ONLINE / DOWNLOADABLE CORPORA

- British National Corpus (BNC): http://www.natcorp.ox.ac.uk/
- BNCWeb: http://bncweb.lancs.ac.uk/cgi-binbncXML/BNCquery.pl?theQuery=search&urlTest=yes
- The Brigham-Young version of the BNC, http://corpus.byu.edu/
- Corpus of Contemporary American English (COCA), http://corpus.byu.edu/
- Corpus of Historical American English (COHA), http://corpus.byu.edu/
- Time Magazine Corpus, http://corpus.byu.edu/
- Links to various learner corpora around the world (some of them are not freely available): http://www.uclouvain.be/en-cecl-lcworld.html
- CORPS: A Corpus of tagged Political Speeches http://hlt.fbk.eu/corps
- American National Corpus (comparable across genres to the BNC): http://www.americannationalcorpus.org/OANC/index.html
- Junk email corpus: http://clg.wlv.ac.uk/resources/junk-emails/index.php
- MICASE (Michigan Corpus of Academic Spoken English): http://quod.lib.umich.edu/m/micase/
- Michigan Corpus of Upper-Level Student Papers: http://search-micusp.elicorpora.info/simple/
- British Academic Spoken English corpus (BASE): http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/
- The Santa Barbara Corpus of Spoken American English: http://www.linguistics.ucsb.edu/research/sbcorpus.html
- Parliament Archive: http://www.parliamentlive.tv/Main/Archive.aspx
- ELFA: English as a Lingua Franca in Academic Settings. Corpus of spoken academic English as a Lingua Franca (not free, more information available online) http://www.helsinki.fi/englanti/elfa/elfacorpus.html#licence

German language Corpora:

- Korpora des gesprochenen Deutsch (not freely available): http://dsav-wiss.idswissmannheim.de/KORPORAS.HTM
• Datenbank für Gesprochenes Deutsch: http://dgd.ids-mannheim.de:8080/dgd/pragdb.dgdExtern.welcome

• Negra-corpus: A Syntactically Annotated Corpus of German Newspaper Texts http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/

**Learner Corpora (data produced by foreign language learners)**

• **ICLE**, the *International Corpus of Learner English* contains argumentative essays written by higher intermediate to advanced learners of English from several mother tongue backgrounds (not available online) http://www.uclouvain.be/en-cecl-icle.html

• **FRIDA**, The *French Interlanguage Database* contains texts written by learners of French as a foreign language (not available online) http://www.uclouvain.be/en-cecl-frida.html


• **LONGDALE**, *Longitudinal Database of Learner English*. a large longitudinal database of learner English containing data from learners from a wide range of mother tongue backgrounds (not available online) http://www.uclouvain.be/en-cecl-longdale.html

• **VESPA**); The Varieties of English for Specific Purposes dAtabase (VESPA) learner corpus. a large corpus of *English for Specific Purposes texts* written by L2 writers from various mother tongue backgrounds. (not available online) http://www.uclouvain.be/en-cecl-vespa.html

**Pedagogical Corpora**

• **TeMa**: pedagogical corpus, which contains pedagogical materials, for instance textbook materials http://www.uclouvain.be/en-cecl-tema.html

**Multilingual Corpora**

• **PLECI**. The *Poitiers-Louvain Échange de Corpus Informatisés*. a large bidirectional English-French translation corpus that includes literary prose and newspaper articles. (not available online) http://www.uclouvain.be/en-cecl-pleci.html

• **MULT-ED**. The *Multilingual Editorial Corpus* is a multilingual comparable corpus of newspaper editorials written in English, Dutch, French and Swedish. (not available online) http://www.uclouvain.be/en-cecl-multed.html

**TV Scripts**

• Scripts of different American TV shows: http://www.script-o-rama.com/tvscript.shtml

• Scripts of the American TV show *Seinfeld*: http://www.seinfeldscripts.com/
• Scripts of the TV show Roswell, German translation: http://www.tv-scripte.de/roswell/index.html