

Valentina Bambini[#] & Marco Trevisan[♦]

EsploraCoLFIS: Un'interfaccia web per le ricerche sul Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*

Presentiamo qui EsploraCoLFIS, una nuova interfaccia Web per la consultazione del Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS, Bertinetto et al. 2005). L'interfaccia consente di effettuare ricerche sulle statistiche di forme o lemmi, visualizzarne le concordanze e ricercare coricorrenze di coppie di parole. È inoltre possibile effettuare ricerche con vincoli, estraendo forme o lemmi che rispettino criteri definiti di parte del discorso, lunghezza e distribuzione. La ricerca può essere modulata per tipologia testuale e per sintagmaticità di forme o lemmi. Dopo una breve presentazione generale di CoLFIS (linee guida per la costruzione del corpus di testi, criteri per la lemmatizzazione, statistiche adottate), il contributo introduce EsploraCoLFIS, illustrandone le funzionalità principali con alcuni esempi di utilizzo. Data la rapidità di risposta e agilità dell'interfaccia, essa si presenta come un utile supporto per le ricerche distribuzionali e in particolare per la ricerca in psico- e neurolinguistica. L'interfaccia EsploraCoLFIS è gratuitamente consultabile sul sito del Laboratorio di Linguistica della Scuola Normale Superiore, all'indirizzo <http://linguistica.sns.it/esploracolfis/home.htm>

[#] Centro di ricerca in Neurolinguistica e Sintassi Teorica (NeTS), Istituto Universitario di Studi Superiori, IUSS, Pavia & Laboratorio di Linguistica "G. Nencioni", Scuola Normale Superiore, Pisa

[♦] CELI S.R.L., Torino

^{*} L'interfaccia web EsploraCoLFIS è stata realizzata nell'ambito di una collaborazione tra il Laboratorio di Linguistica della Scuola Normale Superiore, diretto da Pier Marco Bertinetto, e CELI S.R.L. di Torino, diretto da Luca Dini. Lo sviluppo dell'interfaccia è stato curato da Valentina Bambini (Centro di Neurolinguistica e Sintassi Teorica, IUSS Pavia & Laboratorio di Linguistica, Scuola Normale Superiore, Pisa) e Marco Trevisan (CELI, Torino). Si ringraziano Irene Ricci e Lucio Pileggi per il supporto tecnico. Per gli scopi del presente articolo, la sezione 1 si deve a V. Bambini mentre la sezione 2 è stata scritta congiuntamente dai due autori.

1 CoLFIS

1.1 Il corpus

L'intento alla base della costruzione di CoLFIS¹ era quello di costruire una banca di dati che riproducesse nel modo più fedele possibile la frequenza reale delle parole nell'esperienza del parlante, e che dunque si avvicinasse il più possibile al lessico mentale di un parlante di media cultura, piuttosto che ad un dizionario esaustivo della lingua italiana. A tal fine, è stata devoluta particolare cura nella scelta dei criteri di rappresentatività del campione, basandosi sull'indagine ISTAT sulle letture degli italiani, effettuata su una popolazione di individui di età superiore agli 11 anni, componenti di 24.000 famiglie (ISTAT, 1993).

Quanto alla scelta delle dimensioni del campione, è stato ritenuto che 3 milioni di occorrenze lessicali costituissero la dimensione minima capace di garantire al contempo l'inclusione nel corpus di un numero sufficientemente alto di lemmi a bassa frequenza e la rappresentatività del campione. Nello specifico, le ricorrenze lessicali ammontano a 3191137, salendo a 3798296 se si includono punteggiatura, numeri e altri diacritici. I materiali del corpus provengono per il 48,34% da quotidiani, per il 34,40% da periodici, e per il 17,26% da libri. Rimandiamo a Laudanna et al. (1995) per la descrizione dettagliata della costruzione del corpus.

