (il *Giornale di guerra e di prigionia*) dove è stato opportuno distinguere il giornale (G) dal diario (D), ciò che Gadda ha scritto nel 1916 da ciò che ha annotato nel 1918, etc.
- '.
- una numerazione progressiva, che tiene conto di capitoli, paragrafi così come sono tra l'altro registrati nell'Indice generale, numerazione soggetta ad espansioni in cui:
 - '.' evidenzia il passaggio ad un nuovo livello,
 - un'abbreviazione permette di individuare una particolare caratteristica dell'unità logica (*pref* per prefazione, *nt* per nota, etc).

Il primo riferimento logico registrato in ogni archivio è quello che identifica tutto l'archivio e che comprende solitamente il titolo, l'indice, una dedica; diamo di seguito alcuni esempi.

Nell'archivio ADA (sigla per *L'Adalgisa*) il primo riferimento logico è:

```
%RR1 - SF . ADA     e contiene:
{T} L'Adalgisa
Disegni milanesi
{W} INDICE
Notte di luna    289
Quando il Girolamo ha smesso ....    299
Claudio disimpara a vivere    343
Quattro figlie ebbe e ciascuna regina    353
Strane dicerie contristano i Bertoloni    379
I ritagli di tempo    407
Navi approdano al Parapagàl    425
Un «concerto» di centoventi professori    441
Al Parco, in una sera di maggio    481
L'Adalgisa    509
{P} A Giuseppe De Robertis
```

Seguono gli altri 20, a coppie, dell'unità logica di un *Disegno milanese* e della successiva unità logica contenente le note relative:

2) %RR1-SF.ADA.1
 3) %RR1-SF.ADA.1 (nt)
 ...
 20) %RR1-SF.ADA.10
 21) %RR1-SF.ADA.10 (nt)

Nell'archivio PLF (sigla per *Il Primo Libro delle Favole*) è stato deciso che ogni favola era da considerarsi un'unità logica e quindi oltre al primo riferimento logico e all'ultimo, quello della Nota bibliografica, sono stati inseriti 186 codici tanti quante sono le favole del libro.

1) %SGF2-PLF
 2) %SGF2-PLF.1
 ...
 187) %SGF2-PLF.186
 188) %SGF2-PLF.nt

Nell'archivio TVR (sigla per *Il Tevere*) nella parte seconda 'Parlato del documentario', Gadda raggruppa alcuni paragrafi dopo l'indicazione, nel centro del rigo, dei minuti e/o secondi previsti per girare le scene descritte successivamente; è stato quindi deciso di considerare quel gruppo di paragrafi come una unità logica, come si vede nel caso delle unità logiche 30) e sgg.:

1) %S (V) P-TVR	30) %S (V) P-TVR.II.14/20
2) %S (V) P-TVR.I	...
3) %S (V) P-TVR.I.I	35) %S (V) P-TVR.II.25/26
4) %S (V) P-TVR.I.I.1	...
...	46) %S (V) P-TVR.II.37/38
15) %S (V) P-TVR.I.I.12	...
16) %S (V) P-TVR.II	53) %S (V) P-TVR.II.45/46
17) %S (V) P-TVR.II.1	...
...	71) %S (V) P-TVR.II.64
29) %S (V) P-TVR.II.13	

E' importante strutturare nel modo migliore i riferimenti logici in un archivio, riferimenti che hanno un altro importante compito oltre a quello di localizzare ogni parola indicizzata del testo: pur interrogando l'intero corpus, si può, ponendo delle condizioni, come vedremo in seguito, selezionare solo le concordanze degli archivi, delle unità logiche che interessano.

Codice riferimento topografico

\$xxxx\$

E' una stringa, di 4 caratteri alfanumerici preceduti e seguiti dal segno speciale '\$', che, inserita nel testo input in un record tutto suo, segnala che dal rigo (o dal record) successivo comincia una nuova pagina. Tale stringa,

- che non è obbligatorio ma opportuno inserire,
- priva del segno speciale '\$' iniziale e finale,
- preceduta dall'indicazione 'Pag.'
- accresciuta dopo un '.' del numero della riga della pagina contenente la parola indicizzata, numero calcolato da una funzione del DBT,

consente quindi la localizzazione fisica dell'occorrenza di una parola nel testo ed è visualizzata, dopo il contesto di ogni occorrenza della parola, insieme con il Riferimento logico.

Codice corsivo

&C...&c, &K...&k -

Queste due coppie di codici, coppie formate dal primo elemento che segnala l'inizio, dal secondo che segnala la fine, permettono di 'informare' il DBT della presenza nel testo cartaceo di una parola o di un insieme di parole in corsivo, la prima coppia di codici, la presenza di uno o più caratteri in corsivo all' interno di una parola, la seconda coppia di codici.

Codice segno speciale

&...&

Precede quei segni speciali, quali '-', '!', che non sono da valutare come segni di punteggiatura ma trait d'union, segno d'abbreviazione. Quando il trait d'union è stato utilizzato da Gadda per creare stringhe eccessivamente lunghe è stata adottata la decisione di trasformarlo in semplice trattino, e quindi in segno d'interpunzione, introducendo prima e dopo di esso lo spazio bianco.

Codice maiuscola

&M...&

Precede la prima lettera di tutte quelle parole inizianti nel testo cartaceo con la maiuscola quali:

- i nomi propri di persona, di luogo
- il titolo di un'opera letteraria, del capitolo, del paragrafo di un'opera letteraria
- il titolo di un quadro, di una rivista, di un giornale
- il nome di una nave, di una società finanziaria

- i nomi comuni che per decisione gaddiana iniziano in siffatto modo.

Questo codice permette quindi di poter distinguere la parola iniziante con maiuscola dopo il punto dalla parola iniziante con maiuscola per norma o per volontà gaddiana, non nella visualizzazione della pagina elettronica, dove la maiuscola compare sempre tale, ma nella ricerca della parola.

Codice legame

&_ . . .

Ad_hoc, a_gratis, a_latere, ante_litteram, a_priori, best_sellers, coram_populo, enfant_terrible, en_passant, in_primis, marrons_glacés sono esempi di utilizzo di questo codice che permette quindi di trattare come unità lessicale più stringhe di caratteri (sic_et_simpliciter).

Codice personaggio

\nome personaggio\...

Il DBT permette di inserire nel testo input il codice personaggio, particolarmente utile nei testi teatrali. Questa indicazione, da racchiudere tra due '\', permette di attribuire a un personaggio la sua parte di testo, fino a quando un codice analogo venga utilizzato per un altro personaggio o fino a quando si ritorna al testo narrativo con '\\\'.

Codice poesia

&V . . . &P

L'utilizzo nel file input di questa coppia di codici (il primo segnala l'inizio di un testo poetico, il secondo il ritorno alla prosa) produce nel contesto di una parola che fa parte del testo poetico la segnalazione di fine verso rappresentata dal segno speciale '/\'.

Codice linguaggio

&Lx . . .

Il sistema automaticamente riconosce un solo linguaggio, quello di base, se in un archivio non si utilizza questo codice; mentre, se si decide di distinguere nel testo il linguaggio base da altri tipi di linguaggio, deve essere inizializzato con **&L0** o con un altro codice linguaggio, subito dopo il primo riferimento logico. E'

importante precisare che quando nel testo vi è un brano o una parola, che appartiene ad esempio al francese, si introduce il codice stabilito (**&Lf**) mentre il ripristino del linguaggio base o di un altro linguaggio, ad esempio spagnolo (**&Le**), si ottiene con l'inserimento del codice corrispondente. Il testo, così codificato, in fase di consultazione può essere interrogato per un linguaggio alla volta ma nulla vieta che possano essere attivati tutti i linguaggi insieme, quindi l'intero testo, in cui in ogni caso il DBT distingue parole francesi, tedesche, spagnole etc.

Opportuno far presente che questo codice può essere utilizzato per costruire non solo sottoinsiemi di lingue naturali ma anche di linguaggi speciali o di parole che si vogliono raggruppare per una particolare caratteristica.

Questo codice è stato utilizzato:

- per le lingue naturali (mediante una lettera minuscola o mediante una cifra),
- per linguaggi speciali (mediante una lettera maiuscola, T[itolo], P[erigrafico], A (note), W (editoriale).
- per raggruppare sottoinsiemi 'particolari' di parole (mediante un segno speciale, il segno '|', e quindi mediante l'inserimento del codice **&L|**, ad esempio, si possono selezionare i nomi etichettati come 'personaggi' nel corpus).

Per maggiori dettagli si veda in Appendice (a pag. 41) l'elenco dei codici di linguaggio utilizzati.

Codici Data, Numero, Abbreviazione

&{D...}, &{N...}, &{A...}

E' già stato detto che il DBT permette di indicizzare numeri, date, abbreviazioni. E' obbligatorio che ogni codice sia aperto e chiuso sullo stesso record, record che può contenere anche più di un codice di questo tipo. Opportuno far presente che il DBT considera un'unica stringa di caratteri ciò che viene inserito dopo la D, la N, la A e che nel richiedere una data/abbreviazione /numero è bene utilizzare una stringa compresa tra due asterischi per ottenere una risposta esaustiva. Quando vogliamo recuperare ad esempio 'novembre' non solo attraverso la funzione Parola ma anche attraverso la funzione Data, in quest'ultimo caso (sarà più chiaro in seguito), gli asterischi permetteranno di recuperare date quali '14 novembre 1893', 'novembre '88'.

L'utilizzo di questi codici è stato particolarmente problematico per due importanti opere di Gadda: *Giornale di guerra e di prigionia* e *Meditazione Milanese* per la quantità consistente di date, numeri, abbreviazioni. E' da tenere presente che alla richiesta del sottoinsieme delle date, l'ordine non sarà 'cronologico' ma 'alfabetico'.

Il DBT permette con gli stessi codici 'normalizzati' di costruire anche un elenco di date ordinate cronologicamente, un elenco di sigle o di numeri che nel testo

cartaceo presentano differenze tipografiche (ad esempio la presenza o la mancanza di punti in una sigla, in un numero), rilevanti per il computer. Si prevede di realizzare questa 'normalizzazione' prossimamente.

