

Siamo lieti di annunciare un nuovo repertorio lessicale:

CoLFIS (Corpus e Lessico di Frequenza dell'Italiano Scritto)

realizzato da

Pier Marco Bertinetto[°], Cristina Burani*, Alessandro Laudanna^{^*},
Lucia Marconi⁺, Daniela Ratti⁺, Claudia Rolando⁺, Anna Maria Thornton[§]

[°] Scuola Normale Superiore, Pisa

* Istituto di Scienze e Tecnologie della Cognizione, CNR, Roma

[^] Università di Salerno

⁺ Istituto di Linguistica Computazionale, Unità Staccata di Genova, CNR, Genova

[§] Università de L'Aquila

Il **corpus**, estratto da quotidiani, periodici e libri, consta di **3.150.075** ricorrenze lessicali. E' stato concepito con l'intento di approssimare il più esattamente possibile il quadro delle letture preferite degli Italiani, secondo i dati statistici ufficiali.

Il **lessico di frequenza** include due sottoparti: **formario** e **lemmario**. Quest'ultimo disambigua le forme identiche (es., *porti*) riportandole ai diversi lemmi di provenienza (es., *porti* come forma plurale del sostantivo *porto*, oppure seconda persona singolare dell'indicativo del verbo *portare*) ed inoltre tratta come singole entrate lessicali le parole sintagmatiche (es., *gamba del tavolo*).

Le liste lessicali sono attualmente disponibili per il **download** gratuito presso:

<http://www.istc.cnr.it/material/database/colfis/>

<http://alphalinguistica.sns.it/BancheDati.htm>

Le liste sono organizzate secondo varie possibilità di ricerca: rango di frequenza, ordine alfabetico inverso, con / senza la distinzione tra caratteri minuscoli e maiuscoli, etc.

Il **corpus** non è ancora disponibile, ma ci si propone di metterlo in linea appena possibile, compatibilmente con l'ottenimento delle autorizzazioni.

Quest'opera è stata realizzata col finanziamento del CNR.

Gli autori contano sulla collaborazione degli utenti per allargare e migliorare il servizio offerto.