**Corpuses**

*Spoken and Written - From different time eras*

***Tools***Corpus research, extraction and management tools  
  
Corpus-Extraction & Compilation Sites

Many : <http://martinweisser.org/corpora_site/web_corpus.html>

1. Concordance Web extraction tool: <http://glossa.fltr.ucl.ac.be/>
2. Concordancer for web (kwic) - <http://www.kwicfinder.com/KWiCFinder.html>

Accessing Text Corpora and Lexical Resources

1. Nltk Corpus - <https://www.nltk.org/book/ch02.html>
2. Corpus Readers - <https://www.nltk.org/howto/corpus.html>

Programming Historian Necessary Lessons (Many)

<https://programminghistorian.org/en/lessons/>   
  
  
Concordance

<https://en.wikipedia.org/wiki/Concordance_(publishing)>   
*Its Diminishing Importance*  
Today, the ability to combine the result of queries concerning multiple terms (such as searching for words near other words) has reduced interest in concordance publishing. In addition, mathematical techniques such as [latent semantic indexing](https://en.wikipedia.org/wiki/Latent_semantic_indexing) have been proposed as a means of automatically identifying linguistic information based on word context. **English***General Useful List of Corpora (Spoken & Written)**Spoken*

1. Old Bailey’s Court hearings  
   <https://www.oldbaileyonline.org/>   
   Old Bailey’s Corpus:  
   <http://fedora.clarin-d.uni-saarland.de/oldbailey/downloads.html>   
   OBC Manual: <http://fedora.clarin-d.uni-saarland.de/oldbailey/downloads/OBC_2.0_Manual%202016-07-13.pdf>
2. Opus (Open **Parallel** (multilingual)Corpus) *Ted Talks* - <http://opus.nlpl.eu/index.php>
3. American English (Non-downloadable)  
   Michigan Cospus of Academic Spoken English: <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>
4. *Word Neighbours* (Non-downloadable) - Search Engine: mix of spoken & written English genres (user-selectable); some texts are from the BNC.   
   <http://wordneighbors.ust.hk/>
5. Business Letter Corpus (concordancer/ word-search) Non-downloadable.  
   <http://www.someya-net.com/concordancer/>
6. *Useful - Phrases in English PIE -* [*http://phrasesinenglish.org/*](http://phrasesinenglish.org/)

*Written*

1. Wikipedia -
   1. <https://en.wikipedia.org/wiki/Wikipedia:Database_download>
2. Gutenberg Project -
   1. <http://www.gutenberg.org/>

*Mixed*

1. Brown University Corpus (American English) -
   1. <https://www.kaggle.com/nltkdata/brown-corpus>
   2. NLTK - <https://www.nltk.org/data.html>
   3. <http://www.sls.hawaii.edu/bley-vroman/brown_corpus.html>
2. British National Corpus
   1. <http://www.natcorp.ox.ac.uk/>
   2. <https://www.english-corpora.org/bnc/>
   3. <http://www.natcorp.ox.ac.uk/getting/index.xml>

**Yoruba**  
Yoruba Zipf research Paper  
Diacritized text statistical analysis  
<http://www.dline.info/jcl/fulltext/v6n3/v6n3_2.pdf>