Codice Nota

Indicizzata **&N(A) testo nota**

Non indicizzata **&N testo nota**

Abbiamo già accennato a questo codice che, con il *codice immagine*, è stato l'ultimo ad essere inserito. Nei volumi Garzanti ci sono a) note raccolte a fine testo), b) note a piè di pagina, che talvolta continuano a piè della/e pagina/e successiva/e. L'indicatore-nota è un numero (nota esplicativa-storica) o una lettera in corsivo (variante alternativa o evolutiva) in esponente a una parola. La numerazione può essere per pagina (note a piè di pagina), per uno o più testi (note a fine testo). Il DBT permette, durante la lettura di un'unità logica o di un contesto allargato, la visualizzazione delle note, segnalate da icone (links). Tali note, contenenti un testo dinamico in cui si può col mouse catturare una parola da ricercare nell'intero archivio, possono essere indicizzate o non indicizzate. Ciò è reso possibile dal codice nota, che apre e chiude un testo-nota, fisicamente collocato nel file input subito dopo la parola che lo 'chiama'.

Disporre di questo codice ha determinato quindi un cambiamento importante per le pagine elettroniche contenenti, come le cartacee, le note a piè di pagina. Queste note infatti sono state trasferite nei file come impone il DBT ed è stato deciso di

- indicizzare quelle esplicative-storiche, contrassegnandole con il linguaggio A,
- non indicizzare le note che propongono varianti alternative o evolutive, con la prospettiva di indicizzarle con l'inserimento di un altro linguaggio qualora sia ritenuto utile dagli studiosi (vedi le note ai testi di Dante Isella per "La meccanica").

A questo punto si poneva il problema importante di rendere omogeneo il linguaggio di queste note esplicative-storiche con quelle a fine testo, unità logiche autonome. Ciò è stato ottenuto assegnando anche alle note esplicative-storiche a fine testo, il linguaggio A (sia chiaro, il linguaggio A non il *codice di nota indicizzata*). Poi, per far sì che anche le note a fine testo potessero essere visualizzate come le altre, è stata adottata la soluzione di duplicare il testo di tali note e di inserirle come note non indicizzate là dove nel testo sono richiamate. Gadda ha utilizzato le note per scrivere dei piccoli trattati ed in questi casi è stato possibile inserirne solo la prima parte con il rinvio, da nota, alla unità logica che contiene il testo completo.

Un problema meno importante è quello della numerazione delle note: quelle a piè di pagina solitamente hanno una numerazione per pagina, quelle a fine testo una numerazione progressiva per l'intero testo o per uno o più capitoli. Il DBT dà una numerazione assoluta per archivio. E' stato allora deciso di conservare nella

pagina elettronica il numero o la lettera che nella pagina cartacea individuano una nota. Per tali numeri e lettere, preceduti e seguiti da una griglia, è stato utilizzato il *codice numero* e quindi sono numeri e lettere indicizzate, tramite le quali si possono recuperare, ad esempio, tutti i contesti in cui compare una nota n.1 o n. 2 o n. 3 nel testo Garzanti.

Riassumendo:

Nei testi gaddiani

- le note esplicative/storiche sono contrassegnate da un numero, lo stesso che compare nel testo, e sono in alcune opere a piè di pagina, in altre a fine testo,
- le varianti alternative/evolutive sono contrassegnate da una lettera in corsivo, la stessa che compare nel testo, e sono a piè di pagina.

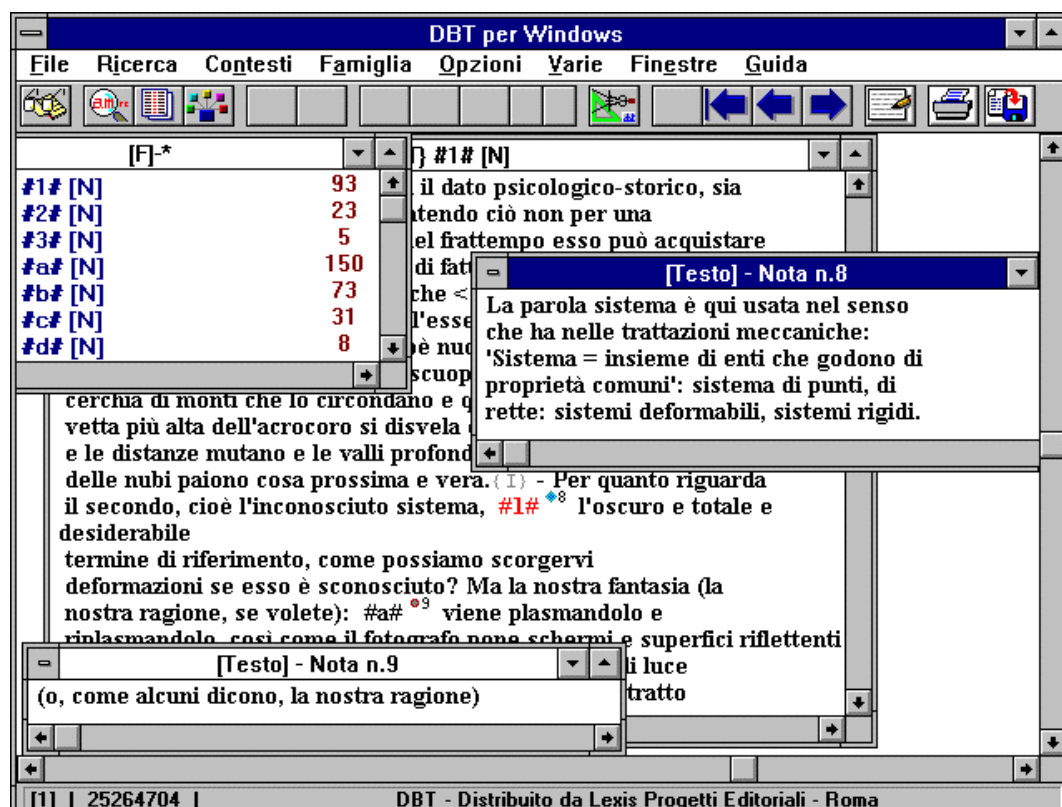
Negli archivi elettronici gaddiani

- le note di un archivio ricevono dal sistema una numerazione progressiva,
- tutte le note possono essere visualizzate durante la lettura di un'unità logica o di un contesto allargato.
- le note esplicative/storiche
 - a fine testo sono unità logiche caratterizzate dal linguaggio A (**&LA**) e, duplicate, sono state inserite nel testo, corredate del *codice nota non indicizzata*,
 - a piè di pagina sono state inserite nel testo, corredate del *codice nota indicizzata*
- le note varianti alternative o evolutive sono state inserite nel testo, corredate del *codice nota non indicizzata*
- il contesto allargato di una parola presente in una nota indicizzata visualizza anche il testo che richiama tale nota.
- tutti i numeri e le lettere che nel testo richiamano una nota sono stati racchiusi tra griglie ed è stato utilizzato il *codice numero* per renderli indicizzati.

NOTE: In pochi casi vi è anche il richiamo di una o più note da un'altra nota. Questi casi sono stati risolti inserendo queste note nel testo-nota chiamante in doppie parentesi tonde.

Pochi archivi (*La Madonna dei Filosofi*, *La meccanica*) sono corredate di brevi note esplicative-storiche alla fine di ogni unità logica e per questo motivo tali note sono state trasferite nel testo, senza creare una duplicazione.

Segue la riproduzione di una schermata con la visualizzazione del testo dei due tipi di note appena descritte tratto da 'Meditazione milanese', pagg. 628-29.



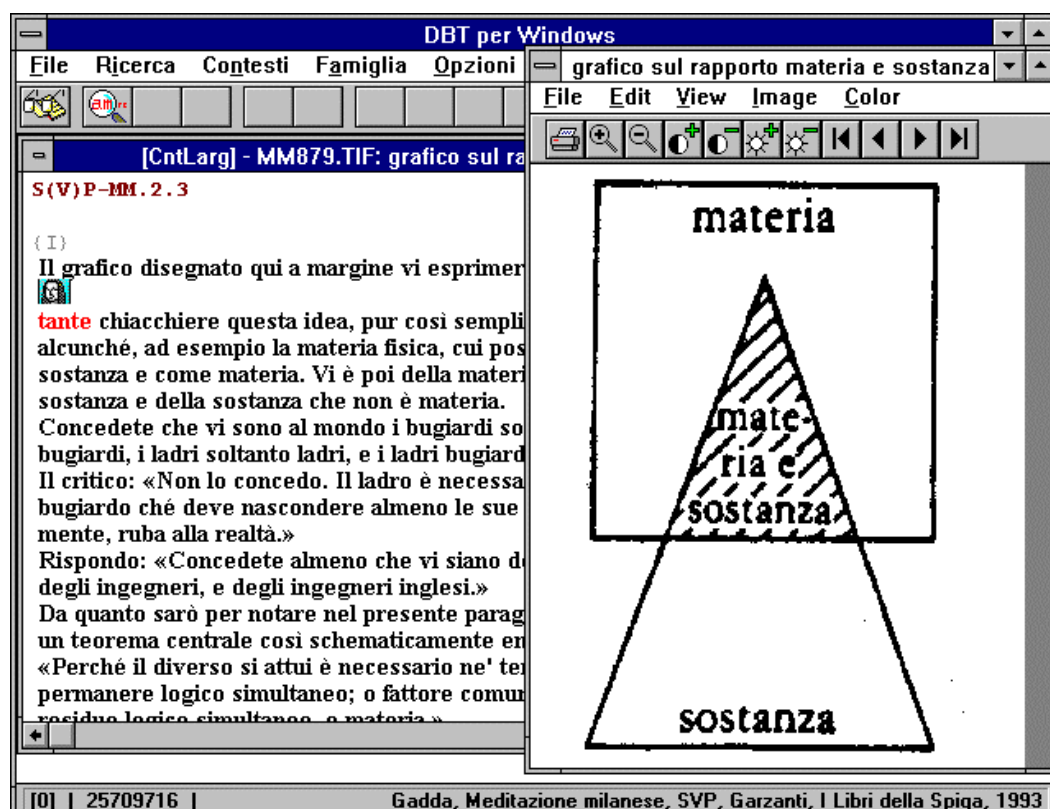
Codice Immagine

&≤ **identificatore immagine** ≥

Il DBT permette di visualizzare delle immagini inserite nel testo tramite la precedente creazione di tanti file-album quanti sono gli archivi contenenti immagini. Un'icona segnala la presenza di un'immagine quando viene attivata la lettura di un'unità logica o di un contesto allargato. Cliccando sull'icona, appare sullo schermo l'immagine che può essere manipolata usando le funzioni elementari dei programmi di gestione grafica (ingrandimento, riduzione, adeguamento della finestra, contrasto e luminosità).

