

Free corpora

ENGLISH

- <http://www.natcorp.ox.ac.uk/> (British National Corpus -BNC-)
- <http://corpus.byu.edu/coca/> (Corpus of Contemporary American English -COCA-)
- <http://ronaldo.cs.tcd.ie/tec2/jnlp/> (Translational English Corpus -TEC-)
- <http://www.ucl.ac.uk/english-usage/ice/index.htm> (International Corpus of English -ICE-)

SPANISH

- <http://www.rae.es/rae.html> (Corpus de Referencia del Español Actual -CREA- and Corpus Diacrónico del Español -CORDE-)
- <http://www.corpusdelespanol.org>

GERMAN

- <http://wortschatz.uni-leipzig.de/>
- <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html> (-NEGRA-). Approx. 355,000 words.
- http://www.ids-mannheim.de/kl/projekte/cosmas_1/

FRENCH

- http://www.lextutor.ca/concordancers/concord_f.html . Various corpora, among which, 1 M words from *Le Monde* 1998.
- <http://www.lib.uchicago.edu/efts/ARTFL/newhome/texts/> . Literary.
- <http://www.swarthmore.edu/Humanities/clicnet/> . Literary resource searcher.

ARABIC

- <http://arabicorpus.byu.edu/> . Literary and journalistic texts.
- * <http://www.andy-roberts.net/software/aConCorde/index.html> . Multi-lingual concordance tool.

CHINESE

- <http://www.sinica.edu.tw/ftms-bin/kiwi.sh> . 5 M words.

GREEK

- <http://hnc.ilsp.gr/en/> (Hellenic National Corpus -HNC-). More than 46 M words of written text.

SOUTH AFRICAN LANGUAGES

- <http://tshwanedje.com/> . Commercial site. Various language resources.

TURKISH

- <http://www.ii.metu.edu.tr/corpus> (METU Turkish Corpus). Written.
- <http://std.metu.edu.tr/en/> . METU Spoken Turkish Corpus Project.

DUTCH

- <http://lands.let.kun.nl/cgn/ehome.htm> (Spoken Dutch Corpus).

POLISH

- <http://www.nkjp.pl/> (National Corpus of Polish).
- <http://korpus.pl/> (IPI PAN Corpus).

VARIOUS LANGUAGES

- <http://bwananet.iula.upf.edu/index.htm> . Various fields.
- <http://sli.uvigo.es/CLUVI/index.html> . Contains texts in Galician, Spanish, English, French and Basque. Various combinations in legal and literary texts.
- <http://www.statmt.org/europarl/> (European Parliament Proceedings Parallel Corpus -EUROPARL-). Aligned texts from 1996-2009.
- <http://corpora.informatik.uni-leipzig.de/download.html> (Leipzig Corpora Collection). Various languages: Catalan, Danish, Dutch, English, Estonian, Finnish, French, German, Italian, Japanese, Korean, Norwegian, Sorbian, Swedish, Turkish.



- <http://www.uow.edu.au/~dlee/CBLLinks.htm> (Bookmarks for Corpus-based Linguists). Wide range of corpus-related resources (online corpora, software, publications, etc.).