

Online Corpora

Corpus	Description	Availability / Accessibility	Link
1 Glottolog	Catalogue of varieties. With bibliographies and links to other websites.		http://glottolog.org/
2 Ethnologue	Catalogue of varieties. With bibliographies and links to other websites.		http://www.ethnologue.com/
3 IDEA	Database of English speech samples from speakers of various L1 languages. With annotations.		http://www.dialectsarchive.com/
4 Speech Accent Archive	Database of English speech samples from speakers of various L1 languages. With annotations.		http://accent.gmu.edu/
5 UPSID	Analysis of phonetic inventories (919) and languages (451)	Online version (simplified): http://web.phonetik.uni-frankfurt.de/upsid_info.html	
6 corpus.byu.edu	Various corpora of spoken and written English (and other languages).	Free registration for online access (daily limit on queries).	Offline version: http://www.linguistics.ucla.edu/faciliti/sales/software.htm#upsid http://corpus.byu.edu/
	Hansard Corpus (British Parliament) (1.6 billion words) Wikipedia Corpus (with virtual corpora) (1.9 billion words) Global Web-Based English (GloWbE) (1.9 billion words) Corpus of Contemporary American English (COCA) (450 million words) Corpus of Historical American English (COHA) (400 million words) TIME Magazine Corpus (100 million words) Corpus of American Soap Operas (100 million words) British National Corpus (BYU-BNC) (100 million words) Strathy Corpus (Canada) (50 million words) Corpus del Español (100 million words) Corpus do Português (45 million words) Google Books: American English (155 billion words) Google Books: British English (34 billion words) Google Books: One Million Books (89 billion words) Google Books: Spanish	Downloadable corpora can be purchased, but database management tools and skills are required.	
7 TIME Magazine	Archive of the US magazine TIME from 1923 onwards.		http://www.time.com/time/archive
8 VOICE Vienna Oxford International Corpus of English	A computer corpus of English as a Lingua Franca (ELF) consisting of transcripts of naturally occurring interactions.	Free registration for online access and a free download version.	http://www.univie.ac.at/voice/
9 LexisNexis	Database providing access to a large number of newspapers, magazines, periodicals etc. from around the world.	Login via University of Bern library database.	http://www.unibe.ch/university/services/university_library/research/databases/index_eng.html
10 Factiva	Database providing access to a large number of newspaper archives around the world.	Login via University of Bern library database.	http://www.unibe.ch/university/services/university_library/research/databases/index_eng.html
11 FRED Freiburg English Dialect Corpus	FRED is a monolingual spoken-language dialect corpus consisting of oral history interviews conducted in the 1970s and 1980s, with speakers from nine larger dialect areas in England, Scotland, Wales, the Hebrides, and the Isle of Man. The corpus consists of sound recordings and orthographic transcripts, spanning approximately 2.5 million words and 300 hours of speech. The FRED sampler (FRED-S) comprises 1 million words (ca. 123 hours) from five dialect areas: the Southwest of England, the Southeast of England, the English Midlands, the North of England, and the Scottish Lowlands. On the whole, FRED-S offers access to 121 interviews (139 recordings) with 144 dialect speakers born between the 1870s and the 1940s. Most of the speakers were born before 1920 (mean date of birth is 1905) and were aged 60 or older when interviewed (mean age is 74.5 years at recording date).		https://freidok.uni-freiburg.de/proj/1 https://fred.ub.uni-freiburg.de/ https://freidok.uni-freiburg.de/data/10845