Una volta attivato un file-album si può passare dall'immagine corrente a quella precedente o a quella successiva, con la contemporanea visualizzazione della finestra contenente il testo in cui vi è l'immagine selezionata.

Segnaliamo inoltre che l'immagine può essere identificata con una breve descrizione, come appare anche dall'esempio seguente.



2.3) La generazione degli archivi-Gadda-DBT

Dopo l'inserimento di codici in un'opera di Gadda, è stato creato il corrispondente archivio-DBT.

Difficilmente la generazione di un archivio viene effettuata alla prima esecuzione senza errori ortografici (caratteri che il sistema non accetta), sintattici (ad esempio il riferimento logico immediatamente preceduto dal riferimento topografico), errori che il DBT raccoglie in un file consultabile. L'archivio viene ugualmente generato e può essere 'testato', in questa fase, alla ricerca, soprattutto, degli errori semantici che il DBT non è in grado di riconoscere. Se ad esempio nel testo al **&Lf**, codice che precede la prima parola in francese di un brano, non segue il ritorno a **&Lo** quando il testo è nuovamente in italiano, il DBT continuerà a considerare attivo sempre il codice **&Lf**.

E' già stato detto (*codice rifer. logico*) che gli archivi-Gadda-DBT sono obbligatoriamente etichettati con un nome non più lungo di tre caratteri. E' il redattore a decidere questa etichetta che corrisponde all'etichetta del file TITOLO in cui il DBT troverà il nome del file input da cui generare l'archivio-DBT. Per questi archivi-Gadda i tre, a volte due caratteri, corrispondono alla sigla utilizzata per la stessa opera nell'Indice dei Titoli che fa parte del volume Bibliografia e Indici. In pochi casi questa corrispondenza non è stata rispettata. Ad esempio non è stata applicata per *L'Adalgisa*, che nell'Indice Garzanti ha come sigla L'A, per *Gli anni* che ha A, per *La meccanica* che ha M. L'elenco delle etichette dei file-Gadda-DBT è riportato in Appendice nella **Tabella 1** (a pag. 39) secondo l'ordine alfabetico relativo alle sigle, come appare quando si vuole consultare un 'testo singolo'.

Nella composizione del Corpus, invece, i singoli archivi sono inseriti nell'ordine in cui i testi sono nei volumi Garzanti, ordine che è visibile quando lo si consulta attraverso il menù 'Selezione testi da corpus'; compare una mascherina che dà la possibilità di attivare tutti gli archivi o un sottoinsieme di essi (ad esempio quelli del volume SGF1 (Saggi Giornali Favole I).

Nella **Tabella 2** (a pag. 40 dell'Appendice) sono riportati alcuni dati quantitativi:

- nella prima colonna vi sono i dati relativi alle occorrenze presenti in ciascun archivio,
- nella seconda colonna quelli relativi alle parole,
- nella terza i bytes del file-input,
- nella quarta i bytes dei file-DBT, relativi a ciascun archivio.

2.4) L'individuazione di 'problemi' nei file-input tramite il DBT

Durante la fase avanzata di inserimento nel file-input dei codici DBT, abbiamo spesso usufruito dell'archivio-DBT provvisorio per verificare se le scelte fatte fossero ottimali e coerenti e se fosse preferibile qualche cambiamento.

Un esempio:

nell'opera di Gadda è abbastanza frequente l'utilizzo di forestierismi. In *Scritti dispersi* (la scelta dell'archivio non è casuale in quanto è il più grosso archivio, costituito da testi variamente etichettabili, recensioni, saggi...) sono stati individuati un centinaio di parole o locuzioni che nel dizionario Garzanti sono registrate come parole o locuzioni che fanno parte del patrimonio lessicale dell'italiano d'uso generale. In questo caso tre soluzioni ci sembravano applicabili:

- codificare con il Linguaggio d'origine la parola o locuzione in modo tale che potesse essere selezionata con le altre. Tale soluzione è stata adottata per quei forestierismi che Gadda non sottopone alla norma che li vuole in italiano invariabili o che Gadda utilizza trascrivendoli correttamente nella lingua d'origine mentre in un dizionario di uso generale sono registrati diversamente (vd. *corvéé*, *corvéés* mentre il Garzanti e lo Zanichelli danno *corvé*);

- codificare con un Linguaggio ad hoc questo gruppo di parole/locuzioni, senza essere certi che Gadda le considerasse parole straniere (a volte le mette in evidenza con le virgolette, a volte utilizzando il carattere corsivo, a volte no) o patrimonio della lingua italiana;

- lasciare nel linguaggio base questi forestierismi, soluzione adottata in considerazione soprattutto del fatto che l'utilizzatore 'medio' dell'archivio non cerca in un dizionario di uso comune humour o sui generis o passe-partout, comprensibile a differenza dell'italianissimo comunella. E' pur vero che Gadda ha scritto qualche decennio fa, ma è anche vero che il lettore a cui si è fatto idealmente riferimento è quello di oggi, quello di domani, quindi tale archivio nella attuale versione è 'general purpose'.

3) Come consultare l'Archivio-Gadda

Il sistema testuale DBT genera da un file opportunamente codificato un archivio-DBT e mette a disposizione le funzioni per consultarlo. Non tutte le funzioni sono attivate automaticamente ma alcune lo sono solo se il redattore ha inserito nel testo input i codici che le attivano. Se, ad esempio, il linguaggio dell'archivio-DBT è solo quello di base, il DBT non propone neppure nel menu a disposizione dell'utente l'opzione 'linguaggio'. L'apertura del sistema visualizza innanzitutto una finestra in cui si può scegliere tra la consultazione di un unico **archivio** e la consultazione del **corpus**. Nel primo caso compare l'elenco degli archivi e, attivata la riga di quello prescelto e cliccando su OK, si rende possibile la sua consultazione. Nel secondo caso il sistema permette di consultare tutto il corpus o di selezionarne una parte.

Nella descrizione seguente il riferimento esplicito è alla consultazione di un archivio. Quando tale consultazione è valida anche per il corpus non lo si è evidenziato. Si è fatto il contrario quando certe funzioni di ricerca sono permesse consultando un singolo archivio e non lo sono consultando il corpus.

3.1) L'ambiente WinDBT

L'ambiente WinDBT gestisce, mediante finestre, le varie fasi di interrogazione di un testo e la visualizzazione degli elementi via via generati. La finestra 'chiave' di tutta la seduta è QUADRO, che contiene la lista di tutte le parole ricercate e selezionate dall'utente nella sessione corrente. Volendo conservare il risultato di una serie di ricerche per una seduta successiva, è possibile salvare il contenuto di Quadro per riutilizzarlo in seguito.

Quando un archivio o un corpus viene aperto per la consultazione, il sistema propone automaticamente all'utente, mediante una finestra di dialogo, RICERCA PAROLE, in cui la barra di scorrimento propone un asterisco, la possibilità di ottenere la visualizzazione di tutte le parole del linguaggio base, con l'attivazione degli ultimi parametri registrati. L'utente ha tre scelte:

- ⇒ cliccare sul tasto OK
- ⇒ cliccare sul tasto ANNULLA e attivare una diversa tattica di interrogazione (cambiando ad esempio linguaggio)
- ⇒ sostituire l'asterisco con una stringa di caratteri e cliccare sull'OK

Scegliamo per ora di cliccare su ANNULLA, facendo sparire la finestra di dialogo ed attivando la BARRA MENU che appare in alto.

Da sinistra a destra si legge:

FILE	RICERCA	CONTESTI	FAMIGLIA	OPZIONI	VARIE	FINESTRE	GUIDA
------	---------	----------	----------	---------	-------	----------	-------

I vari menu non sono tutti quanti attivi in questa fase iniziale della seduta:

FILE	offre comandi per salvare i dati raccolti o per uscire dal DBT.
RICERCA	racchiude l'insieme delle funzioni fondamentali per la ricerca e l'uso successivo degli altri menu.
CONTESTI	sarà attivato dal sistema dopo una prima consultazione positiva dell'archivio.
FAMIGLIA	sarà attivato quando dall'archivio siano state 'estratte' almeno due parole.
OPZIONI	consente di modificare le caratteristiche di visualizzazione dei risultati e verrà descritto per primo.
VARIE	permette ricerche statistiche e l'apertura di un blocknotes per memorizzare appunti.
FINESTRE	fornisce la possibilità di segmentare il video.
GUIDA	propone dati sull'archivio, informazioni sul DBT e una guida in linea all'utilizzo dei comandi.

3.2) Menu Opzioni

Il menu OPZIONI non serve alla consultazione ma ad intervenire sulle modalità di visualizzazione dei dati che successivamente chiederemo.

Tramite OPZIONI si può:

- modificare l'immagine del video e/o della stampa mediante le funzioni Stili Video, Stili Stampa, Colori Finestra di Dialogo. In breve chi consulta l'archivio ha la possibilità di cambiare il colore, la dimensione, il tipo, lo stile dei vari dati, da visualizzare e/o da stampare.
- fissare delle restrizioni relativamente a Riferimento, Personaggio, Parola in fine verso. Non si tratta in questo caso di scelte 'estetiche' e converrà presentare queste funzioni anche con l'ausilio di esempi tratti dagli archivi DBT.

Restrizioni su Riferimento

Quando l'archivio viene interrogato e la risposta è positiva, si possono chiedere tutti i contesti delle occorrenze della forma che è stata cercata. Chi interroga può porre delle limitazioni, se, ad esempio, vuole solo i contesti delle occorrenze che si trovano in unità logiche contenenti note e lo può fare con questa opzione scegliendo tra le seguenti funzioni: contiene / uguale / inizia / finisce (il cui

soggetto è il riferimento logico) e tramite gli operatori *and / or / not*. Il tasto 'Sensibile al Maiuscolo' consente di differenziare il carattere maiuscolo da quello minuscolo nel porre la richiesta.

Con queste funzioni e questi operatori possono essere creati comandi tipo:

'contiene RR and not OP',

che permetterà di consultare tutte le opere comprese nei primi due volumi dell'edizione Garzanti, ad eccezione del '*Pasticciaccio*' nelle due versioni.

'contiene OP and PDO and VM.1.9',

che permetterà di confrontare le due versioni del *Pasticciaccio*, insieme con la sceneggiatura *Il palazzo degli ori* e con '*Il pasticciccio*' in *I viaggi la morte* (corrispondente al saggio n.9 della parte n.1).