1.2 La lemmatizzazione

Il corpus è stato compiutamente lemmatizzato, riconducendo ogni parola alla forma base (o entrata del dizionario). In prima istanza è stata effettuata una lemmatizzazione automatica, utilizzando un lemmatizzatore IBM già utilizzato per il LIP. Sulla lemmatizzazione automatica sono poi stati applicati alcuni interventi di correzione manuale, che hanno riguardato in particolare le parole sintagmatiche, ovvero le parole

¹ Il Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS) è stato realizzato da un gruppo di lavoro coordinato da Pier Marco Bertinetto (Scuola Normale Superiore, Pisa), Cristina Burani (Istituto di Scienze e Tecnologie della Cognizione, CNR, Roma), Alessandro Laudanna (Università di Salerno), Lucia Marconi (Istituto di Linguistica Computazionale, Unità Staccata di Genova, CNR, Genova), Daniela Ratti (Istituto di Linguistica Computazionale, Unità Staccata di Genova, CNR, Genova), Claudia Rolando (Istituto di Linguistica Computazionale, Unità Staccata di Genova, CNR, Genova), Anna Maria Thornton (Università de L'Aquila), e con il sostegno economico del CNR (Comitati Nazionali "Scienze e Tecnologie dell'Informazione" e "Scienze Storiche, Filosofiche e Filologiche". Finanziamento: Banca dati lessicale dell'Italiano scritto contemporaneo).

al cui interno compaiono altri elementi, non riconosciute dal lemmatizzatore automatico.

A ciascuna parola è stato assegnato un codice che la assegna alla classe grammaticale di appartenenza. Riportiamo qui i codice delle classi grammaticali, utilizzati anche nell'interfaccia EsploraCoLFIS.

Aggettivo	G	Numerale	NU
Articolo	R	Preposizione	P
Ausiliare	VA	Pronome	N
Avverbio	B	Punteggiatura	K
Congiunzione	C	Simbolo	Z
Interiezione	I	Sostantivo	S
Lingua non ben identificata	X	Tempo composto di un verbo	TC
Nome proprio	E	Verbo	V

Per le parole sintagmatiche è stato usato il simbolo @, in combinazione con altre categorie grammaticali, ad esempio @E. Rimandiamo alle note di Cristina Burani per i criteri di lemmatizzazione, disponibili sul sito <http://linguistica.sns.it/CoLFIS/Home.htm>.

1.3 Le statistiche

Le statistiche adottate riflettono l'intento di disporre di dati sulla frequenza del ricevente. Oltre alle misure di frequenza assoluta, sono state calcolate misure di dispersione e di correzione, ulteriormente calibrate per riflettere la reale esposizione del lessico. Per ogni lemma e per ogni forma sono state calcolate le seguenti statistiche:

- Frequenza assoluta, definita come il numero di ricorrenze nel corpus;
- Indice di dispersione, definito come la misura della presenza del lemma nei diversi sottosettori, compreso tra 0 e 1: più un lemma è diffuso più il valore è elevato;
- Frequenza corretta (o relativa), definita come la frequenza corretta secondo l'indice di dispersione e secondo le dimensioni dei sottosettori;
- Rango, definito come la posizione della parola nell'ordine di frequenza relativa totale.

I valori di frequenza sono disponibili sia sul corpus totale, sia sui sottocorpora. Il modello statistico adottato è quello lognormal. Si rimanda alle note di Lucia Marconi per la descrizione dei calcoli statistici <http://linguistica.sns.it/CoLFIS/Home.htm>.

I valori statistici ottenuti sono riportati nel formario (per le forme) e nel lemmario (per i lemmi), disponibili per consultazione in formato .txt, .mdb e .dbf sul sito <http://linguistica.sns.it/CoLFIS/Home.htm>, o sui siti dell'Istituto di Linguistica Computazionale di Genova e di Roma.

2 EsploraCoLFIS

Al fine di potenziare l'utilizzabilità di CoLFIS, contemplando differenti necessità e capacità, abbiamo sviluppato EsploraCoLFIS, un'interfaccia Web per la ricerca su CoLFIS. L'interfaccia si basa su un motore di ricerca di testo (Apache Lucene; <http://lucene.apache.org/java>) e fornisce funzionalità di ricerca avanzate con tempi di risposta rapidi e collegamento agile da una ricerca all'altra.

