

LINGUISTICA DEI CORPORA

Strumenti, risorse e applicazioni

Isabella Chiari (isabella.chiari@uniroma1.it)

Home page: www.alphabit.net

CORPORA DI RIFERIMENTO DELLA LINGUA ITALIANA¹

Corpora della lingua scritta

» CORIS / CODIS (**Corpus di Italiano Scritto contemporaneo**), diretto da R. Rossini Favretti (1998): 100 milioni di parole

URL: http://corpus.cilta.unibo.it:8080/coris_ita.html

» COLFIS (**Corpus e Lessico di Frequenza dell'Italiano Scritto**), di Bertinetto, Burani, Laudanna, Marconi, Ratti, Rolando e Thornton (testi anni Novanta): 3.150.075 parole

URL: <http://www.ge.ilc.cnr.it/page.php?ID=archCoLFIS&lingua=it>

» TLIO (**Tesoro della lingua italiana delle origini**), a cura dell'Istituto Opera del Vocabolario Italiano: 21.779.245 parole

URL: <http://www.ovi.cnr.it/index.php?page=la-banca-dati>

» Corpus **LA REPUBBLICA**, a cura di M. Baroni, S. Bernardini, F. Comastri, L. Piccioni, A. Volpi, G. Aston, M. Mazzoleni: 380.823.725 parole

URL: <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>

Corpora della lingua parlata

» LIP (**Lessico di frequenza dell'italiano parlato**), curato da De Mauro, Mancini, Vedovelli e Voghera (1993): 57 ore di parlato

URL: http://languageserver.uni-graz.at/badip/badip/20_corpusLip.php

» CLIPS (**Corpora Linguistici per l'Italiano Parlato e Scritto**), diretto da F. Albano Leoni (2007): 100 ore di parlato

Un corpus elettronico è "A corpus which is encoded in a standardized and homogeneous way for open-ended retrieval tasks"

(Eagles, 1996a: 3).

¹ Per una lista più completa si veda: http://www.alphabit.net/Risorse/Corpora_ita.html

URL: <http://www.clips.unina.it/>

» **Corpus di italiano Parlato (LABLITA)**, diretto da E. Cresti, corpora multipli costruiti in vari momenti.

URL: <http://lablita.dit.unifi.it/>

BIBLIOGRAFIA ESSENZIALE

- Barnbrook G., *Language and Computers. A Practical Introduction to the Computer Analysis of Language*, Edinburgh University Press, Edinburgh 1996.
- Biber D., S. Conrad e R. Reppen (a cura di), *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge University Press, Cambridge 2000.
- Chiari I., *Linguistica e informatica: la linguistica dei corpora in Italia*, in «Bollettino di Italianistica», 4, 2005, pp. 101-118.
- Chiari I., *Introduzione alla linguistica computazionale*, Bari: Laterza 2007.
- De Mauro T. e I. Chiari (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*. Aracne, Roma 2005.
- Gigliozzi G., *Il testo e il computer*, Mondadori, Milano 1997.
- Giuliano L., *L'analisi automatica dei dati testuali*, LED, Milano 2004.
- Kennedy G., *An Introduction to Corpus Linguistics*, Longman, London-New York 1998.
- Lana M., *Il testo nel computer. Dal web all'analisi dei testi*, Bollati Boringhieri, Milano 2004.
- Lawler J. e H. Aristar-Dry (a cura di), *Using Computers in Linguistics. A Practical Guide*, Routledge, London 1998.
- Lenci A. et alii, *Testo e computer. Elementi di linguistica computazionale*, Carocci, Roma 2005.
- McEnery T. e A. Wilson, *Corpus Linguistics*, Edinburgh University Press, Edinburgh 1996.
- Oakes M., *Statistics for Corpus Linguistics*, Edinburgh University Press, Edinburgh 1998.
- Rossini Favretti R. (a cura di), *Linguistica e informatica. Multimedialità, corpora e percorsi di apprendimento*, Bulzoni, Roma 2000.
- Sinclair J., *Corpus, Concordance, Collocation*, Oxford University Press, Oxford 1991.
- Spina S., *Fare i conti con le parole*, Guerra, Perugia 2001.
- Tognini-Bonelli E., *Corpus Linguistics at Work*, in «Studies in Corpus Linguistics», 6, John Benjamins, Amsterdam-Philadelphia 2001.

SITI DI INTERESSE

Corpus linguistics, di T. McEnery e A. Wilson

<http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/contents.htm>

Centre for Corpus Research dell'Università di Birmingham

<http://www.corpus.bham.ac.uk/>

Corpora and Corpus-based Computational Linguistics, di M. Barbera

<http://www.bmanuel.org/>

UCREL (University Center for Computer Corpus Research on Language)

<http://www.comp.lancs.ac.uk/computing/research/ucrel/>

Gateway to Corpus linguistics

<http://www.corpus-linguistics.de/>

Michael Barlow Corpus Page

<http://www.michaelbarlow.com/>

RACCOLTE E DISTRIBUZIONE DI CORPORA E RISORSE

LDC – Linguistic Data Consortium

<http://ldc.upenn.edu/>

ELRA (European Language Resources Association)

<http://www.icp.grenet.fr/ELRA>

Progetto Gutenberg

<http://www.gutenberg.net/>

The Oxford Text Archive (OTA)

<http://ota.ox.ac.uk/>

Internet Corpora (Sharoff)

<http://corpus.leeds.ac.uk/internet.html>

TUTORIAL DI LINGUISTICA DEI CORPORA

Corpus linguistics tutorial, di T. McEnery e A. Wilson

<http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/corpus2/2fra1.htm>

Tutorial Concordances and Corpora, di Ch. N. Ball

<http://www.georgetown.edu/faculty/ballc/corpora/tutorial.html>

W3C Corpus Linguistics tutorial

http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/introduction.html

Il testo di questo hand-out è scaricabile anche al sito Alphabit.net, insieme alle slides Powerpoint di questa lezione all'indirizzo:

<http://www.alphabit.net/Docente/congresses.html>