'finisce con (nt)'

che permetterà la consultazione di tutte le unità logiche contenenti note gaddiane.

Restrizioni su Personaggio

Le restrizioni su Personaggio sono possibili con le stesse funzioni e gli stessi operatori utilizzabili per il Riferimento. Posso ad esempio chiedere al sistema di visualizzare solo i contesti delle occorrenze pronunciate da un personaggio o dai personaggi il cui nome inizia o finisce o contiene una lettera, un gruppo di lettere. Se ad esempio chi consulta l'archivio ricorda vagamente la presenza in un testo di Gadda di un personaggio 'giornalista' che parla di 'linguaggio' può:

- costruire una restrizione Personaggio con il comando *and contiene giornalista*,
- richiedere la parola 'linguaggio' al corpus ed ottenere la visualizzazione dei soli contesti in cui una sua occorrenza sia pronunciata da un personaggio 'giornalista'.

La risposta è:

```
1) *linguaggio           1 (20)
2) {A}linguaggio        -- (6)
```

in quanto la parola *linguaggio* è presente nel linguaggio base di 20 archivi (1 forma nell'archivio attivo che è '*La Madonna dei Filosofi*'), e nel linguaggio A, linguaggio nota, (ed i trattini avvertono che nell'archivio attivo non è presente) di un altro sottoinsieme del corpus.

Alla prima richiesta dei contesti il sistema risponde elencando gli archivi in cui è la forma 'linguaggio' corredata della frequenza fra parentesi quadre e visualizzando solo i due contesti in cui compare il personaggio 'giornalista':

* Restrizione su Personaggio *

```
Gadda, La Madonna dei Filosofi, RR1, Garzanti [1]
Gadda, La cognizione del dolore, RR1, Garzanti [1]
Gadda, Quer pasticciccio., RR2, Garzanti [1]
Gadda, Quer pasticciccio..(Letteratura), RR2, Garzanti [1]
```

Gadda, La meccanica, RR2, Garzanti [2]
 Gadda, Accoppiamenti giudiziosi, RR2, Garzanti [1]
 Gadda, Gli anni, SGF1, Garzanti [1]
 Gadda, Verso la Certosa, SGF1, Garzanti [1]
 Gadda, I viaggi la morte, SGF1, Garzanti [24]
 Gadda, Scritti dispersi, SGF1, Garzanti [50]

giornalista\..." che sia suscettibile di costituire la base omogenea del*linguaggio* letterario?» \gadda\Posso ammettere che l'italiano corrente, - SGF1-SD.114.5.Pag.1190.15

giornalista\... televisione sono strumenti di unificazione culturale anche sul piano del *linguaggio*, e in che misura?» \gadda\Alla quarta domanda - SGF1-SD.114.159.Pag.1194.17

Gadda, I Luigi di Francia, SGF2, Garzanti [3]
 Gadda, Eros e Priapo, SGF2, Garzanti [5]
 Gadda, Il guerriero..Foscolo, SGF2, Garzanti [1]
 Gadda, Giornale di guerra e di prigionia, SGF2, Garzanti [1]
 Gadda, Pagine di divulgazione tecnica, SVP, Garzanti [3]
 Gadda, TRADUZIONI, SVP, Garzanti [1]
 Gadda, Racconto..del novecento, SVP, Garzanti [2]
 Gadda, Meditazione milanese, SVP, Garzanti [8]
 Gadda, Il palazzo degli ori, SVP, Garzanti [1]
 Gadda, Altri scritti, SVP, Garzanti [6]

Alla richiesta di contesti relativi alla seconda forma non viene visualizzato alcun contesto sempre a causa della Restrizione su Personaggio:

1) linguaggio 1 (20)
 2)*{A}linguaggio -- (6)

* Restrizione su Personaggio *

Gadda, L'Adalgisa (disegni milanesi), RR1, Garzanti [1]
 Gadda, La cognizione del dolore, RR1, Garzanti [1]
 Gadda, Accoppiamenti giudiziosi, RR2, Garzanti [2]
 Gadda, Racconti incompiuti, RR2, Garzanti [1]
 Gadda, I viaggi la morte, SGF1, Garzanti [1]
 Gadda, Racconto..del novecento, SVP, Garzanti [1]

NOTA: Da non sottovalutare che è possibile attivare insieme e Restrizioni su Riferimento e Restrizioni su Personaggio. Se questa possibilità non è particolarmente significativa per l'opera di Gadda può esserlo ad esempio per corpora di opere teatrali.

Restrizione su Parola in fine verso

In questo caso la restrizione possibile è una: visualizzare solo i contesti di versi che terminano con la parola che si chiede. Vengono quindi ignorati i contesti che non rispondono a tale condizione. Si ricorda che la barra indica il fine verso.

[Cnt] - {0}questa
 * Restrizione Parola in fine verso *

Una specie di sussulto, stavolta, i prodromi d'una scarica epilettica: per me alla terra non fecondi *questa/ bella d'erbe famiglia e d'animali

E fu, nella vastità infinita del Possibile, fu l'unica cesura ch'_ - RR2-AG19.6.192. Pag.0917.6

come chi titillasse la nostra fede nel domani recitando: ove più il Sole/
per me alla Terra fecondando *questa/ bella d'erbe famiglia e d'animali/
porcaccioni, mi dia unica la prole/ che possa ereditare il capitale/
senza ch'_ - RR2-AG19.6.214. Pag.0917.28

Linguaggio

L'importanza di questa funzione DBT ci sembra che sia stata già sufficientemente descritta (vedi codice Linguaggio). Nel menu OPZIONI, LINGUAGGIO apre una finestra di dialogo, nella quale si ha la possibilità di confermare la condizione di partenza, che corrisponde il più delle volte al linguaggio T[titolo], oppure di accettare la richiesta per 'Tutti i linguaggi' o di scegliere uno tra i codici linguaggio inseriti, scelta che rimane attiva fino al successivo cambiamento.

Parametri DBT

Di questa opzione si sono già date testimonianze attraverso alcuni esempi già proposti. Anche in questo caso cliccando su Parametri DBT (o utilizzando un comando di tastiera composto Ctrl+F10) si attiva una finestra di dialogo che permette di variare interattivamente:

<u>Dimensione Contesto</u>	La lunghezza del contesto da visualizzare, che di partenza è di 10 elementi del testo prima e dopo la parola.
<u>Parametri per Famiglie</u>	La distanza massima che ci può essere nel testo tra le parole che formano una famiglia; viene proposta una distanza standard che è possibile modificare in qualunque momento. Se la distanza è = 1, si cercheranno due parole contigue.
<u>Ampiezza Riga</u>	La lunghezza massima della riga dell'unità testo; la lunghezza proposta è 80 caratteri.
<u>Testo impaginato</u>	Pulsante, che attivato, permette di visualizzare il testo così come è il testo cartaceo. Attivare questo pulsante non esclude l'attivazione della funzione precedente in quanto la lunghezza della riga deve essere $\geq$ al rigo più lungo del testo cartaceo, normalmente tra 60 e 80 caratteri.
<u>Modalità KWIC</u>	Visualizza i contesti con la 'parola' centrale. Richiedendo la modalità KWIC ("Key Word In Context") ogni contesto è presentato su un'unica riga, con la parola in esponente in posizione centrale nella finestra; non viene considerato il valore di Dimensione Contesti.
<u>Visualizza Linguaggio</u>	Visualizza nell'unità testo la presenza dei codici

linguaggio presenti.

Una precisazione importante: queste funzioni possono essere utilizzate all'inizio della seduta ma anche durante la consultazione dell'archivio. Se una richiesta è stata fatta con l'attivazione di certi parametri e ci si accorge dalla risposta che i parametri vanno cambiati, effettuato questo cambiamento, non bisogna dimenticare di rilanciare la richiesta. Se alcuni parametri sono stati cambiati, quando si chiude l'archivio consultato, il sistema chiede se i nuovi parametri debbono essere memorizzati. Se la risposta è positiva nella seduta successiva saranno attivati i nuovi parametri, validi anche se sarà aperto un altro archivio.

3.3) Menu Ricerca

Il menu RICERCA, attivo all'apertura dell'archivio, ci permette di consultarlo.

Se clicco su RICERCA mi viene proposto:

- la ricerca 'Parola'
- [la ricerca 'Data']
- [la ricerca 'Numero']
- [la ricerca 'Abbreviazione']
- la 'Lettura Testo'
- l'attivazione di ricerca 'Parola' con la funzione 'Lista di Parole'
- la 'Distribuzione Frequenze', funzione attiva solo dopo che l'utente ha

ottenuto una prima risposta positiva.

La ricerca 'Data' o 'Numero' o 'Abbreviazione' sarà proposta solo se nell'archivio consultato sono stati inseriti i rispettivi codici. Si può verificare il caso che per alcuni archivi appaia la possibilità di cercare una data mentre non appaia quella di cercare una sigla. Non è detto, si badi bene, che date o sigle non ci siano in quell'archivio mentre è sicuro che il redattore non ha utilizzato questi codici.

Proviamo a consultare un archivio DBT, ad esempio SA (*Schede autobiografiche*, SGF2) ed attiviamo la ricerca 'Parola'.

1.a domanda: *milano*. La risposta, che è una, attiva automaticamente la funzione QUADRO, in cui compare la parola Milano con il numero delle occorrenze riscontrate dal programma nell'archivio corrente.