EsploraCoLFIS mira a soddisfare le necessità della maggior parte degli utenti, i cui bisogni possono essere soddisfatti da necessità predefinite (ad esempio la ricerca della frequenza di una specifica forma) ma possono anche riguardare richieste più specifiche, come ad esempio l'estrazione di lemmi che rispettino determinate caratteristiche. EsploraCoLFIS mira specialmente a soddisfare le necessità della ricerca in psicolinguistica e neurolinguistica, consentendo di modulare le principali variabili psicolinguistiche quali frequenza, lunghezza e classe di parola. Alla luce di tali obiettivi, l'interfaccia permette di esplorare i dati contenuti nel corpus lemmatizzato, le statistiche riportate nel lemmario e nel formario, e di accedere ad alcuni dei testi originali che costituiscono il corpus. È anche possibile effettuare ricerche con vincoli e ricerche di coricorrenze. La struttura dell'interfaccia consente di velocizzare la consultazione attraverso la ricerche di liste e la possibilità di muoversi da una scheda all'altra senza ripetere le ricerche.

L'interfaccia, mostrata in Figura 1, si presenta con 6 schede: *Forme*, *Lemmi*, *Concordanze*, *Coricorrenze*, *Lista* e *Opzioni*. Ciascuna scheda contiene una maschera per la ricerca, in cui è possibile specificare vari parametri lessicali e di distribuzione. I risultati sono mostrati in un pannello sottostante.

Figura 1: L'interfaccia EsploraCoLFIS

A partire dalla scheda *Forme*, l'utente può ricercare le informazioni relative alle forme, e specificamente le statistiche descritte in 1.3, ovvero: frequenza assoluta totale, frequenza relativa totale, indice di dispersione e rango. In aggiunta, in EsploraCoLFIS sono stati introdotti i logaritmi della frequenza. Nel caso di forme omografe, i risultati sono ordinati in base alla parte del discorso e successivamente al rango. Le statistiche possono essere ulteriormente dettagliate per sottocorpora, specificando i criteri di interesse nella scheda *Opzioni*. Ad esempio, l'utente può decidere di visualizzare le statistiche relative ai quotidiani o ai periodici. Nella scheda *Opzioni* è possibile configurare l'interfaccia in modo che siano visualizzati anche i codici delle parti del discorso, in aggiunta alle statistiche di base, al codice della parte del discorso ed alla struttura sintagmatica, che sono visualizzati di default. Un esempio di ricerca di forme è rappresentato in Figura 2. In Figura 3 viene presentato il pannello della scheda *Opzioni*.

Forme Lemmi Concordanze Coricorrenze Lista Opzioni

Forma:

Numero di lettere: minimo: massimo:

▼ Dizionario

Lemma:

Parte del discorso:

Sintagmaticità:

Sintagmaticità della forma:

► Statistiche

Risultati per pagina:

4 forme trovate.

Forma	Lemma	Codice parte del discorso	Parte del discorso	Contesto sintagmatico: codice della parte del discorso	Contesto sintagmatico: parte del discorso	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
quadro	QUADRO	E	nome proprio	-	-	98248	1	0.00	0.02	0.00	Concordanze Lemma
quadro	QUADRO	G	aggettivo	-	-	19174	13	1.11	1.87	0.51	Concordanze Lemma
quadro	QUADRO	S	sostantivo	-	-	1563	221	2.34	51.54	0.88	Concordanze Lemma
quadro	QUADRO	S	sostantivo	in P	in preposizione	10810	25	1.40	4.66	0.69	Concordanze* Lemma

Figura 2: Esempio di ricerca della forma “quadro”. Si noti la presenza di forme omografe ricondotte a diverse parti del discorso

Opzioni

Mostra in forme e lemmi:

- Parte del Discorso
- Contesto sintagmatico
- Numero di lettere
- Rango
- Frequenza assoluta totale
- Frequenza assoluta nei quotidiani
- Frequenza assoluta nei libri
- Frequenza assoluta nei periodici
- Logaritmo della frequenza assoluta totale
- Logaritmo della frequenza assoluta nei quotidiani
- Logaritmo della frequenza assoluta nei libri
- Logaritmo della frequenza assoluta nei periodici
- Logaritmo della frequenza relativa totale
- Frequenza relativa nei quotidiani
- Frequenza relativa nei libri
- Frequenza relativa nei periodici
- Dispersione totale
- Dispersione nei quotidiani
- Dispersione nei libri
- Dispersione nei periodici

Mostra in concordanze e coricorrenze:

- Lemma
- Codice parte del discorso
- Parte del discorso
- Numero di lettere del lemma
- Frequenza assoluta della forma
- Frequenza assoluta del lemma
- Posizione

Azzera Applica

Figura 3: Il pannello *Opzioni*, con le possibili specifiche da adottare nella ricerca sulle altre schede.