Salva Finestra - Gadda, Schede autobiografiche, SGF2, Garzanti	
Quadro	

1) * Milano	9

L'asterisco che compare prima di Milano indica che quella forma è attiva e cliccando su di essa si ottengono i relativi contesti:

[Cnt] - {0}Milano

si perdono nella notte dei tempi: le schede anagrafiche dicono *Milano*
14 novembre 1893 ore 6, via Alessandro Manzoni numero 3, terzo_ -
SGF4biograf1.2. Pag.0871.2

Ginnasio e liceo di stato: il 'Parini'. Politecnico di *Milano*,
felicitemente rotto in due dalla guerra, quella di Trento_ -
SGF4biograf1.8. Pag.0871.8

'Carlo Erba'. Corso di filosofia alla Università di *Milano*, negli anni
dal 1924 al 1928. Incarichi ingegnereschi vari_ - SGF4biograf1.12.
Pag.0871.12

{I} E' nato a *Milano* quattordici giorni avanti la caduta del Ministero
Giolitti,_ - SGF4biograf2.1. Pag.0872.1

{I} Nato a *Milano* il 14 novembre 1893 da genitori lombardi (nonno
materno ungherese, Giovanni_ - SGF4biograf3.1. Pag.0873.1

1924-25 ebbero importanza psichicoimmaginativa per il Nostro le località
di *Milano*, Monza: zona di campagna, e, in Pegli_ - SGF4biograf3.27.
Pag.0873.27

stupendo sui laghi brianzoli, Monte Resegone. Educazione elementare a
Milano: (scuole del Comune): ottima se pure «catechistica_ -
SGF4biograf3.34. Pag.0874.5

anche per intervento della madre. Ginnasio e Liceo Parini (*Milano*) con
buoni insegnanti e buoni risultati; tempi e termini_ - SGF4biograf3.39.
Pag.0874.10

liceali) e preparò gli esami di corso in filosofia (*Milano*) rinunciando
nei susseguenti anni a conseguire la tesi di laurea_ - SGF4biograf3.54.
Pag.0874.25

Se la domanda è: *scuol**, la risposta compare in una finestra Forme [F] come la
seguente:

ed in questo caso è l'utente a decidere se inserire o no in QUADRO le due forme o

[F]-scuol*	

scuola	1
scuole	3

una delle due. Deciso per l'inserimento di entrambe, il QUADRO si presenta:

Quadro	

1) scuola	1
2) *scuole	3

con l'ultima parola inserita preceduta dall'asterisco; cliccando sulla parola di cui si desiderano i contesti, con il primo clic vi si posiziona l'asterisco e con il secondo clic si ottengono i contesti secondo gli ultimi parametri fissati.

Una precisazione importante va fatta a questo punto prima di procedere, relativamente alla posizione dell'asterisco nel QUADRO. Non è infatti solo la funzione CONTESTI che viene attivata per la parola preceduta dall'asterisco ma anche la funzione 'Distribuzione Frequenze' [DF], che, a differenza di CONTESTI, può essere attivata solo usando la BARRA MENU. Questa funzione permette di conoscere la distribuzione delle occorrenze nelle unità logiche dell'archivio e quindi se la attiviamo appare una finestra [DF] con questi dati:

Salva Finestra - Gadda, <i>I miti del somaro</i> , S(V)P, Garzanti	
Quadro	

1) *mito	51

[DF]- S(V)P-MDS	
17	S(V)P-MDS.1
25	S(V)P-MDS.2
9	S(V)P-MDS.3

e cliccando sulla riga dell'unità logica che ci interessa, ad esempio S(V)P-MDS.3, otteniamo una Finestra Contesti come:

[Cnt] - {0}mito in S(V)P-MDS.3

-
- 1) li spirocheti paterni e i suoi proprî. Il *mito* del coltello, dopo il mito energetico del randello e S(V)P-MDS.3.61. Pag.0920.27
 - 2) suoi proprî. Il mito del coltello, dopo il *mito* energetico del randello e dell'olio. Il coltello S(V)P-MDS.3.62. Pag.0920.28
 - 3) divieto lo spazio. Qui mi contento negare che il *mito* del coltello sul ventrone abbî mai e comporti un S(V)P-MDS.3.81. Pag.0921.7
 - 4) tronfio, vent'anni, a tagliare baionette. Il *mito* qualunque, il mito falso. Cinque milioni di baionette S(V)P-MDS.3.91. Pag.0921.17
 - 5) a tagliare baionette. Il mito qualunque, il *mito* falso. Cinque milioni di baionette. Otto milioni di S(V)P-MDS.3.92. Pag.0921.18
 - 6) nel clearing delle smargiassate maramaldesche. No, il *mito* non pol essere un mito qualunque, neppure nel S(V)P-MDS.3.111. Pag.0921.37
 - 7) maramaldesche. No, il mito non pol essere un *mito* qualunque, neppure nel cervello di un criminale. Il S(V)P-MDS.3.111. Pag.0921.37
 - 8) che a trascinar le folle gli bisognava infoiarle d'un *mito*, che fu quel funerario carnovale di coltella, S(V)P-MDS.3.139. Pag.0922.25
 - 9) Bagascia ladra, pescò su dal letamaio dei miti un *mito* qualunque: lo spulizzî e grattò quanto gli venne S(V)P-MDS.3.144. Pag.0922.30

RICERCA offre altre due possibilità: Lettura Testo e Lista di parole.

- Lettura Testo apre la Finestra 'Indice Testo' in cui sono elencati tutti i riferimenti logici dell'archivio, che cliccati permettono la lettura del testo nella/e finestra/e TESTO. Ogni finestra visualizza fino a circa 200 righe di testo; se tale spazio non è sufficiente a contenere l'intera unità logica, si attiva

un pulsante con una freccia rossa rivolta verso il basso che, cliccata, apre una o più finestre successive, fino a esaurimento del testo.

- Lista di Parole attiva una finestra 'File Open' che mostra un elenco di file preparati precedentemente, nei quali siano state inserite, una per riga, tutte le forme che si vogliono cercare negli archivi (possono essere utilizzati anche dei caratteri "jolly" quali l'asterisco ed il dollaro, spiegati dettagliatamente in 'ricerca per sottostringhe' a pag. 34). Tutte le forme presenti nel file sono ricercate automaticamente e, se trovate, inserite direttamente nella finestra Quadro, oppure nel caso di risposte multiple sono inserite in proprie finestre. Forme per eventuali selezioni da parte dell'utente. Il file contenente la lista delle forme dovrà avere una estensione determinata (*.WRD) ed essere creato preventivamente dall'utente in formato ASCII. Il vantaggio è che l'esistenza di questi file è indipendente dall'archivio consultato e quindi utilizzabile su ciascun archivio separatamente o sull'intero corpus. D'altra parte questi file, volendo, possono essere modificati per nuove esigenze di ricerca.

3.4) Menu Famiglia

Partendo dal seguente QUADRO:

Gadda, Schede autobiografiche, SGF2, Garzanti		
Quadro		

1) Milano		9
2) scuola		1
3) *scuole		3

ci cimentiamo a definire una 'Famiglia di ricerca', un tipo di ricerca nei testi che riguarda due o più forme, associate usando gli operatori logici: OR (presenza di almeno una delle parole indicate), AND (co-occorrenza di più parole), (AND NOT) non presenza contemporanea di più parole.

La sequenza di operazioni per definire una nuova famiglia inizia attivando la funzione DEFINISCI. Le forme di QUADRO vengono mostrate nella finestra di dialogo denominata 'Definizione Famiglia' e trasformate in un elenco di identificatori formati da W+n, il numero corrispondente alla posizione della parola in QUADRO. Nel nostro caso Milano=W1, scuola=W2, scuole=W3. Le parole della lista sono selezionabili cliccando sul record della tabella che le contiene per mezzo del mouse; il tasto "Famiglia in OR" genera automaticamente una famiglia in OR (dopo aver attivato un'altra finestra di dialogo che permette di selezionare le parole) ma la stessa richiesta può essere digitata come quelle in cui si utilizzano gli altri operatori logici. Le espressioni sono costruite quindi con gli identificatori selezionati, con gli operatori logici '&'(AND), '|'(OR) e '\'(AND

NOT), con le parentesi '()' che fissano delle gerarchie logiche all'interno dell'espressione.

Il valore standard di distanza massima tra le parole unite in Famiglia di ricerca con l'operatore AND, affinché il contesto venga preso in considerazione, si calcola sulla base degli elementi della frase, che possono essere parole, segni di interpunzione, accapo-riga, riferimenti di vario genere, ecc. Questo valore può essere preventivamente modificato dall'utente (funzione Parametri DBT); per es., se si dà il valore 1, la ricerca otterrà solo quei contesti in cui le parole sono contigue. Un valore troppo elevato prenderebbe in considerazione parole a distanza eccessiva, rendendo non significativa la ricerca. Si faccia comunque attenzione a coordinare il valore della distanza a cui si cercano le parole cooccorrenti con il parametro della dimensione dei contesti, che deve essere sempre uguale o maggiore a quello della distanza massima impostata.

Il tasto Sequenza, se attivato, fa sì che nel corso della ricerca di una famiglia di parole, i contesti che richiedono co-occorrenze debbono essere accettati solamente se le parole che co-occorrono si presentano nella stessa sequenza secondo la quale appaiono nella definizione della famiglia.

Esempi di Definizione di famiglia:

1. W1|W2 => definisce una famiglia che individua i contesti in cui compare la forma 1 oppure la forma 2; in tal caso il parametro Distanza non viene preso in considerazione.
2. W1&W2 => definisce una famiglia di ricerca che individua i contesti in cui compaiono, entro la distanza stabilita, entrambe le forme contrassegnate dai numeri 1 e 2 nel Quadro.
3. W1&(W2|W3) => definisce una famiglia di ricerca che individua i contesti in cui compare la forma 1 accompagnata o dalla forma 2 o dalla forma 3: vedi l'Esempio 1, con il risultato della ricerca *Milano&(scuola|scuole)*.
4. W1&W2\W3 => definisce una famiglia di ricerca che individua i contesti in cui compaiono entrambe le forme contrassegnate dai numeri 1 e 2 e non compare la forma contrassegnata nel Quadro dal numero 3: vedi l'Esempio 2, con il risultato della ricerca *mito&qualunque\un*.

Esempio 1:

[CntFam] - Famiglia n. 1

```
-----
N.Famiglie definite = 1 [20,15]
*Milano* & (*scuola* | *scuole*)
N. Contesti associati : 3
```

da un par d'anni a questa parte ha preso a sorvolare sulla questione. Battezzato a San Fedele, fu cresimato a San Simpliciano. *Scuole* elementari comunali. Ginnasio e liceo di stato: il 'Parini'. Politecnico

di *Milano*, felicemente rotto in due dalla guerra, quella di Trento e di Trieste. Un anno e due mesi di prigionia: a Rastat: e a Celle nello Hannover. Scuola di elettrotecnica 'Carlo Erba'. Corso di filosofia_ - SGF4biograf1.8. Pag.0871.8

dalla guerra, quella di Trento e di Trieste. Un anno e due mesi di prigionia: a Rastat: e a Celle nello Hannover. *Scuola* di elettrotecnica 'Carlo Erba'. Corso di filosofia alla Università di *Milano*, negli anni dal 1924 al 1928. Incarichi ingegnereschi vari, talora gravi o gravissimi. Lavorò in Lombardia, in Argentina, a Roma, in Francia, in Germania: e di nuovo a Roma._ - SGF4biograf1.12. Pag.0871.12

Longone, Lago del Segrino). Suo padre costruì una fottuta casa di campagna a Longone nel '99-'900 e questa strampalata casa gli rimase appiccicata fino al 1937. Panorama stupendo sui laghi brianzoli, Monte Resegone. Educazione elementare a *Milano*: (*scuole* del Comune): ottima se pure «catechistica», che il N. chiama borromeismo laico: (Borromeo-Cattaneo). Ottimi maestri elementari: geometria, italiano. Forti immagini risorgimentali, senso nazionale fin dall'infanzia,_ - SGF4biograf3.34. Pag.0874.5

Esempio 2:

Gadda, <i>I miti del somaro</i> , S(V)P, Garzanti	
Quadro	

1) *mito	51
2) qualunque	15
3) un	88

N.Famiglie definite = 8 [4,10]

(*mito & *qualunque*) \ *un*

N. Contesti associati : 5

- 1) disegno. Ma il mito d'accatto, il *mito *qualunque* non crea della storia, non può essere annoverato tra - S(V)P-MDS.1.154. Pag.0905.3
- 2) viso la focaccia di segatura di pioppo del «mito *qualunque*». - S(V)P-MDS.1.185. Pag.0905.34
- 3) contrario di quel che fecero gli zelatori del «mito *qualunque*». Sono laboriosamente pervenuti al mito della Ragione o del - S(V)P-MDS.2.9. Pag.0909.6
- 4) del mito travolgente. Il mito teatrale ossia «*mito *qualunque* purché mito» arraffa a caso una orecchiata (e - S(V)P-MDS.2.19. Pag.0909.16
- 5) vent'anni, a tagliare baionette. Il *mito *qualunque*, il mito falso. Cinque milioni di baionette. - S(V)P-MDS.3.92. Pag.0921.18

La funzione così definita viene a far parte del patrimonio della funzione FAMIGLIE e, essendo la prima, identificata tramite F1. Le successive saranno numerate con F2, F3, Fn... e potranno essere costruite non solo con le parole inserite nel QUADRO ma anche con le Famiglie precedentemente create. Si potranno ad esempio costruire richieste quali

- 1) (F1|F2)&W5 => che individua i contesti in cui compaiono le parole della F[amiglia] 1 o quelle della F[amiglia] 2 in cooccorrenza con la Forma numero 5.

- 2) (F1&W5)\W6 => che individua i contesti in cui compaiono, entro la distanza stabilita, le forme della F[amiglia] 1 e la Forma numero 5 e in cui non compaia la Forma numero 6.

Dopo la definizione di almeno una famiglia,

RICERCA permette di attivare la finestra di dialogo Ricerca di famiglia e di far scorrere verso l'alto o verso il basso la lista delle famiglie già costruite alla ricerca di quella che vogliamo riutilizzare e per la quale, volendo, possiamo variare, come abbiamo già detto, e la dimensione dei contesti e il parametro distanza tra le parole;

VISUALIZZA invece permette solamente di scorrere lo stesso elenco senza la possibilità di attivare alcuna famiglia.

3.5) Strategie di consultazione

Torniamo a quella che è la funzione fondamentale di un sistema di analisi testuale e proviamo a descrivere una ricerca di PAROLA molto 'varia' che l'attuale versione del DBT permette: si clicca in un punto qualsiasi di un contesto e si ottiene un contesto più ampio (200 righe) e, cliccando su una determinata parola del testo, si seleziona la parola per ottenerne le concordanze. Viene cioè offerta la possibilità di esaudire la 'curiosità' di sapere in quali altri punti dell'archivio (si faccia attenzione, dell'intero archivio o del corpus che è stato aperto) c'è quella parola che ha attirato la nostra attenzione. Si ottiene così, in maniera molto naturale, l'inserimento di tale forma nel QUADRO o in una Finestra Forme con il numero delle occorrenze che la riguardano.

Prima della richiesta dei contesti di una parola o di una famiglia di parole, può essere posta in essere una delle restrizioni, che già sappiamo come porre attraverso OPZIONI.

Proviamo a questo punto a riassumere alcuni percorsi offerti dal DBT che abbiano come risultato finale un insieme di contesti:

1° tipo di percorso

(scelta LINGUAGGIO), richiesta PAROLA <o DATA o NUMERO o ABBREVIAZIONE>, *inserimento automatico di parola* <o.....> in QUADRO, (scelta PARAMETRI), richiesta CONTESTI, richiesta contesto largo;

(scelta LINGUAGGIO), richiesta PAROLA <o....>, scelta PAROLA <o....>, (scelta PARAMETRI), *inserimento in QUADRO*, richiesta DISTRIBUZIONE DELLE FREQUENZE, visualizzazione contesti dall'unità logica scelta;

(scelta LINGUAGGIO), richiesta PAROLA <o...>, scelta PAROLA <o...> , (scelta PARAMETRI), *inserimenti in QUADRO*, definizione FAMIGLIA, visualizzazione contesti, richiesta contesto largo.

2° tipo di percorso

Richiesta di contesto largo (vedi primo percorso) o di unità logica tramite 'Lettura Testo', clic su parola, *inserimento in QUADRO*,..... (ossia si prosegue la ricerca tramite la selezione automatica di parole presenti nei contesti allargati o nell'unità logica aperta, senza passare dal menu PAROLA)

Questo 2° percorso non è attuabile se la ricerca è mirata ad ottenere date, numeri, sigle, etichettate in quanto tali, ma è comunque possibile utilizzare le parole in esse contenute per la ricerca in tutto l'archivio o corpus delle occorrenze della stringa cliccata.

3° tipo di percorso

Soltanto nell'interrogazione del singolo testo, a partire da almeno una forma presente in QUADRO, può essere usata la funzione 'CoOccorrenze statistiche' del menu VARIE. Tale funzione permette di ottenere una lista di parole che, in base ad un calcolo probabilistico che utilizza la formula della 'mutual information', possono essere associate con la/e parola/e data/e. Il risultato consiste in un elenco di parole in ordine decrescente rispetto al peso assegnato a ciascuna di esse in relazione sia alla distanza che alla quantità riscontrate in prossimità della/le parola/e data/e. Ognuna di esse dà accesso ai contesti relativi.

Ricerca per sottostringhe

La ricerca ovviamente più importante che il DBT permette è la ricerca di una parola. Quando si chiede una parola si possono indifferentemente utilizzare le lettere maiuscole o quelle minuscole, le vocali accentate o non accentate.

Il programma risponde distinguendo non solo tra parole inizianti per maiuscola e per minuscola ma anche tra parole con accenti e parole prive di accenti. Provando ad esempio a cercare la parola 'cantò', il programma mi risponde distinguendo 'canto' e cantò', se, è ovvio, sono ambedue nel testo. E' quindi più 'economico' utilizzare le lettere minuscole non accentate, mentre se si vuole selezionare già con la richiesta una parola iniziante con la maiuscola o una parola accentata si ha a disposizione dei segni speciali (° per la maiuscola, ^ per l'accento) che si inseriscono nella richiesta come di seguito elencato.

In maniera schematica e con l'ausilio di esempi si presentano le varie possibilità offerte dal DBT nella ricerca di una parola, passando dai casi più generici a quelli più specifici:

Il carattere '\$' permette di selezionare tutte le stringhe composte di un solo carattere, il carattere '*' di avere l'elenco di tutte le parole dell'archivio consultato, tenendo conto in tutte e due i casi delle restrizioni precedentemente fissate (ad es. un linguaggio, le date etc.). Questi due caratteri possono essere utilizzati singolarmente oppure essere inseriti in stringhe variamente composte.

Elenchiamo allora i vari elementi che possono integrare le stringhe o unire più sottostringhe con cui si consulta l'archivio-DBT:

\$	è un carattere jolly che sostituisce una sola lettera, iniziale, centrale, o finale,
*	è un carattere jolly che può essere $= > 0$,
°	è un carattere che si può utilizzare solo all'inizio della stringa. per richiedere parole inizianti con la maiuscola, in singolo testo, se viene consultato l'intero corpus la stessa richiesta deve iniziare con asterisco,
^	è un carattere, iniziale, centrale o finale che permette di selezionare le parole provviste di vocali accentate,
?	è un carattere che permette la cosiddetta ricerca 'fuzzy', la ricerca cioè di parole, della cui grafia siamo incerti. Se la stringa è preceduta da questo carattere in fase di consultazione, il DBT risponde elencando la stessa parola, se la grafia era effettivamente corretta, e tutte le parole che presentano in qualunque posizione un carattere differente. La differenza può riguardare anche un carattere in più o in meno in qualsiasi posizione della forma. Possibile effettuare la stessa ricerca proponendo il doppio:
??	a cui il DBT risponde elencando tutte le parole che differiscono da quella data, che potrebbe essere 'restituita' in quanto corretta, non solamente per un carattere ma anche per due caratteri purché adiacenti. L'utilizzo di questo carattere, singolo o doppio, è incompatibile con i caratteri '\$' e '*', mentre è compatibile con '^' e con '°'.

Si elencano alcune richieste permesse dal DBT tramite la funzione 'Parola' e la relativa spiegazione:

<i>pi</i> \$	parole di tre caratteri inizianti con 'pi'
<i>pi</i> *	parole di almeno due caratteri inizianti con 'pi'
<i>\$pi</i>	parole di tre caratteri terminanti con 'pi'
<i>*pi</i>	parole di almeno due caratteri terminanti con 'pi'
<i>\$pi</i> \$	parole di quattro caratteri contenenti 'pi'
<i>*pi</i> *	parole di almeno due caratteri contenenti 'pi'
<i>pi</i> \$sa	parole di cinque caratteri inizianti con 'pi' e terminanti con 'sa'
<i>pi</i> *sa	parole inizianti con 'pi' e terminanti con 'sa'.

<i>pi*sa\$\$</i>	parole inizianti per 'pi', contenenti 'sa' seguita da due lettere
<i>pi*sa*</i>	parole inizianti per 'pi', contenenti 'sa'
<i>*pi*sa*</i>	parole che contengono 'pi' e 'sa'

Se si vogliono selezionare parole inizianti con la lettera maiuscola, si formuleranno richieste di questo tipo, in caso di interrogazione di archivio singolo, mentre durante l'interrogazione del corpus è sempre necessario anteporre un '*' in tutti i casi in cui viene usato il simbolo per la maiuscola.