La ricerca di lemmi, effettuabile dalla scheda *Lemmi*, funziona in modo analogo a quella della scheda *Forme*, come illustrato in Figura 4.

The screenshot shows the 'Lemmi' tab of the EsploraCoLFIS interface. The search criteria are: Lemma: quadro; Parte del discorso: aggettivo; Numero di lettere: minimo: [] Azzera massimo: [] Azzera; Sintagmaticità: [] Azzera; Risultati per pagina: 10. Below the search form, it indicates '1 lemma trovato' and displays a table with the following data:

Lemma	Codice parte del discorso	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
QUADRO	G	6404	36	1.56	6.07	0.61	Concordanze Forme

Figura 4: Esempio di ricerca del lemma aggettivale “quadro”.

Le funzionalità di EsploraCoLFIS consentono di osservare anche le concordanze, visualizzando i contesti d’uso di una forma o di un lemma, interrogando la scheda *Concordanze*. È possibile modulare l’ampiezza del contesto mostrato, indicando il numero di elementi a destra e a sinistra della ricorrenza trovata. Di default il pannello dei risultati mostra anche la posizione delle parole del contesto nel testo di origine, e la loro frequenza assoluta totale. Ulteriori informazioni sono ottenibili dalla scheda *Opzioni*, come già visto per la ricerca di forme e lemmi. Se è stata concessa l’autorizzazione per la pubblicazione del testo di origine, questo è accessibile cliccando su un link a margine. La Figura 5 contiene un esempio di ricerca di concordanze.

Forme Lemmi **Concordanze** Coricorrenze Lista Opzioni

Forma: squalo

Lemma:

Parte del discorso: sostantivo

Sintagmaticità della forma: Non sintagmatica

Contesto sintagmatico: Non parte di forma sintagmatica

Elementi nel contesto: a sinistra: 2 a destra: 3

Tipo del testo:

Risultati per pagina: 10

5 elementi trovati

Titolo: SULLE SECCHIE DI FIUMICINO Autore: RICCARDO LUNA Genere: cronaca locale Classe: quotidiano visualizza il testo completo	Posizione nel testo di origine:	319	320	321	322	323	324
	Forma:	dello	-lo	squalo	sembrava	irrimediabilmente	in
	Frequenza assoluta del lemma:	197126	408845	16	2447	20	65568

Titolo: A TAVOLA CON I POTENTI E CONTO A TANGENTOPOLI Autore: Genere: cronaca locale Classe: quotidiano visualizza il testo completo	Posizione nel testo di origine:	1011	1012	1013	1014	1015	1016
	Forma:	.	lo	squalo	e'	un	tipico
	Frequenza assoluta del lemma:	163116	408845	16	54838	78322	170

Titolo: Zio Paperone e il "bacalao" Autore: Gentina Carlo Genere: bambini, ragazzi Classe: periodico	Posizione nel testo di origine:	50	51	52	53	54	55
	Forma:	il	nostro	squalo	e'	molto	piu'
	Frequenza assoluta del lemma:	408845	2894	16	54838	3004	13825

Titolo: Parla l'amante della contessa Autore: Haver Flavio Genere: cronaca nera Classe: quotidiano visualizza il testo completo	Posizione nel testo di origine:	193	194	195	196	197	198
	Forma:	-lo	.	squalo	.	vittorio	sbardella
	Frequenza assoluta del lemma:	408845	44809	16	44809	162	19

Titolo: Tutte le notti divento la ragazza dei camionisti Autore: Cenni Rita Genere: informazione generale Classe: periodico	Posizione nel testo di origine:	728	729	730	731	732	733
	Forma:	.	uno	squalo	tigre	?	se
	Frequenza assoluta del lemma:	222233	78322	16	34	11054	6824

Figura 5: La figura riporta le concordanze della forma “squalo” sostantivo.

La scheda *Coricorrenze* consente di cercare i contesti in cui compare una coppia di parole, adiacenti oppure ad una distanza predefinita in termini di numero di parole intercorrenti. Si noti che non è possibile tuttavia ottenere informazioni sulla forza statistica delle coricorrenze, né ricercare le collocazioni data una forma base. La Figura 6 mostra un esempio di ricerca della coppia “primavera” e “estate”.



Figura 6: La figura riporta i primi quattro risultati della ricerca della coppia data da “primavera” e “estate”, dove “estate” ricorre ad una distanza di massimo tre posizioni prima o dopo “primavera”.

Tra gli aspetti innovativi e particolarmente utili per la ricerca in psicolinguistica, vi è la possibilità di effettuare ricerche con vincoli. I vincoli sono attivabili sulla parte del discorso, la lunghezza, le statistiche. Quando si effettua una ricerca con vincoli, il sistema restituisce tutte le forme o i lemmi che le rispettano, ordinando i lemmi per rango e le forme per lemma, parte del discorso e rango. La Figura 7 mostra un esempio di ricerca di lemmi verbali con vincoli di frequenza e lunghezza.

Forme Lemmi Concordanze Coricorrenze Lista Opzioni

Lemma: Azzera

Parte del discorso: verbo Azzera

Numero di lettere: minimo: 7 Azzera massimo: 7 Azzera

Sintagmaticità: Azzera

▼ Statistiche

Frequenza assoluta: Minima 2 Azzera Massima 2.097 Azzera

Frequenza assoluta nei periodici: Azzera Azzera

Frequenza assoluta nei quotidiani: Azzera Azzera

Frequenza assoluta nei libri: Azzera Azzera

Usa i logaritmi dei valori

Risultati per pagina: 25

Cerca Azzera tutto

12 lemmi trovati.

Lemma	Codice parte del discorso	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
OPPORRE	V	2387	119	2.08	28.96	0.92	Concordanze Forme Forme dei tempi composti
GIURARE	V	2445	123	2.09	28.11	0.86	Concordanze Forme Forme dei tempi composti
FALLIRE	V	2462	121	2.08	27.97	0.87	Concordanze Forme Forme dei tempi composti
STUPIRE	V	2551	117	2.07	26.59	0.85	Concordanze Forme Forme dei tempi composti

Figura 7: La figura mostra i primi 4 risultati della ricerca di lemmi che siano verbi, con lunghezza pari a 7 lettere e con frequenza assoluta espressa in logaritmo tra 2 e 2.1.

Di notevole utilità è anche la possibilità di ricercare liste di elementi. A questo scopo occorre caricare nella scheda *Liste* un file di testo contenente la lista delle parole da ricercare. Le parole interamente in maiuscolo saranno interpretate come lemmi, quelle in minuscolo come forme: nel secondo caso, il sistema restituirà tutte le forme omografe ritrovate nel corpus, ricondotte al lemma. La Figura 8 illustra un esempio di ricerca di una lista di forme.

The screenshot shows a web-based linguistic search interface. At the top, there are tabs for 'Forme', 'Lemmi', 'Concordanze', 'Coricorrenze', 'Lista', and 'Opzioni'. Below the tabs is a search box with a 'Browse...' button. The search results are displayed in a list box containing the following words: acciaio, acqua, ambra, fumo, labbra, nebbia, ombra, pace, perla, pioggia, ramo, sonno, stella. Below the list box is a 'Cerca' button.

Below the search interface is a detailed table of results. The table has the following columns: Richiesta, Tipo della richiesta, Forma, Lemma, Codice parte del discorso, Codice struttura sintagmatica, Struttura sintagmatica, Rango, Freq. ass. totale, Log. freq. ass. totale, Freq. rel. totale, Dispersione totale, and Cerca concordanze. The table lists 18 rows of data for the word 'ambra' and other related words.