<i>°*</i>	parole inizianti con la maiuscola
<i>°p*</i>	parole inizianti con la lettera 'P'
<i>°p\$\$\$</i>	parole inizianti con la lettera 'P' e lunghe 4 caratteri

Se la volontà è quella di selezionare solo parole accentate si comporranno richieste simili:

<i>*^</i>	parole che finiscono con una vocale accentata;
<i>*a^*</i>	parole che contengano la vocale 'a' comunque accentata.

Se si vogliono selezionare parole inizianti con la maiuscola e contenenti vocali accentate si comporranno richieste quali:

<i>°*^</i>	parole con la maiuscola e con vocali accentate
<i>°p*\$^</i>	parole inizianti con la 'P' e con l'ultima lettera accentata

Richieste ancora più complicate possono essere

<i>\$(pi/sa)</i>	parole di tre lettere che terminano con 'pi' o con 'sa'
<i>\$(pi/sa)\$</i>	parole di quattro lettere contenenti 'pi' o 'sa'
<i>(pi/sa)\$</i>	parole di tre lettere inizianti con 'pi' o con 'sa'
<i>*(pi/sa)</i>	parole di n (con $n \geq 2$) lettere terminanti con 'pi' o con 'sa'
<i>*(pi/sa)*</i>	parole di n (con $n \geq 2$) lettere contenenti 'pi' o con 'sa'
<i>(pi/sa)*</i>	parole di n (con $n \geq 2$) lettere inizianti con 'pi' o con 'sa'
<i>*(pi/sa)\$</i>	parole di almeno tre lettere inizianti con m ($m \geq 0$) lettere + 'pi' o 'sa' + 1 lettera
<i>*(pi/sa)*(na/no)*</i>	parole di almeno 4 lettere, inizianti, contenenti e terminanti con m ($m \geq 0$) lettere + 'pi' o 'sa' + 'na' o 'no'.

NOTE: Alla richiesta di parole composte da un preciso numero di caratteri, per es. 5, il DBT risponde proponendo un elenco in cui compaiono anche parole di 4 caratteri se quelle parole contengono caratteri accentati o iniziano con una lettera maiuscola, di 3 caratteri se quelle parole contengono caratteri accentati e iniziano con una lettera maiuscola.

Appendice

Utility del sistema

BlockNotes

Durante la seduta di lavoro è possibile utilizzare, sotto il menu VARIE, la funzione Apri BlockNotes, che permette di rendere attiva una finestra per annotarvi direttamente a video le osservazioni che emergono nel corso dell'analisi. Oltre a questo l'utente può trasferire in BlockNotes tutti i contesti o liste di parole che risultano dalla ricerca nel testo. Alla fine i materiali così costituiti saranno memorizzati in un file (formato ASCII) a disposizione per successive elaborazioni.

Il BlockNotes è costituito da una finestra di Editor nella quale sono espletabili tutte le funzioni tipiche di un word processor: Taglia, Copia e Incolla.

Per trasferire dati dalle finestre di DBT in BlockNotes si utilizza il comando Trasferisci in BlocNotes, che a partire dalle varie finestre di interrogazione (Contesti, Frequenze, Quadro, ecc. ma non dalla finestra Testo) trasferisce dati in BlockNotes.

Il trasferimento si ottiene selezionando l'apposito menu locale col pulsante destro del mouse. Attraverso la funzione "Trasferisci in BlockNotes" può essere trasferito l'elemento della finestra sul quale era posizionato il cursore, mentre in alcuni casi il programma prevede anche la funzione "Trasferisci Finestra in BlockNotes", che permette il trasferimento dell'intero contenuto della finestra selezionata, compresa la parte che non viene visualizzata per motivi di spazio.

Stampa Finestra

Si può stampare l'intero contenuto di una finestra sulla stampante collegata attraverso la funzione omonima nel menu FILE oppure tramite il clic destro del mouse all'interno della finestra attiva. La stampa avviene con le modalità di stampa proprie di Windows ed utilizza il file di selezione dei caratteri personalizzabili dall'utente con la funzione Modifica Stili Stampa nel menu OPZIONI.

Salva Finestra.

Questa funzione ricopia l'intero contenuto di una finestra in un file esterno con le stesse modalità della funzione precedente. Il file viene registrato nella directory del DBT e il nome di tale file, generato automaticamente dal programma, è comunicato in un apposito messaggio. Più salvataggi di finestre vengono accodati

nello stesso file, a meno che non si provveda, prima del comando di salvataggio, a chiudere il file precedente. La chiusura del file, che avviene normalmente al momento della fine della esecuzione del programma, può essere espressamente richiesta per mezzo dell'apposito comando che è posto nel menu FILE.

Con questa funzione è possibile trasferire in file su disco i dati presenti in una finestra, così da renderli disponibili per successive elaborazioni. Tutto il contenuto della finestra selezionata, e non soltanto ciò che è visibile sullo schermo al momento, è trasferito su file. La copia avviene in codice ASCII, rispettando l'impaginazione che la finestra assume sullo schermo al momento della richiesta di salvataggio.

Alcune Sigle DBT

A	Abbreviazione
Cnt	Contesto/i
Cntfam	Contesto/i di famiglia
Co_occ	Cooccorrenze
D	Data
DF	Distribuzione delle Frequenze
F	Forma
N	Numero
\$	Segnala l'attivazione di Tutti i Linguaggi

Alcuni Comandi da tastiera

F1	Guida DBT
F3	Continua visualizzazione
Alt + L	Lista di Parole
Alt + P	Parola
Alt + Q	Distribuzione delle Frequenze
Alt + T	Lettura testo
Alt + B	Apri BlockNotes
Alt + S	Salva Finestra
Alt + C	Contesti senza parametri
Alt + K	Contesti con parametri
Alt + D	Definizione di Famiglia
Alt + R	Ricerca di Famiglia
Ctrl + F5	Finestra precedente
Ctrl + F6	Finestra successiva
Ctrl + F7	Chiude tutte le finestre
Ctrl + F10	Parametri DBT

Tabella 1

ADA	L'Adalgisa (disegni milanesi)	RR1	1988
AG	Accoppiamenti giudiziosi	RR2	1989
AN	Gli anni	SGF1	1991
AS	Altri scritti	SVP	1993
CD	La cognizione del dolore	RR1	1988
CDU	Il castello di Udine	RR1	1988
EP	Eros e Priapo (Da furore a cenere)	SGF2	1992
GB	Gonnella buffone	SVP	1993
GAS	Il guerriero, l'amazzone, lo spiritoFoscolo	SGF2	1992
GGP	Giornale di guerra e di prigionia	SGF2	1992
HJ	Háry János	SVP	1993
LDF	I Luigi di Francia	SGF2	1992
MDF	La Madonna dei Filosofi	RR1	1988
MDI	Le meraviglie d'Italia	SGF1	1991
MDS	I miti del somaro	SVP	1993
MEC	La meccanica	RR2	1989
MM	Meditazione milanese	SVP	1993
PDO	Il palazzo degli ori	SVP	1993
PDT	PAGINE DI DIVULGAZIONE TECNICA	SVP	1993
PLF	Il primo libro delle Favole	SGF2	1992
PS	Poesie	SGF2	1992
QP	Quer pasticciaccio brutto de via Merulana	RR2	1989
QPL	Quer pasticciaccio... («Letteratura», 1946-47)	RR2	1989
RAI	Racconti incompiuti	RR2	1989
RD	Racconti dispersi	RR2	1989
RI	Racconto italiano di ignoto del novecento	SVP	1993
SA	Schede autobiografiche	SGF2	1992
SD	Scritti dispersi	SGF1	1991
TRE	TRADUZIONI	SVP	1993
TVR	Il Tevere	SVP	1993
UI	Ultimi inediti	SVP	1993
VLC	Verso la Certosa	SGF1	1991
VM	I viaggi la morte	SGF1	1991

Tabella 2

N.occorr.	N.forme	byte file input	byte file DBT	testi
28204	8950	183984	349542	Gadda, La Madonna dei Filosofi, RR1, Garzanti
62980	15819	415773	702417	Gadda, Il castello di Udine, RR1, Garzanti
109840	25852	738770	1222110	Gadda, L'Adalgisa (disegni milanesi), RR1, Garzanti
70737	15898	455902	758899	Gadda, La cognizione del dolore, RR1, Garzanti
100427	19121	625042	1014548	Gadda, Quer pasticciaccio..., RR2, Garzanti
71397	15171	448084	734693	Gadda, Quer pasticciaccio...(Letter.),RR2, Garzanti
40593	10664	263814	470532	Gadda, La meccanica, RR2, Garzanti
98295	19838	626563	1015199	Gadda, Accoppiamenti giudiziosi, RR2, Garzanti
29818	8576	186977	343327	Gadda, Racconti dispersi, RR2, Garzanti
33412	9438	218707	391037	Gadda, Racconti incompiuti, RR2, Garzanti
57154	16191	373743	983943	Gadda, Le meraviglie d'Italia, SGF1, Garzanti
20691	7099	133846	263975	Gadda, Gli anni, SGF1, Garzanti
43581	13148	283916	525441	Gadda, Verso la Certosa, SGF1, Garzanti
78434	20047	520418	907452	Gadda, I viaggi la morte, SGF1, Garzanti
194762	33953	1301592	2006408	Gadda, Scritti dispersi, SGF1, Garzanti
15665	5972	98260	228050	Gadda, Il primo libro delle Favole, SGF2, Garzanti
40317	9977	272638	455334	Gadda, I Luigi di Francia, SGF2, Garzanti
56678	15152	359188	643952	Gadda, Eros e Priapo, SGF2, Garzanti
9374	3215	68912	118604	Gadda, Il guerriero ... Foscolo, SGF2, Garzanti
136485	18923	893496	1273896	Gadda, Giornale di guerra e ..., SGF2, Garzanti
1327	754	9879	24295	Gadda, Schede autobiografiche, SGF2, Garzanti
3187	1267	20020	42160	Gadda, Poesie, SGF2, Garzanti
60474	12755	411904	639412	Gadda, Pagine di divulg. tecnica, SVP, Garzanti
49575	10264	313006	504495	Gadda, TRADUZIONI, SVP, Garzanti
66986	13412	444790	712148	Gadda, Racconto .. del novecento, SVP, Garzanti
83558	14872	552105	839746	Gadda, Meditazione milanese, SVP, Garzanti
5758	2456	37878	82606	Gadda, I miti del somaro, SVP, Garzanti
19863	4786	135153	220527	Gadda, Il palazzo degli ori, SVP, Garzanti
8469	2476	57324	97204	Gadda, Gonnella buffone, SVP, Garzanti
11267	3051	81144	129359	Gadda, Hary Janos, SVP, Garzanti
8460	3122	59903	124053	Gadda, Il Tevere, SVP, Garzanti
2255	1028	14760	32698	Gadda, Ultimi inediti, SVP, Garzanti
9389	3616	61453	151954	Gadda, Altri scritti, SVP, Garzanti
1629412	366863	10668944	18010016	

NOTE: Ai 18 Mb occupati dagli archivi DBT bisogna aggiungere altri 6 Mb occupati dagli indici fusi che permettono la consultazione del corpus.