Richiesta	Tipo della richiesta	Forma	Lemma	Codice parte del discorso	Codice struttura sintagmatica	Struttura sintagmatica	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Cerca concordanze
acciaio	forma	acciaio	ACCIAIO	G	-	-	129648	1	0.00	0.01	0.00	Concordanze Lemma
acciaio	forma	acciaio	ACCIAIO	S	-	-	4596	76	1.88	14.99	0.74	Concordanze Lemma
acciaio	forma	acciaio	ACCIAIO	S	in E	in nome proprio	62082	3	0.48	0.14	0.12	Concordanze Lemma
acqua	forma	acqua	ACQUA	S	-	-	369	945	2.98	223.14	0.89	Concordanze Lemma
acqua	forma	acqua	ACQUA	S	in E	in nome proprio	17615	15	1.18	2.16	0.51	Concordanze Lemma
ambra	forma	ambra	AMBRA	E	-	-	17659	18	1.26	2.15	0.42	Concordanze Lemma
ambra	forma	ambra	AMBRA	S	-	-	31034	7	0.85	0.78	0.40	Concordanze Lemma
fumo	forma	fumo	FUMO	E	-	-	88160	1	0.00	0.03	0.00	Concordanze Lemma
fumo	forma	fumo	FUMO	S	-	-	3163	109	2.04	23.44	0.80	Concordanze Lemma
fumo	forma	fumo	FUMO	S	in B	in avverbio	23042	11	1.04	1.37	0.44	Concordanze Lemma
fumo	forma	fumo	FUMARE	V	-	-	36850	5	0.70	0.55	0.38	Concordanze Lemma
labbra	forma	labbra	LABBRO	S	-	-	2663	143	2.16	28.30	0.74	Concordanze Lemma
nebbia	forma	nebbia	NEBBIA	S	-	-	4225	77	1.89	16.56	0.80	Concordanze Lemma
ombra	forma	ombra	OMBRA	S	-	-	1813	184	2.26	43.47	0.89	Concordanze Lemma
ombra	forma	ombra	OMBRA	S	in P	in preposizione	8945	34	1.53	6.14	0.66	Concordanze Lemma
ombra	forma	ombra	OMBRA	S	in F	in nome	39330	6	0.78	0.48	0.27	Concordanze

Figura 8: Esempio di ricerca di una lista di forme. Si noti ad esempio la forma “ambra”, ricondotta a due diversi lemmi: nome proprio (E) e sostantivo (S).

Le ricerche con vincoli menzionate sopra possono essere ulteriormente affinate prendendo in considerazione la struttura sintagmatica. Si possono estrarre le statistiche relative a lemmi sintagmatici specifici (ad esempio, “di fianco”). È inoltre possibile estrarre tutti i lemmi sintagmatici che rispettino criteri definiti (es. tutti gli avverbi sintagmatici), come illustrato in Figura 9.

Lemma	Codice parte del discorso	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
PER ESEMPIO	B	450	635	2.80	155.05	0.92	Concordanze Forme
IN PARTICOLARE	B	980	327	2.51	79.06	0.91	Concordanze Forme
AD ESEMPIO	B	1027	331	2.52	76.49	0.87	Concordanze Forme
PIU' O MENO	B	1518	215	2.33	51.15	0.89	Concordanze Forme

Figura 9: La figura mostra i risultati relativi alla ricerca di lemmi avverbiali sintagmatici (vincolo imposto tramite la tendina “sintagmaticità”) con lunghezza minima di 10 lettere e frequenza assoluta minima di 200.

All’interno dell’interfaccia, è stata poi introdotta la nozione di forma sintagmatica, che ricalca quella di lemma sintagmatico definita in CoLFIS. Una forma è sintagmatica quando contiene altri elementi: essa può far capo sia ad un lemma sintagmatico sia ad un lemma non sintagmatico. La ricerca di forme sintagmatiche è di particolare utilità nel caso dei verbi, consentendo di distinguere tra tempi semplici e tempi composti (questi ultimi considerati come forme sintagmatiche). Si veda l’esempio in Figura 10.

Forme Lemmi Concordanze Coricorrenze Lista Opzioni

Forma:

Numero di lettere: minimo: massimo:

▼ Dizionario

Lemma:

Parte del discorso:

Sintagmaticità:

Sintagmaticità della forma:

► Statistiche

Risultati per pagina:

6 forme trovate.