Si prevede di effettuare ulteriori controlli dopo la raccolta di questi dati e in seguito si potranno riscontrare dati quantitativi leggermente differenti.

Elenco dei linguaggi utilizzati

Si riporta l'elenco dei codici (tra apici e seguiti dalla relativa spiegazione) dei linguaggi attualmente presenti nel corpus.

'0'	è per il DBT il linguaggio base
' '	precede il nome di personaggi
':'	Identifica quelle parole che sono divise in due sottostringhe da una lineetta verticale (RI)
'A'	è utilizzato per tutte le note esplicative a piè di pagina ed anche per le unità logiche che contengono note
'T'	è inserito come prova per alcuni nomi propri che non si possono definire solamente o tedeschi, francesi, etc. (Chopin, Falstaff) o che di cui non si conosce l'origine (Clodig, Cilli)
'P'	è utilizzato per etichettare parole che si trovano nelle dediche, nelle introduzioni ad un'opera, ecc.
'W'	è usato per le parti inserite nei volumi dall'editore
'X'	etichetta un sottoinsieme molto limitato di termini facenti parte di brani gaddiani in cui sono mescolate forme dialettali, gergali con forme italiane (è sufficiente per Gadda indicare un solo linguaggio-base?)
'a'	arabo
'd'	tedesco
'e'	spagnolo
'f'	francese
'g'	greco traslitterato
'i'	inglese
'j'	genovese
'l'	latino
'm'	milanese, lombardo
'p'	piemontese
'r'	romanesco
's'	siciliano
't'	toscano
'z'	nizzardo
'6'	friulano

Unità logiche del Corpus

1)%RR1-SF	1)%RR1-CD
2)%RR1-SF.MDF	2)%RR1-CD.1
3)%RR1-SF.MDF.1	3)%RR1-CD.1.I
4)%RR1-SF.MDF.2(a)	4)%RR1-CD.1.II
5)%RR1-SF.MDF.2(b)	5)%RR1-CD.1.III
6)%RR1-SF.MDF.3	6)%RR1-CD.1.IV
7)%RR1-SF.MDF.3.1	7)%RR1-CD.2
...	8)%RR1-CD.2.V
14)%RR1-SF.MDF.3.8	9)%RR1-CD.2.VI
15)%RR1-SF.MDF.4	10)%RR1-CD.2.VII
16)%RR1-SF.MDF.5	11)%RR1-CD.2.VIII
...	12)%RR1-CD.2.IX
20)%RR1-SF.MDF.5.IV	13)%RR1-CD.app(ed-aut)
1)%RR1-SF.CDU	14)%RR1-CD.app(poesia)
2)%RR1-SF.CDU.pref	15)%RR1-CD.app(chiarim.)
3)%RR1-SF.CDU.sinossi	1)%RR2-QP
4)%RR1-SF.CDU.intr.	2)%RR2-QP.1
5)%RR1-SF.CDU.intr(nt)	...
6)%RR1-SF.CDU.1	11)%RR2-QP.10
7)%RR1-SF.CDU.1.a	1)%RR2-QPL
8)%RR1-SF.CDU.1.a(nt)	2)%RR2-QPL.1
...	...
15)%RR1-SF.CDU.1.e	6)%RR2-QPL.5
16)%RR1-SF.CDU.1.e(nt)	1)%RR2-MEC
17)%RR1-SF.CDU.2	2)%RR2-MEC.1
18)%RR1-SF.CDU.2.a	...
...	6)%RR2-MEC.5
22)%RR1-SF.CDU.2.e	7)%RR2-MEC.abbozzo
23)%RR1-SF.CDU.2(nt)	8)%RR2-MEC.abbozzo.6
24)%RR1-SF.CDU.3	9)%RR2-MEC.abbozzo.7
25)%RR1-SF.CDU.3.a	10)%RR2-MEC.abbozzo.8
26)%RR1-SF.CDU.3.a(nt)	1)%RR2-AG
27)%RR1-SF.CDU.3.b	2)%RR2-AG.1
28)%RR1-SF.CDU.3.c	...
29)%RR1-SF.CDU.3.c(nt)	28)%RR2-AG.19.6
30)%RR1-SF.CDU.4	1)%RR2-RD
31)%RR1-SF.CDU.4.I	2)%RR2-RD.1
32)%RR1-SF.CDU.4.II	...
33)%RR1-SF.CDU.4.III	8)%RR2-RD.7
34)%RR1-SF.CDU.4(nt)	1)%RR2-RAI
1)%RR1-SF.ADA	2)%RR2-RAI.1
2)%RR1-SF.ADA.1	...
3)%RR1-SF.ADA.1(nt)	1)%RR2-RAI
...	2)%RR2-RAI.1
20)%RR1-SF.ADA.10	
21)%RR1-SF.ADA.10(nt)	

3)%RR2-RAI.1.I	187)%SGF2-PLF.nt
4)%RR2-RAI.1.II	+++++
5)%RR2-RAI.2	1)%SGF2-LDF
6)%RR2-RAI.2.1	2)%SGF2-LDF.XIII
7)%RR2-RAI.2.2	3)%SGF2-LDF.XIII.1
8)%RR2-RAI.3	...
+++++	14)%SGF2-LDF.XIII.12
1)%SGF1-MDI	15)%SGF2-LDF.XIV
2)%SGF1-MDI.1	16)%SGF2-LDF.XIV.1
3)%SGF1-MDI.1.1	...
...	21)%SGF2-LDF.XIV.6
14)%SGF1-MDI.1.12	22)%SGF2-LDF.XV
15)%SGF1-MDI.1.12.nt	23)%SGF2-LDF.XV.1
16)%SGF1-MDI.1.13	...
17)%SGF1-MDI.2	35)%SGF2-LDF.XV.13
18)%SGF1-MDI.2.1	36)%SGF2-LDF.Bg
...	+++++
33)%SGF1-MDI.4.5	1)%SGF2-EP
+++++	2)%SGF2-EP.pr
1)%SGF1-AN	3)%SGF2-EP.1
2)%SGF1-AN.1	...
...	16)%SGF2-EP.12
11)%SGF1-AN.10	+++++
+++++	1)%SGF2-GAS
1)%SGF1-VLC	2)%SGF2-GAS.1
...	3)%SGF2-GAS.2
19)%SGF1-VLC.18	+++++
20)%SGF1-VLC.app.	1)%SGF2-GGP.C15.1
+++++	2)%SGF2-GGP.C15.2
1)%SGF1-VM	3)%SGF2-GGP.C16
2)%SGF1-VM.1	4)%SGF2-GGP.G16
3)%SGF1-VM.1.1	5)%SGF2-GGP.G16.1
...	6)%SGF2-GGP.G16.2
11)%SGF1-VM.1.9	7)%SGF2-GGP.G16.3
12)%SGF1-VM.2	8)%SGF2-GGP.G16.4
13)%SGF1-VM.2.1	9)%SGF2-GGP.G16.5
...	10)%SGF2-GGP.D17
25)%SGF1-VM.2.13	11)%SGF2-GGP.M
26)%SGF1-VM.3	12)%SGF2-GGP.D18.a
27)%SGF1-VM.3.1	13)%SGF2-GGP.D18.b
28)%SGF1-VM.3.1.1	14)%GSF2-GGP.Vita
29)%SGF1-VM.3.1.2	+++++
30)%SGF1-VM.3.2	1)%SGF2-SA
+++++	2)%SGF2-SA.1
1)%SGF1-SD	3)%SGF2-SA.2
2)%SGF1-SD.1	4)%SGF2-SA.3
...	+++++
123)%SGF1-SD.122	1)%SGF2-PS
+++++	2)%SGF2-PS.1
1)%SGF2-PLF	...
2)%SGF2-PLF.1	9)%SGF2-PS.8
...	+++++
186)%SGF2-PLF.186	1)%SV(P)-PDT

Indice

PREMESSA	5
1) IL DATA BASE TESTUALE	7
2) GLI ARCHIVI-GADDA-DBT	9
2.1) Il materiale Garzanti e la preparazione dei file-Gadda-input	9
2.2) I codici DBT	10
Codice riferimento logico	10
Codice riferimento topografico	12
Codice corsivo	13
Codice segno speciale	13
Codice maiuscola	13
Codice legame	14
Codice personaggio	14
Codice poesia	14
Codice linguaggio	14
Codici Data, Numero, Abbreviazione	15
Codice Nota	16
Codice Immagine	19
2.3) La generazione degli archivi-Gadda-DBT	20
2.4) L'individuazione di 'problemi' nei file-input tramite il DBT	21
3) COME CONSULTARE L'ARCHIVIO-GADDA	22
3.1) L'ambiente WinDBT	22
3.2) Menu Opzioni	23
Restrizioni su Riferimento	23
Restrizioni su Personaggio	24
Restrizione su Parola in fine verso	25
Linguaggio	26
Parametri DBT	26
3.3) Menu Ricerca	27
3.4) Menu Famiglia	30
3.5) Strategie di consultazione	33
1° tipo di percorso	33
2° tipo di percorso	34
3° tipo di percorso	34
Ricerca per sottostringhe	34

APPENDICE	37
Utility del sistema	37
BlockNotes	37
Stampa Finestra	37
Salva Finestra.	37
Alcune Sigle DBT	38
Alcuni Comandi da tastiera	38
Tabella 1	39
Tabella 2	40
Elenco dei linguaggi utilizzati	41
Unità logiche del Corpus	42
Indice	45