Forma	Lemma	Codice parte del discorso	Codice struttura sintagmatica	Struttura sintagmatica	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
affilare	AFFILARE	V	-	-	42758	4	0.60	0.40	0.34	Concordanze Lemma
affilano	AFFILARE	V	-	-	43735	4	0.60	0.38	0.31	Concordanze Lemma
affiliando	AFFILARE	V	-	-	62082	2	0.30	0.14	0.19	Concordanze Lemma
affilato	AFFILARE	V	in TC	in tempo composto	63761	2	0.30	0.13	0.19	Concordanze* Lemma
affilava	AFFILARE	V	-	-	67570	2	0.30	0.12	0.19	Concordanze Lemma
affiliato	AFFILARE	V	-	-	129648	1	0.00	0.01	0.00	Concordanze Lemma

Forme Lemmi Concordanze Coricorrenze Lista Opzioni

Forma:

Numero di lettere: minimo: massimo:

▼ Dizionario

Lemma:

Parte del discorso:

Sintagmaticità:

Sintagmaticità della forma:

► Statistiche

Risultati per pagina:

2 forme trovate.

Forma	Lemma	Codice parte del discorso	Codice struttura sintagmatica	Struttura sintagmatica	Rango	Freq. ass. totale	Log. freq. ass. totale	Freq. rel. totale	Dispersione totale	Ricerche correlate
ha affilato	AFFILARE	V	-	-	98248	1	0.00	0.02	0.00	Concordanze Lemma
hanno affilato	AFFILARE	V	-	-	98248	1	0.00	0.02	0.00	Concordanze Lemma

Figura 10: Il pannello superiore mostra i risultati della ricerca di forme non sintagmatiche del lemma verbale affilare (vincolo imposto tramite la tendina “Sintagmaticità della forma”). Il pannello inferiore mostra i risultati della ricerca di forme sintagmatiche sul medesimo lemma. Si noti che la ricerca di forme non sintagmatiche restituisce anche la forma “affilato” in TC, con due ricorrenze: si tratta dei due casi che ricorrono come elemento non sintagmatico all’interno di strutture sintagmatiche nel pannello in alto.

Una caratteristica saliente di tutta l'interfaccia EsploraCoLFIS è la sua agilità, ovvero il facile accesso alle informazioni ottenibili dai diversi tipi di ricerche. Nel caso delle schede *Forme*, *Lemmi* e *Liste*, la maschera dei risultati contiene una colonna adiacente *Ricerche correlate*, che consente di spostarsi ai risultati corrispondenti delle schede *Concordanze* e *Corricorrenze*, senza dover effettuare una ricerca *ex novo*. Similmente, nelle schede *Concordanze* e *Corricorrenze* è possibile cliccare sulle forme e sui lemmi che compaiono nei risultati corrispondenti spostandosi sulle schede relative.

3 Note conclusive

Nel panorama dei corpora disponibili per la lingua italiana, CoLFIS si presenta come uno dei riferimenti più validi quando si vogliono ottenere valori rappresentativi della distribuzione del lessico della lingua italiana (Baroni, 2010). Rispetto ad altri corpora esistenti, se da un lato CoLFIS presenta dimensioni ridotte, dall'altro lato si distingue per i criteri di campionamento, che riflettono la proporzione di tipologie testuali diverse (quotidiani, periodici, libri) riscontrata nelle letture degli italiani. Per questo CoLFIS rappresenta ormai il corpus di riferimento per la frequenza nella letteratura psico e neurolinguistica sull'italiano. Ci auguriamo che l'interfaccia EsploraCoLFIS possa migliorare l'esperienza degli utenti nella consultazione, ed espandere i possibili utilizzi di CoLFIS.

Riferimenti Bibliografici

- Baroni, M. (2010). Corpora di italiano. In R. Simone (ed.), *Enciclopedia dell'italiano*, Vol 1, pp. 300-303. Roma: Istituto della Enciclopedia Italiana.
- Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thornton, A.M. 2005. Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS). <http://linguistica.sns.it/CoLFIS/Home.htm>
- Burani, C. Criteri di lemmatizzazione. <http://linguistica.sns.it/CoLFIS/Home.htm>
- Laudanna, A., Thornton, A.M., Brown, G., Burani, C. e Marconi, L. (1995). Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente. In S. Bolasco, L. Lebart e A. Salem (a cura di), *III Giornate internazionali di Analisi Statistica dei Dati Testuali*. Volume I, pp. 103-109. Roma: Cisu.
- Marconi, L. Analisi statistiche. <http://linguistica.sns.it/CoLFIS/Home.htm>